



**UNIVERSIDAD MICHOACANA DE  
SAN NICOLÁS DE HIDALGO**

**DIVISIÓN DE ESTUDIOS DE POSGRADO EN  
INGENIERÍA MECÁNICA**



**PRONÓSTICO DE SERIES DE TIEMPO DE  
VIENTO MULTIFRACTALES**

**TESIS**

Que para obtener el grado de:  
**DOCTORA EN CIENCIAS EN INGENIERÍA MECÁNICA**

PRESENTA:  
**M.C. ALMA ROSA MÉNDEZ GORDILLO**

DIRECTOR DE TESIS:  
**DR. ERASMO CADENAS CALDERÓN**

CO- DIRECTOR DE TESIS:  
**DR. GILBERTO GONZÁLES AVALOS**

MORELIA, MICHOACÁN, SEPTIEMBRE DE 2022



**CONACYT**

Consejo Nacional de Ciencia y Tecnología

**Dedicado:**

*A mis padres,*

*Hermanos,*

*Abuelos,*

*y para Sixtos*

*Hay formas muy complejas que serían iguales desde cerca y desde lejos.*

***Benoit Mandelbrot***

# Agradecimientos

Primeramente a mis asesores los Doctores. Erasmo Cadenas Calderón y Gilberto González Avalos por la disposición, asesorías y ayuda brindada en este camino, por creer en mi persona desde el comienzo de este difícil sendero y lograr una marca total en mi vida profesional.

Gracias Universidad Michoacana de San Nicolás de Hidalgo (UMSNH) por acogerme como mi segunda casa, agradezco a mis profesores de la División del Posgrado de la Facultad de Ingeniería Mecánica, por ser académicos con un gran compromiso por la educación. Las palabras para demostrar mi sentimiento no existen; por hacer mis metas y sueños posibles, forjarme como una profesional de bien. Gracias a esta gran Institución por su calidad y humanidad por ser mi hogar durante algunos años, siempre le deberé lo que soy profesionalmente.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por la beca (No 249609) para realizar mis estudios de Doctorado.

Agradezco a los miembros de mi comité tutorial, los Doctores Laura Alicia Ibarra Bracamontes, Alicia Aguilar Corona, y Homero Geovani Díaz Marín, reconociendo su colaboración en mi trabajo de tesis para obtener el grado de Doctora, desde el inicio del proceso y hasta el día de hoy. Gracias por la ayuda en mi formación profesional y humana, por sus comentarios y revisiones, durante el periodo de estudios.

Finalmente, pero no por ello menos importante agradezco a Dios, familiares, amigos y compañeros por el apoyo durante este difícil camino.

# Resumen

En este trabajo, se analizaron los errores de pronóstico producidos por los modelos ARIMA, Aprendizaje Profundo y Persistencia, modelos aplicados a dos series de tiempo para predecir la velocidad del viento un paso adelante. El análisis consistió en verificar, con la ayuda del MF-DFA, si existían patrones multifractales en los errores de pronóstico. Luego, los patrones multifractales de los errores fueron extraídos por un modelo de Cascada Multiplicativa. La suma de los patrones extraídos de los modelos generó nuevos modelos híbridos como la Cascada Multiplicativa-ARIMA, la Cascada Multiplicativa-Aprendizaje Profundo y la Cascada Multiplicativa-Persistencia. Esos nuevos modelos mostraron una disminución del error de pronóstico final. Las series temporales de velocidad del viento utilizadas para este estudio se midieron en dos parques eólicos del Estado de Oaxaca, México, a saber, La Mata y La Venta. La primera serie de tiempo tiene cinco años de datos promediados de diez minutos, mientras que la segunda tiene seis años y once meses de datos por hora. Ambas series de tiempo fueron proporcionadas por la Comisión Federal de Electricidad de México. Las herramientas típicas de métricas de rendimiento se utilizaron para el análisis cuantitativo de los resultados de pronóstico, donde los modelos simples se compararon con los modelos híbridos utilizando el error cuadrático medio, el error absoluto medio, la raíz del error cuadrático medio, el error porcentual absoluto medio, y el índice evaluador. Los resultados obtenidos mostraron, en el error porcentual absoluto medio, las diferencias de rendimiento entre los modelos híbridos y los convencionales para el conjunto prueba fueron: 2.364, 1.210 y 1.635 puntos porcentuales para la serie temporal de diez minutos. En cambio, estas diferencias son de 0.751, 0.192 y 0.331 puntos porcentuales para la serie temporal horaria, demostrando que es conveniente el análisis de los errores de pronóstico generados por los modelos convencionales, con la finalidad de mejorar los pronósticos que ayuden en la planeación y despacho de la energía.

**Palabras claves:** Series de tiempo multifractales, ARIMA, Aprendizaje Profundo, Errores de pronóstico, y Cascada multiplicativa.

# Abstract

In this work, the forecast errors produced by the ARIMA, Deep Learning and Persistence models, models applied to two time series to predict the wind speed one step ahead, were analyzed. The analysis consisted of verifying, with the help of MF-DFA, if there were multifractal patterns in the forecast errors. Then, the multifractal patterns of the errors were extracted by a Multiplicative Cascade model. The sum of the patterns extracted from the models generated new hybrid models such as the Multiplicative Cascade-ARIMA, the Multiplicative Cascade-Deep Learning and the Multiplicative Cascade-Persistence. Those new models showed a decrease in the final forecast error. The wind speed time series used for this study were measured at two wind farms in the State of Oaxaca, Mexico, namely La Mata and La Venta. The first time series has five years of ten-minute averaged data, while the second has six years and eleven months of hourly data. Both time series were provided by the Federal Electricity Commission of Mexico. Typical performance metrics tools were used for quantitative analysis of forecast results, where simple models were compared to hybrid models using mean square error, mean absolute error, root mean square error, percentage error mean absolute, and the evaluator index. The results obtained showed, in the mean absolute percentage error, the performance differences between the hybrid and conventional models for the test set were: 2,364, 1,210 and 1,635 percentage points for the ten-minute time series. On the other hand, these differences are 0.751, 0.192 and 0.331 percentage points for the hourly time series, demonstrating that it is convenient to analyze the forecast errors generated by conventional models, in order to improve the forecasts that help in the planning and dispatch of the energy.

**keywords:** Multifractal time series, ARIMA, Deep Learning, Forecasting errors and Multiplicative Cascade.

# Índice general

<b>Dedicatoria</b>	<b>II</b>
<b>Agradecimientos</b>	<b>III</b>
<b>Resumen</b>	<b>IV</b>
<b>Abstract</b>	<b>V</b>
<b>1. Introducción</b>	<b>2</b>
1.1. Antecedentes	5
1.2. Objetivos	10
1.3. Justificación	10
1.4. Hipótesis	11
1.5. Contribución	11
<b>2. Fractales</b>	<b>12</b>
2.1. Series de Tiempo	12
2.1.1. Clasificación descriptiva de las series temporales	13
2.1.2. Componentes de una serie temporal	13
2.1.3. Series de tiempo de viento	15
2.2. Fractales	15
2.2.1. La importancia de los fractales	16
2.2.2. Principales características de los fractales	17
2.2.3. Tipos de fractales	18
2.2.4. Fractales Mandelbrot	22
2.2.5. Fractales Newton-Raphson	23
2.3. Series de tiempo fractales y multifractales	24
<b>3. Pronóstico y revisión de técnicas de pronóstico</b>	<b>26</b>
3.1. Horizonte de predicción	27
3.2. Tipos de pronóstico	28
3.3. Errores de los Pronósticos	28
3.4. Persistencia	29
3.5. Modelo autorregresivo integrado de promedio móvil (ARIMA)	29
3.5.1. Modelos autorregresivos, $AR(p)$	33
3.5.2. Modelos media móvil, $MA(q)$	33
3.5.3. Modelos mixtos autorregresivo de media móvil, $ARMA(p, q)$	33
3.5.4. Modelos media móvil integrada autorregresiva, $ARIMA(p, d, q)$	34

3.5.5.	Modelos media móvil integrada autorregresiva estacional, SARIMA $(p, d, q)(P, D, Q)_s$	34
3.6.	Deep Learning (Aprendizaje Profundo)	36
3.6.1.	Redes neuronales (Modelo biológico)	37
3.6.2.	Redes neuronales (Modelo artificial)	39
3.6.3.	Funciones de activación	43
3.6.4.	Diseño de redes neuronales artificiales	48
3.6.5.	Entrenamiento de una red neuronal	51
3.6.6.	Sobreaprendizaje en una red neuronal	54
3.6.7.	Datos de entrenamiento, validación y prueba	55
3.6.8.	Ventajas y limitaciones del aprendizaje profundo	56
3.7.	Cascada Multiplicativa	57
3.7.1.	Construcción de Multifractales Multiplicativos	58
3.7.2.	Cascada Multiplicativa Determinística	59
3.7.3.	Cascada Multiplicativa Conservadora Binomial	60
3.7.4.	Cascada de Poisson compuestas	62
3.8.	Dimensión Fractal ( $D_f$ ) y Coeficiente de Hurst ( $H$ )	65
3.8.1.	Métodos para calcular el Coeficiente de Hurst ( $H$ )	65
3.8.2.	Conteo de Cajas	65
3.8.3.	Rango Reescalado ( $R/S$ )	66
3.8.4.	Espectro de Potencia	67
3.8.5.	Wavelet	68
3.9.	Análisis de Fluctación sin Tendencia ( $DFA$ )	68
3.10.	Multifractalidad	70
3.10.1.	Técnicas de Multifractalidad	70
3.10.2.	Análisis Multifractal de Fluctuación sin Tendencia (MF-DFA)	71
3.11.	Métricas de rendimiento	72
<b>4.</b>	<b>Análisis de las Series de Tiempo de viento de La Mata y La Venta</b>	<b>75</b>
4.1.	Tratamiento de datos utilizados en el estudio	75
4.2.	Modelos matemáticos	78
4.2.1.	Modelo de Persistencia	78
4.2.2.	ARIMA (Autoregressive Integrated Moving Average)	78
4.2.3.	Media móvil integrada autorregresiva estacional (SARIMA)	78
4.2.4.	Aprendizaje Profundo	79
4.2.5.	Cascada Multiplicativa	80
4.2.6.	Conformación de los modelos híbridos	84
4.3.	Proceso de Modelado	85
<b>5.</b>	<b>Análisis de Resultados</b>	<b>88</b>
5.1.	Análisis cualitativo del comportamiento de los modelos generados	88
5.2.	Características multifractales de los errores de los modelos	90
5.3.	Extracción del patrón multifractal en los errores de pronóstico de los modelos híbridos	94
5.4.	Comparación de las métricas de rendimiento entre los modelos simples y modelos híbridos	95
<b>6.</b>	<b>Conclusiones</b>	<b>98</b>
6.1.	Trabajos Futuros	99
6.2.	Expectativas de posibles desarrollos	99

---

<b>A. Glosario</b>	<b>100</b>
<b>B. Anexo: Procesos no estacionarios</b>	<b>103</b>
B.1. Procesos ARIMA . . . . .	103
B.2. Procesos SARIMA . . . . .	104
<b>C. Anexo: Conceptos con formalismo matemático</b>	<b>105</b>
C.0.1. Formalismo multifractal . . . . .	105
C.0.2. log-cumulantes . . . . .	106
C.0.3. Procesos Monofractales y Multifractales . . . . .	108
C.0.4. Invarianza a la escala . . . . .	109
C.0.5. Relación entre las clasificaciones . . . . .	110
C.0.6. Atributos Multifractales . . . . .	111
<b>D. Anexo: Muestra de datos utilizados</b>	<b>113</b>
<b>E. Anexo: Publicaciones y Aportes</b>	<b>115</b>

# Índice de figuras

1.1. Diagrama de la energía eólica en la República Mexicana, representación de la zona sur del país, técnicas utilizadas en la predicción de la velocidad de viento y el método propuesto Cascada Multiplicativa para modelar los errores de pronóstico primarios. . . . .	9
2.1. Diferentes áreas donde se aplican pronósticos. . . . .	13
2.2. Diferentes componentes de una serie temporal. . . . .	14
2.3. Conjuntos representativos de los fractales no lineales. . . . .	16
2.4. Estructuras fractales encontradas en la naturaleza. . . . .	17
2.5. La construcción del Polvo de Cantor establece $F$ , mediante la eliminación repetida del tercio medio de los intervalos. Tenga en cuenta que $F_L$ y $F_R$ , las partes izquierda y derecha de $F$ , son copias de $F$ escaladas por un factor $\frac{1}{3}$ . . . . .	19
2.6. Iteración del conjunto de Cantor. . . . .	20
2.7. (a) Construcción de la curva de von Koch $F$ . En cada etapa, el tercio medio de cada intervalo es reemplazado por los otros dos lados de un triángulo equilátero. (b) Tres curvas de Von Koch unidas para formar una curva de copo de nieve. . . . .	20
2.8. Isla o copo de nieve de Koch. . . . .	21
2.9. Iteraciones del Triángulo de Sierpinski. . . . .	22
2.10. Conjunto de Mandelbrot. . . . .	23
2.11. Fractal de Newton basado en tres brazos. . . . .	24
2.12. Fractal de Newton basado en dos raíces. . . . .	24
3.1. Clasificación de los modelos de predicción basados en resolución espacial de los datos de entrada y resolución temporal. . . . .	27
3.2. Representación esquemática de la metodología de Box-Jenkins para el tiempo. . . . .	31
3.3. Deep Learning es un subconjunto de Machine Learning, que es solo una parte de la inteligencia artificial, aunque en estos momentos quizás es más dinámica y la que está haciendo que la inteligencia artificial esté nuevamente en pleno auge. . . . .	37
3.4. Esquema de una red neuronal biológica. . . . .	39
3.5. Modelo simplificado de neurona con integración y activación. . . . .	41
3.6. Modelo estándar de neurona con función de activación no lineal. . . . .	41
3.7. Función de activación lineal. . . . .	44
3.8. Función de activación TLU. . . . .	44
3.9. Función de activación TLU simétrica o bipolar. . . . .	45
3.10. Función de activación lineal (a) con saturación. (b) bipolar con saturación. . . . .	45
3.11. Función de activación sigmoidal: Función logística. . . . .	46
3.12. Función de activación sigmoidal bipolar: Tangente hiperbólica. . . . .	47
3.13. La función gudemanniana: Otra función de activación sigmoidal bipolar con rango de $-\frac{1}{2}$ a $\frac{1}{2}$ . . . . .	47

3.14. Función de activación lineal rectificada, utilizada por las unidades ReLU, muy habituales en AP. . . . .	48
3.15. Función de activación RBF de tipo gaussiano. . . . .	48
3.16. Red neuronal simple, sólo tiene una capa de neuronas de entrada y una capa de neuronas de salida. . . . .	49
3.17. Red neuronal multicapa, con una única capa oculta. . . . .	49
3.18. Red neuronal profunda, con múltiples capas ocultas sólo dos. . . . .	50
3.19. Red neuronal competitiva, con conexiones inhibitorias entre neuronas de la misma capa. . . . .	50
3.20. Red neuronal recurrente. . . . .	51
3.21. Representación gráfica del efecto de la tasa de aprendizaje sobre la convergencia del algoritmo de entrenamiento de una red neuronal multicapa. . . . .	53
3.22. Representación gráfica del problema de sobreajuste y el subajuste encontrado en modelos lineales. . . . .	54
3.23. Repartición de los datos disponibles en tres grupos (entrenamiento, validación y prueba) y el propósito de cada uno de ellos. . . . .	56
3.24. Esquema que ilustra la regla de construcción de un multifractal multiplicativo. . . . .	58
3.25. Construcción clásica en cascada binomial conservadora. . . . .	60
4.1. a) Ubicación geográfica de las torres anemométricas en La Mata y La Venta. . . . .	75
4.2. b) Diagrama de flujo del preprocesamiento de datos. . . . .	76
4.3. a) y b) Series temporales originales de La Mata y La Venta, respectivamente. . . . .	76
4.4. Series temporales estabilizadas en la varianza a) La Mata y b) La Venta, seleccionando los años a tratar. . . . .	77
4.5. Series temporales estabilizadas en la media a) La Mata y b) La Venta. . . . .	77
4.6. Diagrama de la red neuronal utilizando Aprendizaje Profundo para la serie La Mata. . . . .	79
4.7. Diagrama de la red neuronal utilizando Aprendizaje Profundo para la serie La Venta. . . . .	80
4.8. Modelo de Cascada Multiplicativa. . . . .	81
4.9. Cascada en la forma en que la turbulencia ocurre en el aire. . . . .	82
4.10. Construcción de la cascadas conservadoras correspondientes al instante actual y al instante de tiempo futuro. . . . .	84
4.11. Diagrama de flujo del proceso de modelado. . . . .	86
5.1. Comportamiento de los modelos generados. . . . .	88
5.2. Resultados de los modelos simples, modelos híbridos y errores de pronóstico para la serie La Mata. . . . .	89
5.3. Resultados de los modelos simples, modelos híbridos y errores de pronóstico para la serie La Venta. . . . .	90
5.4. Caminata aleatoria para los errores primarios en cada uno de los modelos. . . . .	91
5.5. Función de fluctuación $F_q(s)$ para cada uno de los modelos. . . . .	91
5.6. Exponentes generalizados de Hurst de los tres modelos simples. . . . .	91
5.7. Espectro multifractal $f(\alpha)$ para cada uno de los modelos. . . . .	92
5.8. Caminata aleatoria para los errores primarios en cada uno de los modelos. . . . .	92
5.9. Muestran la función de fluctuación $F_q(s)$ para cada uno de los modelos. . . . .	93
5.10. Exponentes generalizados de Hurst de los tres modelos simples. . . . .	93
5.11. Espectro multifractal $f(\alpha)$ para cada uno de los modelos. . . . .	94
5.12. (a) Patrones multifractales extraídos $P_{fe}$ en los diferentes modelos simples mediante la técnica de la cascada multiplicativa para (a) la serie La Mata y (b) la serie La Venta. . . . .	94

5.13. Resultados del pronóstico de los diferentes modelos y las métricas de pronóstico de MAE, RMSE y MAPE para cada modelo de la serie La Mata. . . . .	96
5.14. Resultados del pronóstico de los diferentes modelos y las métricas de pronóstico de MAE, RMSE y MAPE para cada modelo de la serie La Venta. . . . .	97

# Índice de tablas

3.1. Modelos cuantitativos de pronóstico. . . . .	26
3.2. Características FAS y FAP Teóricas. . . . .	33
3.3. Características de los modelos estacionales vs. FAS y FAP. . . . .	34
3.4. Condiciones de estacionariedad e invertibilidad. . . . .	35
4.1. Especificaciones de los sensores de velocidad y dirección del viento. . . . .	75
5.1. Comparación estadística de los datos de entrenamiento, validación y prueba entre los modelos simples y los modelos híbridos para la serie de tiempo La Mata. . . . .	95
5.2. Comparación estadística de los datos de entrenamiento, validación y prueba entre los modelos simples y los modelos híbridos para la serie de tiempo La Venta. . . . .	96

# 1

## Introducción

### Capítulo

---

El continuo crecimiento de la energía eólica en el mundo, favorecido por los acuerdos universales de las Naciones Unidas sobre cambio climático, genera un precedente positivo que alienta a la comunidad mundial para seguir impulsando los esfuerzos orientados a limitar el aumento del calentamiento global.

Según datos publicados por el Consejo Mundial de Energía Eólica (GWEC) [1] y la Asociación Mexicana de Energía Eólica (AMDEE) [2], la capacidad instalada acumulada del mercado mundial de energía eólica alcanzó los 743 GW, con una capacidad instalada en México a 6.681 GW a finales de 2020, respectivamente. Debido al abundante recurso eólico, la región del Istmo de Tehuantepec en el estado de Oaxaca, México, concentra el 76.8 % de la energía eólica total instalada [2].

Por lo anterior, se puede determinar que la tendencia del uso de esta energía, sigue siendo favorable y los resultados alientan el estudio de temas relacionados con el pronóstico de energía eólica. El pronóstico de la velocidad de viento puede realizarse mediante métodos físicos y métodos estocásticos. Los métodos físicos están basados en consideraciones físicas del terreno, como rugosidad, orografía y obstáculos, así como de la atmósfera, donde modelan los perfiles locales del viento, tienen ventajas en la predicción a largo plazo. En los modelos estocásticos la predicción se basa en el análisis de series de datos y se realizan mediante series de tiempo, técnicas estadísticas, y técnicas de inteligencia artificial, teniendo ventajas principalmente en predicción a corto plazo.

El problema de la predicción en las series de tiempo de la velocidad del viento no es algo nuevo. Es una tarea complicada, ya que hay una amplia gama de patrones de comportamiento diferentes en los datos a modelar. Hodge et al. [3], menciona que al viento no es posible controlarlo, y la mejor solución es poder pronosticar la producción de viento. Para Slini [4], al modelar series de tiempo de viento se observan distintos comportamientos, debido a su naturaleza intermitente e incierta, por lo que, es necesario generar un modelo para cada caso. Por lo tanto, el análisis de series de tiempo se encuentra entre las principales herramientas de investigación y desarrollo en este campo.

Un pronóstico requiere de una planeación la cual puede ser una predicción a largo plazo (5 a 10 años), predicción a mediano plazo (de varios meses a 5 años) ó una predicción a corto plazo (de minutos, horas, días a semanas). Actualmente existen fenómenos que pueden pronosticarse, por ejemplo la lluvia, la velocidad del viento, radiación solar, etcétera. Mientras que la calidad de los pronósticos está relacionada con la información del pasado. Las herramientas estadísticas, el software, el hardware junto con la técnica adecuada de predicción, son la base para obtener éxito en el pronóstico.

---

Para alcanzar los objetivos propuestos se ha llevado a cabo una profunda labor de recopilación de datos, obtención e interpretación de resultados que se muestran de manera organizada a través de los siete capítulos en los que se subdivide esta memoria de tesis. A continuación se realiza una breve descripción de cada una de ellos para facilitar al lector el seguimiento del presente documento.

- **Capítulo 1** - Introducción. En este capítulo se aborda la revisión de antecedentes y el estado del arte, se exponen y justifican los objetivos de la presente investigación. Finalmente, se propone la hipótesis de la investigación.
- **Capítulo 2** – Series de tiempo. Este capítulo es el resultado de una profunda recopilación y revisión bibliográfica sobre las series de tiempo. Donde se da una clasificación descriptiva de los tipos de series temporales. Así como las componentes de estas mismas. Finalizando con un apartado de las series de tiempo de viento mostrando algunas de sus características.
- **Capítulo 3** – Fractales. Este capítulo describe la importancia del estudio de los fractales, sus principales características, los tipos de fractales. Además, se ha entendido conveniente mostrar un apartado relacionando las series de tiempo con los fractales y multifractales.
- **Capítulo 4** – Pronóstico y revisión de técnicas de pronóstico. Este capítulo muestra dos partes importantes los tipos de horizontes de predicción que pueden llevarse a cabo en las series de tiempo de viento. Y la revisión de las técnicas para realizar un pronóstico en el área de las series de tiempo. Para ello es necesario estudiar los tipos y errores de pronóstico. Así como los tipos de modelos, técnicas que se pueden realizar, sus características, ventajas y limitaciones. Esto anterior se llevo a cabo para los modelos de la Persistencia, Deep Learning y Cascada Multiplicativa. Y para las técnicas de Análisis de fluctuación sin tendencia (DFA) y Análisis Multifractal de fluctuación sin tendencia (MF-DFA). Finalmente, se explican las métricas de rendimiento utilizadas en la presente investigación.
- **Capítulo 5** – Análisis de las series de tiempo de viento La Mata y La Venta. En este capítulo se describe el tratamiento de los datos utilizados. Modelos matemáticos y numéricos, la conformación de ellos y el proceso de modelado que se llevo a cabo.
- **Capítulo 6** – Análisis de resultados. Ciñéndonos a los objetivos planteados, este capítulo expone los resultados obtenidos durante la realización de esta investigación. El capítulo se subdivide en cuatro apartados: La primera parte se corresponde con un análisis cualitativo del comportamiento de los modelos generados. La segunda muestra los resultados característicos multifractales de los errores de los diferentes modelos. La tercera parte expone la extracción del patrón multifractal en los errores de pronóstico de los modelos híbridos como área objeto de estudio. La cuarta se basa en la comparación de las métricas de rendimiento entre los modelos simples y los modelos híbridos.
- **Capítulo 7** – Conclusiones. En este capítulo se enfatizan los principales hallazgos obtenidos en relación a los objetivos planteados.
- **Bibliografía** – La memoria concluye con un listado de las fuentes consultadas y citadas en el texto.
- **Anexo A: Glosario** – Presenta un catálogo de palabras relativas al campo de estudio presente, con su definición y/o comentarios.
- **Anexo B: Procesos estacionarios** – Desarrollo para la obtención de los modelos ARIMA y SARIMA.

- **Anexo C: Conceptos con formalismo matemático** – Se detallan conceptos utilizados durante la presente investigación.
- **Anexo D: Datos utilizados** – Listado representativo de datos utilizados por las series tratadas.
- **Anexo E: Publicaciones y Aportes** – Muestra un listado de las publicaciones y divulgación científica.

## 1.1. Antecedentes

La predicción de la velocidad del viento ha ido ganando popularidad debido al aumento de la demanda y el incremento de generación de electricidad mediante energía de viento, aumentando la competitividad de la energía eólica [5]. Sin embargo, para Moreno y dos Santos Coelho [6] uno de los grandes problemas con el proceso de generación de energía eólica en gran medida es la dependencia con la velocidad de viento, ya que puede verse afectada por muchos factores complejos, la velocidad del viento se considera uno de los parámetros meteorológicos más difíciles de predecir, y la predicción precisa de la velocidad del viento sigue siendo un tema crítico.

En los últimos años se han desarrollado numerosos modelos de predicción de la velocidad del viento, clasificándose en diferentes categorías conforme a diferentes estándares. Por un lado, en función del horizonte temporal de predicción dividiéndose en tres grupos: modelos de predicción a largo plazo [7–9], modelos de predicción a medio plazo [10–12] y modelos de predicción a corto plazo [13–15]. Por ejemplo, Cadenas y Rivera [16] desarrollaron un pronóstico de velocidad del viento a corto plazo, aplicando la técnica de red neuronal artificial (ANN) a series de tiempo horarias encontrando buena precisión. Por otro lado, en función de las características típicas de los modelos, los modelos de predicción de la velocidad del viento se pueden clasificar en cinco categorías: modelos de inteligencia artificial, modelos de correlación espacial, modelos estadísticos convencionales, modelos físicos y modelos híbridos [17].

Existe una gran diversidad de técnicas y modelos para generar pronósticos. Algunos autores como Wang et al. [18] y Singh et al. [19], mencionan que los modelos de series temporales ampliamente utilizados son la media móvil integrada autorregresiva (ARIMA), que consta de tres componentes, cada uno de los cuales ayuda a modelar un cierto tipo de patrón. El componente AR (Autorregresivo), que especifican los valores previos utilizados para predecir el comportamiento de la serie, MA (Media Móviles), especifica el modo en que se utilizan las desviaciones de la media de la serie y finalmente el componente I (Integrados), especifica las diferencias realizadas a la serie para conseguir su estacionariedad [20]. Además de las componentes anteriores para el modelo de media móvil estacional autorregresiva integrada (SARIMA), se añade el componente S (Periodicidad) que especifican si la periodicidad es inferior al año, si suele darse el factor estacional, ó deberán analizarse las correlaciones entre los mismos meses, trimestres, etc. de años sucesivos. Entre otras técnicas de suavizado exponencial y predictores grises. Además, dichos modelos derivan de sólidas teorías de la probabilidad clásica y de la estadística matemática.

Para Brockwell et al. [21] y Koutroumanidis et al. [22], los modelos ARIMA tienen la ventaja de generar un grado de ajuste respecto de las series de tiempo de mediciones reales comparados con el modelado clásica, donde el modelo matemático ya establecido se ajusta a una serie, utilizando el enfoque estadístico clásico. Pero siguiendo la metodología propuesta por Box-Jenkins, se puede encontrar el modelo ARIMA de velocidad del viento apropiado. Construyendo los modelos como funciones lineales de valores pasados de la serie y/o choques aleatorios anteriores (o errores) [23]. Es decir, debe contener todos los elementos necesarios para describir el fenómeno. Para lo anterior, Box y Jenkins recomiendan tener como mínimo 50 observaciones en la serie temporal. Otra ventaja que muestran estos modelos es que son buenos para la predicción en el corto plazo, debido a su capacidad de aprender de los cambios en la dinámica de la serie.

Por otro lado, dentro de las dificultades que se pueden presentar en el modelado ARIMA, se encuentra la correcta elección de un modelo adecuado, siendo el principal reto de un analista recién iniciado en

la teoría y la ejecución de la metodología Box-Jenkins. Es por ello que el modelado de un proceso ARIMA ha sido referida por algunos autores como un arte. La habilidad se obtiene tras un aprendizaje relativamente mediano, debido a que las técnicas básicas en el modelado de los procesos ARIMA son fácilmente accesibles y ajustables.

Por otro lado, aunque en la actualidad se utilizan programas de cómputo en el cálculo de los coeficientes de un modelo ARIMA probable, la aplicación de la metodología Box-Jenkins sigue siendo una labor manual, es decir, requiere ser ejecutada por el analista experto, siendo este hecho una desventaja más que limita la utilización correcta del modelo. Hay que recordar que no cualquier modelo propuesto proporciona necesariamente el ajuste adecuado. Un modelo incorrecto será pobre para predecir en momentos de cambios estructurales y puntos de giro.

Otros tipos de modelos utilizados para la predicción de viento se basan en la inteligencia artificial, técnica que nace en la búsqueda constante del hombre de nuevas vías para mejorar sus condiciones de vida. La utilización de la inteligencia artificial suma esfuerzos para reducir el trabajo en aquellas operaciones en las que la fuerza juega un papel primordial. Los progresos obtenidos han permitido dirigir estos esfuerzos a otros campos, como por ejemplo, a la construcción de máquinas calculadoras (servidores, computadoras y supercomputadoras, con procesadores gráficos especializados) que ayuden a resolver de forma automática y rápida determinadas operaciones que resultan tediosas cuando se realizan a mano [24]. Así, los desarrollos actuales, se dirigen al estudio de las capacidades humanas como una fuente de nuevas ideas para el diseño de las nuevas máquinas. Por lo tanto, la inteligencia artificial es un intento por descubrir y describir aspectos de la inteligencia humana que pueden ser simulados mediante máquinas. Esta disciplina se ha desarrollado fuertemente en los últimos años teniendo aplicación en diversos campos.

Hoy en día, además de las redes neuronales artificiales (RNA), el aprendizaje profundo (AP) es una excelente herramienta con interés particular para minería de datos, puesto que ofrece los medios para modelar de manera efectiva y eficiente problemas grandes y complejos [25,26]. Ambas técnicas están inspiradas en las redes neuronales biológicas del cerebro humano.

Debido a su constitución y a sus fundamentos, se comportan de forma similar a la neurona biológica en sus funciones más comunes [27]. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en múltiples áreas. Entre las ventajas que menciona Matich et al. [24,27] se incluyen :

- Aprender: adquirir el conocimiento de una cosa por medio del estudio, ejercicio o experiencia. Las RNA y AP pueden cambiar su comportamiento en función del entorno. Se les muestra un conjunto de entradas y ellas mismas se ajustan para producir unas salidas consistentes.
- Generalizar: extender o ampliar una cosa. Las RNA y AP generalizan automáticamente debido a su propia estructura y naturaleza. Estas redes pueden ofrecer, dentro de un margen, respuestas correctas a entradas que presentan pequeñas variaciones debido a los efectos de ruido o distorsión.
- Abstractar: aislar mentalmente o considerar por separado las cualidades de un objeto. Algunas RNA y/o AP son capaces de abstraer la esencia de un conjunto de entradas que aparentemente no presentan aspectos comunes o relativos.

Así, Berzal [25] y DasGupta [28], consideran que ambos métodos de forma individual o combinadas con otros métodos, pueden resolver aquellas tareas de clasificación, identificación, diagnóstico, optimización o predicción. Cadenas et al. [16], menciona que los modelos RNA son utilizados para manejar la naturaleza no lineal de la velocidad del viento, lo cual es determinado a partir de los datos, es decir, son capaces de encontrar relaciones (patrones) de forma inductiva por medio de los algoritmos de aprendizaje automático, basado en los datos existentes, en lugar de requerir la ayuda de un modelador para especificar la forma funcional y sus interacciones.

En el presente trabajo, se propone la utilización de modelos híbridos, que constan de las técnicas de predicción anteriores, junto con la utilización de la cascada multiplicativa. Los modelos de cascada multiplicativa (CM) se desarrollaron inicialmente para comprender las características intermitentes de la turbulencia [29, 30]. La teoría moderna de la turbulencia comenzó con la teoría de Kolmogorov [30, 31]. En la actualidad, los modelos de cascada multiplicativa se ha aplicado al estudio de varios fenómenos como la convección tropical [32], la distribución de agua líquida dentro de los estratocúmulos marinos [33, 34], las finanzas [35] y la lluvia [36]. Alzate [37] considera en su libro, la distribución de la lluvia dada por un proceso o cascada multiplicativa de transferencia de energía desde grandes hacia pequeñas escalas. El movimiento turbulento de la atmósfera es la razón principal para emplear la fenomenología de la cascada de la turbulencia para el estudio multifractal de la lluvia.

Algunos autores como de Godoy et al. [38], definen a estos modelos como construcciones matemáticas adecuadas para capturar comportamientos intermitentes y altamente irregulares. Para Yang et al. [39], una cascada multiplicativa, es un proceso iterativo que fragmenta un conjunto dado en piezas cada vez más pequeñas de acuerdo con alguna regla geométrica. Por otra parte, Molnar y Burlando [40], definen estos modelos como estructuras de dependencia entre escalas sucesivas, que simplemente transcriben las propiedades multifractales de la distribución de observaciones, es decir, si el soporte de CM es representado por una superficie o el tiempo; y, además si es posible dividir el soporte en múltiples conjuntos fractales, conlleva a la multifractalidad de dichos modelos.

Así Faleiro [41], menciona que a partir del desarrollo de distintas técnicas basadas en conceptos multifractales, los modelos CM se han aplicado a una gama de problemas que abarca con éxito sistemas de laboratorio, geofísicos y astrofísicos. Por tanto, actualmente el modelo de cascada multiplicativa ha encontrado aplicaciones en varias áreas que necesitan describir fenómenos no lineales que tienen una estructura multiplicativa, como el modelado de tráfico, los tipos de cambio, los fenómenos geofísicos, evolución del ADN, y otros, pero careciendo aún de información del uso de la cascada multiplicativa en la predicción de la velocidad del viento.

A pesar del gran desarrollo en sistemas avanzados de pronóstico durante las últimas dos décadas, vale la pena considerar el sistema de pronóstico más simple, la persistencia [42], ya que siempre proporciona información útil sobre las características generales del viento en un sitio determinado. Y es utilizado como referencia de comparación para los sistemas de pronóstico más avanzados y elaborados [43, 44].

En cuanto a los modelos híbridos, en la actualidad han cobrado relevancia debido a su estructura, como menciona Liu et al. [45], los modelos híbridos son combinaciones de métodos físicos y métodos estadísticos. Modelos que generalmente puede alcanzar mayor precisión en la predicción. Además, de utilizar diversas tecnologías inteligentes para el pronóstico. Como mencionan Hu y Zeng [9], los modelos híbridos pueden integrar las ventajas de otros modelos individuales y así contribuir a la capacidad de pronóstico del modelo y mejorar la eficiencia del mismo. En la misma línea, Shi et al. [46] consi-

dera potencialmente viable pronosticar series temporales mediante modelos híbridos, en comparación con los enfoques de modelado de pronóstico convencionales. También, refiere que generalmente un modelo híbrido consta de dos modelos de predicción para la serie, uno para la componente lineal y uno para la no lineal.

Debido a que las series de tiempo de viento presentan patrones irregulares pero con repetición a diferentes escalas, da cabida al estudio de la fractalidad y multifractalidad; permitiendo reconocer, clasificar y extraer distintas características de la serie, ya sea en el sentido estadístico como geométrico.

El concepto de conjunto fractal fue creado para dar oportunidad en la estructura formal de la geometría a diversos objetos, denominados monstruos geométricos, cuyo comportamiento se apartaba radicalmente del mostrado por los conjuntos que estaban en la base conceptual de esta disciplina [41]. Mientras que la geometría fractal, adaptada al estudio de series temporales está fuertemente relacionado con la invarianza de escalas y la existencia de dependencias a largo plazo, además de proporcionar un marco atractivo para el análisis de fenómenos naturales en numerosas ramas de la ciencia [47]. Es importante tener en cuenta que aunque para caracterizar el comportamiento de determinados procesos es suficiente con un único exponente de escala, en cuyo caso se habla de proceso monofractal, en las dos últimas décadas se ha puesto de manifiesto que para describir unívocamente las características de escala de series temporales de numerosos procesos naturales se requiere el uso de múltiples exponentes de escala, indicando una naturaleza multifractal del proceso [47]. De ahí el concepto multifractal, que es considerado como una generalización de la teoría fractal, donde una única dimensión es insuficiente para describir la complejidad total de la señal, es decir, surge para incluir en el esquema a conjuntos aún más complejos y requiere que la dimensión fractal sea sustituida por una distribución espectral de funciones de dimensión, esto es que presentan ley de escalado múltiple y en su versión más actual, la de los multifractales universales estocásticos, permite modelar una gran variedad de procesos no lineales [41, 48].

Para lo anterior, se emplean metodologías como la de Kantelhardt [49] el análisis MF-DFA, que se considera como una herramienta adecuada para procesar y modelar series de datos que muestran una fuerte variabilidad espacio-temporal, especialmente en términos de las características dinámicas, no uniformes de los fenómenos y su comportamiento extremo. Los métodos primarios y avanzados son el análisis wavelet [50] y el MF-DFA [51–53] respectivamente. El MF-DFA se basa en el análisis de fluctuación estándar (FA) y es una generalización multifractal del análisis de fluctuación de tendencia (DFA) [54, 55]. En comparación con FA, el método se puede utilizar para realizar datos no estacionarios para su función de tendencia descendente. Puede mencionarse que, la ventaja indiscutible del MF-DFA es menos sensible a la duración de la serie temporal que otros métodos de análisis multifractal, por ejemplo, el enfoque de Wavelet Transform Modulus Maxima [56]. El MF-DFA se ha aplicado a diferentes campos de la ciencia, por ejemplo, para describir los registros diarios [51, 57] y horarios [51, 58] de velocidad del viento.

Entre tanto, los pronósticos por lo general son intentos de predecir acontecimientos futuros, un pronóstico bien realizado siempre va acompañado de su error. El cual, visto como una serie de datos, se le puede aplicar alguno de los métodos mencionados anteriormente, determinando con ello el Coeficiente de Hurst que es conocido como el índice  $H$ , y resulta ser una medida de la independencia de las series de tiempo y una forma de distinguir series fractales [59]. Donde el valor de  $H$  estará en el intervalo  $[0, 1]$ , y según el valor que tome el índice es posible determinar algunas características particulares del proceso al que corresponde dicho coeficiente. Los valores del coeficiente de Hurst para Ihlen et al. [60], Rodríguez [61] y Ruiz et al. [62] indican tres casos:

- Si  $0 \leq H < \frac{1}{2}$  denominado como “ruido rosa”, corresponde a un comportamiento antipersistente en la serie de tiempo, a este se le designa reversión a la media, es decir incremento de acontecimientos del pasado.
- $H = \frac{1}{2}$  denominado como “ruido blanco”, implica un proceso independiente y aleatorio con ausencia de correlación, sin ningún tipo de memoria.
- Si  $\frac{1}{2} < H \leq 1$  denominado como “ruido negro”, implican procesos persistentes al continuar con comportamientos del valor medio de los datos, sobresale por memoria a largo plazo, es decir, los sucesos de hoy afectaran los sucesos del futuro. Este contrasta la dependencia a largo plazo en series de tiempo.

En la presente investigación se propone utilizar la cascada multiplicativa para modelar los errores de pronóstico primarios entendiéndose éstos, en el presente trabajo como los errores de pronóstico obtenidos al generar un primer modelo de pronóstico, con lo cual al modelo convencional se le sumó el patrón extraído generando los modelos híbridos para predecir la velocidad de viento con mayor exactitud. Los modelos de cascada son fundamentados en la geometría fractal y han sido utilizados en la investigación de otros fenómenos naturales, transformándose en pocos años en una herramienta multidisciplinaria.

La Fig. 1.1 es una representación de lo que se realizó en la presente tesis. Lugar de estudio, zona de los aerogeneradores, métodos utilizados y resultados gráficos de aplicar los métodos mencionados a una de las series que se utilizó.

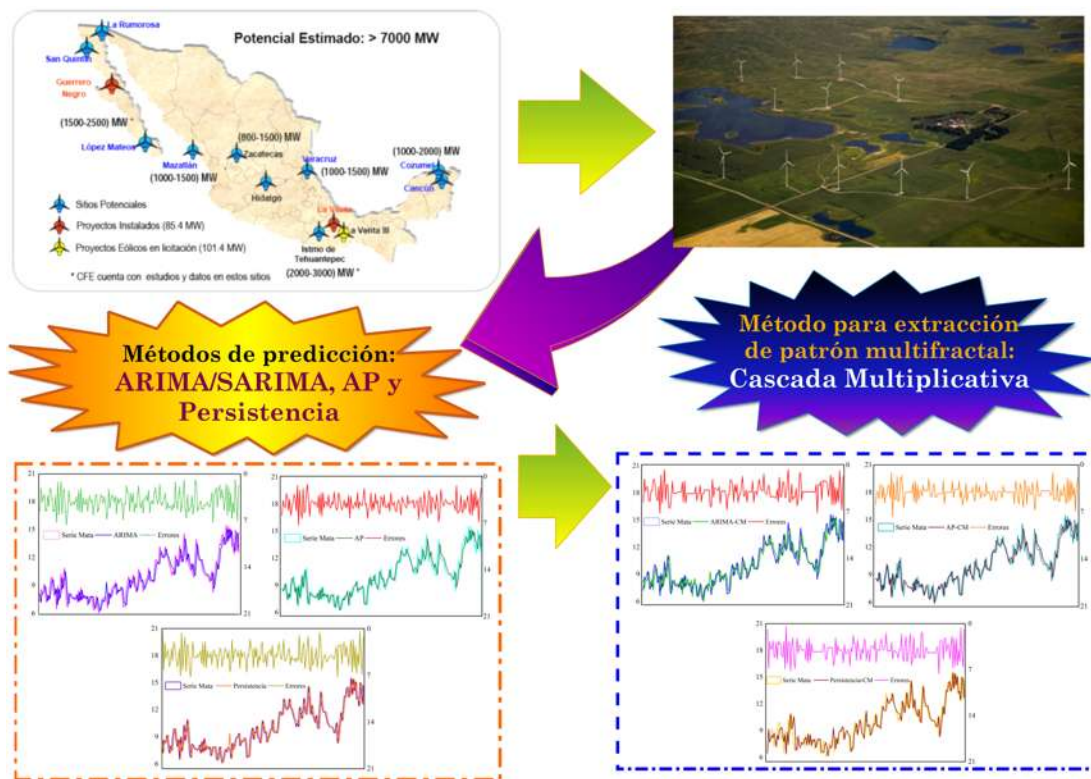


Figura 1.1: Diagrama de la energía eólica en la República Mexicana, representación de la zona sur del país, técnicas utilizadas en la predicción de la velocidad de viento y el método propuesto Cascada Multiplicativa para modelar los errores de pronóstico primarios.

## 1.2. Objetivos

**GENERAL** Extraer un patrón multifractal a los errores de pronóstico de los modelos (ARIMA, AP y Persistencia) mediante la técnica de la Cascada Multiplicativa, con la finalidad de mejorar la predicción un paso adelante con modelos híbridos (ARIMA-CM, AP-CM y Persistencia-CM).

### ESPECÍFICOS

- Realizar un estudio a profundidad de las Matemáticas Fractales.
- Buscar de la información de velocidad de viento cada 10 minutos y una hora en los lugares de La Mata y La Venta, Oaxaca, respectivamente.
- Analizar las series de tiempo obtenidas.
- Revisar las técnicas de predicción como lo son:
  1. Técnica de ARIMA.
  2. Técnica de Aprendizaje Profundo.
  3. Técnica de Cascada Multiplicativa.
- Generar modelos de predicciones utilizando matemáticas fractales, así como la validación de la propuesta.
- Comparar el modelo con otras técnicas.

## 1.3. Justificación

El desarrollo de la energía eólica en escala mundial es notable, representando ya una parte importante de la capacidad eléctrica global con fuentes renovables. La industria eólica mundial enfoca su atención hacia países en desarrollo con locaciones idóneas para la instalación de centrales eólicas [63]. Así el istmo de Tehuantepec, el cual se encuentra al sur de México, es la región con mayor potencial eoloenergético y donde se concentra el despliegue de proyectos eólicos en México.

La presente investigación busca mejorar la predicción de la velocidad del viento y por consecuencia tener un mejor control de la cantidad de energía eólica que se produciría, en una región específica dentro de la República Mexicana o de algún otro lugar en el mundo. Algunas técnicas utilizadas en la predicción de la velocidad de viento son: el método ARIMA, AP y Persistencia. Sin embargo, sus aproximaciones presentan diferencias notorias en comparación con los datos históricos. Debido que la velocidad del viento es una señal caótica cuyas características son difíciles de modelar.

Por lo anterior, el lugar geográfico de interés para este estudio fue en México en la zona sur y sureste del país, debido que cuenta con condiciones territoriales para producir vientos aprovechables, por ejemplo en el Istmo de Tehuantepec (en el estado de Oaxaca), la velocidad llega a los 25  $m/s$  [64], siendo una de las regiones con mayor potencial de energía eólica del mundo y económicamente competitiva como tipo de fuente de energía limpia [65].

## 1.4. Hipótesis

Es posible reducir los errores de pronóstico de los modelos ARIMA, AP y Persistencia utilizando la técnica de Cascada Multiplicativa para mejorar la predicción de la velocidad de viento mediante modelos híbridos.

## 1.5. Contribución

Actualmente la forma en la cual se establecen los modelos de pronóstico del viento proporciona ventajas que brindan conocimiento a las empresas dedicadas al aprovechamiento del recurso eólico. Sin embargo, aun estos modelos no cuentan con la precisión deseada. Es difícil encontrar mecanismos que permitan tal exactitud, por ello la utilización de modelos matemáticos y numéricos permiten acercarnos más al objetivo, lograr validar y garantizar que la información obtenida del modelo pronostica con un error mínimo.

Por otro lado, el uso de la técnica MF-DFA garantizan una correcta información de las características fractales/multifractales de las series de tiempo de viento. Permitiendo la creación de conocimiento nuevo en base a los conocimientos de esta técnica. Asimismo, el estudio de la técnica de Cascada Multiplicativa permiten extraer un patrón presente en los errores de pronóstico de diferentes modelos utilizados.

Esta investigación aporta:

- Modelos convencionales de pronóstico para la velocidad de viento ARIMA/SARIMA, Aprendizaje Profundo y Persistencia generados para las series de tiempo de viento La Mata y La Venta.
- Mediante la técnica MF-DFA se garantizó la multifractalidad de los errores de las series de tiempo de viento La Mata y La Venta.
- La técnica de Cascada Multiplicativa utilizada para modelar los errores de pronóstico primarios entendiéndose éstos, en el presente trabajo como los errores de pronóstico obtenidos al generar un primer modelo de pronóstico (ARIMA/SARIMA, Aprendizaje Profundo y Persistencia).
- La obtención de los modelos híbridos mediante los modelos convencionales a los que se le sumó el patrón extraído de los errores para predecir la velocidad de viento con mayor exactitud.

# 2 Fractales

## Capítulo

---

### 2.1. Series de Tiempo

La predicción del futuro ha sido considerado como un trabajo complejo y absolutamente crítico para el hombre. Evidentemente, ese trabajo parte de evitar tomar decisiones bajo una situación de incertidumbre, teniendo como ventaja el conocer de antemano las consecuencias que las acciones o decisiones tomadas implicarán y más cuando entran en juego cuestiones económicas. De hecho, ha sido en el campo económico donde las técnicas de predicción de valores futuros de una serie temporal han sido más intensamente utilizadas [66].

Una serie de tiempo (también conocida como serie histórica o cronológica) es un conjunto de valores, correspondientes a una variable, ordenados según el instante en que fueron tomados, y registrados a intervalos regulares de tiempo [66]. En otras palabras, refiriéndose también a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral, anual, entre otros). Donde los sucesivos valores contenidos en una serie temporal se designan por un número de índice, correspondiente al número de observación (número del instante en que fue tomado el valor). Suponiendo una serie temporal  $X$ , está vendría expresada como indica la Ec. (2.1).

$$X = X_1, X_2, X_3, \dots \quad (2.1)$$

Para Molugaram y Rao [67] la importancia de hacer pronósticos para el futuro, está en analizar los datos de las series de tiempo que representan el cambio a lo largo de un período de tiempo. Es decir, una serie de tiempo es un conjunto de observaciones  $x$ , cada una registrada en un tiempo específico  $t$  [68].

Hoy en día diversas organizaciones requieren conocer el comportamiento futuro de ciertos fenómenos con el fin de planificar o prevenir, utilizando estos estudios para predecir lo que ocurrirá con una variable en el futuro, basados en el comportamiento de esa variable en el pasado [69].

Algunas de las áreas de aplicación de Series de Tiempo según Rios [69] son las siguientes:

- **Economía:** Precios de un artículo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etc.
- **Meteorología:** Cantidad de agua caída, temperatura máxima diaria, *Velocidad del viento (energía eólica)*, energía solar, etc.
- **Geofísica:** Series sismológicas.
- **Química:** Viscosidad de un proceso, temperatura de un proceso.
- **Medicina:** Electrocardiograma, electroencefalograma.

La Fig. 2.1 muestra pronósticos de las diferentes áreas en las cuales existe una preocupación por tener la mayor precisión al momento de tomar decisiones, puesto que el hecho futuro por su naturaleza en las diferentes áreas pueden ser inciertas y aleatorias. Por ejemplo en la economía, la tasa de inflación es variacional en los diferentes años; en la geofísica los movimientos de la tierra generan lecturas sísmológicas variadas, así como la realización de estudios médicos oscilan dependiendo el paciente.

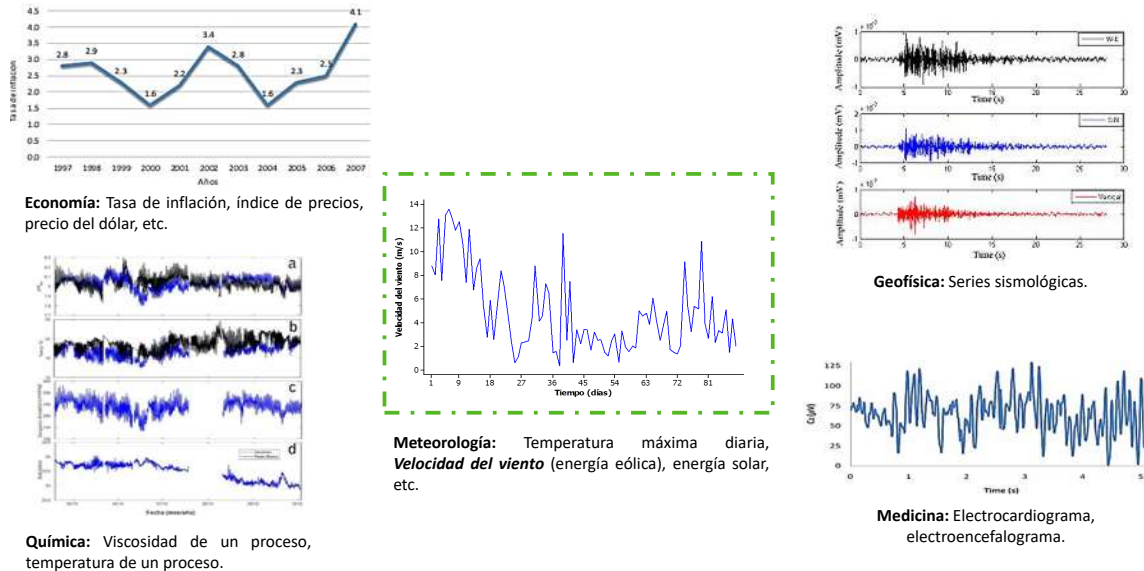


Figura 2.1: Diferentes áreas donde se aplican pronósticos.

### 2.1.1. Clasificación descriptiva de las series temporales

Para Rios [69] las series temporales se pueden clasificar en:

- **No estacionarias:** son series estables a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo.
- **Estacionarias:** son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

### 2.1.2. Componentes de una serie temporal

El análisis clásico de las series temporales se basa en la suposición de los valores que toma la variable de observación es la consecuencia de tres componentes, cuya actuación conjunta da como resultado los valores medidos, estos componentes son: [67, 69]:

1. **Tendencia ( $T_t$  o  $T$ ):** o tendencia a largo plazo, significa el movimiento a largo plazo, midiendo los cambios que ocurren en una serie de tiempo sin preocuparse por las fluctuaciones a corto

plazo que se producen en el medio, es decir, representa el comportamiento predominante de la serie. Además, la tendencia se relaciona con el crecimiento o disminución regular.

2. *Variación estacional* ( $S_t$  o  $S$ ): describe las fluctuaciones en las series de tiempo que se repiten durante períodos de tiempo específicos. En ese sentido podría decirse que es un movimiento periódico que se produce dentro de un período corto y conocido. Este componente está determinado, por ejemplo, por factores institucionales y climáticos.
3. *Componente aleatoria*: esta componente no responde a ningún patrón de comportamiento, sino que es el resultado de factores fortuitos o aleatorios que inciden de forma aislada en una serie de tiempo.

De las tres componentes el primero y el segundo componente son determinísticos, mientras que la última es aleatoria (véase en la Fig. 2.2).

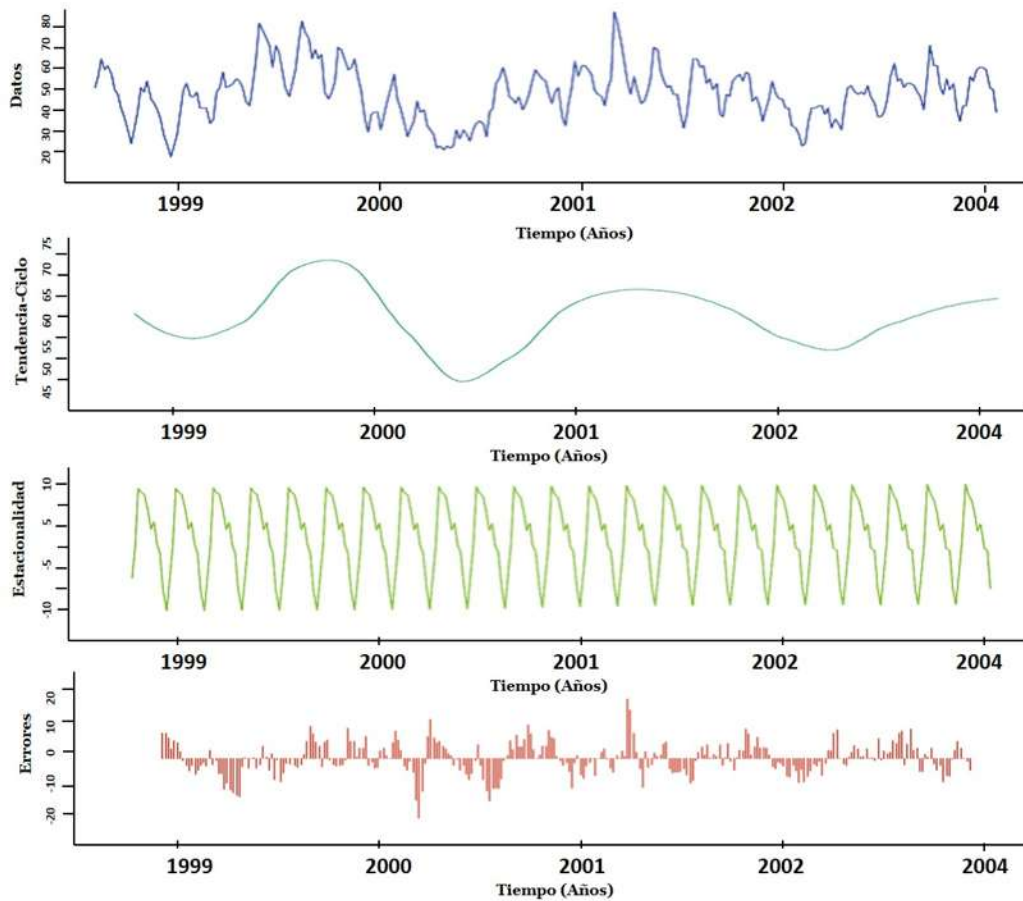


Figura 2.2: Diferentes componentes de una serie temporal.

La representación matemática general del método de descomposición es [70]:

$$Y_t = f(S_t, T_t, E_t) \quad (2.2)$$

donde  $Y_t$  es el valor de la serie temporal (datos reales) en el período  $t$ ,  $S_t$  es el componente estacional (o índice) en el período  $t$ ,  $T_t$  es el componente del ciclo de tendencia en el período  $t$ , y  $E_t$  es el com-

ponente irregular (o resto) en el período  $t$ .

La forma funcional exacta de Ec. (2.2) depende del método de descomposición realmente utilizado. Un enfoque común es asumir que la Ec. (2.2) tiene la forma aditiva.

$$Y_t = S_t + T_t + E_t \quad (2.3)$$

Es decir, los componentes estacional, tendencia-ciclo y errores simplemente se suman para dar la serie observada.

Alternativamente, la descomposición multiplicativa tiene la forma:

$$Y_t = S_t \times T_t \times E_t \quad (2.4)$$

Es decir, los componentes estacional, de tendencia-ciclo y errores se multiplican para dar la serie observada.

Un modelo aditivo es apropiado si la magnitud de las fluctuaciones estacionales no varía con el nivel de la serie. Pero si las fluctuaciones estacionales aumentan y disminuyen proporcionalmente con los aumentos y disminuciones en el nivel de la serie, entonces es apropiado un modelo multiplicativo. La descomposición multiplicativa es más frecuente con las series económicas porque la mayoría de las series económicas estacionales tienen la variación estacional que aumenta con el nivel de la serie [70].

### 2.1.3. Series de tiempo de viento

El viento puede verse representado de acuerdo a sus movimientos en dos escalas diferentes de tiempo:

- 1) Los movimientos relacionados con las variaciones de los sistemas climáticos planetarios.
- 2) Las ráfagas.

Este comportamiento fue explicado por Van der Hoven en 1957 [71], quien identificó dos fases dentro del espectro de potencia del viento. La primera fase está asociada con el primer y mayor pico dentro del espectro, al que denomina pico macro o meso-meteorológico que muestra una concentración de energía por un período aproximado de cuatro días, correspondiendo con el tiempo de tránsito típico de un sistema climático plenamente desarrollado. El segundo pico importante, micro-meteorológico, contiene energías con periodicidades del orden de 1 min hasta décimas de segundos y es causado por turbulencias generadas por los efectos topográficos: rugosidad del terreno u obstáculos alrededor del sitio [72].

Debido a la naturaleza intermitente del viento, en la literatura se encuentra disponible una gran cantidad de métodos y técnicas empleadas para efectuar la modelización de la magnitud de la velocidad del viento en el dominio del tiempo.

## 2.2. Fractales

En la actualidad, no se cuenta con una única definición de un fractal, ni Mandelbrot pionero de la geometría fractal desarrolló una definición precisa y formal, aún cuando es fácil reconocer cuando se

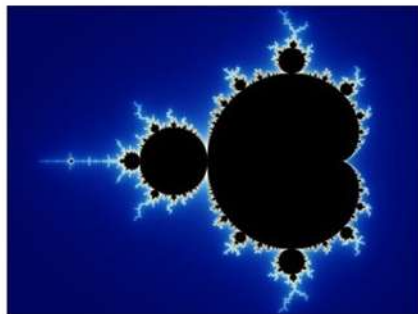
encuentra alguno. En términos prácticos los fractales tienen propiedades para propósitos de modelado ideal y características que los hacen medibles. Entre la variedad de definiciones encontradas sobre fractales se pueden mencionar las siguientes [73]:

- a) Proviene del latín que significa roto, debido a que se asocia con las discontinuidades de las funciones.
- b) Son objetos matemáticos que posee esencialmente dos características: auto-similaridad y dimensión fractal.
- c) Son objetos matemáticos cuya dimensión es fraccionaria.
- d) Son objetos matemáticos que conforman la geometría del Caos.

Existen dos tipos muy bien definidos de fractales: a) lineales y b) no lineales. Los primeros se construyen con un cambio en la variación de escala, en otras palabras los fractales lineales son idénticos hasta aumentar la escala a infinito.

Ejemplos de fractales lineales son el conjunto de Cantor (ver Fig. 2.5), el conjunto von Koch (o copo de nieve) (véase Fig. 2.7) y el triángulo de Sierpinski (véase Fig. 2.9).

Por otro lado, los fractales no lineales se generan a partir de distorsiones complejas o no lineales. Ejemplo de ello es el conjunto de Mandelbrot (véase Fig. 2.3 a)) y el conjunto de Julia (véase Fig. 2.3 b)).



(a) Conjunto de Mandelbrot



(b) Conjunto de Julia

Figura 2.3: Conjuntos representativos de los fractales no lineales.

### 2.2.1. La importancia de los fractales

Los fractales pueden observarse casi en cualquier área, por ejemplo: la turbulencia en la atmósfera hace que sea difícil para los telescopios, desde la tierra producir imágenes precisas de estrellas; una atmósfera turbulenta está bien modelada por una distribución fractal del índice de refracción. En otro caso, la luz que rebota en el océano, se parece mucho al reflejo de un espejo fractal. Con los ejemplos anteriores, es tangible que el mundo natural proporciona una fuente inagotable de problemas importantes en la física fractal.

El hecho de que los objetos fractales poseen un carácter y una estructura distintiva, y que no son solo irregulares o aleatorios, resulta lógico considerar que el universo está lleno de fractales. Además el mismo universo, puede ser estudiado como uno. Los fractales enseñan a no confundir la complejidad con la irregularidad, abriendo nuevas posibilidades y formas de estudio, representando un nuevo régimen completo de modelado matemático, que la ciencia está comenzando a explorar [74].

### 2.2.2. Principales características de los fractales

Los fractales se consideran *objetos auto-similares*, esto implica que el objeto se parece a su parte ampliada. [75]. En otras palabras, significa que todas sus partes están relacionadas de alguna forma con el sistema, además de tener una relación estadística con sus propiedades globales [73].

La estructura fractal tiene las siguientes características en general:

- Auto-similitud: también categorizado en completa auto-similitud, semi-auto- semejanza, auto-similitud estadística.
- Estructura fina o detallada en escalas arbitrariamente pequeñas.
- Irregularidad local y global que no se puede describir fácilmente en el lenguaje tradicional geométrico euclidiano.
- Definiciones simples y “quizás recursivas”.

La relación entre la naturaleza y el fractal está presente en todas partes, y son varios objetos naturales con características fractales, como ríos, animales, patrones de piel, ADN, etc. Algunos ejemplos se muestran en la Fig. 2.4.

Sin embargo, la importancia de los fractales, va más allá de su atractivo visual, está en los nuevos campos de estudio para modelar la naturaleza, ya que dan una herramienta poderosa para entender los procesos y estructuras hasta ahora descritos simplemente como “irregulares”, “intermitentes”, “rudos” o “complicados” [74, 76].



Figura 2.4: Estructuras fractales encontradas en la naturaleza.

### 2.2.3. Tipos de fractales

Los fractales son, sin duda alguna, más que interesantes curiosidades matemáticas. A diferencia de la geometría euclidiana, en donde los elementos básicos pueden generarse de manera directa (líneas, círculos, planos, etcétera), en la geometría fractal las formas primarias son conjuntos de procedimientos matemáticos (algoritmos) que al ejecutarse dentro de un rango de valores, dan como resultado las extraordinarias formas de los fractales. Pero tras todos esas formas siempre se esconden la realimentación y la iteración. La clasificación fue realizada por orden cronológico.

#### Fractales clásicos

Los fractales geométricos clásicos tiene su origen a finales del siglo XIX o a comienzos del siglo XX. Tachados de monstruos matemáticos por algunos famosos matemáticos de la época como Poincaré, sirvieron para alentar la búsqueda rigurosa de conceptos como infinito, curva continua o dimensión. Estos estudios recopilados por Benoît Mandelbrot a mediados del siglo XX, dando origen a una nueva teoría geométrica: los fractales.

El origen de la geometría fractal y de los fractales, habría que establecerlo hacia 1875-1925, cuando se produce una crisis en la definición de dimensión, y se discrepa con la idea que se había aceptado hasta entonces, según la cual se concebía la dimensión como número de coordenadas. Probablemente, el primer objeto fractal puro en la historia, es el conjunto o polvo de Cantor.

A continuación se muestran los fractales geométricos clásicos más relevantes.

- i) **Polvo de Cantor** [77] es uno de los fractales más conocidos y fáciles de construir; sin embargo, muestra muchas características fractales típicas. El polvo de Cantor es históricamente el primer objeto fractal puro. Fue descrito por el matemático alemán *Georg Ferdinand Ludwig Philipp Cantor* alrededor de 1872. Su dimensión Hausdorff-Besicovitch es  $D = \log(2)/\log(3) \approx 0.6309297$ . Este fractal es una de las excepciones (junto con el triángulo de Sierpinski y la curva de Peano) a la definición de Mandelbrot, ya que la dimensión Hausdorff-Besicovitch es menor que la dimensión topológica, de una recta es 1. El polvo de Cantor se inicia con un segmento lineal conocido como el iniciador. Este se divide en tres segmentos menores de la misma longitud, el central de los cuales se extrae y se elimina. Este proceso (denominado, usualmente, como el generador) se repite indefinidamente (véase en la Fig. 2.5), y al final se habrá producido el polvo de Cantor (puntos que no tienen tamaño, por lo que se le asigna una dimensión nula o de cero).

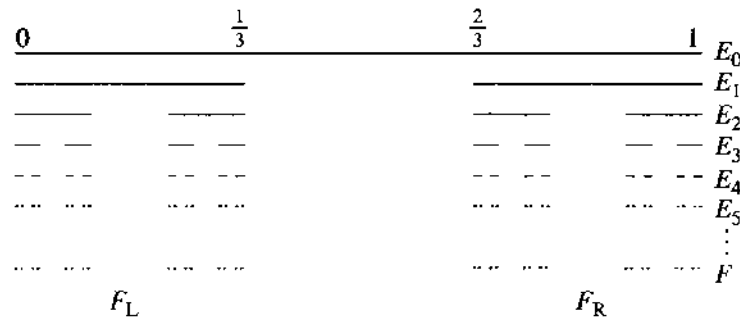


Figura 2.5: La construcción del Polvo de Cantor establece  $F$ , mediante la eliminación repetida del tercio medio de los intervalos. Tenga en cuenta que  $F_L$  y  $F_R$ , las partes izquierda y derecha de  $F$ , son copias de  $F$  escaladas por un factor  $\frac{1}{3}$ .

Algunas de las características del Conjunto de Cantor  $F$ , son:

- (i)  $F$  es auto-similar. Está claro que la parte de  $F$  en el intervalo  $[0, \frac{1}{3}]$  y la parte de  $F$  en  $[\frac{2}{3}, 1]$  son geoméricamente similares a  $F$ , escaladas por un factor  $\frac{1}{3}$ . Nuevamente, las partes de  $F$  en cada uno de los cuatro intervalos de la segunda iteración ( $E_2$ ) son similares a  $F$  pero escaladas por un factor de  $\frac{1}{9}$ , y así sucesivamente. El conjunto de Cantor contiene copias de sí mismo en muchas escalas diferentes.
  - (ii) El conjunto  $F$  tiene una “estructura fina”; es decir, contiene detalles a escalas arbitrariamente pequeñas. Cuanto más aumentemos la imagen del conjunto de Cantor, más vacíos se harán evidentes a simple vista.
  - (iii) Aunque  $F$  tiene una estructura detallada compleja, la definición real de  $F$  es muy sencilla.
  - (iv)  $F$  se obtiene mediante un procedimiento recursivo. La construcción consistió en eliminar repetidamente los tercios medios de los intervalos. Los pasos sucesivos dan cada vez mejores aproximaciones en la  $k$ -ésima iteración ( $E_k$ ) al conjunto  $F$ .
  - (v) La geometría de  $F$  no se describe fácilmente en términos clásicos: no es el lugar de los puntos que satisface alguna condición geométrica simple, ni es el conjunto de soluciones de una ecuación simple.
  - (vi) Es incómodo describir la geometría local de  $F$ , cerca de cada uno de sus puntos hay una gran cantidad de otros puntos, separados por espacios de diferentes longitudes.
  - (vii) Aunque  $F$  es, de alguna manera, un conjunto bastante grande (es infinitamente incontable), su tamaño no se cuantifica mediante las medidas usuales, como la longitud; en cualquier definición razonable,  $F$  tiene la longitud cero.
- ii) **Iteración del polvo de Cantor** [77]. El interés de Cantor por este conjunto ha de inscribirse en el debate sobre los fundamentos de la matemática en el siglo XIX. En concreto sobre el concepto de número. Cantor presentó su monstruo a la comunidad matemática. Había creado un conjunto de números del intervalo  $[0, 1]$  de medida cero (al lanzar un “dardo matemático” al intervalo, la probabilidad de dar en un punto del conjunto es nula), y al mismo tiempo el conjunto era incontable, su cardinal era idéntico al del intervalo  $[0, 1]$ .

A la función de Cantor se la suele denominar escalera del diablo (Devil's Staircase) pues posee un número infinito de escalones. Cada escalón corresponde a un intervalo eliminado en el proceso iterativo de construcción del conjunto de Cantor.

También se pueden construir conjuntos de Cantor planos comenzando con un cuadrado de lado 1, reemplazándolo por 4 cuadrados de lado  $1/2$ , y continuando este proceso indefinidamente (véase en la Fig. 2.6).



Figura 2.6: Iteración del conjunto de Cantor.

- iii) **La curva de Von Koch** [77]. En 1904 *Niels Helge von Koch* define la curva que lleva su nombre. Se forma partiendo de un segmento el cual es dividido en tres partes iguales. La parte central se sustituye por dos segmentos del mismo tamaño que el eliminado. Sucesivamente se repite el mismo proceso por cada segmento formado. La longitud de esta curva evoluciona de acuerdo a la siguiente sucesión:  $1, 4/3, 16/9, 64/27, 256/81 \dots$  (véase la Fig. 2.7). Dado que la sucesión anteriormente indicada no converge hacia ningún valor, estamos ante una curva de longitud infinita.

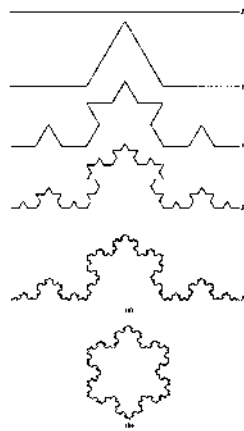


Figura 2.7: (a) Construcción de la curva de von Koch  $F$ . En cada etapa, el tercio medio de cada intervalo es reemplazado por los otros dos lados de un triángulo equilátero. (b) Tres curvas de Von Koch unidas para formar una curva de copo de nieve.

Las curvas de von Koch y el conjunto de Cantor tiene características en muchos aspectos similares. Está formado por cuatro “cuartos”, cada uno similar al conjunto, pero escalado por un factor  $1/3$ . La estructura fina se refleja en las irregularidades en todas las escalas. Si bien es razonable llamar a  $F$  una curva, es demasiado irregular para tener tangentes en el sentido clásico. Un simple cálculo muestra que  $E_k$  es de longitud  $(4/3)^k$ ; que  $k$  tienda al infinito implica que  $F$  tiene una longitud infinita. Por otro lado,  $F$  ocupa un área cero en el plano, por lo que ni la longitud ni el área proporcionan una descripción muy útil del tamaño de  $F$ .

La dimensión de Hausdorff-Besicovitch para la curva de Koch. Dado que la dimensión es representada por la siguiente fórmula:  $D = \log S / \log L$ , donde  $S$  es el tamaño del Fractal y  $L$  la escala de medición y  $D$  es la dimensión fractal que se busca. En el caso de la curva de Koch  $S = 4$  y  $L = 1/3$  y por tanto:  $D = \log 4 / \log(1/3) = \log 4 / \log 3 = 1.262$ .

Por tanto, es mayor que su dimensión topológica, que en el caso de una línea es 1 y además, tiene valor fraccionario. Cumple por tanto todos los requisitos para clasificarla como fractal, de acuerdo con la definición de Benoît B. Mandelbrot.

La versión que introdujo Von Koch en 1904 fue la denominada isla de Koch (también se denomina copo de nieve) (véase en la Fig. 2.8) cuya construcción comienza con un triángulo equilátero, aplicando luego a cada uno de sus lados un algoritmo análogo al descrito para la curva.

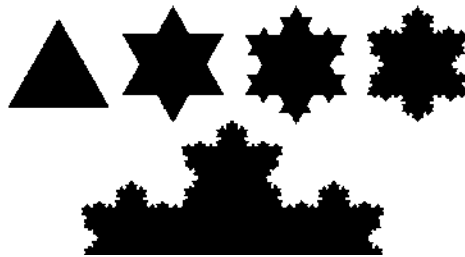


Figura 2.8: Isla o copo de nieve de Koch.

- iv) **Triángulo de Sierpinski** [77], fue descrito por matemático polaco *Waclaw Sierpinski* en 1915. El Triángulo de Sierpinski es un objeto fractal de dimensión Hausdorff-Besicovitch  $D = \log(3) / \log(2) \cong 1.58496$ . Este fractal es una de las excepciones a la definición de Mandelbrot ya que la dimensión Hausdorff-Besicovitch es menor que la dimensión topológica, ya que un triángulo es 2.

Se define como Triángulo de Sierpinski en la iteración  $n = 0$  un triángulo equilátero de lado  $x$ . En iteraciones sucesivas  $n = 1, 2, 3, \dots$ , donde se recorta un triángulo equilátero con la base invertida de lado mitad al de la iteración anterior del centro del triángulo de la iteración anterior. Tal como se puede observar en la Fig. 2.9 en la primera iteración, se traza un triángulo equilátero, cuyos vértices, deben coincidir con los puntos medios de cada lado del triángulo mayor. Esta nueva figura tendrá una orientación invertida con respecto a la primera. Seguidamente, se retira o se elimina de la figura ese nuevo triángulo invertido, tal que solamente se conserven los tres triángulos equiláteros menores—y similares—que se observan dentro del grande. En las sucesivas iteraciones se aplica el mismo procedimiento de iteración para cada triángulo pequeño, obteniéndose, como resultado, un triángulo de Sierpinski.

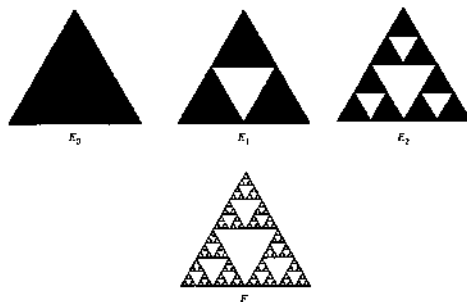


Figura 2.9: Iteraciones del Triángulo de Sierpinski.

El modelo teórico de Triángulo de Sierpinski necesitaría un número infinito de iteraciones para construirse. Conforme las iteraciones se suceden, muestra como el área total se va desvaneciendo hasta llegar a cero, al mismo tiempo el perímetro de los infinitos triángulos tenderá a infinito. Por tanto, el área se va desvaneciendo porque cuando en las sucesivas iteraciones se eliminan triángulos, esta eliminación no solamente implica quitar ese triángulo de la figura y olvidarse de él, sino que los puntos contenidos en ese área específicamente, no pertenecen al conjunto de puntos comprendidos en el triángulo de Sierpinski; o dicho de otro modo, esa sección no pertenece al conjunto.

#### 2.2.4. Fractales Mandelbrot

El conjunto de Mandelbrot fue descubierto por Benoît Mandelbrot en la década de 1970, y nombrado en su honor por Adrien Douady y J. Hubbard en 1982 [78, 79].

Benoît Mandelbrot, con ayuda de los recursos informáticos puestos a su disposición por IBM a partir de 1957 en el Centro de investigación Thomas J. Watson, se dedicó al estudio de series temporales relacionadas con precios y posteriormente con el ruido de las líneas telefónicas para interconexión de ordenadores. En el análisis de este ruido, relacionó el conjunto de Cantor con los errores aparentemente desordenados de las líneas de datos, estableciendo que era una muestra de tiempo fractal [78].

En 1975 creó las ilustraciones de su ensayo “Les objets fractales: Forme, hasard et dimension” [80]. En 1980, con ayuda de un ordenador VAX, pantalla Tektronix y plotter Versatec, sorprendió a la comunidad científica con el primer dibujo detallado sobre papel de impresora, de un gráfico deducido de la evolución del sistema dinámico en el campo complejo [81].

En el año 1982 en un nuevo libro titulado “The Fractal Geometry of Nature” [82], que estaba ilustrado con gráficos espectaculares creados con la tecnología informática que, por aquel tiempo, estaba a su disposición. Al referirse al impacto de los fractales en la concepción e interpretación de los objetos que se encuentran en la naturaleza dice: “*La Geometría Fractal cambiará a fondo su visión de las cosas. Seguir leyendo es peligroso. Se arriesga a perder definitivamente la visión inofensiva de las nubes, bosques, galaxias, hojas, flores, rocas, montañas, tapices y otras cosas. Jamás volverá a recuperar las interpretaciones de estos objetos que hasta ahora le eran familiares.*”

El conjunto de Mandelbrot [83] es un fractal que se obtiene aplicando una función iterativa sobre el conjunto de los números complejos. La función matemática que define al conjunto de Mandelbrot puede expresarse como el conjunto de todos los valores posibles de  $c$  (siendo  $c$  un número complejo)

tal que la iteración de  $Z \rightarrow Z^2 + c$ , comenzando con  $z = 0$ , no va al infinito.

En esta fórmula:  $z$  es la variable y  $c$  el valor de las coordenadas del punto analizado. Con cada punto,  $z$  comienza siendo  $(0, 0)$ , y se va aplicando reiteradamente esa fórmula. Si el módulo de  $z$  se hace en algún momento mayor que 2, significará que el punto no pertenece al conjunto de Mandelbrot. Dicho de otra forma, Mandelbrot es el conjunto de puntos cuya órbita generada con la fórmula dada nunca escapa de un círculo de radio 2.

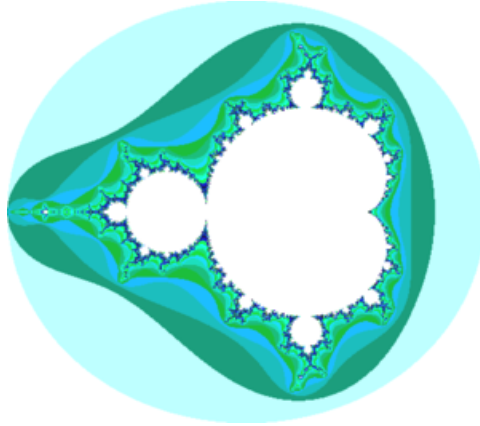


Figura 2.10: Conjunto de Mandelbrot.

### 2.2.5. Fractales Newton-Raphson

Este fractal está basado en el “Método de Newton-Raphson” para hallar las raíces de una ecuación. Isaac Newton (1642-1727) describe en su obra *Methodus fluxionum et serierum infinitarum*, un procedimiento denominado como “El método de las fluxiones” para la determinación aproximada de las raíces de una ecuación. Este método fue modificado ligeramente por Joseph Raphson en 1690, y después por Thomas Simpson en 1740, para dar la forma actual [84].

El fractal de Newton nace de una indecisión, basado en un algoritmo para calcular por aproximación las raíces complejas de un polinomio. Aplicado a una imagen fractal resulta que cuando la estimación inicial se encuentra a mitad de camino de dos raíces se produce una situación caótica, en la cual el método “no sabe” por cual de ellas decidirse [85].

$$f(z) = z^n - 1 = 0 \quad (2.5)$$

Al establecer los parámetros de la fórmula, podemos elegir el número de raíces complejas del polinomio. La versión clásica del Newton tiene tres raíces, y el fractal resultante tiene tres “brazos”, tal como se muestra en la Fig. 2.11.

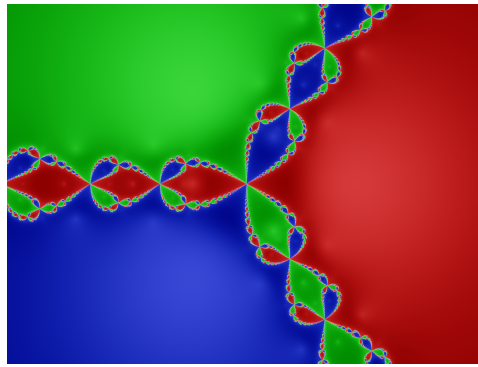


Figura 2.11: Fractal de Newton basado en tres brazos.

Es de destacar el carácter autorrepetitivo del fractal Newton-Raphson, véase en la Fig. 2.12.

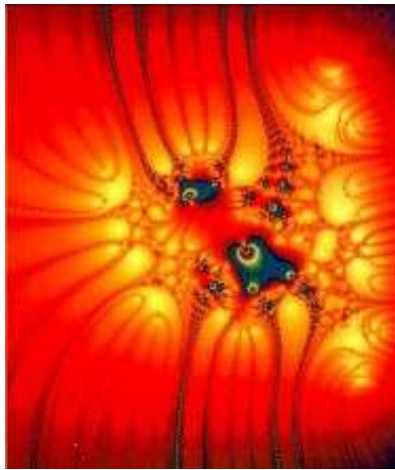


Figura 2.12: Fractal de Newton basado en dos raíces.

## 2.3. Series de tiempo fractales y multifractales

Los fractales han tenido un gran impacto en el análisis estadístico, que no es suficientemente apreciado. La naturaleza no es una serie de patrones repetitivos, está caracterizada por aleatoriedad local y orden global. Cada ser vivo es diferente en detalle pero similar en concepto de otro de su misma especie. Los fractales en la naturaleza están gobernados por una estructura estadística global, mientras que mantienen una aleatoriedad local [86].

Algunas características para distinguir series de tiempo fractales son:

- 1) Los fractales son un excelente medio para modelar formas y procesos complejos, tales como las series de tiempo de diversa índole.
- 2) Las series de tiempo completamente aleatorias llenan por completo el espacio continente, es decir su dimensión fractal es 2. La presencia de correlación hace que la serie de tiempo tienda a confinarse en espacio de dimensión fractal menor a 2.

La fractalidad se ocupa principalmente de los espacios, la multifractalidad por otro lado se ocupa de describir las medidas. Incluso si la idea detrás es también estudiar la complejidad y describir las propiedades de escala de un objeto, esos dos conceptos son distintos, de hecho, una medida puede ser multifractal a pesar de que su soporte no es fractal [87].

Las series de tiempo de viento exhiben correlación temporal, es decir, fluctuaciones aleatorias que ocurren con determinada tendencia, en que es posible observar distribuciones de probabilidad de colas pesadas. Es difícil observar la estructura multifractal en gráficos de la velocidad de viento a simple vista, ya que poseen ciertas características [88]:

- 1) Son autosimilares, sin leyenda no es posible identificar tiempo de análisis.
- 2) Poseen alta sensibilidad a las fluctuaciones del viento.
- 3) Dimensión fraccional entre 1 y 2.
- 4) No se cumplen los supuestos de normalidad, ni independencia de rendimientos en los modelos. Esto ha originado que se utilice el Coeficiente de Hurst para determinar la correlación entre el valor de la velocidad de viento en tiempos posteriores, el cual está relacionado con la dimensión fractal.

# 3

Capítulo

## Pronóstico y revisión de técnicas de pronóstico

Los pronósticos, o predicciones, son una herramienta esencial en cualquier proceso de toma de decisiones. Sus aplicaciones varían desde la determinación de los requerimientos de inventario de una pequeña empresa hasta la estimación de la energía generada a partir de las velocidades de viento en un cierto lugar. La calidad de los pronósticos está estrechamente relacionada con la información que puede extraerse y utilizarse a partir de los datos históricos. El análisis de series de tiempo es un método cuantitativo, utilizado para determinar patrones de comportamiento en los datos recolectados a través del tiempo. Los modelos cuantitativos de pronóstico se basan en datos históricos y los suponen relevantes para el futuro [89]. Algunos modelos cuantitativos de pronóstico se muestran en la Tabla 3.1. Dado que los pronósticos se preparan antes de conocer los datos reales, la precisión de los pronósticos sólo se puede determinar después de que haya transcurrido el tiempo y es mediante una cuenta acumulada de lo que se han equivocado los pronósticos en relación con los datos reales a través del tiempo. Si la precisión de un modelo es baja, se debe modificar el método o escoger uno diferente.

Tabla 3.1: Modelos cuantitativos de pronóstico.

Tipos de modelos	Características
Regresión lineal	Emplea el método de mínimos cuadrados para determinar las relaciones entre una variable dependiente y las variables independientes, presentes en un conjunto de observaciones históricas. Por lo general, este modelo es utilizado en pronósticos a largo plazo como a corto plazo.
Promedios móviles	Es el promedio aritmético de la serie de tiempo real para un determinado número de los periodos pasados más recientes es el pronóstico para el siguiente periodo. Modelo para pronósticos a corto plazo.
Promedio móvil ponderado	Modelo parecido al modelo de promedio móvil arriba descrito, excepto, que el pronóstico para el siguiente periodo es un promedio ponderado, de la serie de tiempo pasada, en lugar del promedio aritmético.
Suavización exponencial	La serie de tiempo pronosticada para el último periodo se modifica utilizando la información correspondiente al error de pronóstico del último periodo. Esta modificación del pronóstico del último periodo se utiliza como pronóstico para el siguiente periodo. Y pronostica a corto plazo.
Suavización exponencial con tendencia	Para este modelo se toma en consideración datos con un patrón de tendencia. Estos patrones pueden estar presentes en datos a mediano plazo. También se conoce como suavización exponencial doble, ya que se suavizan tanto la estimación del promedio como la estimación de la tendencia utilizando dos constantes de suavización.

### 3.1. Horizonte de predicción

El desarrollo de cualquier modelo de predicción debe contemplar, desde el primer momento de diseño, el horizonte de predicción, esto es, el periodo de tiempo (contabilizado desde un instante determinado en que realizamos la predicción), que determina el momento futuro para el que realizamos las predicciones. En otros problemas de predicción relacionados con el sistema eléctrico, como es el caso de la predicción de la demanda eléctrica, según el valor del horizonte la predicción puede clasificarse como corto plazo, medio plazo o largo plazo [66].

El horizonte de predicción y el interés de la predicción para cada grupo es el indicado (véase en la Sección 3.2). Además, Diagne et al. [90] revisaron varias propuestas, donde una de los grupos utiliza valores pronosticados de un modelo numérico de predicción del tiempo (NWP). Dentro del grupo se tienen:

- Modelo físico (Predicción numérica del tiempo)
- Modelos estadístico (Persistencia, Modelos ARIMA, Modelos SARIMA, Redes Neuronales Artificiales)
- Modelo Híbrido

La clasificación de los modelos de pronóstico basados en la resolución espacial de salida o datos previstos se ilustran en la Fig. 3.1.

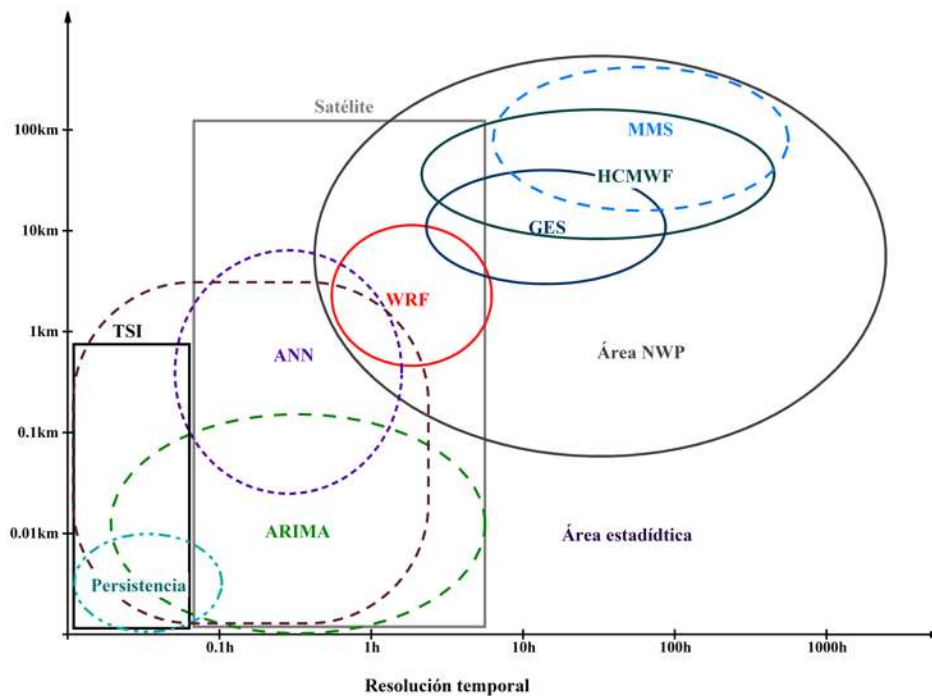


Figura 3.1: Clasificación de los modelos de predicción basados en resolución espacial de los datos de entrada y resolución temporal.

## 3.2. Tipos de pronóstico

Existe una variedad de tipos de pronóstico que va desde largo plazo a corto plazo. En primer término, se deben clasificar los procedimientos de pronóstico de largo o corto plazos. Los pronósticos a largo plazo son necesarios para establecer el curso general de alguna organización para un largo periodo; de ahí que se conviertan en el enfoque particular del alto mando. Los pronósticos a corto plazo se utilizan para diseñar estrategias inmediatas y que usan los administradores de rango medio y de primera línea para enfrentar las necesidades del futuro inmediato [91].

**Pronóstico a largo plazo** incorpora la estimación de condiciones futuras a lo largo de lapsos que por lo general son mayores a un año. Los pronósticos a largo plazo son necesarios en la administración de la producción y de las operaciones para dar apoyo a decisiones estratégicas sobre planeación de productos, procesos, tecnologías e instalaciones. Estas decisiones son de tal importancia para el éxito a largo plazo de los sistemas de producción que para el desarrollo de estos pronósticos se aplica un intenso esfuerzo organizacional [89].

**Pronóstico a mediano plazo** un pronóstico de rango mediano, o intermedio, generalmente con un lapso de tres meses a tres años. Es valioso en la planeación de producción y presupuestos, planeación de ventas, presupuestos de efectivo, y el análisis de varios planes de operación [89].

**Pronósticos a corto plazo** son estimaciones de situaciones futuras sobre lapsos que van desde unos cuantos días hasta varias semanas. Estos pronósticos pueden abarcar períodos tan cortos de tiempo que los ciclos, la estacionalidad y los patrones de tendencia surten muy poco efecto. El patrón principal de datos que afecta a estos pronósticos es la fluctuación aleatoria [89].

## 3.3. Errores de los Pronósticos

El error en el pronóstico es la diferencia numérica entre el pronóstico y el valor real de la serie de tiempo. Medida que indica la efectividad al utilizar alguno de los métodos de pronóstico. Si el pronóstico está funcionando adecuadamente quiere decir que los errores de predicción están distribuidos normalmente. Es decir, en cada uno de los periodos se compara la serie de tiempo real contra la pronosticada; y si la predicción fue perfecta lo que significa que la serie de tiempo real es igual a la predicción el error es nulo, entonces de lo contrario el pronóstico sigue el grado de error acumulándose y registrándose período a período.

La mayoría de estas mediciones implican promediar alguna función de la diferencia entre el valor real y su valor de pronóstico. Utilizándose diversos métodos para resumir los errores generados por una técnica particular de pronóstico. A menudo se denominan residuales a estas diferencias entre valores observados y los valores de pronóstico.

Así, resulta más útil calcular los errores de pronóstico en términos de porcentaje y no de cantidades; siempre y cuando el enfoque este en el tamaño o magnitud de la variable de pronóstico en la evaluación de la precisión del pronóstico.

Las técnicas de pronósticos disminuyen la incertidumbre sobre el futuro, permitiendo estructurar planes y acciones congruentes con los objetivos propuestos y permiten también tomar acciones correctivas apropiadas y a tiempo cuando ocurren situaciones fuera de lo pronosticado.

Cada tipo de pronóstico debe ajustarse al tipo de datos o serie de tiempo que desea ser analizada. Por ejemplo, para una serie de alta densidad de datos y baja variabilidad, la media móvil puede ser un buen método de pronóstico, dado que construye el pronóstico a partir del promedio de los datos anteriores. Si se quiere realizar un pronóstico con series de tiempo de viento a corto plazo se pueden utilizar técnicas a continuación descritas.

### 3.4. Persistencia

Una técnica muy simple de pronóstico de corto plazo está basada en la tendencia del tiempo de un lugar dado a permanecer con poca variación por varias horas, o aún por un día. Entonces para predecir el tiempo futuro se usa un pronóstico de persistencia, que supone será igual al de las condiciones presentes. Por ejemplo, si está a punto de llover o lloviendo ahora, podemos predecir que lloverá o continuará lloviendo las próximas horas. Los pronósticos de persistencia no tienen en cuenta los cambios que podrían tener los sistemas de tiempo en su dirección e intensidad, ni pueden predecir la formación o disipación de ciclones. Por esta limitación y por la rapidez con la cual pueden cambiar los sistemas de tiempo, este pronóstico se puede validar hasta 6 o 12 horas [92].

Se conoce como el predictor naïve [70]. En términos de velocidad del viento, el modelo de persistencia estima que la velocidad del viento en el momento  $t - 1$  es igual a la velocidad del viento en el momento  $t$ :

$$y_t = y_{t-1} \quad (3.1)$$

Este método es más preciso que la mayoría de los modelos físicos y estadísticos que se utilizan para pronosticar horizontes temporales cortos [165]. Por lo tanto, cualquier método de pronóstico propuesto debe probarse primero con el modelo de persistencia como referencia típica [166].

### 3.5. Modelo autorregresivo integrado de promedio móvil (ARIMA)

Existe una variedad de metodologías para elegir un modelo el cual sea el más representativo a la serie de tiempo de entre los diversos modelos existentes. La metodología más usada y difundida es la que propusieron los profesores G.E.P Box y G.M Jenkins en la década de los años 1970, en la cual lograron grandes avances en identificar, ajustar y verificar los modelos ARIMA adecuados [93]. Comúnmente se le conoce como *Metodología Box-Jenkins*. Esta metodología se basa en tratar de determinar cuál es el modelo probabilístico que rige el comportamiento del proceso a lo largo del tiempo. Cabe notar la diferencia entre proceso y modelo. Un proceso es lo real, es decir, el fenómeno en sí, del cual se desconoce su mecanismo generador. Por el otro lado un modelo es solo la imitación o representación del proceso [94].

El enfoque de Box-Jenkins es una de las metodologías de uso más amplio para el modelado estocástico de series de tiempo. Es popular debido a su generalidad, ya que puede manejar cualquier serie, esta-

cionaria o no estacionaria, y por haber sido implementado en numerosos programas computacionales.

Algunas consideraciones importantes acerca de la naturaleza de los datos en el modelado ARIMA son [94]:

- 1) Los modelos ARIMA se emplean tanto para datos discretos o continuos. Los datos discretos son aquellos que son medidos solamente en números enteros, nunca con cifras decimales. En tanto los datos continuos son medidos en intervalos fraccionarios, es decir, con cifras decimales.
- 2) Aún cuando la metodología ARIMA trata con datos discretos y continuos, sólo se puede aplicar a datos espaciados equidistantemente en el tiempo, en intervalos discretos de tiempo.

Los datos medidos en intervalos discretos de tiempo pueden clasificarse en dos tipos:

- a) Datos que son producto de la acumulación durante un periodo de tiempo.
  - b) Datos que son producto de la medición instantánea periódicamente
- 3) Para elaborar un modelo ARIMA se requiere una cierta cantidad de datos mínimo. Los profesores Box y Jenkins sugieren un mínimo de 50 observaciones. Un modelo ARIMA se puede aplicar a una serie con menor tamaño, realizando con especial cautela su interpretación. Para series con patrones estacionales se aconseja una serie con gran número de muestras observadas.
  - 4) Los modelos ARIMA son especialmente útiles en el tratamiento de series que presentan patrones estacionales.

**Pasos básicos de la metodología de Box-Jenkins** El método Box Jenkins de pronóstico es diferente de la mayoría de los métodos. Esta técnica no asume ningún patrón particular en los datos históricos de la serie a pronosticar. Utilizan un enfoque iterativo de identificación de un modelo útil a partir de modelos de tipo general.

El modelo elegido se verifica contra los datos históricos para ver si describe la serie con precisión. El modelo se ajusta bien si los residuos entre el modelo de pronóstico y los puntos de datos históricos son reducidos, distribuidos de manera aleatoria e independientes. Si el modelo especificado no es satisfactorio, se repite el proceso utilizando otro modelo diseñado para mejorar el original. Este proceso se repite hasta encontrar un modelo satisfactorio. La Figura 3.2 ilustra el enfoque.

Se destacan en especial 5 momentos principales [95]:

1. **Verificar la estacionariedad de la serie.** La idea de estacionariedad está relacionada con la estabilidad de la serie. Un proceso estacionario se describe como una secuencia de datos o valores que no presentan cambio sistemático en la media, ni cambio sistemático alguno en la varianza, se dice entonces que la serie es estable. Si ésta no es estacionaria, es necesario diferenciar la serie de tiempo hasta alcanzar estacionariedad.

De forma parecida, un proceso no estacionario es aquel que presenta cambios sistemáticos en la media y/o varianza; la serie de datos es inestable en el tiempo. Existen dos formas de conocer si una serie es estacionaria: por medio del gráfico de la serie y mediante la exploración de función de autocorrelación simple.

- **Inspección gráfica.**- Visualmente se puede advertir si la serie es estacionaria si se detectan elevaciones o inclinaciones en los datos. Cualquier patrón de este tipo expresa que la serie es inestable.

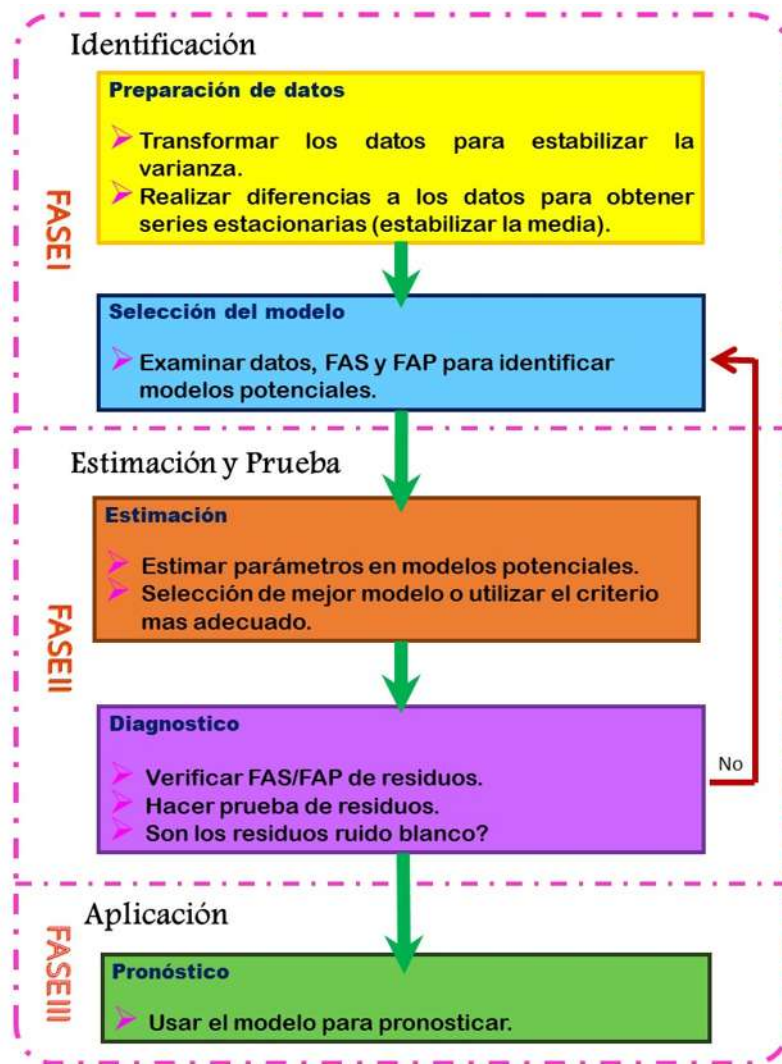


Figura 3.2: Representación esquemática de la metodología de Box-Jenkins para el tiempo.

- Funciones de autocorrelación.**- Cuando visualmente no es factible determinar si la serie es estacionaria, usualmente se recurre a la función de autocorrelación simple (FAS), en especial en aquellas series con tendencia poco remarcada. El reconocimiento de la estabilidad se logra a partir de la variedad de comportamientos que esta función que puede mostrar. Cuando se mide una variable a través del tiempo, con frecuencia está correlacionada consigo misma cuando se desfasa uno o más periodos. Los patrones de datos que incluyen componentes como tendencia, estacionalidad e irregularidad se pueden estudiar usando el enfoque del análisis de autocorrelación.

Los coeficientes de autocorrelación para diferentes desfases de tiempo de una variable se emplean para identificar patrones en las series de tiempo de datos. Comúnmente se usa la Ec. (3.2) para calcular el coeficiente de autocorrelación de primer orden ( $r_1$ ) o la correlación entre  $Y_t$ , y  $Y_1$  [93].

$$r_1 = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (3.2)$$

donde:  $r_1$  es el coeficiente de autocorrelación de primer orden,  $\bar{Y}$  la media de los valores de la serie,  $Y_t$  la observación en el periodo  $t$  y  $Y_{t-1}$  la observación en un periodo anterior o en el periodo  $t - 1$ .

La Ec. (3.3) es la fórmula para calcular el coeficiente de autocorrelación de orden  $k$  ( $r_k$ ), entre observaciones separadas por  $k$  periodos:  $Y_t$  y  $Y_{t-k}$ .

$$r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (3.3)$$

donde:  $r_k$  coeficiente de autocorrelación para un desfase de  $k$  periodos,  $\bar{Y}$  media de los valores de la serie y  $Y_{t-k}$  observación en  $k$  periodos anteriores o en el periodo  $t - k$ .

Si una serie tiene una tendencia, existe una relación significativa entre los valores sucesivos de la serie de tiempo. Es típico que los coeficientes de autocorrelación sean significativamente diferentes de cero para varios de los primeros periodos de desfase y caigan después gradualmente hacia cero al incrementarse el número de periodos. Generalmente, estos coeficientes de autocorrelación caen a cero después del segundo o tercer periodo de desfase, mientras que, en las series no estacionarias son significativamente diferentes de cero durante varios periodos. Así, estos coeficientes se emplean para ayudar a identificar el grado de relación entre los valores reales de una variable y valores anteriores de la misma, mientras que se mantienen constantes los efectos de las otras variables (periodos retrasados).

2. **Identificar un modelo tentativo.**- Una vez que se asegura que la serie es estable en el tiempo. El siguiente paso en la metodología Box- Jenkins es la identificación del modelo probable que rige el proceso de la serie de tiempo. Las ideas básicas de esta fase son las siguientes:

- La serie de tiempo que se encuentre en proceso de estudio cuenta con sus respectivas funciones de autocorrelación simple y parcial (FAS y FAP), que se denominan prácticas o calculadas.
- Por otra parte, cada una de las distintas configuraciones ARIMA posee su FAS y FAP teóricas asociadas al modelo.
- Si la FAS y FAP calculadas de la serie a la que deseamos ajustar un modelo se asemeja a alguna o varias FAS y FAP teóricas, entonces podemos decir que el modelo ARIMA teórico es un modelo tentativo para la serie.

Entonces la identificación del modelo probable se realiza por medio de la comparación de las funciones de autocorrelación calculadas contra las teóricas (tanto simples como parciales).

Las funciones de autocorrelación calculadas sólo se usan como guías para escoger uno o varios modelos ARIMA que parezcan apropiados, ya que tan sólo brindan una aproximación a la estructura más adecuada que debe considerarse.

Las funciones de autocorrelación teóricas por su parte se derivan de una familia de modelos ARIMA propuestos por Box y Jenkins. Las principales características que se observan de las funciones de autocorrelación teóricas se muestran en la Tabla 3.2 [96]:

Tabla 3.2: Características FAS y FAP Teóricas.

Modelo	FAS	FAP
AR ( $p$ )	Decae a cero	Se trunca o se corta (después del retraso $p$ )
MA ( $q$ )	Se trunca o se corta (después del retraso $q$ )	Decae a cero
ARMA ( $p, q$ )	Decae a cero	Decae a cero

### 3.5.1. Modelos autorregresivos, AR( $p$ )

Los procesos autorregresivos forman una familia de procesos tales que una observación depende de las observaciones anteriores. Sea una serie temporal  $Y_t$  estacionaria que sigue un proceso autorregresivo de orden  $p$  si:

$$Y_t = \delta + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t. \quad (3.4)$$

En los modelos autorregresivos, el proceso se representa como una suma ponderada de observaciones pasadas de la variable. El número de rezagos  $p$  determina el orden del modelo autorregresivo [97].

### 3.5.2. Modelos media móvil, MA( $q$ )

Las familias de modelos que representan los procesos de memoria corta se denominan genéricamente procesos de media móvil. Se denomina al proceso de media móvil de orden  $q$ , MA( $q$ ) con la Ec. (3.5).

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (3.5)$$

En los modelos de media móvil, el proceso se representa como una suma ponderada de errores actuales y anteriores. El número de rezagos del error considerados  $q$  determina el orden del modelo de media móvil [97].

### 3.5.3. Modelos mixtos autorregresivo de media móvil, ARMA( $p, q$ )

En la práctica se observa que las series no son puras AR o MA, sino que presentan parte AR y parte MA. Los procesos ARMA( $p, q$ ) son combinación de estructuras autorregresivas y de media móvil que tienen una parte AR( $p$ ) y una parte MA( $q$ ).

El proceso ARMA( $p, q$ ) será:

$$Y_t = \delta + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (3.6)$$

En estos modelos, el proceso se representa en función de observaciones pasadas de la variable y de los valores actuales y rezagados del error. El número de rezagos de la variable de interés  $p$  y el número de rezagos del error  $q$  determinan el orden del modelo mixto [97].

### 3.5.4. Modelos media móvil integrada autorregresiva, ARIMA( $p, d, q$ )

Antes de comenzar, un elemento llamado *diferencia regular*, siendo  $B$  el operador retardo, llamamos operador diferencia regular,  $\nabla$  a:

$$\nabla = 1 - B,$$

que actúa de manera que, al aplicarlo sobre un término de un proceso estocástico, se obtiene:

$$\nabla Y_t = (1 - B)Y_t = Y_t - Y_{t-1}.$$

Este operador puede también aplicarse sobre un término de un proceso estocástico de segundo orden y sucesivamente.

Un modelo ARIMA( $p,d,q$ ) es un proceso estocástico integrado de orden  $d$  cuando es necesario diferenciarlo  $d$ -veces para conseguir un proceso estacionario, en la práctica no suele ser necesario aplicar una diferencia regular superior a 2. Además  $p$  es el orden de la parte autorregresiva y  $q$  el orden de la parte de medias móviles [97].

### 3.5.5. Modelos media móvil integrada autorregresiva estacional, SARIMA ( $p, d, q$ )( $P, D, Q$ ) $_s$

Es el modelo más general de la familia de modelos ARIMA cuya expresión es:

$$\phi_p(B)\Phi_p(B^S)\nabla^d\nabla_s^D y_t = c + \theta_q(B)\Theta_Q(B^s)\varepsilon_t,$$

donde  $s$  es la longitud del periodo estacional [97].

Como resumen de los modelos estacionales puede construirse el siguiente cuadro (véase la Tabla 3.3)

Tabla 3.3: Características de los modelos estacionales vs. FAS y FAP.

Modelos	FAS	FAP
$AR(p) \times AR(P)_s$	Decaimiento exponencial y/u oscilaciones sinusoidales atenuadas	Se anula para retardos mayores a $p + sP$
$MA(q) \times MA(Q)_s$	Se anula para retardos mayores a $q + sQ$	Decaimiento exponencial y/u oscilaciones sinusoidales atenuadas.

3. **Estimar el modelo.**- En esta etapa se estiman los coeficientes del modelo seleccionado tentativamente en el paso anterior, existen distintos criterios para dicho propósito. Usualmente se realiza por medio de paquetes informáticos. Además en esta etapa se puede pre-visualizar el modelo adecuado a la serie de tiempo. Particularmente, si los coeficientes no satisfacen ciertas condiciones matemáticas, el modelo es rechazado.

Finalmente obtenidos los parámetros  $\theta$ ,  $\varphi$ ,  $\Theta$  y  $\Phi$  del modelo ARIMA propuesto, deben de cubrir en conjunto los siguientes requerimientos [95]:

- 1.- *El modelo debe tener parsimonia:* este principio es importante ya que en la práctica un modelo con esta característica produce mejores predicciones. El objetivo en el modelado ARIMA no es encontrar el modelo exacto que represente al proceso generador de las observaciones, es más bien, encontrar el modelo que se aproxime al verdadero proceso y que explique el comportamiento de la variable en estudio de forma adecuada y práctica. Es decir, se opta siempre por el modelo con el menor número de coeficientes y que satisfaga el hecho de un buen ajuste.
- 2.- *El modelo debe ser estacionario e invertible:* la metodología Box-Jenkins requiere que el modelo que se utiliza para la descripción y pronóstico de una serie de tiempo sea tanto estacionario como invertible. La invertibilidad se refiere a que cualquier modelo ARIMA puede expresar a la serie  $Y_t$  en función de las  $Y$  observaciones pasadas, es decir  $Y_{t-1}$ ,  $Y_{t-2}$ ,  $Y_{t-3}$ , etc.

Las condiciones para modelos generales son complicadas, se puede decir que si un modelo utiliza parámetros autorregresivos, una condición de estacionariedad necesaria (pero no suficiente) es que la suma de los valores de los parámetros autorregresivos es menor que 1. La tabla 3.4 [95] muestra las condiciones de estacionariedad e invertibilidad para los modelos ARMA más comunes.

Tabla 3.4: Condiciones de estacionariedad e invertibilidad.

Modelo	Condiciones de estacionariedad	Condiciones de invertibilidad
AR(1)	$ \varphi_1  < 1$	Ninguna
AR(2)	$\varphi_1 + \varphi_2 < 1$ $\varphi_2 + \varphi_1 < 1$ $ \varphi_2  < 1$	Ninguna
MA(1)	Ninguna	$ \theta_1  < 1$
MA(2)	Ninguna	$\theta_1 + \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $ \theta_2  < 1$
ARMA(1,1)	$ \varphi_1  < 1$	$ \theta_1  < 1$

4. **Evaluación del modelo.**- En este paso se comprueba la eficiencia del modelo y se decide si es estadísticamente adecuado. Un modelo estadísticamente adecuado es aquel cuyos residuales son independientes entre sí. Es decir, si los residuales son completamente aleatorios.

5. *Pronóstico.*- La fase final de la metodología Box-Jenkins es pronosticar valores futuros de la serie de tiempo. En primer lugar se debe comprender cómo realizar predicciones puntuales a partir de un modelo ARIMA ya estimado. En segundo lugar se debe estudiar cómo estimar límites de probabilidad alrededor de una estimación puntual, conocidos particularmente como intervalos de confianza. Y por último se debe conocer y probar la suficiencia de los pronósticos.

## 3.6. Deep Learning (Aprendizaje Profundo)

La ciencia se ha inspirado de la naturaleza en una gran variedad de entornos. Podemos observar por ejemplo, como las aeronaves utilizan principios de aerodinámica tomados de la forma en que las aves se mantienen en el aire, el funcionamiento de las cámaras se ha ido perfeccionado a través de la imitación de los ojos, entre otros ejemplos. La naturaleza provee una potente herramienta para aprender de sus errores a través de la selección natural, y llega a diseños que generan gran eficiencia y adaptación a su entorno. Esa característica hace que muchos de ellos se vuelvan grandes ideas tecnológicamente hablando.

La inteligencia artificial se está considerando como la nueva revolución industrial, corazón de lo que algunos llaman industria 4.0, mientras Deep Learning es el motor de este proceso [98]. En este campo tenemos ejemplos de cómo se han adaptado modelos naturales para crear algoritmos que permitan aprender de sus errores y volverse más eficientes en su tarea. Estos algoritmos son particularmente útiles para resolver tareas en las que no es posible hallar una solución matemáticamente pura. Un ejemplo de lo anterior son los sistemas informáticos actuales ya traducen textos en cualquier idioma, responden automáticamente correos electrónicos o crean *fake news*. Esto se debe en gran medida, a una parte de la inteligencia artificial que se denomina Deep Learning, término que agrupa una parte de técnicas de aprendizaje automático (Machine Learning en inglés) que se basan en modelos de redes neuronales y cuya gran aplicabilidad se ha mostrado recientemente en multitud de usos por parte de la industria. Aunque se espera que nuevas vías se abran para la inteligencia artificial, y las nuevas tecnologías sean desarrolladas, así como la capacidad computacional se incrementa en ordenes de magnitud. Sin embargo, antes de esto, será necesario encontrar qué se quiere computar. Hay al menos una nueva dimensión de computación que ha sido muy difícil de alcanzar con los ordenadores digitales [98, 99].

Hoy en día el uso de la expresión aprendizaje profundo (*Deep Learning*, siglas en inglés (DL)) en redes neuronales artificiales tiene su origen en el año 2000, aunque fue otro artículo publicado en 2006 por la revista Science, el que sirvió para rebautizar las redes neuronales artificiales y despertar de nuevo el interés en ellas, más allá de los investigadores irreductibles que habían seguido trabajando en ellas cuando casi nadie les prestaba atención [24]. Desde entonces, se hace referencia a las redes neuronales artificiales con un apelativo mucho más comercial: *Aprendizaje profundo* [25], sigue siendo una excelente herramienta con interés particular para minería de datos, puesto que ofrece los medios para modelar de manera efectiva y eficiente problemas grandes y complejos como la visión por computadora y el reconocimiento de voz, que eran difíciles con otros enfoques [25, 100]. El aprendizaje profundo es visto ahora como un subconjunto de técnicas de aprendizaje automático, que a su vez es una de las disciplinas englobadas dentro del campo de la Informática conocido como Inteligencia Artificial (véase la Fig. 3.3).

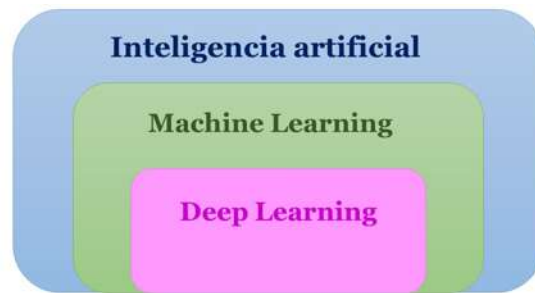


Figura 3.3: Deep Learning es un subconjunto de Machine Learning, que es solo una parte de la inteligencia artificial, aunque en estos momentos quizás es más dinámica y la que está haciendo que la inteligencia artificial esté nuevamente en pleno auge.

Autores como Berzal [25] y DasGupta [28] consideran el aprendizaje profundo como un método para resolver problemas, de forma individual o combinadas con otros métodos, para aquellas tareas de clasificación, identificación, diagnóstico, optimización o predicción, donde el balance datos/conocimiento se inclina hacia los datos y adicionalmente, puede haber la necesidad de aprendizaje en tiempo de ejecución con cierta tolerancia a fallos. En estos casos las RNA se adaptan dinámicamente reajustando constantemente los “pesos” de sus interconexiones. Además, proporciona uno de los mecanismos mediante los que se puede conseguir que un ordenador aprenda. Cadenas et al. [16] menciona que los modelos RNA son utilizados para manejar la naturaleza no lineal de la velocidad del viento, los cuales son determinados a partir de los datos, es decir, son capaces de encontrar relaciones (patrones) de forma inductiva por medio de los algoritmos de aprendizaje, basado en los datos existentes, en lugar de requerir la ayuda de un modelador para especificar la forma funcional y sus interacciones.

### 3.6.1. Redes neuronales (Modelo biológico)

La teoría y modelado de redes neuronales artificiales está inspirada en la estructura y funcionamiento de los sistemas nerviosos, donde la neurona es el elemento fundamental. Existen neuronas de diferentes formas tamaños y longitudes. Estos atributos son importantes para determinar la función y utilidad de la neurona. La clasificación de estas células en tipos estándar ha sido realizada por muchos neuroanatomistas [101]. Algunas de las peculiaridades de las redes neuronales biológicas es su tamaño en todo el sistema nervioso central, teniendo en cuenta además que el tamaño y la estructura de la red están cambiando radicalmente durante y después de la niñez, cuando está ya en uso. En otras palabras, la distribución de los recursos y los caminos de comunicación más importantes están formados de acuerdo con un plan genético, mientras que el resto de la programación en especial la memoria, debe ser adquirido después de nacer. Programar tal red puede significar dos cosas: a) las estructuras de interconexión entre las células son alteradas, y b) las fuerzas de estas interconexiones son cambiadas. Parece que existen bastantes estrategias claras de cómo cambiar las fuerzas en la dirección correcta, mientras que cambios en las interconexiones son más difíciles de definir, porque suelen tener efectos radicales en el comportamiento de la red, especialmente en lo concerniente a la operación secuencial y las funciones jerárquicas [101, 102].

Otra función importante del sistema nervioso es definir acciones que son parte del comportamiento, y controlar el estado del organismo en relación con su entorno. Mientras que las representaciones internas en las cuales está basado el comportamiento pueden ser derivadas de las entradas de forma bastante directa, las definiciones de las salidas deben estar basadas en estrategias completamente di-

ferentes [103]. De hecho, apenas existe otra posibilidad para programar las acciones que aplicar el principio de retroceso y empuje para alterar los mecanismos que son responsables de ellas.

Por otro lado, el significado y la calidad de las acciones deben ser juzgados no desde los movimientos inmediatos, sino desde la realización de criterios que tienen en cuenta el resultado requerido, a veces bastante indirecto. A menudo, las acciones son solamente corregidas si se realizan con cierta frecuencia, donde el mecanismo que las provoque debe contener circuitos que definan tales frecuencias y que se cambien en relación con los resultados aprendidos. Está claro que programar las acciones es un proceso mucho más indirecto que programar las representaciones internas; el factor aleatorio no puede ser evitado [104]. Al programar las funciones de entrada y de salida lleva sólo a una operación de comportamiento en la que el estímulo y la respuesta son consideradas muy relevantes. Ciertamente, es posible realizar autómatas bastante complejos y comportamientos necesidad-conducta de este modo. Sin embargo, se mantiene la expectativa sobre las posibilidades de que las redes neuronales puedan actuar como ordenadores para algunos problemas abstractos y también donde la computación se realice en el estado interno de la red.

### Estructura de la neurona

Una neurona es una célula viva y, como tal, contiene los mismos elementos que forman parte de todas las células biológicas. Además, contienen elementos característicos que las diferencian. En general, una neurona consta de un cuerpo celular más o menos esférico, de 5 a 10 micras de diámetro, del que salen una rama principal, el axón, y varias ramas más cortas, llamadas dendritas (véase la Fig. 3.4). A su vez, el axón puede producir ramas en torno a su punto de arranque, y con frecuencia se ramifica extensamente cerca de su extremo [101, 105].

Una de las características que diferencia a las neuronas del resto de las células vivas, es su capacidad de comunicarse. En términos generales, las dendritas y el cuerpo celular reciben señales de entrada; el cuerpo celular las combina e integra y emite señales de salida. El axón transporta esas señales a los terminales axónicos, que se encargan de distribuir información a un nuevo conjunto de neuronas [106]. Las señales que se utilizan, son de dos tipos distintos de naturaleza: eléctrica y química. La señal generada por la neurona y transportado a lo largo del axón es un impulso eléctrico, mientras que la señal que se transmite entre los terminales axónicas de una neurona y las dendritas de las neuronas siguientes es de origen químico; concretamente, se realiza mediante moléculas de sustancias transmisoras (neurotransmisores) que fluyen a través de unos contactos especiales, llamados sinapsis, que tienen la función del receptor y están localizados entre los terminales axónicos y las dendritas de la neurona siguiente (véase la Fig. 3.4).

Por lo general, una neurona recibe información de miles de otras neuronas y, a su vez, envía información a miles de neuronas más. Se calcula que en el cerebro humano existen del orden de 10<sup>15</sup> conexiones [101].

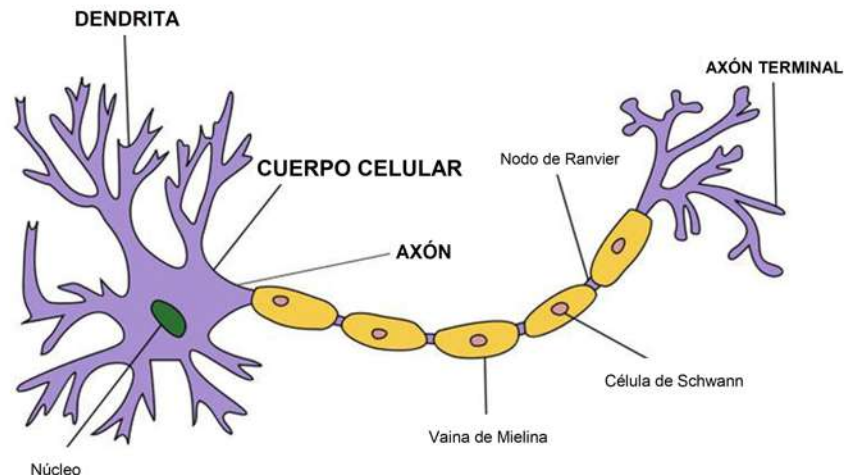


Figura 3.4: Esquema de una red neuronal biológica.

### 3.6.2. Redes neuronales (Modelo artificial)

Para tratar de entender cómo funciona el cerebro biológico, para diseñar la Inteligencia Artificial, Warren McCullock y Walter Pitts publicaron, en 1943, el primer concepto de una célula cerebral simplificada, la denominada neurona **McCullock-Pitts (MCP)**, recogido en su libro *A Logical Calculus of the Ideas Immanent in Nervous Activity* [107]. Un caso especial de algoritmos de Machine Learning son las redes neuronales artificiales, que en cierta manera son un intento de imitar la actividad en capas de neuronas en la neocorteza, que es la parte del cerebro humano donde ocurre el pensamiento. Estas redes neuronales aprenden estructuras jerárquicas y niveles de representación y abstracción para comprender los patrones de datos que provienen de varios tipos de fuentes, como imágenes, videos, sonido o texto. Una de las mayores ventajas del Deep Learning es su capacidad de aprender automáticamente, la representación de características en múltiples niveles de abstracción. Esto permite que un sistema aprenda funciones complejas asignadas desde el espacio de entrada al espacio de salida sin muchas dependencias de las funciones creadas por humanos. Aunque, precisamente, esto implica que a menudo la explicabilidad del modelo o más bien de los resultados de esto no sea fácilmente explicable, lo cual se convierte en un problema en ciertas áreas de aplicación [98].

Existen numerosas formas de definir lo que son las redes neuronales, desde las definiciones cortas y genéricas hasta las que intentan explicar más detalladamente lo que significa red neuronal o computación neuronal. Algunos ejemplos son:

- La red neuronal es vista como sistema de computación hecho por un gran número de elementos simples, elementos de proceso muy interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas [108]. Es decir, una nueva forma de computación, inspirada en modelos biológicos. Y un modelo matemático compuesto por un gran número de elementos procesales organizados en niveles.
- Redes neuronales artificiales son redes interconectadas masivamente en paralelo de elementos simples (usualmente adaptativos) y como organización jerárquica, las cuales intentan interactuar con los objetos del mundo real del mismo modo que lo hace el sistema biológico [99]. Es necesario destacar que tales computadores neuronales no ejecutan las típicas instrucciones de

máquina de los computadores digitales, al menos de que estén hechos para anular el comportamiento de las redes neuronales físicas. En principio, la operación de proceso básico realizada por todos los procesadores elementales es una operación análoga de transformación de sus señales de entrada. A diferencia del proceso en las redes neuronales biológicas, las células neuronales (neuronas) corresponden a los elementos de proceso anteriores. Las interconexiones se realizan por medio de las ramas de salida (axones) que producen un número variable de conexiones (sinapsis) con otras neuronas (o quizá con otras partes, como músculos y glándulas). Las redes neuronales son sistemas de simples elementos de proceso muy interconectados [101].

Por lo que, la compleja operación de las redes neuronales es el resultado de abundantes lazos de realimentación junto con no linealidades de los elementos de proceso y cambios adaptativos de sus parámetros, que pueden definir incluso fenómenos dinámicos muy complicados.

Además, es importante mencionar que existen diferentes aspectos de una neurona biológica que podemos descartar cuando construimos un modelo computacional con el que construir una red neuronal artificial [25], estos son:

- **Abstracción espacial:** aunque las neuronas biológicas ocupan un espacio tridimensional en el cerebro de un animal, lo que les impone restricciones con respecto a su configuración e interconectividad. Desde el punto de vista computacional, las neuronas son consideradas como nodos en los modelos de neuronas artificiales. Y las conexiones entre neuronas se modelan mediante enlaces. Para la efectividad de una sinapsis a la hora de desencadenar una reacción en la neurona postsináptica depende de distintos factores, como su localización espacial (proximal cuando está cerca del soma de la neurona, distal cuando está lejos) o su propia eficiencia (dependiendo de la densidad y tipos de vesículas presinápticas y receptores postsinápticos). Los enlaces, con pesos, dan lugar a un grafo ponderado, generalmente dirigido, aunque en algunos modelos de redes neuronales también se admiten enlaces no dirigidos o, incluso, enlaces no simétricos.
- **Abstracción temporal:** en una neurona biológica, la transmisión de una señal a lo largo del axón lleva asociado un retardo que depende de la longitud del axón, su grosor y de si está mielinado o no. El retardo de propagación de la señal a lo largo de un axón, generalmente, no se tiene en cuenta en los modelos computacionales de redes neuronales.
- **Abstracción funcional:** los modelos computacionales que se utilizan en Inteligencia Artificial suelen prescindir de detalles espaciotemporales que sí son relevantes para la neurociencia [109].

El modelo que normalmente se utiliza como base para construir redes neuronales artificiales es el que se muestra en la Fig. 3.5, dicho modelo consta de dos etapas. En una primera etapa, se combinan las entradas provenientes de otras neuronas teniendo en cuenta los pesos de las sinapsis. El resultado de esta primera etapa es la entrada neta o excitación de la neurona. En la segunda etapa, la entrada neta se utiliza directamente para determinar el valor de salida de la neurona, que se propagará a otras neuronas. Es decir, se ha supuesto que la combinación de entradas y pesos generan directamente la salida, en el modelo general de neurona artificial, la generación de la salida de la neurona es responsabilidad de una fase de activación independiente de la fase de integración de entradas [25, 98, 107].

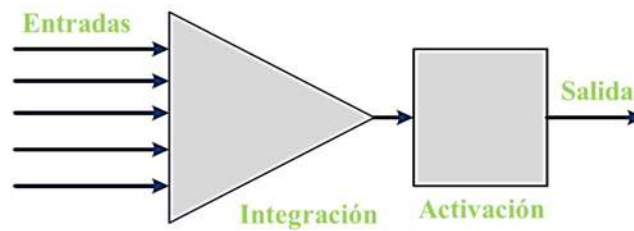


Figura 3.5: Modelo simplificado de neurona con integración y activación.

Generalmente, la activación de una neurona artificial no reproduce sin más su entrada neta, sino que aplica algún tipo de transformación no lineal a esa entrada neta. Ya que, la mayor parte de los problemas reales incorporan algún tipo de no linealidad, interesa que las redes sean capaces de representar esas no linealidades (véase Fig. 3.6). Obteniéndose así el modelo estándar de neurona artificial utilizado en redes neuronales artificiales.

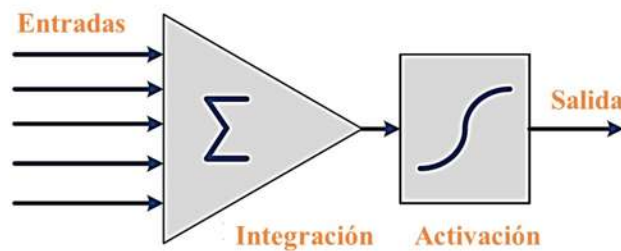


Figura 3.6: Modelo estándar de neurona con función de activación no lineal.

Habitualmente, las redes neuronales artificiales están formadas por conjuntos de neuronas que son agrupadas en capas, de forma que todas las neuronas de una misma capa compartan ciertas características. Por convención, se denota por  $x_i$  a cada una de las  $n$  entradas recibidas por una capa de neuronas formada por  $m$  neuronas. La salida de cada neurona estará representada por  $y_j$ , habitualmente un valor binario o real, que puede ser considerado análogo al estado de activación (binario) o a la frecuencia de activación (real) de una neurona biológica.

Los pesos sinápticos  $w_{ij}$  son utilizados para modelar las conexiones de entrada a las neuronas, siendo el peso  $w_{ij}$  el peso asociado a la sinapsis que conecta la entrada  $i$ -ésima con la neurona  $j$ -ésima. Los pesos, por regla general, tiene valores reales, positivos para modelar conexiones excitatorias y negativos para modelar conexiones inhibitorias. Se considera, en ocasiones que todas las neuronas reciben una entrada especial permanentemente fija,  $x_0 = 1$ . Esa entrada permite modelar el sesgo o bias  $b_j$  de la neurona  $j$  como un peso más:  $w_{0j} = b_j$ . En determinados modelos de neuronas binarias, el sesgo sirve para determinar el umbral de activación de la neurona, el punto a partir del cual la neurona activa su salida. Si el nivel de excitación de la neurona queda por debajo del umbral, se mantiene inactiva su salida. En cuanto su entrada neta supera el umbral de activación, se activa la salida de la neurona.

En la etapa de integración de entradas, una neurona artificial combina las diferentes entradas  $x_i$  con sus pesos para determinar su entrada neta  $z_j$  o  $net_j$ , como en la Ec. (3.7).

$$z_j = net_j = \sum_i w_{ij}x_i \quad (3.7)$$

En la etapa de activación, una neurona artificial utiliza el valor asociado a su entrada neta para generar una salida  $y_j$ . El modelo más habitual genera una salida de tipo numérico a partir de la entrada neta de la neurona (véase la Ec. (3.8)).

$$y_j = f(z_j) = f(net_j) = f\left(\sum_i w_{ij}x_i\right) \quad (3.8)$$

donde la función  $f$  es la función de activación de la neurona, usualmente no lineal.

En otros modelos de redes neuronales, la salida puede depender del valor de activación previo de la neurona. En ese caso, la salida tiene una componente temporal, véase en la Ec. (3.9).

$$y_j(t) = F(y_j(t-1), z_j(t)) = f(net_j) = F(y_j(t-1), net_j(t)) \quad (3.9)$$

También existen modelos de redes neuronales estocásticas. En ese tipo de modelos, la entrada neta de la neurona se utiliza para representar la probabilidad de activación de una neurona. Si se tiene una neurona estocástica binaria, la misma expresión que se utiliza antes para la salidas posible emplearla para establecer una probabilidad de activación  $p_j$ , véase en la Ec. (3.10)

$$p_j(t) = f(z_j) = f(net_j) = f\left(\sum_i^k w_{ij}x_i\right) \quad (3.10)$$

En este caso, una neurona estocástica binaria generaría un pulso (*salida* 1) con probabilidad  $p_j$  y se mantendría inactiva (*salida* 0) con probabilidad  $1 - p_j$ .

### Características del aprendizaje profundo

Lo que diferencia al aprendizaje profundo de otras familias de técnicas de aprendizaje es esencialmente, su capacidad de abstracción y su capacidad de formar jerarquías de conceptos utilizando múltiples niveles de representación [25]. De modo que casi todos los algoritmos de aprendizaje profundo se pueden describir como instancias particulares de una receta bastante simple, que aplicamos para resolver un problema concreto de aprendizaje automático:

- Recopilar un conjunto de datos asociado al problema (enorme, si es posible).
- Diseñar una función de coste apropiada para el problema, también conocida como función de pérdida.
- Seleccionar un modelo de red neuronal y establecer sus hiperparámetros (tamaño, características,...).
- Aplicar un algoritmo de optimización para minimizar la función de coste ajustando los parámetros de la red.

Por lo que, las técnicas de aprendizaje profundo parten de la idea base *si somos capaces de aprender con éxito múltiples niveles de representación, podremos generalizar correctamente*. El aprendizaje profundo proporciona herramientas que permiten construir características a partir de otras características. Repitiendo el proceso, es posible construir modelos jerárquicos de características. Esos niveles de

características proporcionan múltiples niveles de representación, que corresponden a múltiples niveles de abstracción. En esencia, las redes neuronales utilizadas en la actualidad no difieren tanto de las que se usaban hace 30 años, se ha popularizado el aprendizaje profundo y ha contribuido a revitalizar las redes neuronales artificiales precisamente ahora por tres causas principales [25]:

- 1.- Datos.- la disponibilidad de conjuntos de datos enormes.
- 2.- Potencia de cálculo.- es proporcionada por las GPU [Graphical Processing Unit] y el desarrollo de nuevos algoritmos.
- 3.- Algoritmos.- se logra con éxito el entrenar redes neuronales con múltiples capas ocultas.

### 3.6.3. Funciones de activación

Los modelos de neuronas utilizados en redes neuronales artificiales combinan sus entradas usando pesos que modelan sus conexiones sinápticas y, a continuación, le aplican a la entrada neta de la neurona una función de activación o transferencia. La entrada neta de la neurona recoge el nivel de estímulo que la neurona recibe de sus entradas y es la función de activación la que determina cuál es la salida de la neurona. A grandes rasgos, las funciones de activación se pueden clasificar en [25]:

1. *Funciones de activación discretas.*- es donde la salida de la neurona sólo puede tomar un conjunto finito de valores. Normalmente, se utilizan dos valores, por lo que se habla de neuronas binarias cuando la salida es 0 ó 1, mientras se trate de neuronas bipolares cuando su salida puede ser  $-1$  ó  $+1$ . Los primeros modelos de redes neuronales artificiales eran discretas, usadas por el algoritmo de aprendizaje del perceptrón.
2. *Funciones de activación continuas.*- en este caso, la salida de la neurona puede tomar cualquier valor dentro de un intervalo. Dada una función de activación  $f_{[0,1]}$  con un rango definido, el intervalo  $[0, 1]$ , es fácil crear una función de activación con el rango deseado  $[a, b]$ . Basta con aplicar una transformación lineal del tipo de la Ec. (3.11).

$$f_{[a,b]}(z) = a + (b - a)f_{[0,1]}(z) \quad (3.11)$$

Para neuronas bipolares de cualquier función de activación que sólo tome valores positivos (véase la Ec. (3.12)).

$$f_{[-1,1]}(z) = 2f_{[0,1]}(z) - 1 \quad (3.12)$$

Dentro de las funciones de activación continuas, se pueden utilizar tanto funciones de activación lineales (p.ej. la función identidad) como funciones no lineales. Éstas últimas son las más utilizadas habitualmente, ya que son las que dotan a una red neuronal multicapa de su capacidad de aproximador universal.

3. *Función de activación lineal.*- la función más sencilla para una neurona artificial es la función identidad, como se muestra en la Ec. (3.13).

$$y = f_{lin}(z) = z \quad (3.13)$$

Sobre esta función de activación es posible aplicar cualquier transformación lineal de las mencionadas arriba y seguirá teniendo una neurona lineal. La característica principal de las neuronas

lineales es que, si se conectan varias capas de neuronas lineales en serie, el resultado final será siempre equivalente a una única capa de neuronas lineales (véase la Fig. 3.7).

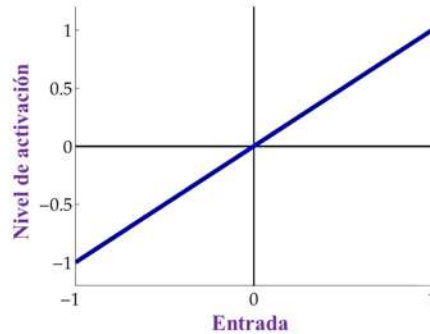


Figura 3.7: Función de activación lineal.

Desde un punto de vista más formal, las capas de neuronas lineales actúan como simples clasificadores lineales, por lo que no resultan adecuadas para resolver problemas complejos por sí mismas.

4. *Función escalón.*- primera función no lineal que se utilizó para modelar neuronas artificiales, también conocida como función umbral, limitador estricto o función de Heaviside. La función se muestra en la Fig. 3.8 y la ecuación del modelo neuronal parte de McCulloch y Pitts (Ec. (3.14)).

$$y = f_{tlu}(z) = u(z) = 1_{z \geq 0} = \begin{cases} 1 & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases} \quad (3.14)$$

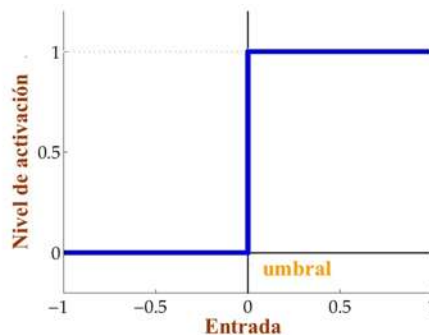


Figura 3.8: Función de activación TLU.

Como se trata de una función de activación binaria de la que se puede derivar una versión bipolar: la función signo, también conocida como limitador estricto simétrico (Ec. (3.15)).

$$y = f_{sgn}(z) = sgn(z) = \begin{cases} 1 & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases} \quad (3.15)$$

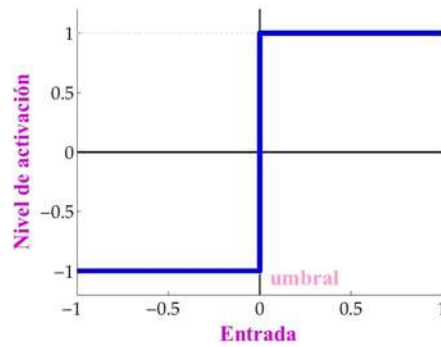


Figura 3.9: Función de activación TLU simétrica o bipolar.

Esta función de activación es simétrica con respecto al origen (véase la Fig. 3.9). Se utilizó, en algunos modelos primitivos de redes neuronales artificiales, porque su implementación electrónica resultaba más sencilla que en su versión binaria.

5. *Función lineal con saturación.*- es una función lineal de activación a trozos que limite el rango de salida de la neurona (véase Fig. 3.10):

- En su versión binaria:

$$y = f_{sl}(z) = \begin{cases} 1 & \text{si } z > +\frac{1}{2} \\ z + \frac{1}{2} & \text{si } -\frac{1}{2} \leq z \leq +\frac{1}{2} \\ 0 & \text{si } z < -\frac{1}{2} \end{cases} \quad (3.16)$$

- En su versión bipolar o simétrica:

$$y = f_{ssl}(z) = \begin{cases} 1 & \text{si } z > 1 \\ z & \text{si } -1 \geq z \geq -1 \\ -1 & \text{si } z < -1 \end{cases} \quad (3.17)$$

- Una forma alternativa de definir esta función es la siguiente:

$$y = f_{ssl}(z) = \max(-1, \min(1, z)) \quad (3.18)$$

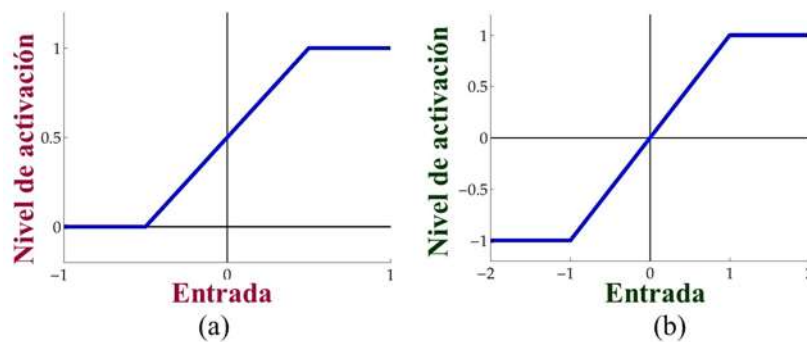


Figura 3.10: Función de activación lineal (a) con saturación. (b) bipolar con saturación.

Por su similitud con la tangente hiperbólica, Ronan Collobert la denominó tangente hiperbólica estricta. Su derivada, aunque no está formalmente definida para los puntos  $z = 1$  y  $z = -1$ , puede calcularse en la práctica mediante la Ec. (3.19).

$$y = \frac{d}{dz} f_{ssl}(z) = \begin{cases} 1 & \text{si } |z| \leq 1 \\ 0 & \text{en otro caso} \end{cases} \quad (3.19)$$

6. *Función de activación sigmoidal.*- es una función no lineal, estrictamente creciente, continua y derivable, también satisface que se entrena usando backpropagation, es decir dados los errores observados en la capa de salida, propaga esos errores hacia atrás en la red para ir ajustando sus parámetros internos. La forma en que se ajustan esos parámetros depende de la derivada de la función de activación de las neuronas, de ahí el interés en que la función sea derivable. Existen distintas funciones sigmoideas, aunque todas preservan en común su característica forma de S. Algunos ejemplos de funciones sigmoideas: la función logística [25], la tangente hiperbólica [25] y la función gudermanniana [110].

■ *Función logística*

$$y = f_{logistic}(z) = \sigma(z) = \frac{1}{1 + e^{-z}} \quad (3.20)$$

La función logística es simétrica (veáse Fig. 3.11), en el sentido de la Ec. (3.21).

$$\sigma(-z) = 1 - \sigma(z) \quad (3.21)$$

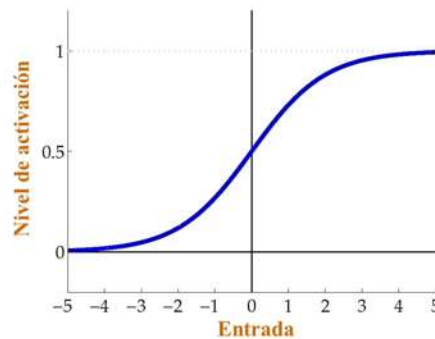


Figura 3.11: Función de activación sigmoidal: Función logística.

- *Tangente hiperbólica.*- se utiliza en trigonometría esférica y se define como en la Ec. (3.22). Según algunos autores, funciona mejor que la función logística al parecerse más a la función identidad para valores de  $z$  cercanos a cero (veáse Fig. 3.12).

$$y = f_{tanh}(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^{-2z}}{1 + e^{-2z}} \quad (3.22)$$

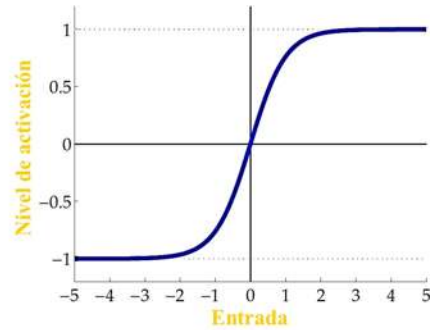


Figura 3.12: Función de activación sigmoide bipolar: Tangente hiperbólica.

- *Función gudermanniana*.- Relaciona las funciones trigonométricas, también conocidas como funciones circulares, con las funciones hiperbólicas (Ec. (3.88)) sin recurrir a los números complejos. Tal como se ve en la Fig. 3.13, toma valores de  $-\frac{1}{2}$  a  $\frac{1}{2}$ .

$$y = f_{gd}(z) = gd(z) = \int_0^z \frac{1}{\cosh t} dt \quad (3.23)$$

Su cálculo involucra varias funciones trascendentales, por lo que no tiene demasiado sentido emplearla en redes neuronales artificiales cuando ya existen otras funciones con mismas características y propiedades que las hacen adecuadas para conseguir implementaciones altamente eficientes.

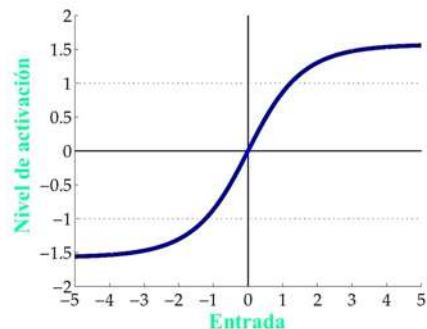


Figura 3.13: La función gudermanniana: Otra función de activación sigmoide bipolar con rango de  $-\frac{1}{2}$  a  $\frac{1}{2}$ .

7. *Función de activación lineal rectificadora* [111].- función que no requiera la utilización de funciones trascendentales. Por lo que resulta extremadamente eficiente. Es precisamente lo que hace las unidades lineales rectificadas o ReLU (Rectified Linear Units). Estas unidades utilizan una función de activación lineal rectificadora, que se define en la Ec. (3.24).

$$y = f_{relu}(z) = \begin{cases} z & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases} \quad (3.24)$$

Igual que sucede con otras funciones de activación, existen formas alternativas de definir la función de activación lineal rectificadora (véase Ec. (3.25) y Fig. 3.14):

$$y = f_{relu}(z) = z^+ = \max(0, z) \quad (3.25)$$

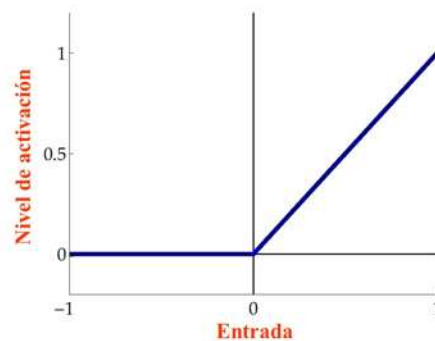


Figura 3.14: Función de activación lineal rectificada, utilizada por las unidades ReLU, muy habituales en AP.

8. *Funciones de base radial* [25].- miden la similitud entre el vector de entradas y el vector de pesos es calcular su distancia euclídea, también conocida como norma  $L^2$  de la diferencia entre ambos vectores. Es lo que hacen, por ejemplo, las funciones de base radial de tipo gaussiano (véase Ec. (3.26)).

$$y = f_{rbf}(w, x) = \phi(\|w - x\|) = e^{-\frac{1}{\sigma^2} \|w - x\|^2} \quad (3.26)$$

La salida de esa función es cero para la mayor parte de los valores de  $x$ , por lo que una red con elementos de este tipo puede ser difícil de entrenar utilizando técnicas basadas en el gradiente descendente (véase Fig. 3.15).

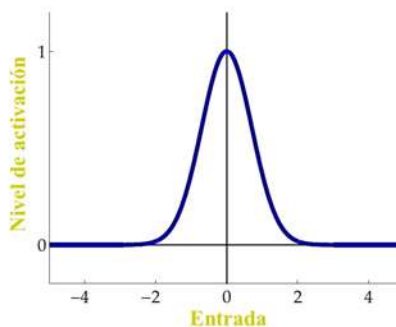


Figura 3.15: Función de activación RBF de tipo gaussiano.

### 3.6.4. Diseño de redes neuronales artificiales

Una vez estudiado algunos de los modelos y funciones de activación más utilizados en el aprendizaje profundo, es posible combinar colecciones de neuronas para resolver problemas. Algunas de ellas se muestran a continuación.

- *Redes feed-forward* [25].- carecen de mecanismo alguno de re-alimentación, son útiles ya que la terminología es lo habitual en los sistemas de control. Además, se les denominan redes recurrentes si incluye algún tipo de re-alimentación de las salidas a las entradas. Finalmente, este tipo de red se puede encontrar con varias situaciones dependiendo del número de capas ocultas (estás son todas las capas intermedias diferentes a las capas de entrada y de salida, por no ser observables directamente desde el exterior). Algunas redes multicapa son [25]:
  - \* *Redes simples, con una única capa.*- es el caso más simple de red neuronal. Las neuronas de la capa de entrada, que se limitan a recibir las señales de entrada provenientes del exterior, redistribuyen esas entradas a las neuronas de la capa de salida, que es la única capa de la red que realmente hace algo. De ahí que, aunque se vean dos capas, una de entrada y una de salida, este tipo de redes se considere unicapa (véase Fig. 3.16).

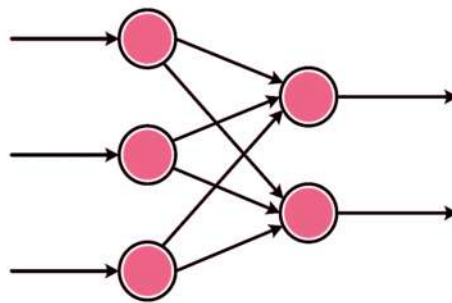


Figura 3.16: Red neuronal simple, sólo tiene una capa de neuronas de entrada y una capa de neuronas de salida.

- \* *Redes multicapa, con una capa oculta.*- son redes que incluyen aparte de la capa de entrada como la capa de salida nuevas capas intermedias, estas capas ya no serán visibles desde el exterior, lo que obligará a utilizar algoritmos como backpropagation para ajustar sus parámetros internos (véase Fig. 3.17).

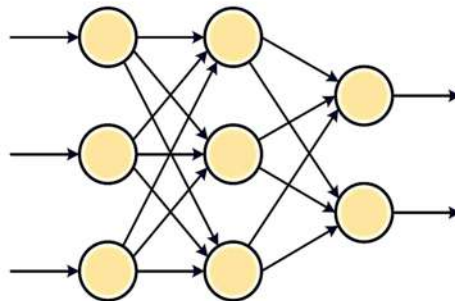


Figura 3.17: Red neuronal multicapa, con una única capa oculta.

La presencia de una única capa oculta ya dota a la red multicapa de su capacidad de aproximador universal (informalmente, su capacidad de aprender cualquier cosa que se pueda aprender). Tipo de red que fue muy popular desde los años 80 hasta finales del siglo XX.

- \* *Redes profundas (deep networks), con varias capas ocultas.*- las redes neuronales actuales suelen incluir múltiples capas ocultas. En todas las redes, las conexiones pueden ser excitatorias o inhibitorias, dependiendo del signo del peso asociado a ellas. Las funciones de activación bipolares, como la tangente hiperbólica, pueden resultar útiles en situaciones en las que un cambio de signo en la salida de la neurona influya en el comportamiento de las neuronas de las siguientes capas (véase Fig. 3.18).

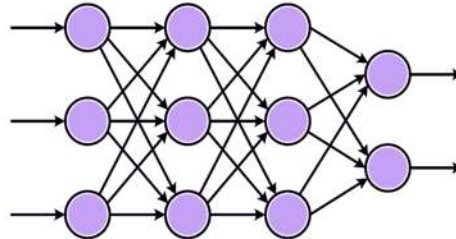


Figura 3.18: Red neuronal profunda, con múltiples capas ocultas sólo dos.

- \* *Redes competitivas.*- presenta conexiones inhibitorias entre las neuronas de una misma capa. Esas conexiones inhibitorias hacen que, cuando una neurona se active, ejerza una influencia inhibitoria sobre todas las demás. Son útiles para modelar situaciones en las que una capa deba responder en función del contexto, como sucede en algunos modelos de memorias asociativas (véase Fig. 3.19).

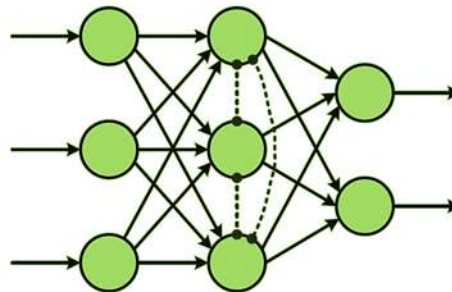


Figura 3.19: Red neuronal competitiva, con conexiones inhibitorias entre neuronas de la misma capa.

- \* *Redes recurrentes.*- carecen por completo de memoria. Cuando se añaden conexiones inhibitorias entre neuronas de una misma capa, se consiguen redes cuyo funcionamiento depende del contexto. Si las conexiones entre neuronas de la misma capa se generalizan (o se admiten conexiones de salida de una capa como entradas de una capa anterior), se obtienen redes neuronales con memoria: las redes recurrentes (véase Fig. 3.20).

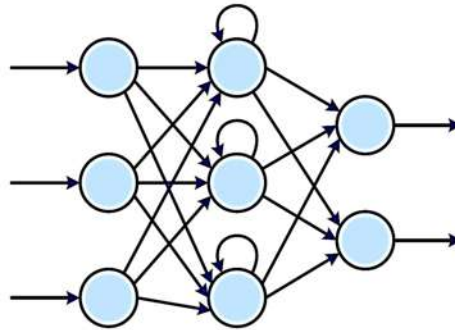


Figura 3.20: Red neuronal recurrente.

### 3.6.5. Entrenamiento de una red neuronal

Una red neuronal multicapa es un aproximador universal, capaz de aprender cualquier cosa, siempre que la red sea lo suficientemente grande como para representar las características de la función que se pretende aprender. Además, si se dispone de un número suficiente de ejemplos de entrenamiento, una red multicapa de la capacidad suficiente será capaz de construir un modelo de los datos con los que es entrenada. En la práctica lo máximo que se pretende es descubrir una serie de criterios heurísticos que ayuden a tomar decisiones con respecto al diseño de una red neuronal y de su proceso de entrenamiento.

A la hora de entrenar una red neuronal multicapa, se debe tener en cuenta los hiperparámetros (como la toma de decisiones con respecto tanto a los parámetros de diseño de la red como a los del algoritmo de entrenamiento) para distinguirlos de los propios parámetros de la red, los pesos de sus sinapsis y los sesgos de sus neuronas que se ajustan durante el proceso de entrenamiento de la red. Así pues, Berzal [25] menciona que para el desarrollo de un proyecto de aprendizaje automático en general, o de aprendizaje profundo en particular, la estrategia recomendada suele consistir en:

1. Establecer cuáles son las métricas mediante las que se evaluará el rendimiento del sistema. La más habitual es la precisión, puede ser de interés fijar un mínimo aceptable en cuanto a precisión y evaluar el progreso en función del grado de cobertura del sistema; esto es, el porcentaje de casos que se resuelven de forma automática, sin requerir la intervención manual de un operador humano. También es posible garantizar unos mínimos con respecto a la escalabilidad, con el objetivo de dar soporte a una carga de trabajo determinada.
2. Poner en marcha un prototipo funcional completo del sistema lo antes posible. Que abarque desde la adquisición de datos de entrenamiento hasta el uso del sistema en su entorno de producción. Además, debe incluir los mecanismos de instrumentación adecuados que faciliten la monitorización del sistema y su diagnóstico para corregir errores e introducir mejoras.
3. Realizar cambios graduales en el sistema, sobre la marcha, en función de los análisis que se realicen para determinar dónde se encuentran los principales problemas del sistema y cuáles son sus áreas potenciales de mejora.

Una vez planteada la estrategia general para resolver problemas de aprendizaje automático, se debe analizar a detalle los distintos aspectos del entrenamiento de una red neuronal sobre los que se puede tomar decisiones de diseño: su topología (capas y conexiones entre capas), sus elementos de procesamiento (p.ej. las funciones de activación de sus neuronas), el modo de entrenamiento de la red (por

lotes, online o por minilotes), el preprocesamiento de los datos de entrenamiento, la inicialización de los parámetros de la red y el ajuste de las tasas de aprendizaje [25].

- \* *Topología de la red.*- Existe una gran variedad de bloques que se pueden utilizar en su construcción, por ejemplo las capas *softmax* utilizadas para problemas de clasificación, *capas convolutivas* utilizadas en el procesamiento de imágenes, *redes recurrentes* que doten de memoria a una red neuronal, construyendo redes LSTM [Long Short-Term Memory]. Una vez establecida la construcción general de la red, el diseñador debe decidir del abanico de posibilidades el número de capas de la red, sus patrones de interconexión y el tamaño de cada una de las capas (esto es, el número de neuronas que forman parte de cada capa), lo que proporciona al diseñador de redes neuronales múltiples grados de libertad a la hora de abordar un nuevo problema y una flexibilidad de la que carecen otras técnicas de Inteligencia Artificial.
- \* *Funciones de activación.*- En el entrenamiento de una red neuronal se utiliza el gradiente del error y en el cálculo de dicho gradiente interviene la derivada de la función de activación, se suelen emplear funciones que sean derivables. Sin embargo, no es necesario que las funciones sean estrictamente derivables en todos los puntos, basta con que las funciones tengan definidas sus derivadas por la izquierda y por la derecha, aunque en algunos puntos no coincidan. Las implementaciones en software, simplemente, devuelven una de sus derivadas laterales y la optimización basada en el gradiente ya está sujeta a errores numéricos en su implementación digital, por lo que la no diferenciabilidad puntual de la función de activación no suele suponer problema alguno en la práctica.
- \* *Modos de entrenamiento.*- La esencia del entrenamiento de redes neuronales consiste en utilizar una señal de error, analizar cómo fluctúa ese error en función de los parámetros de la red y ajustar dichos parámetros con la intención de reducir el error observado, en dirección del gradiente descendente. Una de las decisiones que se han de tomar a la hora de entrenar una red neuronal es con qué frecuencia se ajustaran los pesos durante el entrenamiento de la red o, dicho de otro modo, cómo se estimará el error de la red para ajustar sus parámetros. Se tienen tres opciones:
  - *Entrenamiento por lotes:* se ajustan los pesos una vez por cada época. Se recorre el conjunto de entrenamiento completo, acumulando el error para cada uno de los ejemplos de entrenamiento y, al final del recorrido, se realiza una actualización de los pesos de la red.
  - *Entrenamiento online:* se ajustan los pesos de la red cada vez que le mostramos a la red un ejemplo de entrenamiento.
  - *Entrenamiento por minilotes:* se ajustan los pesos a partir del error estimado para una pequeña muestra del conjunto de entrenamiento.
- \* *El conjunto de entrenamiento.*- Como cualquier otra técnica de aprendizaje automático, el rendimiento de una red neuronal depende de forma crítica del conjunto de datos de entrenamiento que se haya utilizado para ajustar sus parámetros. Es fundamental que el conjunto de entrenamiento sea completo y se parezca a los datos con los que realmente se encontrará la red en la práctica, una vez entrenada y puesta en marcha en un entorno real. Las diferencias que se pueden conseguir ajustando los múltiples parámetros de un algoritmo de aprendizaje automático no son significativas en comparación con las mejoras que se pueden llegar a lograr si se invierte algo más de tiempo y esfuerzo en la recopilación y preparación de un conjunto de datos de entrenamiento adecuado para el problema que se pretende resolver.

- \* *Tasas de aprendizaje.*- Es el parámetro clave cuyo ajuste puede ayudar a conseguir un funcionamiento adecuado del proceso de entrenamiento de la red, ya que determina cuánto se ajustan los pesos en cada iteración del algoritmo. Una vez inicializados correctamente los parámetros de la red neuronal artificial, se tiene que ajustarlos utilizando un algoritmo de entrenamiento. Si se utiliza el gradiente descendente, la actualización de los pesos de la red se hará de acuerdo a una expresión de la forma (véase la Ec. (3.27))

$$\Delta w = -\eta \Delta E \quad (3.27)$$

Esta expresión se acaba traduciendo en la siguiente fórmula de actualización de los pesos de la red (Ec. (3.28))

$$\Delta w_{ij} = -\eta x_i \delta_j \quad (3.28)$$

donde  $\delta_j$  representa la derivada parcial del error con respecto a la entrada neta  $z_j$  de la neurona  $j$  de una capa con entradas  $x_i$ . La magnitud del ajuste de los pesos dependerá, pues, de la magnitud del gradiente de la señal de error, de las entradas recibidas y de un parámetro del algoritmo,  $\eta$ , su tasa de aprendizaje.

La regla de ajuste de los pesos se basa en la estimación de un gradiente, al ser una estimación numérica es inestable, y los errores de aproximación se van multiplicando, lo que puede ocasionar problemas de convergencia en el algoritmo de optimización de los parámetros de la red, basado en el gradiente descendente. En una red multicapa, los gradientes del error que se propagan pueden crecer o desvanecerse relativamente rápido, lo que dificulta el proceso de optimización de los parámetros de la red. Para muchos, se trata del hiperparámetro más importante del algoritmo de entrenamiento de una red. Como criterio heurístico general, se busca elegir un valor para la tasa de aprendizaje  $\eta$  que haga que, en cada iteración, los pesos cambien en una fracción de su valor actual. Suele ser útil comparar la magnitud de los gradientes del error con la magnitud de los parámetros que se pretenden ajustar. Normalmente, se busca que las actualizaciones de los parámetros correspondiesen a un cambio más cerca del 1% del valor del parámetro que al 50% o al 0.001% [112], véase la Fig. 3.21.

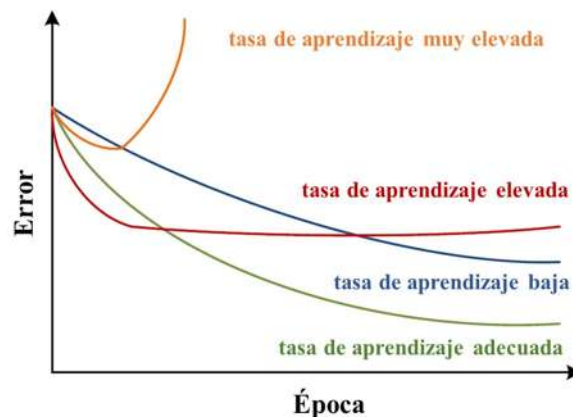


Figura 3.21: Representación gráfica del efecto de la tasa de aprendizaje sobre la convergencia del algoritmo de entrenamiento de una red neuronal multicapa.

### 3.6.6. Sobreajuste en una red neuronal

El **sobreajuste** es un problema común en el aprendizaje automático, donde un modelo funciona bien en el entrenamiento de datos pero no generaliza bien con los datos no vistos (datos de prueba) [107]. Este problema se acentúa en el caso de que los datos tengan ruido o errores. Por sencillo que sea un modelo de aprendizaje automático, el temido sobreajuste siempre realiza su aparición, porque el número de hipótesis crece exponencialmente con el número de atributos. Por tanto, el número de conceptos que un modelo podría aprender es una función exponencial de un número exponencial de patrones que se pueden obtener de un número finito de atributos.

Formalmente, un modelo de red con mayor número de conexiones (es decir, con mayor número de neuronas) puede modelar datos más complejos, por lo que, también es capaz de aprender detalles de los datos de entrenamiento que no son realmente relevantes para la tarea para la que se construyó. Además, de contener ruido, que no será útil en ninguna situación: regularidades accidentales debidas al conjunto particular de datos (lo que los estadísticos denominarían error de muestreo).

Desgraciadamente, cuando se ajustan los parámetros de una red neuronal a los datos del conjunto de entrenamiento, no es posible diferenciar las regularidades realmente útiles de las irrelevantes o de las debidas al muestreo del conjunto de entrenamiento, por lo que siempre se está expuesto al riesgo de sobreajuste (*overfitting*). Cuando sobreajuste, un modelo se ajusta demasiado bien al conjunto de entrenamiento, modelando sus peculiaridades al detalle y su ruido. Como consecuencia, no generaliza bien: su rendimiento se degrada cuando lo utilizamos sobre casos distintos a los del conjunto de entrenamiento, que es realmente donde es interesante que funcione bien. Una señal inequívoca de que un modelo sobreajuste es que, dotado de la capacidad suficiente, su error sobre el conjunto de entrenamiento es mucho menor que su error sobre el conjunto de prueba.

Si un modelo sufre una situación de sobreajuste, el modelo tiene una alta varianza, causada quizás por tener demasiados parámetros que conducen a un modelo demasiado complejo dados los datos de **subajuste** o *underfitting* (un sesgo elevado), que significa que el modelo no es suficientemente complejo como para capturar correctamente el patrón en los datos de entrenamiento y, por lo tanto, sufre por el bajo rendimiento en los datos no vistos. El problema de sobreajuste y el subajuste se puede ilustrar mejor mediante la comparación de un límite de decisión lineal con otros límites de decisión no lineales y más complejos [107], como se muestra en la Fig. 3.22.

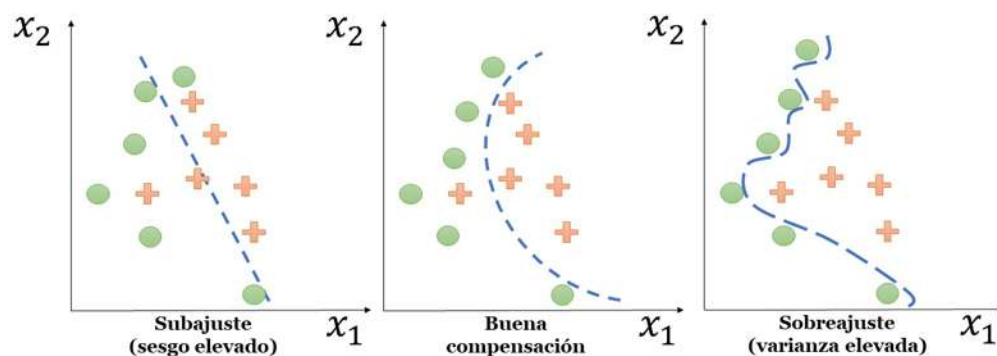


Figura 3.22: Representación gráfica del problema de sobreajuste y el subajuste encontrado en modelos lineales.

Se presentan tres estrategias, para evitar el sobreaprendizaje:

- *Obtener más datos.*- La mejor manera es usar una gran cantidad de información en el aprendizaje [113]. Si se dispone de la capacidad de cálculo necesaria para entrenar la red utilizando un conjunto de datos de entrenamiento más grande. El objetivo es hacer a la red robusta frente a transformaciones habituales en los datos de entrada y también para posibles ataques usando técnicas de entrenamiento con adversario.

El sobreaprendizaje es habitual si los datos incluyen muchas características en proporción al número de observaciones disponible, donde los conjuntos de datos se caracterizan por ser más anchos que altos y no siempre poder obtener más datos con facilidad. A menudo se reduce la dimensionalidad de los datos seleccionando o extrayendo características, aunque entonces se corre el riesgo de estar desperdiciando información potencialmente útil [25].

Requena et. al. [113] menciona que si el número de casos en el entrenamiento es al menos 30 veces superior al número de pesos de la red la probabilidad de que se produzca sobreaprendizaje es mínima. Si la información en el aprendizaje está libre de ruido, es suficiente con que el número de casos sea 5 veces superior al número de pesos.

- *Ajustar la capacidad del modelo.*- Ajustar adecuadamente los parámetros de la red también puede servir para que la red tenga la capacidad adecuada. Al reducir el tamaño de la red a su menor número de parámetros, de modo que la red neuronal es menos flexible y más robusta frente al ruido podría ser suficiente para identificar las regularidades relevantes en los datos de entrenamiento, pero no demasiada para que también se ajuste a las espurias. Obviamente, suponiendo que las espurias serán más débiles que las auténticas, ya que de otro modo el aprendizaje resultaría por completo imposible.

Es, por tanto, conveniente seleccionar de entre todas las redes que hayan sido capaces de adaptarse a los datos de entrenamiento aquella que, produciendo un menor error en el aprendizaje, contiene un menor número de neuronas en su capa intermedia [113].

- *Combinar múltiples modelos.*- Una tercera estrategia para evitar el sobreaprendizaje consiste en asumir que resulta inevitable y combinar múltiples modelos con la intención de que los errores cometidos por algunos de ellos se compensen con los aciertos de los demás. Suele ser apropiado cuando el número de datos de los que se dispone es demasiado pequeño [113]. En las redes neuronales, esta combinación se puede realizar de dos formas diferentes:
  - Promediando múltiples modelos: Se construyen muchos modelos diferentes con diferentes parámetros o el mismo tipo de modelo utilizando distintos subconjuntos del conjunto de entrenamiento. Esto es, se crea un ensamble, tal como se haría con cualquier otra técnica de aprendizaje automático.
  - Aplicando un enfoque bayesiano, se selecciona una única arquitectura de red. A continuación, se combinan las predicciones realizadas por muchos vectores de pesos diferentes, que se obtienen tras repetir el entrenamiento de la red en múltiples ocasiones. Esto es, en vez de entrenar la red con distintos subconjuntos de datos de entrenamiento, aprovechando la naturaleza estocástica del algoritmo de entrenamiento de redes neuronales para entrenar múltiples versiones de una misma red neuronal.

### 3.6.7. Datos de entrenamiento, validación y prueba

Para la configuración y evaluación de un modelo en Machine Learning, habitualmente se dividen los datos disponibles en dos conjuntos: datos de entrenamiento (*training*) y datos de prueba (*test*). A su

vez, una parte de los datos de entrenamiento (*training*) se reservan como datos de validación (*validation*). Los datos de entrenamiento que nos quedan después de sacar los de validación y prueba son los que se usan para que el algoritmo de aprendizaje calcule los parámetros del modelo, mientras que los de validación se usan para afinar los hiperparámetros. Con las métricas -como con la precisión (*accuracy*)- que se pueden obtener de este conjunto de datos de validación para decidir cómo ajustar los hiperparámetros del algoritmo antes de repetir el proceso de entrenamiento [98].

Es importante notar que al esforzarse por mejorar el algoritmo ajustando los hiperparámetros gracias al comportamiento del modelo con los datos de validación, se está incidiendo en el modelo, que puede sesgar los resultados a favor del conjunto de validación. De aquí la importancia de disponer de un conjunto de datos de prueba reservados para una prueba final, con datos que el modelo no ha visto nunca anteriormente durante la etapa de entrenamiento (ni como datos de entrenamiento ni como datos de validación). Esto permite obtener una medida de comportamiento del algoritmo más objetiva y evaluar si el modelo generaliza correctamente [98].

La Fig. 3.23 esquematiza esta división y el propósito del uso de cada subconjunto de datos.

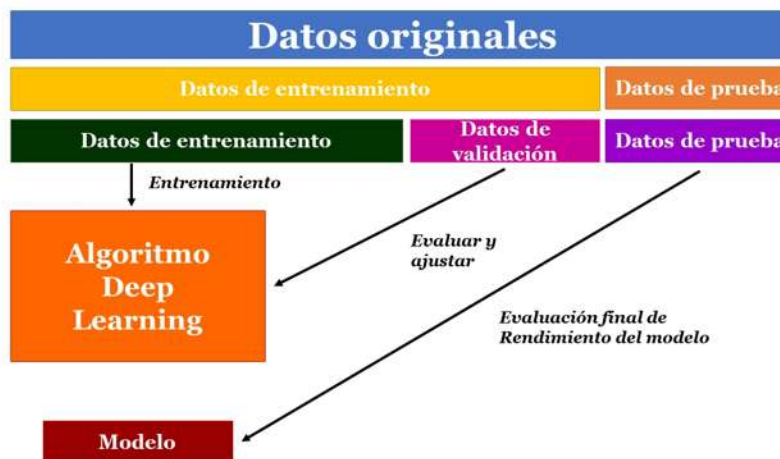


Figura 3.23: Repartición de los datos disponibles en tres grupos (entrenamiento, validación y prueba) y el propósito de cada uno de ellos.

Esta es la estrategia de evaluación más habitual. Pero hay otras estrategias de repartir los datos para validar los modelos cuando tenemos escasez de datos, como por ejemplo la validación cruzada (*cross-validation*) que, básicamente, divide los datos en  $k$  particiones del mismo tamaño y para cada partición  $i$  el modelo es entrenado con las restantes  $k - 1$  particiones, y evaluando en la propia partición  $i$  [98].

### 3.6.8. Ventajas y limitaciones del aprendizaje profundo

Algunas ventajas que se pueden enunciar son:

1. Robustez frente al ruido en datos de entrada y salida.- Ya que el procesado de la información es local, es decir que al estar compuesto por unidades individuales de procesamiento dependiendo de sus entradas y pesos y de que todas las neuronas de una capa trabajan en forma paralela y proporcionan una respuesta al mismo tiempo [114, 115].

2. Independencia entre complejidad del problema y dimensionamiento de la red.- Es decir, pueden reconocer patrones que no han sido aprendidos, sólo deben tener cierto parecido con el conocimiento previo que tenga la red [116, 117].
3. Rapidez de ajuste y simulación.- Debido a que los pesos son ajustados basándose en la experiencia, lo que significa que se le tiene que enseñar a la red lo que necesita saber antes de ponerla en funcionamiento [24].
4. Tolerantes a fallos.- Dado que si parte de la red no trabaja, solo dejará de funcionar la parte para que dicha neurona sea significativa; el resto tendrá su comportamiento normal [27, 118].

Algunos aspectos considerados como limitaciones de las redes neuronales o más problemáticas que existe en el aprendizaje profundo se detallan a continuación.

1. Complejidad en el aprendizaje.- Esto para grandes tareas, cuanto más cosas se necesiten que aprenda una red, mas complicado será enseñarle [119].
2. Tiempo de aprendizaje elevado.- Esto depende de dos factores: primero si se incrementa la cantidad de patrones a identificar o clasificar y segundo si se requiere mayor flexibilidad o capacidad de adaptación de la red neuronal para reconocer patrones que sean sumamente parecidos, se deberá invertir mas tiempo en lograr que la red converja a valores de pesos que representen lo que se quiera enseñar [114, 119].
3. Sobreaprendizaje.- Ya que, para evitarlo se requiere de una elevada cantidad de datos para el entrenamiento, cuanto mas flexible se requiera que sea la red neuronal, mas información tendrá que enseñarle para que realice de forma adecuada la identificación. Mientras que el conjunto de entrenamiento contiene dos tipos de ruido: *errores en los datos* y *errores de muestreo*. Los errores en los datos son inevitables y provienen del proceso de adquisición de datos, especialmente si éste incluye algún tipo de procesamiento manual [25].
4. Vistas como cajas negras.- En una red neuronal, podemos observar la entrada de la red y ver cuál es su salida, pero su funcionamiento interno es algo que no se puede describir de forma simbólica. Puesto que, no permite interpretar lo que se ha aprendido, la red por si sola proporciona una salida, un número, que no puede ser interpretado por ella misma, sino que se requiere de la intervención del programador y de la aplicación en si para encontrarle un significado a la salida proporcionada. En situaciones en la que no sólo sea importante la validez del modelo, sino su credibilidad, puede que el carácter opaco de una red neuronal juegue en contra. La validez de un modelo con redes neuronales, es algo que no se puede avalar con argumentos meramente cuantitativos. Aunque funcionen realmente bien, si son incapaces de ofrecer una explicación, puede que no sea permitido usarlas [120, 121].

### 3.7. Cascada Multiplicativa

Las cascadas multiplicativas fueron originalmente introducidas en hidrodinámica como modelos de turbulencia, y luego fueron reintroducidas por Mandelbrot y unificadas en el paradigma de las martingalas multiplicativas [122]. Una cascada es un proceso que divide un conjunto dado en subconjuntos cada vez más y más pequeños de acuerdo con alguna regla geométrica, preservando alguna característica de la medida del conjunto inicial [123].

La cascada multiplicativa también es definida como un proceso iterativo que fragmenta un conjunto dado en piezas cada vez más pequeñas por medio de una regla geométrica ya establecida y un punto interesante de estos subconjuntos, es que siguen conservando las propiedades del conjunto principal y el proceso multiplicativo se caracteriza por ser la primera descomposición del intervalo unitario. Estas pueden ser aleatorias o con pesos estadísticos ya establecidos [124–126].

### 3.7.1. Construcción de Multifractales Multiplicativos

Para apreciar mejor las reglas de construcción, es importante señalar que estas reglas esencialmente involucran particiones diádicas. Considere un intervalo de unidad. Asociarlo con una unidad de masa. Divida el intervalo de la unidad en dos, digamos, segmentos izquierdo y derecho de igual longitud. Además, divida la masa asociada en dos fracciones,  $r$  y  $1-r$ , y se asigna a los segmentos izquierdo y derecho, respectivamente. El parámetro  $r$  es en general una variable aleatoria, gobernada por el conjunto PDF  $(P(r), 0 \leq r \leq 1)$ . La fracción  $r$  se llama multiplicador. Cada nuevo subintervalo y su peso asociado se dividen en dos partes siguiendo la misma regla. Este procedimiento se muestra esquemáticamente en la Fig. 3.24, donde el multiplicador  $r$  se escribe como  $r_{ij}$ ,  $i$  indicando el número de etapa y  $j$  (suponiendo solo números impares, dejando números pares para  $1-r_{ij}$ ) indicando las posiciones de un peso en ese escenario. Tener en cuenta que la escala (es decir, la longitud del intervalo) asociada con la etapa  $i$  es  $2^{-i}$ . Suponiendo que  $P(r)$  es simétrico respecto de  $r = 1/2$  y tiene momentos sucesivos  $\mu_1, \mu_2, \dots$ . Por lo tanto,  $r_{ij}$  y  $1-r_{ij}$  tienen una distribución marginal  $P(r)$ . Los pesos en la etapa  $N$ ,  $w_n, n = 1, \dots, 2^N$ , se pueden expresar como:

$$w_n = u_1 u_2 \cdots u_N \tag{3.29}$$

donde  $u_l, l = 1, \dots, N$ , son  $r_{ij}$  o  $1 - r_{ij}$ . Por lo tanto,  $\{u_i, i \geq l\}$  son variables aleatorias independientes e idénticamente distribuidas, iid, que tienen PDF  $P(r)$ .

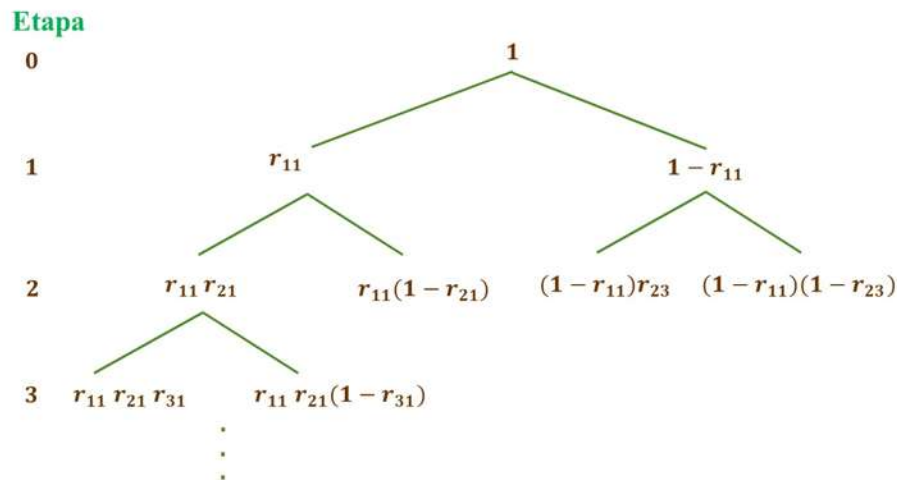


Figura 3.24: Esquema que ilustra la regla de construcción de un multifractal multiplicativo.

Las reglas descritas anteriormente generan un proceso conservador. Para generar un proceso no conservador, simplemente puede requerir que una masa  $m_i$  en la etapa  $i$  se divida en dos partes,  $W_1 m_i$  y  $W_2 m_i$ , donde  $W_i, i = 1, 2$  son independientes, siguen el mismo PDF y ambas tienen media  $1/2$ .

Como este último no garantiza que  $W_1 + W_2 = 1$ , la suma de los pesos en la etapa  $n$  no es necesariamente igual a 1 [29].

A continuación, la construcción de las diferentes Cascadas Multiplicativas así como algunos ejemplos. Finalmente, la propuesta para nuestro trabajo.

### 3.7.2. Cascada Multiplicativa Determinística

Es posible generar una familia de medidas singulares con propiedades de multifractalidad si se utiliza un modelo genérico denominado cascada multiplicativa con reglas deterministas en la secuencia de pasos a efectuar [41]. Considérese, un intervalo  $I$ ,  $\mu(I) = N_0$ . Supongamos que durante la primera etapa,  $I(0)$ , llegan  $\mu(I(0)) = p \cdot N_0$  y durante la segunda etapa,  $I(1)$ , llegaron  $\mu(I(1)) = (1-p) \cdot N_0$ , donde  $p \in (0, 1)$ , de manera que se conserva el número original en el intervalo original,  $I$ . Ahora si se divide cada subintervalo de media,  $I(0)$  e  $I(1)$ , en dos subintervalos de un cuarto,  $[I(0,0), I(0,1)]$  y  $[I(1,0), I(1,1)]$ , y se asigna la misma fracción de etapas a cada subintervalo:  $\mu(I(0,0)) = p^2 \cdot N_0$ ,  $\mu(I(0,1)) = (1-p) \cdot p \cdot N_0$ ,  $\mu(I(1,0)) = p \cdot (1-p) \cdot N_0$ ,  $\mu(I(1,1)) = (1-p)^2 \cdot N_0$ . Iterando este proceso, en la etapa  $l$  se encontrará que el número que llegaron en el intervalo  $I(j_1, j_2, \dots, j_l)$  es  $\mu(j_1, j_2, \dots, j_l) = N_0 \prod_{k=1}^l p_{j_k} = N_0 p^n (1-p)^{l-n}$  donde  $j_k \in (0, 1)$ ,  $p_0 = p$ ,  $p_1 = (1-p)$ ,  $n$  es el número de ceros que hay en la secuencia  $(j_1, j_2, \dots, j_l)$ , y la longitud del intervalo es  $2^{-l}$  horas. Así se ha construido una cascada determinística que preserva el número total que llegan durante el tiempo observado, repartándose en subintervalos diádicos [123].

Un ejemplo general de un modelo multiplicativo que parte de una medida  $\mu$  uniformemente distribuida sobre el soporte, considerado habitualmente como el intervalo  $[0, 1]$ . La medida del soporte completo  $\mu([0, 1]) = 1$  se toma inicialmente normalizada. En la primera etapa de la cascada,  $n = 1$ , se parte el soporte en dos intervalos iguales y se reparte la medida entre ellos con arreglo a una ley determinista, por ejemplo al trozo izquierdo se le da una fracción  $p$  de la medida inicial y por tanto su medida será ahora  $\mu_0 = p$ . Al trozo derecho se le da la fracción restante por lo que su medida será  $\mu_1 = 1 - p$ . El proceso se repite dividiendo cada subintervalo en otros dos, repartiendo la medida con la misma regla determinista. Se obtienen así segmentos cada vez más pequeños que contienen menos cantidad de la medida total. Para la tercera generación, por ejemplo, tendríamos un número  $N(\delta = 2^{-3}) = 2^3$  intervalos de tamaño  $\delta = 2^{-3}$  en los que la medida valdría:

$$\{\mu\}_{i=0}^{2^3-1} = \{\mu_0\mu_0\mu_0, \mu_0\mu_0\mu_1, \mu_0\mu_1\mu_0, \mu_0\mu_1\mu_1, \mu_1\mu_0\mu_0, \mu_1\mu_0\mu_1, \mu_1\mu_1\mu_0, \mu_1\mu_1\mu_1\} \quad (3.30)$$

La medida obtenida por este procedimiento, llamado modelo  $p$  determinista o modelo binomial, posee un espectro de singularidades que puede estimarse si se considera que en la etapa  $n$ ésima hay  $C_n^m$  intervalos de longitud  $2^{-n}$  para los que la medida vale  $\mu = \mu_0^{n-m} \mu_1^m$ , donde  $0 \leq m \leq n$ . Puede probarse entonces que:

$$\alpha = \frac{In\mu_0 + (n/m - 1)In\mu_1}{(n/m)In(1/2)} \quad (3.31)$$

$$f(n/m) = \frac{(n/m - 1)In(n/m - 1) - (n/m)In(n/m)}{(n/m)In(1/2)} \quad (3.32)$$

Eliminando  $n/m$  entre las dos expresiones se obtiene inmediatamente la  $f(\alpha)$ . Puede comprobarse que cualesquiera que sean los valores de  $p$  y  $1 - p$  la función  $f$  toma su valor máximo para  $n/m = 2$  el cual es  $f = 1$ . El máximo de la curva  $f(\alpha)$  corresponde a la dimensión fractal del soporte de la medida [127–129]. Este modelo simple, con reglas deterministas, puede servir de base a modelizaciones

mucho más sofisticadas, si el reparto de medida en cada etapa se realiza mediante un procedimiento aleatorio. Las medidas así obtenidas gozarán de propiedades multifractales en un sentido estadístico [41].

### 3.7.3. Cascada Multiplicativa Conservadora Binomial

La cascada binomial es el método más simple para obtener un proceso multifractal, que consiste en un procedimiento iterativo en el intervalo compacto  $[0,1]$ . Se supone que la medida inicial se conserva en todas las etapas. Sin pérdida de generalidad, se asume la unidad de medida inicial. Sea  $m_0 = r$  y  $m_1 = 1 - r$ , dos multiplicadores para la generación en cascada, posiblemente con  $r$  aleatorio. En el escenario  $n = 0$  de la cascada iteración, tenemos la unidad de medida denotada por  $\mu_0$  distribuido uniformemente en el intervalo  $I_0 = [0, 1]$ . En el escenario  $n = 1$ , esta unidad de medida se divide en dos partes, con medidas  $m_0$  y  $m_1$ , y asignado a dos subintervalos,  $I_{00} = [0, 1/2]$  y  $I_{01} = [1/2, 1]$ , respectivamente. En la siguiente etapa ( $n = 2$ ), este proceso se repite tanto en el conjunto  $I_{00}$  como  $I_{01}$  para generar cuatro subintervalos aún más pequeños  $I_{000}, I_{001}, I_{010}$  y  $I_{011}$ , cada uno asignado respectivamente con las siguientes medidas:

$$\begin{aligned} \mu[0, 1/4] &= m_0 m_0 \\ \mu[1/4, 1/2] &= m_0 m_1 \\ \mu[1/2, 3/4] &= m_1 m_0 \\ \mu[3/4, 1] &= m_1 m_1 \end{aligned}$$

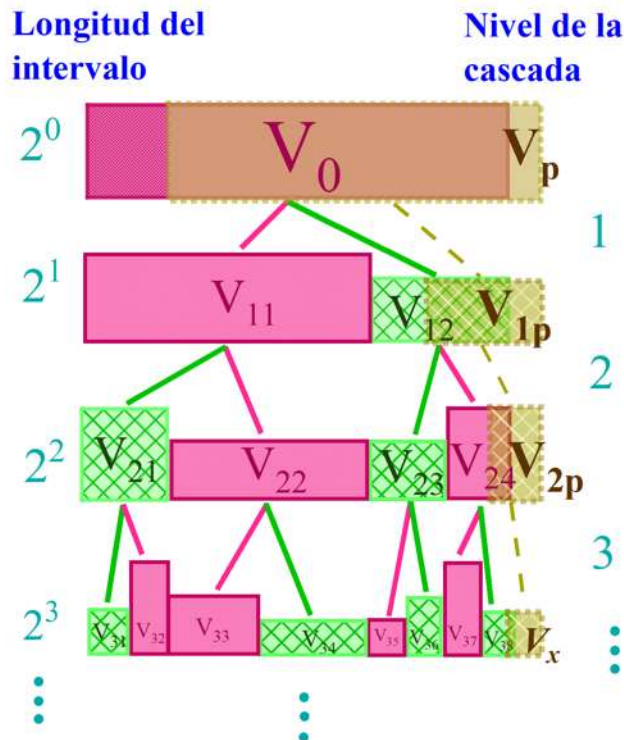


Figura 3.25: Construcción clásica en cascada binomial conservadora.

A través de este proceso iterativo, en la etapa  $k$ , hay un total de  $2^k$  subintervalos diádicos disjuntos de tipo  $[t, t + 2^{-k}]$  con sus medidas correspondientes  $\mu'$ s. Deje que  $\varphi_0$  y  $\varphi_1$  denote las frecuencias relativas de 0's y 1's, respectivamente, en el desarrollo en cascada. La medida en el intervalo diádico está dado por:

$$\mu[t, t + 2^{-k}] = \mu[\Delta_k] = m_0^{k\varphi_0} m_1^{k\varphi_1} \quad (3.33)$$

Para este procedimiento de construcción en cascada se conserva la masa total de cada intervalo diádico, lo que resulta en la cascada conservadora [125].

### Ejemplos de la construcción para el caso de una Cascada Multiplicativa Binaria

Se comienza ubicando una masa unitaria en el intervalo  $[0, 1]$ . A continuación, el intervalo se subdivide en los subintervalos  $[0, 1/2)$  y  $[1/2, 1]$ ; si la cascada fuera  $n$ -aria en lugar de binaria, la subdivisión se haría en  $n$  intervalos. A continuación, la masa de cada subintervalo es multiplicada por las variables aleatorias iid  $W_{1,1}$  y  $W_{1,2}$ , respectivamente. Luego, este proceso es repetido recursivamente en cada uno de los subintervalos.

Sean  $W_{j,k}$ ,  $k = 1, \dots, 2^j$ , los multiplicadores involucrados en la  $j$ -ésima iteración. Nótese también que cada uno de los subintervalos en la construcción binaria es un intervalo diádico a la escala  $j$  y posición  $k$ ,  $\lambda = \lambda_{j,k}$ ; dado que cada punto  $x \in [0, 1]$  está incluido en un único intervalo diádico de ancho  $2^j$ . Finalmente, sea  $J$  el número de iteraciones empleadas en el procedimiento constructivo. Entonces, la cascada canónica de Mandelbrot  $Q_J(x)$ , en el caso binario, se define como:

$$Q_J(x) = \prod_{j=1}^J W_{j,k} \quad (3.34)$$

Los multiplicadores  $W_{j,k}$  deben ser estrictamente positivos. Para garantizar la convergencia de la cascada, deben estar normalizados de forma tal que  $\mathbb{E}[W] = 1$ , el valor esperado de la variable aleatoria  $W$ . La función de escalamiento  $\zeta(q)$  está dada por [125, 126] :

$$\zeta(q) = -\log_2 \mathbb{E}[W^q] \quad (3.35)$$

Es decir, la distribución de los multiplicadores controla la función de escalamiento y, por lo tanto, las propiedades multifractales de la cascada. A continuación, se indican las propiedades multifractales de las CMC habitualmente utilizadas en la práctica [122].

### CCM log-normal

En este trabajo nos referiremos a ellas como cascadas canónicas de Mandelbrot (en inglés, Canonical Mandelbrot Cascades, CMC). Los multiplicadores se eligen como  $W = 2^{-X}$ , en donde  $X$  sigue una distribución normal con media  $\mu$  y varianza  $\sigma^2$ . Se utiliza esta definición no convencional en base 2, debido al logaritmo que debe ser tomado al aplicar la Ec. (3.35). La restricción  $\mathbb{E}[W] = 1$  hace que los parámetros  $\mu$  y  $\sigma^2$  no se puedan escoger independientemente, sino que deban satisfacer  $\sigma^2 = 2\mu/\log(2)$ . A partir de la Ec. (3.35), los atributos multifractales (ver más detalles en Anexo C) son [122]:

$$\zeta(q) = \mu q(1 - q), \quad (3.36)$$

$$D(h) = 1 - \frac{(h - \mu)^2}{4\mu}, \quad (3.37)$$

$$c_1 = \mu, \quad (3.38)$$

$$c_2 = -2\mu, \quad (3.39)$$

$$c_m = 0, \forall m \geq 3 \quad (3.40)$$

Nótese que esta cascada es un modelo sencillo en el que las propiedades multifractales están controladas por un único parámetro  $\mu$ , y en el que  $c_m = 0, \forall m \geq 3$ .

### CCM log-Poison

En este caso, los multiplicadores se eligen como  $W = 2^\gamma \exp(\log(\beta)K)$ , en donde  $K$  sigue una distribución de Poisson con parámetro  $\lambda$ . La restricción  $\mathbb{E}[W] = 1$  hace que el parámetro  $\lambda$  no se pueda escoger libremente, sino que tenga que satisfacer  $\lambda = -\gamma \log(2)/(\beta - 1)$ . A partir de la Ec. (3.35), los atributos multifractales son [122]:

$$\zeta(q) = -\gamma q + \frac{\gamma(\beta^q - 1)}{\beta - 1}, \quad (3.41)$$

$$D(h) = 1 + \frac{\gamma}{\beta - 1} + \frac{h + \gamma}{\log \beta} \left[ \log \left( \frac{(h + \gamma)(\beta - 1)}{\gamma \log \beta} \right) - 1 \right], \quad (3.42)$$

$$c_1 = -\gamma + \frac{\gamma \log \beta}{\beta - 1}, \quad (3.43)$$

$$c_m = \frac{\gamma (\log \beta)^m}{\beta - 1}, \forall m \geq 2 \quad (3.44)$$

#### 3.7.4. Cascada de Poisson compuestas

Las cascadas de Poisson compuestas (en inglés, compound Poisson cascades, CPC) son un caso particular de cascadas infinitamente divisibles. Fueron introducidas originalmente en [122], y luego estudiadas extensivamente en [130–134]. Su construcción es similar a las de CMC (en inglés, Canonical Mandelbrot Cascades), en el sentido en que se basan en la ubicación de multiplicadores aleatorios en el plano tiempo-escala, y en la posterior multiplicación de todos aquellos que estén en una vecindad del punto  $x$ . La diferencia radica en que en las CPC los puntos  $(x_i, a_i)$  en los que se ubican los multiplicadores son a su vez variables aleatorias provenientes de un proceso planar de Poisson con medida de control  $dm(x, a)$ . Esta geometría poissoniana permite la multiplicación continua y, por lo tanto, la obtención de leyes de escalamiento arbitrarias [122]. Los multiplicadores se escogen como variables aleatorias iid (variables aleatorias independientes e idénticamente distribuidas) y positivas, e independientes del proceso de Poisson. Entonces, la cascada  $Q(x)$  se genera mediante el producto de todos los multiplicadores incluidos en un *cono de influencia*  $C_a(x)$  del punto  $x$ :

$$Q_a(x) = \exp[(1 - \mathbb{E}[W])m(C_a(x))] \prod_{C_a(x)} W_i, \quad (3.45)$$

en donde

$$m(C_a(x)) = \int_{C_a(x)} dm(x, a) \quad (3.46)$$

El factor de normalización delante del producto en la Ec. (3.45) hace que el proceso tenga media unitaria [43]. El proceso  $Q_a(x)$  definido de esta forma es estacionario [132]. Si se define un cono de influencia triangular como [132, 133, 135]:

$$C_a(x) = \left\{ (x', a') : a' > a \wedge |x - x'| < \frac{a'}{2} \right\}, \quad (3.47)$$

y una medida de control separable como [132, 133, 135]:

$$dm(x, a) = \begin{cases} \frac{c}{a^2} da dx, & \text{si } 0 < a \leq 1 \\ 0, & \text{en otro caso} \end{cases} \quad (3.48)$$

se obtiene una CPC en la que el decaimiento de los momentos con las escalas toma la forma de una ley de potencias [132, 133, 135]:

$$\mathbb{E}[(Q_a(x))^q] \sim a^\zeta(q), \quad (3.49)$$

La función de escalamiento está dada por [43, 44, 111]:

$$\zeta(q) = c(1 - \mathbb{E}[W^q] - q(1 - \mathbb{E}[W])), \quad (3.50)$$

en donde  $c$  es una constante arbitraria (más detalles en Anexo C, *log-cumulantes*. Es decir, las propiedades de invarianza a la escala están controladas únicamente por la distribución de los multiplicadores, mientras que el proceso de Poisson está involucrado sólo en la determinación de la forma en que se produce el escalamiento.

### CPC log-normal

En este caso, los multiplicadores se eligen como  $W = e^X$ , en donde  $X$  sigue una distribución normal con media  $\mu$  y varianza  $\sigma^2$ . A partir de la Ec. (3.50), la función de escalamiento es:

$$\zeta(q) = c \left[ 1 - \exp\left(\mu q + \frac{\sigma^2}{2} q^2\right) - q + q \exp\left(\mu + \frac{\sigma^2}{2}\right) \right]. \quad (3.51)$$

Mediante la aplicación de la transformación de Legendre se obtiene el espectro multifractal, en forma paramétrica:

$$h(q) = c \left[ \exp\left(\mu + \frac{\sigma^2}{2}\right) - 1 - (\mu + \sigma^2 q) \exp\left(\mu q + \frac{\sigma^2}{2} q^2\right) \right], \quad (3.52)$$

$$D(q) = 1 - c + (1 - \mu q - \sigma^2 q) \exp\left(\mu q + \frac{\sigma^2}{2} q^2\right). \quad (3.53)$$

Los log-cumulantes son:

$$c_1 = -c \left[ \mu + 1 - \exp\left(\mu + \frac{\sigma^2}{2}\right) \right], \quad (3.54)$$

$$c_2 = -c(\sigma^2 + \mu^2), \quad (3.55)$$

$$c_3 = -c(3\mu\sigma^2 + \mu^3), \quad (3.56)$$

$$c_4 = -c \left( \frac{5}{2}\sigma^4 + 6\mu^2\sigma^2 + \mu^4 \right). \quad (3.57)$$

El espectro multifractal se presenta en forma paramétrica debido a la imposibilidad de invertir la Ec. (3.52). Los log-cumulantes para órdenes mayores a  $m = 4$  son no nulos, pero no son reportados aquí [122].

### CPC log-Poisson

En este caso, los multiplicadores se eligen como  $W = 2^\gamma \exp(\log(\beta)K)$ , en donde  $K$  sigue una distribución de Poisson con parámetro  $\lambda$ . A partir de la Ec. (3.50), la función de escalamiento es:

$$\zeta(q) = 2^\gamma e^{\lambda(\beta-1)} q - 2^{\gamma q} e^{\lambda(\beta q-1)} - q + 1. \quad (3.58)$$

Mediante la aplicación de la transformación de Legendre se obtiene el espectro multifractal, en forma paramétrica:

$$h(q) = 2^\gamma e^{\lambda(\beta-1)} - \log(2)\lambda 2^{\gamma q} e^{\lambda(\beta-1)} - \log(\beta)\lambda 2^{\gamma q} \beta q e^{\lambda(\beta q-1)} - 1, \quad (3.59)$$

$$D(q) = -2^\gamma q H e^{\lambda(\beta^q H-1)} (\gamma \log(2) q H + \beta^q \lambda \log(\beta) q H - 1) \quad (3.60)$$

Los log-cumulantes son [122]:

$$c_1 = 2^\gamma e^{\lambda(\beta-1)} - \log(2)\lambda - \log(\beta)\lambda - 1, \quad (3.61)$$

$$c_2 = -\log(2)^2 \gamma^2 - \log(4)\log(\beta)\gamma\lambda - \log(\beta)^2 \lambda, \quad (3.62)$$

$$c_3 = -\log(2)^3 \gamma^3 - 3\log(\beta^{\log(2)^2}) \gamma^2 \lambda - \log(8)\log(\beta)^2 \gamma \lambda (1 + \lambda) - \lambda^3 \log(\beta)^3 - 3\lambda^2 \log(\beta)^3 - \lambda \log(\beta)^3, \quad (3.63)$$

$$c_4 = -\lambda^4 \log(\beta)^4 - \lambda^3 (6\log(\beta)^4 + 4\gamma \log(2)\log(\beta)^3) - \lambda^2 (6\log(2)^2 \gamma^2 \log(\beta)^2 + 12\log(2)\gamma \log(\beta)^2) - \lambda^2 (6\log(2)^2 \gamma^2 \log(\beta)^2 + 12\log(2)\gamma \log(\beta)^2) + 4\log(2)\gamma \log(\beta)^3 + \log(\beta)^4 - \gamma^4 \log(2)^4. \quad (3.64)$$

### 3.8. Dimensión Fractal ( $D_f$ ) y Coeficiente de Hurst ( $H$ )

Los fractales cambiaron la concepción de la Geometría Euclidiana, a una conocida como Geometría Fractal, donde la dimensión entera es un caso particular de la dimensión fractal. La “*dimensión fractal*”  $D_f$  es una magnitud estadística que permite describir matemáticamente los objetos que presentan alto grado de complejidad, de auto-similaridad o caóticos. En el estudio de la geometría fractal, se cuenta con una herramienta para evaluar dos atributos de gran relevancia como son el coeficiente de Hurst y la dimensión fractal Mandelbrot [136] relacionados mediante la Ec. (3.65).

$$D_f = 2 - H, \text{ con } 0 < H < 1 \quad (3.65)$$

Mientras el coeficiente de Hurst fue desarrollado originalmente para estudios hidrológicos, las técnicas modernas lo han utilizado para estimar la multifractalidad que proviene del fenómeno de estudio. Esta estimación proporciona una medida para comprender si los datos son un camino aleatorio puro o tienen tendencias subyacentes. La estimación del coeficiente de Hurst se ha aplicado en áreas que van desde la biofísica a las redes de computadoras.

La interpretación de  $H$  es la siguiente [137–139]:

1.  $0 < H < 0.5$ , implica un comportamiento *antipersistente* en la serie de tiempo, es decir un período de crecimiento es seguido de otro de decrecimiento.
2.  $H = 0.5$ , implica un proceso aleatorio e independiente, es decir el proceso se considera *caminata aleatoria*.
3.  $0.5 < H < 1$ , implica procesos *persistentes*, es decir un período de crecimiento seguido de otro análogo.

#### 3.8.1. Métodos para calcular el Coeficiente de Hurst ( $H$ )

Existen diferentes métodos para la estimación del coeficiente de Hurst  $H$ , se enuncian a continuación.

#### 3.8.2. Conteo de Cajas

El método del conteo de cajas (BC, por sus siglas en inglés *Box Counting*) ayuda a determinar la dimensión fractal de una imagen. Este método ha sido aplicado por ejemplo para biología, neurociencia, mineralogía, geología y física [140].

El método de conteo de cajas está basado en las propiedades de llenado del espacio de una curva, de la siguiente manera [141]:

- Primero la curva se cubre con un conjunto de objetos de la misma área o cajas cuadradas, se determina un tamaño para el área denotado por ( $s$ ) la longitud del lado de la cuadrícula del objeto.
- Segundo se cuenta el número de cajas mínimo necesarias para cubrir a la curva completamente. A medida que el tamaño de las cajas se aproxima a cero, el área total cubierta por las cajas convergerá a la medida deseada de la curva. Esto dará como resultado un número de casillas  $N(s)$ .

- Tercer paso se propone obtener varios tamaños de cajas y hacer un ajuste lineal a una gráfica  $\log - \log$  de  $N(s)$  contra  $s$ . La pendiente  $D_b$  de la recta de mínimos cuadrados se toma como una estimación de la dimensión fractal de la curva. La pendiente de la línea  $D_b$  está dada por la Ec. (3.66):

$$D_b = \frac{[\log(N(s_2)) - \log(N(s_1))]}{\left[ \log\left(\frac{1}{s_2}\right) - \log\left(\frac{1}{s_1}\right) \right]} \quad (3.66)$$

donde  $(1/s)$ , es el número de casillas en la parte inferior de la grilla. Si la Ec. (3.66) es correcta, los puntos deberían estar (aproximadamente) sobre una línea recta, mostrando la evidencia de *fractalidad*.

En resumen, la dimensión de conteo de cajas se usa ampliamente en la práctica para estimar la dimensión de una variedad de objetos fractales. Esta técnica estima las dimensiones de los objetos en el plano como también objetos fractales en espacios multidimensionales.

### 3.8.3. Rango Reescalado ( $R/S$ )

Permite encontrar el parámetro o exponente de Hurst, el cual es un valor numérico que hace posible determinar la auto-correlación en una serie de datos. A continuación se describe la metodología desarrollada por Hurst [136]:

1. Se inicia con una serie de tiempo de tamaño  $M$ , pero como interesa los rendimientos logarítmicos entonces la serie original se reduce a una nueva serie de tiempo de tamaño  $N = M$  donde cada rendimiento está definido por:

$$N_i = \log\left(\frac{M_{i+1}}{M_i}\right), \quad i = 1, 2, 3, \dots, N \quad (3.67)$$

2. Este periodo de tiempo  $N$  se divide en  $A$  subperíodos contiguos de longitud  $n$ , tal que  $A*n = N$ . Se nombra cada uno de los subperíodos o subgrupos  $I_\alpha$ , con  $\alpha = 1, 2, 3, \dots, A$ . Cada elemento de  $I_\alpha$ , es etiquetado por  $N_{\kappa,\alpha}$ , tal que  $\kappa = 1, 2, 3, \dots, n$ . Para cada subperíodo  $I_\alpha$  de longitud  $n$ , el valor promedio está definido por la siguiente expresión:

$$e_\alpha = \frac{1}{n} \sum_{\kappa=1}^n N_{\kappa,\alpha} \quad (3.68)$$

3. Las diferencias de cada elemento  $N_{\kappa,\alpha}$  con respecto a la media  $e_\alpha$  para cada subperíodo  $I_\alpha$  se van sumando para obtener la serie de tiempo acumulada  $y_{\kappa,\alpha}$ , definida como:

$$y_{\kappa,\alpha} = \sum_{i=1}^{\kappa} (N_{i,\alpha} - e_\alpha), \quad \kappa = 1, 2, 3, \dots, n \quad (3.69)$$

4. El rango  $R_{I_\alpha}$  se define como la diferencia entre el valor máximo y el valor mínimo de  $Y_{\kappa,\alpha}$  para cada subperíodo  $I_\alpha$ :

$$R_{I_\alpha} = \max(y_{\kappa,\alpha}) - \min(y_{\kappa,\alpha}), \quad \text{donde } 1 \leq \kappa \leq n \quad (3.70)$$

5. Se calcula la desviación estándar muestral  $S_{I_\alpha}$  de la forma tradicional para cada periodo  $I_\alpha$ :

$$S_{I_\alpha} = \left( \frac{1}{n} \sum_{i=1}^{\kappa} (N_{i,\kappa} - e_\alpha)^2 \right)^2 \quad (3.71)$$

6. Para cada periodo  $I_\alpha$ , su rango  $R_{I_\alpha}$  se normaliza dividiendo por su desviación estándar muestral  $S_{I_\alpha}$  correspondiente. Por lo tanto el rango reescalado para cada subperíodo  $I_\alpha$  es igual a  $R_{I_\alpha}/S_{I_\alpha}$ . Ya que se tienen períodos continuos de longitud, entonces se toma el valor promedio  $R/S$  para períodos de longitud  $n$  y que está definido como:

$$(R/S)_n = \frac{1}{\alpha} \sum_{\alpha=1}^A \left( \frac{R_{I_\alpha}}{S_{I_\alpha}} \right) \quad (3.72)$$

7. La longitud  $n$  o el tamaño del subperíodo se incrementa al siguiente valor posible de tal forma que sea un valor entero. Se inicia con el valor más pequeño de acuerdo a la condición anterior y se repiten los pasos del 1 al 6 hasta  $n = (M - 1)/2$  utilizando siempre la serie completa. Finalmente se aplica una regresión de mínimos cuadrados de  $\log(R/S)_n$  contra  $\log(n)$ . Donde  $\log(R/S)_n = \log c + H \log n$ , así la ordenada al origen es  $\log c$  y la pendiente de la ecuación es la estimación del coeficiente de Hurst  $H$ .

### 3.8.4. Espectro de Potencia

Es un método que describe la distribución de energía de una serie de tiempo en el dominio de la frecuencia [29]. La energía es una cantidad de valor real, por lo que el espectro de potencia no contiene información de fase. Debido a que una serie de tiempo puede contener componentes de señales periódicas, no periódicas o muestreadas de forma asíncrona, el espectro de potencia de una serie de tiempo, típicamente se considera una función continua de la frecuencia. Cuando utiliza una serie de intervalos de frecuencia discreta para representar la frecuencia continua, el valor en un intervalo de frecuencia específico es proporcional al intervalo de frecuencia. Para eliminar la dependencia del tamaño del intervalo de frecuencia, se puede normalizar el espectro de potencia para producir la densidad espectral de potencia (PSD, por sus siglas en inglés), el espectro de potencia dividido por el tamaño del intervalo de frecuencia.

El PSD mide la potencia de señal por unidad de ancho de banda para una serie de tiempo en  $V^2/Hz$ , que asume implícitamente que el PSD representa una señal en voltios que genera una carga de 1 ohmio. Si desea utilizar otras unidades para la PSD estimada de una serie de tiempo, debe escalar la unidad de la serie de tiempo en unidades de ingeniería apropiadas. Después de escalar la unidad de la serie de tiempo, puede obtener la unidad correspondiente para el valor de PSD lineal.

Los métodos de estimación de PSD se clasifican de la siguiente manera [29]:

- *Métodos paramétricos*: basados en modelos como los modelos AR, los modelos de media móvil (MA) y los modelos de media móvil autoregresiva (ARMA). Por lo que, los métodos paramétricos también se conocen como métodos basados en modelos. Para estimar la PSD de una serie de tiempo con métodos paramétricos, primero debe obtener los parámetros del modelo de la serie de tiempo y se debe crear un modelo apropiado que refleje el comportamiento del sistema que genera las series de tiempo.

- *Métodos no paramétricos:* estos métodos son el periodograma, el método de Welch y el método de Capon, se basan en la transformada de Fourier discreta. No es necesario obtener los parámetros de la serie de tiempo antes de utilizar estos métodos.

Una de las ventajas de los métodos paramétricos es que no utilizan ventanas de datos, suponen que una señal se ajusta a un modelo en particular, las PSD estimadas con métodos paramétricos están menos sesgadas y poseen una varianza menor que las PSD estimadas con métodos no paramétricos si el modelo supuesto es correcto. Sin embargo, las magnitudes de las PSD estimadas con métodos paramétricos generalmente son incorrectas. En contraste, los métodos no paramétricos realizan el cálculo utilizando ventanas de datos, como consecuencia produce una distorsión de las PSD resultantes debido a los efectos de las ventanas. El beneficio clave de los métodos no paramétricos es la solidez: los PSD estimados no contienen picos de frecuencia no esenciales.

### 3.8.5. Wavelet

De manera muy general, la Transformada Wavelet de una función  $f(t)$  es la descomposición de  $f(t)$  en un conjunto de funciones  $\psi_{s,\tau}(t)$ , que forman una base y son llamadas las “Wavelets” [142]. La Transformada Wavelet se define como:

$$W_f(s, \tau) = \int f(t) \psi_{s,\tau}^*(t) dt \quad (3.73)$$

Las Wavelets son generadas a partir de la traslación y cambio de escala de una misma función wavelet  $\psi(t)$ , llamada la “Wavelet madre”, y se define como:

$$\psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) \quad (3.74)$$

donde  $s$  es el factor de escala, y  $\tau$  es el factor de traslación.

Las wavelets  $\psi_{s,\tau}(t)$  generadas de la misma función wavelet madre  $\psi(t)$  tienen diferente escala  $s$  y ubicación  $\tau$ , pero tienen todas la misma forma. Se utilizan siempre factores de escala  $s > 0$ . Las Wavelets son dilatadas cuando la escala  $s > 1$ , y son contraídas cuando  $s < 1$ . Así, cambiando el valor de  $s$  se cubren rangos diferentes de frecuencias. Valores grandes del parámetro  $s$  corresponden a frecuencias de menor rango, o una escala grande de  $\psi_{s,\tau}(t)$ . Valores pequeños de  $s$  corresponden a frecuencias de menor rango o una escala muy pequeña de  $\psi_{s,\tau}(t)$  [143].

## 3.9. Análisis de Fluctación sin Tendencia (DFA)

Método introducido por Peng [144]. De acuerdo a Yuval y Broday [145], se conoce por sus siglas en inglés como DFA (Detrended Fluctuation Analysis), el cual se ha convertido en una técnica ampliamente utilizada para determinar las propiedades de escalado (mono) fractales, es decir dentro del método del DFA el coeficiente de hurst es único, debido a que es un proceso monofractal. El algoritmo del DFA consta de los siguientes pasos:

Considere la serie de tiempo  $x_t : t = 1, 2, \dots, N$ .

1. Construir la serie sin tendencia:

$$y_t = \sum_{i=1}^N (x_i - \mu) \quad (3.75)$$

donde  $\mu = \frac{1}{N} \sum_{i=1}^N x_i$ ,  $i = 1, 2, 3, \dots, N$  (media aritmética). Es decir, la manera de contrarrestar el impacto de la no estacionariedad.

2. Dividir el conjunto  $y_t$  en  $s = \lceil \frac{N}{n} \rceil$  cajas o ventanas que no deben superponerse de longitud  $n$ . Y definimos cada subperíodo por los  $I_{\alpha_i}$ .
3. Fraccionar la serie de datos en particiones de tamaño  $n$ , es decir, la tendencia local para cada ventana  $I_{\alpha_i}$  se evalúa mediante el ajuste de mínimos cuadrados de los datos,  $v = I_{\alpha_1}, I_{\alpha_2}, \dots, I_{\alpha_s}$ . Así la tendencia local está sujeta al número de cajas o ventanas que se eligen para la serie temporal.
4. Calcular las fluctuación en la serie integrada en cada ventana.

$$F(n)^2 = \frac{1}{n} \sum_{i=1}^N |y_t - z_t|^2 \quad (3.76)$$

Donde  $y_t$  es la serie sin tendencia y  $z_t$  es el polinomio de ajuste de primer orden o polinomio de orden superior. Es decir, para cada  $I_{\alpha_i}$  de longitud  $n$ , la varianza correspondiente entre  $y_t$  y su valor ajustado  $z_t$  es dado por  $F(n)$ .

5. Variamos la escala  $n$  (longitud de las particiones) y repetimos el algoritmo.
6. Empíricamente se ha evidenciado por Peng [144] que si la serie exhibe fractalidad, ésta sigue una ley de potencia:

$$F(n) = c n^\alpha \quad (3.77)$$

Donde  $c$  es una constante independiente de  $n$ .

Aplicando una regresión lineal a los logaritmos de la Ec. (3.77) y estimando el coeficiente de Hurst:

$$\log(F(n)) = \log(c n^\alpha) \quad (3.78)$$

$$\log(F(n)) = \log(c) + \alpha \log(n) \quad (3.79)$$

El exponente de escala  $\alpha$ , se relaciona con el exponente de Hurst  $H$ , de acuerdo a la Ec. (3.80):

$$\alpha = 2 H \quad (3.80)$$

$$H = \frac{\alpha}{2} \quad (3.81)$$

7. Se determina la dimensión fractal por medio de la Ec. (3.82) planteada por [146]:

$$2H + 1 = 5 - 2D \quad (3.82)$$

Simplificando la Ec. (3.82) se obtiene la relación directa entre la dimensión fractal y el exponente de Hurst:

$$D = 2 - H \quad (3.83)$$

Donde  $D$  es la dimensión fractal y  $H$  es el coeficiente de Hurst.

## 3.10. Multifractalidad

Concepto introducido por Frish y Parisi [147], ya que muchos de los procesos no muestran un comportamiento con un único coeficiente de escala (monofractal), sino que necesitan diferentes coeficientes de escala  $H$  en diferentes partes de las series, para una descripción completa del comportamiento en el mismo rango de escalas de tiempo. Existen dos tipos generales de multifractalidad en series de tiempo [148]:

1. Multifractalidad debido a la amplia distribución de probabilidad (función de densidad) para los valores de las series de tiempo.
2. Multifractalidad debido a las diferentes correlaciones de largo plazo en las pequeñas y grandes fluctuaciones.

En caso de haber dos tipos de multifractalidad, las series reestructuradas mostrarán multifractalidad más débil que las series originales. Un análisis multifractal de series de tiempo también revelará un mayor orden de correlaciones. Además, el escalado multifractal se observa si el comportamiento de escala de las fluctuaciones pequeñas y grandes es diferente.

La teoría fractal realiza una medición y descripción de objetos fractales, caracterizado principalmente por la dimensión fractal que contiene cada objeto, pero como las técnicas fractales no son suficientes para la descripción de los objetos que no son fractales, como ocurre en la naturaleza. Por esta razón, se desarrolló la “*técnica multifractal*” la cual describe y mide los eventos de la naturaleza, sobre la base de la técnica fractal [149]. Análisis de los elementos de un objeto que pueden realizar una descripción más real de las leyes naturales. Esta nueva estadística es descrita por su “*espectro de singularidades*”, dado por los valores de *índice de singularidad* y la dimensión fractal de la singularidad correspondiente.

Existen muchas técnicas multifractales para obtener dicha caracterización. En el presente trabajo de tesis se desarrollará la técnica de *Análisis Multifractal de Fluctuación sin Tendencia (MF-DFA)*, *Cascada Multiplicativa*, *ARIMA*, *Redes Neuronales (RNA)* e *Inteligencia Artificial (IA)*.

### 3.10.1. Técnicas de Multifractalidad

Para determinar cuando una serie de tiempo es multifractal es necesario utilizar *métodos para análisis de series de tiempo multifractales*. Algunos de los análisis más simples se basan en el formalismo

multifractal de la función de partición estándar, para la caracterización multifractal de las medidas normalizadas, estacionarias. Desafortunadamente, éste no da resultados correctos para las series de tiempo no estacionarias afectadas por las tendencias o que no se pueden normalizar. Por otro lado, se desarrolló un formalismo multifractal mejorado, el método de la transformada Wavelet de módulo máximo, basado en el análisis Wavelet y consiste en el seguimiento de las líneas máximas en la transformada continua Wavelet en todas las escalas. Una alternativa importante es el algoritmo *Análisis Multifractal de Fluctuación sin Tendencia (MF-DFA)*, que no requiere el procedimiento de módulo máximo, y entonces implica un poco más de esfuerzo en la programación que la convencional [150].

### 3.10.2. Análisis Multifractal de Fluctuación sin Tendencia (MF-DFA)

Esta técnica se divide en 5 pasos y se describe en el trabajo de Jan W. Kantelhardt [52, 53, 151]:

1. Sea  $x_i$  un conjunto de datos discretos donde  $i = 1, \dots, N$ . Definimos el perfil de la serie  $Y(i)$  como la Ec. (3.84):

$$Y(i) = \sum_{k=1}^i [x_k - \bar{x}], \quad i = 1, 2, 3, \dots, N, \quad (3.84)$$

donde:  $Y(i)$  es la caminata aleatoria,  $x_k$  son los componentes de la serie de tiempo y  $\bar{x}$  es la velocidad promedio del viento.

2. Se divide el perfil en  $N_s = \text{int}(N/s)$  segmentos no traslapados de igual longitud  $s$  de izquierda a derecha. Como la longitud  $N$  de los datos puede no ser un múltiplo entero de la escala  $s$  y pueden sobrar datos en los extremos se repite el procedimiento desde el extremo opuesto. Por lo tanto se tienen  $2N_s$  segmentos.
3. Se calcula la tendencia local de cada uno de los  $2N_s$  segmentos por un ajuste de mínimos cuadrados. Entonces se determina la varianza mediante la relación:

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s [Y[(v-1)s + i] - y_v(i)]^2, \quad (3.85)$$

para cada segmento,  $v, v = 1, \dots, N_s$ ; y,

$$F^2(s, v) = \frac{1}{s} \sum_{i=1}^s [Y[N - (v - N_s)s + i] - y_v(i)]^2, \quad (3.86)$$

para  $v = N_s + 1, \dots, 2N_s$ . Aquí  $y_v(i)$  es el ajuste polinomial en el segmento  $v$ . Dicho ajuste polinomial puede ser lineal, cuadrático, cúbico o superior.

4. Promediando sobre todos los segmentos obtenemos la función de fluctuaciones de orden  $q$ , véase la Ec. (3.87).

$$F_q(s) = \begin{cases} \left\{ \frac{1}{2N_s} \sum_{v=1}^{2N_s} [F^2(s, v)]^{\frac{q}{2}} \right\}^{\frac{1}{q}} & \forall q \in (\delta - \{0\}) \\ \exp \left[ \frac{1}{4N_s} \sum_{v=1}^{2N_s} \ln F^2(v, s) \right] & \text{para } q = 0 \end{cases} \quad (3.87)$$

donde  $\delta$  es cualquier valor real.

5. Se determina la ley de escalamiento de la función de partición analizando el gráfico log-log  $F_q(s)$  versus  $s$  para cada valor de  $q$ :

$$F_q(s) \sim s^{h(q)} \quad (3.88)$$

### 3.11. Métricas de rendimiento

Hasta el día de hoy no se ha desarrollado ningún protocolo de evaluación para modelos de pronóstico, siendo una tarea pendiente y necesaria. En la mayoría de estudios científicos y/o empresariales, dónde se cuantifica, se han aplicado numerosas métricas para evaluar el rendimiento y éxito de los modelos de pronóstico. Sin embargo, aún no existe un estándar universal para esta evaluación [152, 153]. Cuando hay más de un modelo desarrollado, se puede hacer un análisis comparativo calculando la de cada modelo y posteriormente pueden compararse entre ellos. La exactitud de cualquier modelo predictivo sólo puede determinarse eligiendo las medidas de error apropiadas [23].

En este trabajo, se compararon diferentes modelos utilizando cinco criterios de error estadísticos, entre los más conocidos están el error absoluto medio (MAE), error medio cuadrático (MSE), raíz del error cuadrático medio (RMSE), error porcentual absoluto medio (MAPE), índice de acuerdo (IA). Los detalles de estos criterios se presentan a continuación.

1. *El error absoluto medio (MAE)*, es una medida común del error de pronóstico en análisis de series de tiempo. Algunas características que se deben considerar son:
  - La serie de tiempo es homogénea o igualmente espaciada.
  - Las dos series de tiempo deben ser idénticas en tamaño.

El error medio absoluto esta definido como [154]:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (3.89)$$

donde:  $y_t$  son la observaciones actuales de las series de tiempo,  $\hat{y}_t$  es la serie de tiempo estimada o pronosticada y  $n$  es el número total de datos de la serie de tiempo.

2. *El error cuadrático medio (MSE)*, es el criterio de evaluación más usado para problemas de regresión. Se usa principalmente con aprendizaje automático supervisado. Para cada dato histórico podremos indicar el resultado correcto. Algunas características que se deben considerar son [155–158]:
  - La serie de tiempo es homogénea o igualmente espaciada.
  - Las dos series de tiempo deben tener el mismo tamaño.
  - El MSE proporciona una función cuadrática de pérdida en la medida en que eleva al cuadrado y, posteriormente, promedia los diferentes errores; dando más peso a los grandes errores (valores atípicos) que a los más pequeños.

- Es más útil esta métrica cuando se trata de grandes errores que cuando se trata de pequeños errores.
- La Ec. (3.90) es similar a la medida estadística de varianza ( $\sigma^2$ ), que nos permite medir la incertidumbre alrededor del más probable pronóstico  $y_t$ . En otras palabras, esta métrica puede verse como la varianza del error de pronóstico.
- La mayor desventaja del MSE es la escala de dependencia. Si el pronóstico de tareas incluye objetos con diferentes magnitudes entonces la medida MSE no se puede aplicar.

El error cuadrático medio (MSE) se define así:

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (3.90)$$

donde:  $y_t$  es el resultado actual en el tiempo  $t$ ,  $\hat{y}_t$  es el pronóstico de valor en el tiempo  $t$ .

3. *La raíz del error cuadrático medio (RMSE)*, estadístico que permite medir la precisión del pronóstico. Algunas características que se deben considerar son [156, 157]:

- Es usada para comparar diferencias entre dos conjunto de datos, ninguno de los cuales son aceptados como “estándar o actual”.
- La serie de tiempo es homogénea o igualmente espaciada.
- Las dos series de tiempo deben ser idénticas en tamaño.

La función de la desviación de la raíz media cuadrática se define de la siguiente manera:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3.91)$$

donde:  $y_t$  son las observaciones actuales de las series de tiempo,  $\hat{y}_t$  es la serie de tiempo estimada o pronosticada,  $n$  es el número total de datos de la serie de tiempo.

4. *El error porcentual absoluto medio (MAPE)*, mide la exactitud de un método para la construcción ajustada de valores de serie de tiempo en estadísticas. Algunas características que se deben considerar son [156, 157]:

- La serie de tiempo es homogénea o igualmente espaciada.
- Las dos series de tiempo deben ser idénticas en tamaño.
- Cuando se calcula el promedio MAPE para un número de series de tiempo, usted puede encontrar un problema: algunas de las series que tienen un alto MAPE pueden distorsionar una comparación entre el MAPE promedio de una serie de tiempo ajustada con un método en comparación con el MAPE promedio al usar otro método.

El error porcentual absoluto medio (MAPE) se define de la siguiente manera:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100 \% \quad (3.92)$$

donde:  $y_t$  son las observaciones actuales de las series de tiempo,  $\hat{y}_t$  es la serie de tiempo estimada o pronosticada,  $n$  es el número total de datos de la serie de tiempo.

5. *Índice de acuerdo (IA)*, es un índice adimensional que se utiliza para comparar diferentes modelos.

El índice, IA de acuerdo a Wang et al. [159] se define de la siguiente manera:

$$IA = 1 - \frac{\sum_{t=1}^N (\hat{y}_t - y_t)^2}{\sum_{t=1}^N (|\hat{y}_t - \bar{y}| + |y_t - \bar{y}|)} \quad (3.93)$$

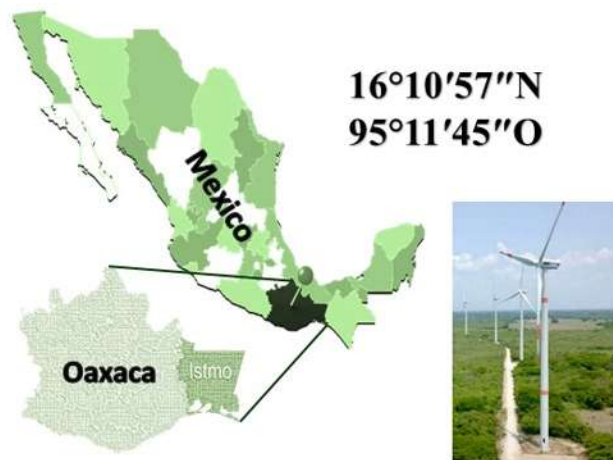
donde:  $y_t$  son las observaciones actuales de las series de tiempo,  $\hat{y}_t$  es la serie de tiempo estimada o pronosticada,  $\bar{y}$  es la serie de tiempo a predecir.

Para Wang et al. [159], MAE y RMSE son medidas clásicas basadas en la precisión para calificar el desempeño de los modelos de pronóstico de la velocidad del viento. Además, RMSE asume que los errores son in-sesgados y siguen una distribución normal, mientras que MAE es adecuado para describir errores distribuidos uniformemente. Sin embargo, tienen varios inconvenientes, como penalizaciones más severas en los errores positivos que en los negativos [160]. Son consideradas como medidas dependientes de la escala que no son adecuadas para comparar series de tiempo con diferentes escalas [161, 162].

Mientras tanto, Li et al. [163] y Liu et al. [164] consideran que las métricas MAPE, MAE, y MSE se pueden utilizar para medir eficazmente el rendimiento de la predicción y su criterio específico. Estas tres métricas miden el efecto promedio del pronóstico y son muy sensibles a los cambios del conjunto de datos. Cuando hay un pequeño cambio en el conjunto de datos, las métricas cambian significativamente. Similar a reducir la cantidad de conjuntos de entrenamiento, agregar un paso de pronóstico aumentará las métricas.

### 4.1. Tratamiento de datos utilizados en el estudio

Los datos utilizados en este estudio fueron dos series de tiempo de velocidad del viento recopiladas por la Comisión Federal de Electricidad de México (CFE) en las plantas de energía eólica La Mata y La Venta. Las bases de datos se construyeron con mediciones de la velocidad del viento a una altura de 30 m, con una frecuencia de muestreo de 1 Hz y un período promedio de 10 minutos y 1 hora, respectivamente. Las principales características de los equipos de medición de las estaciones meteorológicas para ambas bases de datos se muestran en la Tabla 4.1. Los conjuntos de datos de velocidad del viento incluyen desde el 1 de enero de 2006 al 31 de diciembre de 2010 para La Mata, y desde el 1 de febrero de 1994 al 27 de noviembre de 2000 para La Venta. La Fig. 4.1 muestra la ubicación de las plantas de energía eólica.



La Mata y La Venta están en el Istmo de Tehuantepec, Oaxaca, México.

Figura 4.1: a) Ubicación geográfica de las torres anemométricas en La Mata y La Venta.

Tabla 4.1: Especificaciones de los sensores de velocidad y dirección del viento.

Especificaciones	Anemómetro	Veleta de viento
Rango de medición	0-45 m/s	0-360°
Exactitud	± 1.5 %	± 1.5 %
Resolución	<0.1 m/s	<1°

Es conocida la importancia de estabilizar los valores de medias y varianzas de las series de tiempo

para realizar correctamente los modelos ARIMA / SARIMA. Entonces, en primer lugar se realizó un tratamiento de datos en ambas series de tiempo, La Mata ( $S_{t_1}$ ) y La Venta ( $S_{t_2}$ ), según la metodología de tratamiento de datos mostrada en la Fig. 4.2.

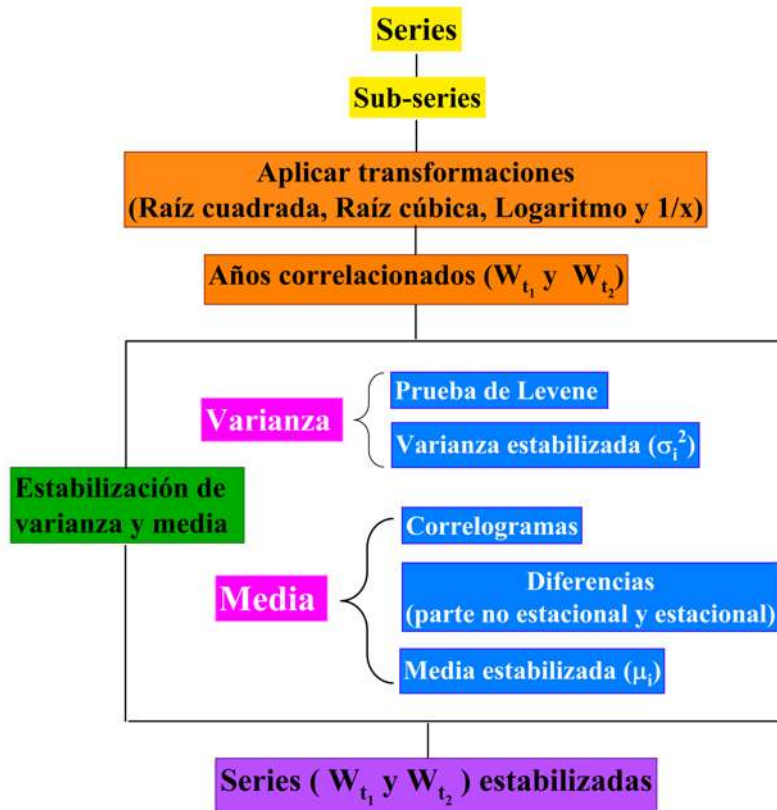


Figura 4.2: b) Diagrama de flujo del preprocesamiento de datos.

1. **Reagrupación de datos.** Del conjunto de datos de La Mata se formaron 5 grupos de observaciones, mientras que del conjunto de datos de La Venta se ordenaron 7 grupos de observación, con base en la disponibilidad de datos anuales, ver Fig. 4.3 a) y b).

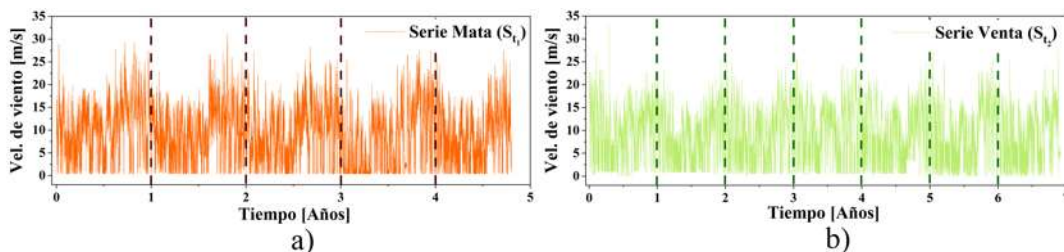


Figura 4.3: a) y b) Series temporales originales de La Mata y La Venta, respectivamente.

2. **Transformaciones de datos.** Luego, para encontrar la transformación que describe mejor la relación entre los datos, se aplicaron las transformaciones más comunes a cada una de las muestras, como raíz cuadrada, raíz cúbica, logaritmo y  $1/x$ . Dada la mayor relación de datos durante

dos años (2006 y 2008) de la serie temporal  $S_{t_1}$ , se eligió la transformación de raíz cúbica, como se muestra en la Fig. 4.4 a); resultando en el conjunto de datos denotado como  $W_{t_1}$ , vea la Fig. 4.5 a). De la misma forma, para la serie de tiempo  $S_{t_2}$ , se determinó que la transformación de la raíz cuadrada tuvo un comportamiento estable durante 3 años (1996, 1997 y 2000), ver Fig. 4.4 b). Los datos resultantes se denotan como  $W_{t_2}$ , ver Fig. 4.5 b). Por tanto, no fue posible considerar todas las bases de datos debido a probables comportamientos irregulares en las mediciones, observados principalmente en los primeros años de la serie temporal.

3. **Estabilización de la varianza.** La prueba de Levene para la homocedasticidad se realizó en ambas series de tiempo  $W_{t_1}$  y  $W_{t_2}$ , lo que nos permitió probar la hipótesis nula: no existen diferencias significativas entre la varianza de la series de tiempo en los grupos propuestos, con un nivel de significancia de 5 %, obteniendo valores de  $p = 0.097$  y  $p = 0.076$ , respectivamente. Al obtener la aceptación de la prueba de Levene para ambas series temporales, es decir, se estabilizó la varianza.

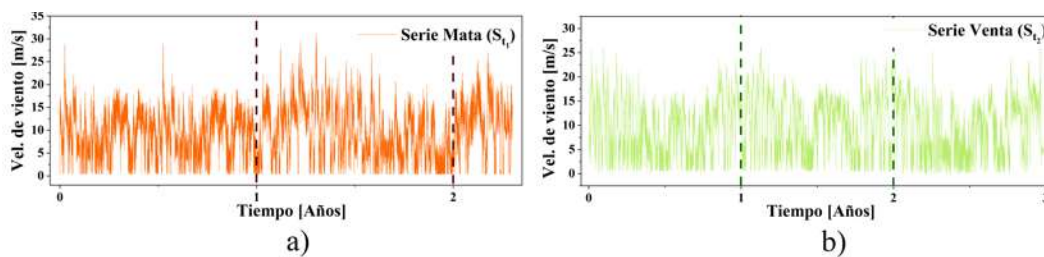


Figura 4.4: Series temporales estabilizadas en la varianza a) La Mata y b) La Venta, seleccionando los años a tratar.

4. **Estabilización de la media.** Las series de tiempo  $W_{t_1}$  y  $W_{t_2}$  también deben estabilizarse en promedio, es decir, cuando fluctúan alrededor de una media única. Para ello, se realizó un correlograma para estabilizar los promedios ( $W_{t_1}$  y  $W_{t_2}$ ), mostrando rezagos en la serie temporal de velocidades del viento y con una disminución gradual a cero, esto sucede fuera de los límites de la confianza. Evidencia de que las series de tiempo de velocidad del viento no son estacionarias en el medio. Esto permitió que las series ( $W_{t_1}$  y  $W_{t_2}$ ) tuvieran diferencias, generalmente haciendo que la serie fuera estacionaria en promedio. Para la serie de tiempo  $W_{t_1}$ , se tomaron las diferencias en el orden de parte no estacional  $d = 1$ , ver Fig. 4.5 e). De manera similar, de la serie de tiempo  $W_{t_2}$ , se tomaron las diferencias del orden  $d = 1$  y  $D = 1$ ; para la parte no estacional y estacional, respectivamente; ver Fig. 4.5 f).

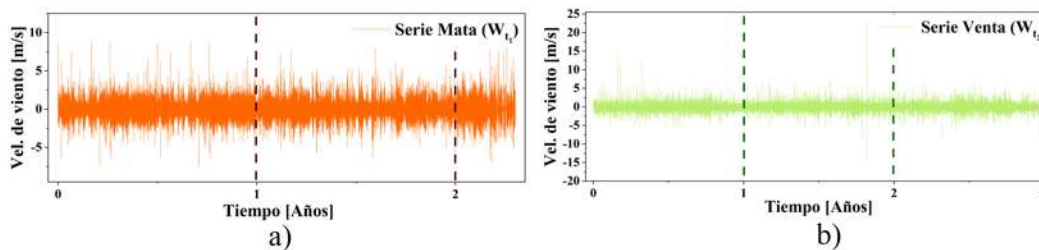


Figura 4.5: Series temporales estabilizadas en la media a) La Mata y b) La Venta.

## 4.2. Modelos matemáticos

Los modelos utilizados en el presente estudio fueron generados con el software matemático MATLAB; utilizado para calcular los coeficientes de los modelos ARIMA y SARIMA. En este mismo programa se adoptó el código de MF-DFA utilizado en [60] para revisar la multifractalidad de las series y errores primarios, de la misma manera se utilizó para calcular el coeficiente de Hurst.

La programación de las Redes Neuronales y de la Cascada Multiplicativa, se realizó en Python. El programa Benoit, se utilizaría en los casos en donde las series y errores primarios hubiesen sido monofractales, ya que permite aplicar diferentes técnicas como Box counting, R/S statistic y Power spectrum. Para lograr el manejo de la gran cantidad de datos disponibles, se utilizó una Workstation Dell Precision T1700 que cuenta con un procesador Xeon (E3-1220v3) y 16 GB de memoria RAM.

### 4.2.1. Modelo de Persistencia

Este método también se conoce como “Predictor Navie” [70]. Supone que la velocidad del viento en el tiempo  $y_t = y_{t-1}$  será la misma que fue en el momento  $t$ . Este método es más preciso que la mayoría de los métodos físicos y estadísticos para corto plazo [165]. Por tanto, cualquier método de pronóstico que sea desarrollado debe, primero, ser probado contra el punto de referencia clásico de método de persistencia para verificar cuánto puede mejorar esta técnica [166].

### 4.2.2. ARIMA (Autoregressive Integrated Moving Average)

El modelo general no estacional se conoce como ARIMA  $(p, d, q)$ , donde:  $p$  es el orden de la parte autorregresiva,  $d$  el grado de primera diferencia involucrada,  $q$  es el orden de parte media móvil.

La expresión lineal que define un modelo ARIMA (notación anterior) [21, 167] es la siguiente:

$$y_t = \sum_{i=1}^{p+d} \phi_i y_{t-1} + a_t - \sum_{j=1}^q \theta_j a_{t-j}, \quad (4.1)$$

donde:  $\phi_i$  es el  $i$ -ésimo parámetro autorregresivo,  $\theta_i$  es el  $i$ -ésimo parámetro de medias móviles, y  $a_t$  el residual del modelo, para mayor detalle ver Anexo B.

Para modelar la serie temporal, se siguió la metodología de Box-Jenkins según el procedimiento descrito por Cadenas y Rivera [168].

La expresión obtenida para la serie de tiempo de viento La Mata es la siguiente ARIMA(1,1,2):

$$y_t = y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}, \quad (4.2)$$

donde:  $\phi_1 = 0.6096$ ,  $\theta_1 = -0.8260$ ,  $\theta_2 = 0.0644$ .

### 4.2.3. Media móvil integrada autorregresiva estacional (SARIMA)

Un modelo SARIMA con términos de orden no-estacionales  $(p, d, q)$  y términos de orden estacionales  $(P, D, Q)$  se abrevia un modelo SARIMA  $(p, d, q)(P, D, Q)_s$  y se puede escribir

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = c + \theta(B)\Theta(B^s)\varepsilon_t, \quad (4.3)$$

donde  $\Phi$ ,  $\Theta$  denotan polinomios en  $B^s$  de orden  $P$ ,  $Q$ , respectivamente. Además,  $B$  es el operador de desplazamiento hacia atrás y  $s$  la ventana de tiempo de patrón estacional, para más detalle ver Anexo B.

La expresión obtenida para la serie de tiempo de viento La Venta es la siguiente:

$$\text{SARIMA}(1,1,1) (0, 1, 1)_{24},$$

$$y_t = (1 - \phi_1)y_{t-1} - \phi_1 y_{t-2} + y_{t-24} - (1 - \phi_1)y_{t-25} + \phi_1 y_{t-26} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \Theta_{24} \varepsilon_{t-24} + \theta_1 \Theta_{24} \varepsilon_{t-25}, \quad (4.4)$$

$$\text{donde: } \phi_1 = -0.0334, \theta_1 = -0.1325, \Theta_{24} = -0.9520.$$

#### 4.2.4. Aprendizaje Profundo

El modelo de redes neuronales y aprendizaje profundo para las series de tiempo de viento La Mata y La Venta fueron bajo el esquema [169], representado en las Figs. 4.6 y 4.7.

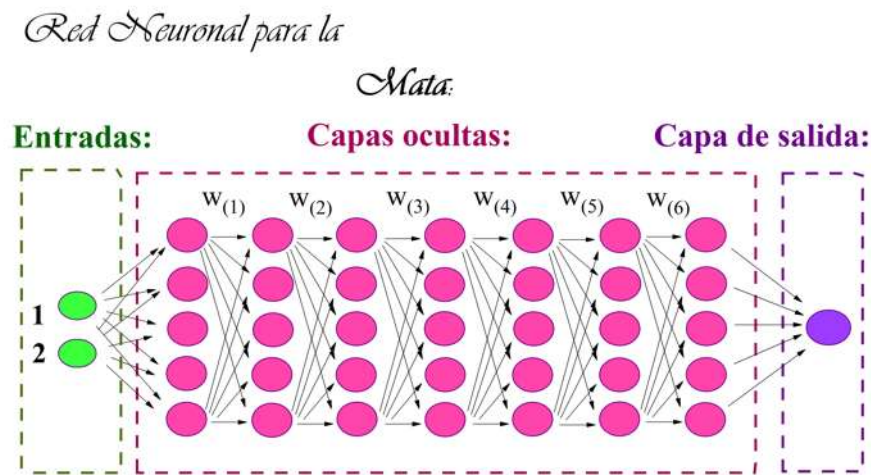


Figura 4.6: Diagrama de la red neuronal utilizando Aprendizaje Profundo para la serie La Mata.

La Fig. 4.6 muestra el esquema representativo de las redes neuronales utilizadas. Para la serie de tiempo  $W_{t_1}$  se utilizó un modelo con 2 entradas, 7 capas ocultas, con 5 neuronas y una capa de salida. Mientras que para la serie  $W_{t_2}$  (ver Fig. 4.7) fue un modelo con 26 entradas, 5 capas ocultas, con 5 neuronas y una capa de salida. Las salidas para ambas series de redes neuronales fueron calculadas mediante la función de activación (Rectified Linear Unit (ReLU)) la cual otorga flexibilidad y permite estimar complejas relaciones no lineales en los datos de las redes neuronales.

*Red Neuronal para la serie:*

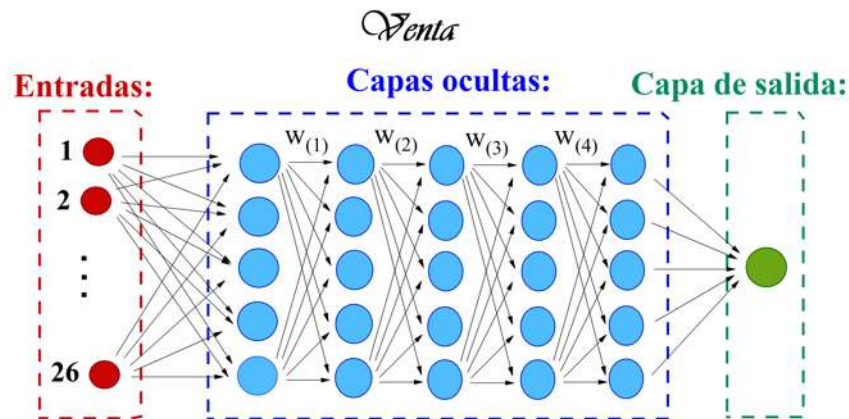


Figura 4.7: Diagrama de la red neuronal utilizando Aprendizaje Profundo para la serie La Venta.

#### 4.2.5. Cascada Multiplicativa

Para el modelado de las series de tiempo de viento, fue necesario que la cascada preserve la medida original, como lo hace la cascada determinista [29], pero que ofrezca flexibilidad para ajustar las estadísticas de la cascada a las cascadas observadas en las series, como lo hace la cascada aleatoria [29]. Con ese propósito se utiliza la cascada conservadora [170]; la cual, utiliza una estructura repetitiva divisoria, que se utiliza en el estudio de la turbulencia y otros procesos complejos. Además, se consideró las rupturas sucesivas de la energía del viento que corresponden al 70 % y 30 % [171]. Obteniéndose lo anterior, cuando la energía del viento supera un umbral, su cohesión interna disminuye y fluye de manera intermitente e irregular. La cascada multiplicativa fue compuesta por diferentes niveles (1, 2, 3...), el nivel inicial 1 representa la velocidad promedio que se considerará para iniciar la descomposición. El nivel inicial tiene un período  $2^n$  como una de sus características, además una velocidad promedio  $V_0$ . En la primera descomposición, la naturaleza misma del viento divide el período en dos partes; uno de ellos con un 70 % de la energía del viento y el otro con el 30 % restante. Dicha descomposición es aleatoria, lo que significa que, a veces irá hacia el lado izquierdo, y en otras ocasiones, hacia el lado derecho, su distribución es aleatoria (véase la Fig. 4.8).

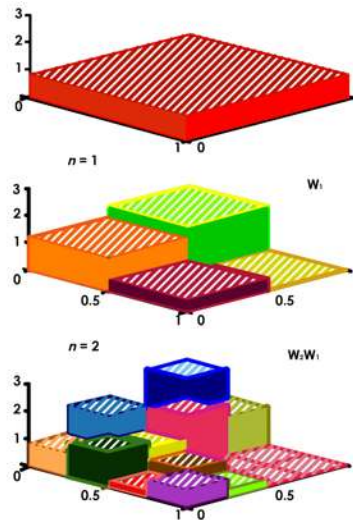


Figura 4.8: Modelo de Cascada Multiplicativa.

### ***Propuesta de la Cascada Multiplicativa (Conservadora Binomial)***

Para el modelo de la velocidad de viento, es necesario que la cascada preserve la medida original, como lo hace la cascada determinística, pero que ofrezca flexibilidad para ajustar las estadísticas de la cascada a las estadísticas observadas en trazas reales, como lo hace la cascada aleatoria. Con ese propósito se define la cascada conservadora, que preserva la medida y ofrece una alta flexibilidad estadística.

El procedimiento iterativo para la construcción de la cascada multiplicativa se basa en los multiplicadores determinados por la expresión binomial de Newton. Además, considerando que una cascada aplicada a la velocidad del viento de una serie forma elementos espirales que llevan por sí mismos distintas cantidades de energía de sitio a sitio. Al continuar el proceso, las energías se ordenan en erupciones intermitentes y violentas que eventualmente se disipan en la forma de calor, una vez que los remolinos se vuelven pequeños. A lo largo de una dimensión son capas entrelazadas de energías las cuales son universalmente consistentes con la primera cascada, cuando los remolinos consecutivos llevan precisamente el 70 y 30 % de la energía de la velocidad inicial. Como lo reportaron Charles Meneveau y Katepalli Sreenivasan en 1987 [172], la naturaleza produce una permutación de la cascada, con los remolinos más energéticos sucediendo no siempre a la izquierda sino a ambos lados, como guiados por el azar [171]. Considerando en nuestro caso  $p$  como la probabilidad de ser el 70 o 30 % de esa energía (véase en la Fig. 4.9).

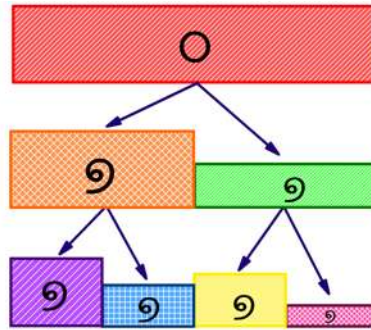


Figura 4.9: Cascada en la forma en que la turbulencia ocurre en el aire.

El proceso de construcción en cascada es el siguiente. Dejar denotar el intervalo de la unidad  $[0, 1]$ . Sea  $x$  una variable aleatoria de valor real con distribución uniforme en el intervalo  $[0, 1]$  y  $V_{prom}$  la velocidad promedio en el intervalo. En la  $n$ -ésima etapa, el intervalo  $I$  se divide en  $2^n$  subintervalos, cada uno con una medida proporcional a una cantidad numérica determinada por la siguiente expresión binomial de Newton:

$$\binom{2^n}{k} (2^{-n}) (x)^{2^n-k} (1-x)^k (p-p^2) (V_{prom}) \quad (4.5)$$

En otras palabras, en la etapa  $n$ , se aplica lo siguiente factor de ponderación para el primer subintervalo:

$$W_{00\dots0} = (2^{-n}) ((x)^{2^n} + (1-x)^{2^n}) (p-p^2) (V_{prom})$$

$n$ -dígitos

mientras que para los subintervalos restantes los factores de ponderación son:

$$W_{b_1 b_2 \dots b_n} = \binom{2^n}{i} (2^{-n}) ((x)^{2^n-i} + (1-x)^i) (p-p^2) (V_{prom}) \mid i = 1, \dots, 2^n - 1 \quad (4.6)$$

donde  $b_1 b_2 \dots b_n$  es la representación binaria del número entero  $i$ . Por ejemplo, la medida asignada al primer intervalo  $[0, 1/2]$  de la etapa  $n = 1$  es:

$$W_0 = \mu_1[0, 1/2] = (2^{-n}) ((x_0)^2 + (1-x_0)^2) (p-p^2) (V_{prom}) \quad (4.7)$$

mientras que la medida para el segundo intervalo es:

$$W_1 = \mu_1[1/2, 1] = \binom{2}{1} (2^{-n}) (x_0 + (1-x_0)) (p-p^2) (V_{prom}) \quad (4.8)$$

Para  $n = 2$ , la segunda etapa de la iteración en cascada, tenemos las siguientes reglas. La medida asignada al primer intervalo es igual a:

$$\begin{aligned} \mu_{00}[0, 1/4] &= \mu_1[0, 1/2] W_{00} \\ &= \mu_1[0, 1/2] (2^{-n}) ((x_1)^2 + (1-x_1)^2) (p-p^2) (V_{prom}) \end{aligned} \quad (4.9)$$

La medida para el segundo, tercer y cuarto intervalos son, respectivamente,

$$\begin{aligned}\mu_{01}[1/4, 1/2] &= \mu_1[0, 1/2]W_{01} \\ &= \mu_1[0, 1/2] \left( \binom{2^2}{1} (2^{-n}) ((x_1)^{2^2-1} + (1-x_1)^1) (p-p^2) (V_{prom}) \right)\end{aligned}\quad (4.10)$$

$$\begin{aligned}\mu_{10}[1/2, 3/4] &= \mu_1[1/2, 1]W_{10} \\ &= \mu_1[1/2, 1] \left( \binom{2^2}{2} (2^{-n}) ((x_1)^{2^2-2} + (1-x_1)^2) (p-p^2) (V_{prom}) \right)\end{aligned}\quad (4.11)$$

y

$$\begin{aligned}\mu_{10}[3/4, 1] &= \mu_1[1/2, 1]W_{11} \\ &= \mu_1[1/2, 1] \left( \binom{2^2}{3} (2^{-n}) ((x_1)^{2^2-3} + (1-x_1)^3) (p-p^2) (V_{prom}) \right)\end{aligned}\quad (4.12)$$

Este proceso iterativo se repite y es lo suficientemente interesante, la medida total de la velocidad en cada etapa se conserva. Considerando la etapa  $k$ -ésima de la cascada, la medida de masa de cada subintervalo de la etapa  $k$ -ésima se deriva de pesar la medida de masa de los subintervalos de la etapa anterior ( $k-1$ ). En otras palabras, la medida del primer intervalo  $I_k$  es igual a:

$$\begin{aligned}\mu[0, 2^{-k}] &= \mu[0, 2^{-k+1}] \left( W_{\underbrace{00\dots 0}_{k\text{-dígitos}}} \right) \\ &= \mu[0, 2^{-k+1}] (2^{-k}) \left[ (x_{k-1})^{2^k} + (1-x_{k-1})^{2^k} \right] (p-p^2) (V_{prom})\end{aligned}\quad (4.13)$$

Para los otros intervalos de la etapa  $k$ -ésima tenemos:

$$\begin{aligned}\mu[I_k] &= \mu[I_{k-1}] (W_{b_1 b_2 \dots b_k}) \\ &= \binom{2^k}{i} (2^{-(k-1)}) (x_{k-1})^{2^k-i} (1-x_{k-1})^i (p-p^2) (V_{prom}) \quad | \quad i = 1, \dots, 2^k - 1\end{aligned}\quad (4.14)$$

donde  $b_1 b_2 \dots b_k$  es la representación binaria del número entero  $i$ . Observe que las variables aleatorias  $x_0, x_1, x_2, x_3, \dots$  son independientes e idénticamente distribuidas (i.i.d.) con distribución uniforme en  $[0, 1]$ .

Para  $t = 0.b_1 b_2 \dots b_k = \sum_{i=1}^k b_i 2^{-i}$ , sea  $R(b_1 \dots b_k) = W_{b_1 b_2 \dots b_k}$ , representando el nuevo miembro multiplicador aleatorio incorporado para el subintervalo  $[t, t + \Delta t_k]$  en la etapa  $k$ -ésima. Por lo tanto, la medida en el intervalo  $[t, t + \Delta t_k]$  puede expresarse como:

$$\mu(\Delta t_k) = \mu[t, t + \Delta t_k] = R(b_1) R(b_1 b_2) \dots R(b_1 \dots b_k) \quad (4.15)$$

Consideremos, por ejemplo, la representación como cascada conservadora de la serie La Mata ( $W_{t_1}$ ), a la cual llegaron 80,460 datos en un período de 2 años (2006, 2008). Estos períodos de tiempo serán nuestro intervalo unitario  $I = [0, 1)$ . La medida total del intervalo es  $\mu(I) = 80,460$ . De los cuales llegaron 56,322 minutos en la primera mitad del intervalo,  $I(0) = [0, 1/2)$ , y 24 138 minutos en la segunda mitad,  $I(1) = [1/2, 1)$ , de manera que en la primera etapa el generador de energía  $p$  toma

el valor de 0.7, y  $q = 1 - p = 0.3$ . De manera que este cálculo es de manera aleatoria para los correspondientes valores del generador de energía durante 20 etapas de descomposición en cascada multiplicativa. El procedimiento implementado, permite predecir el siguiente valor de la velocidad del viento para las series medidas cada 10 minutos, estimando los multiplicadores  $W(j_1, j_2, \dots, j_n)$  correspondiente a un nivel anterior deseado. El hecho de llevar a cabo sólo una predicción por muestra, reduce significativamente el error que se comete en el procedimiento total.

Mediante la Fig. 4.10, se explicará de forma detallada el algoritmo de predicción. a) Con líneas continuas se muestra la cascada conservadora del instante de tiempo actual; con líneas punteadas se ilustra la que corresponde a un instante de tiempo futuro. Así pues, usando la propuesta de la cascada multiplicativa para la predicción de los errores de pronóstico, se construye la cascada conservadora, comenzando desde el nivel a predecir (mínima escala de tiempo) y yendo hacia escalas superiores en la cascada. b) Este proceso se lleva acabo mediante un programa para obtener los multiplicadores de la energía  $p$  de manera aleatoria y los valores de las velocidades en el rango de escalas de interés, es decir, la velocidad promedio  $V_{prom}$ . Posteriormente, se selecciona el tiempo de la predicción deseada y así obtener la velocidad de viento que llegará en el siguiente instante de tiempo.

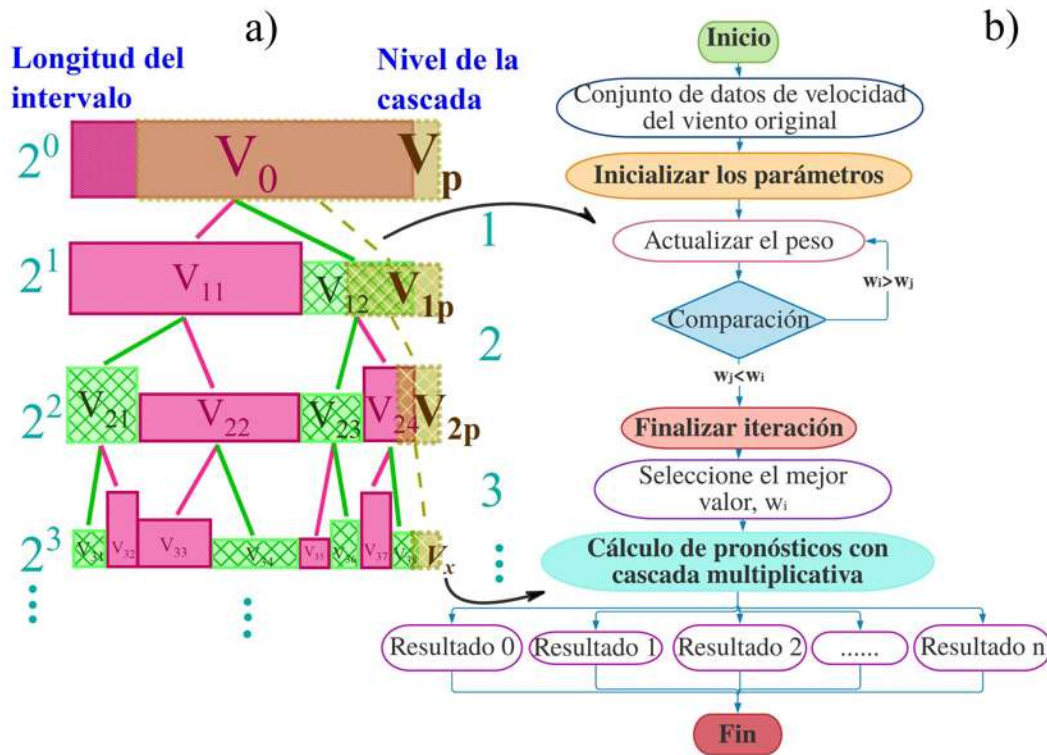


Figura 4.10: Construcción de la cascadas conservadoras correspondientes al instante actual y al instante de tiempo futuro.

### 4.2.6. Conformación de los modelos híbridos

Para conformar los modelos híbridos, se utilizaron los modelos en su forma simple y a éstos se le sumó el patrón multifractal extraído a los errores de pronóstico primarios con la técnica CM, además

de agregar los errores secundarios. Una serie de tiempo completa se expone en la expresión matemática de la Ec. (4.16):

$$y_t = S + CM + P_{f_s} \quad (4.16)$$

donde: *S* son los Modelos Simples (ARIMA, AP y Persistencia), *CM* el Modelo de los errores primarios con Cascada Multiplicativa, *P<sub>f<sub>s</sub></sub>*

### 4.3. Proceso de Modelado

Para desarrollar los modelos de las series transformadas  $W_{t_1}$  y  $W_{t_2}$  con las diferentes técnicas se consideró particionar los datos en un 70 % (entrenamiento), 15 % (validación) y 15 % (prueba), cuya finalidad es la comparación de los resultados de los modelos matemáticos y numéricos.

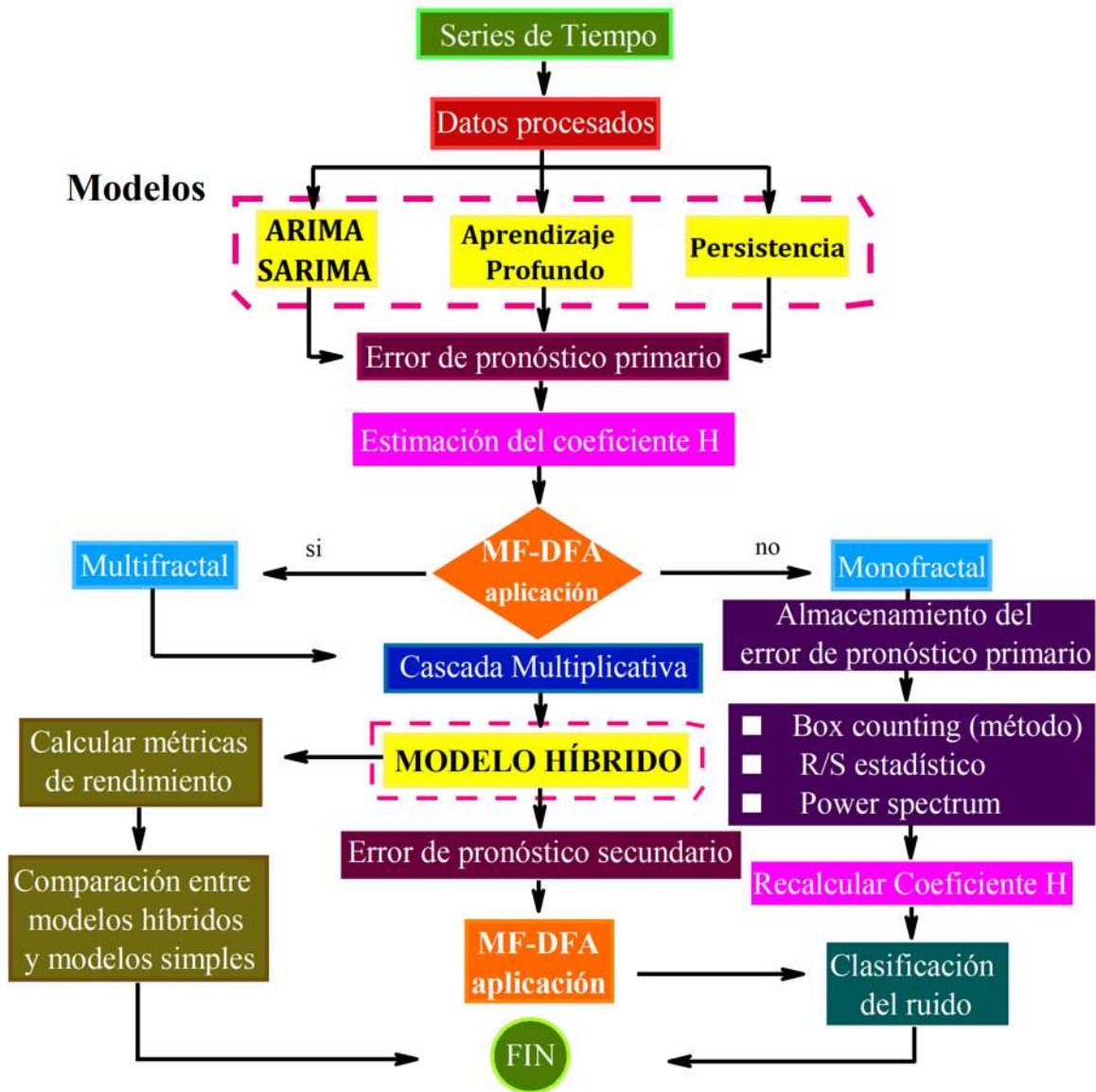


Figura 4.11: Diagrama de flujo del proceso de modelado.

La Fig. 4.11, presenta el diagrama de flujo del proceso de modelado propuesto para las series de tiempo multifractales. En primer lugar, se realizó un tratamiento de los datos a las series  $S_{t_1}$  y  $S_{t_2}$ , determinando estabilidad en varianza y media. Continuando con la aplicación de las técnicas estadísticas (ARIMA para la serie  $W_{t_1}$ , SARIMA para la serie  $W_{t_2}$ ), Aprendizaje Profundo y Persistencia, cuya finalidad fue obtener los distintos modelos. Posteriormente, se obtienen los errores de pronóstico primarios ( $P_{f_e}$ ) para cada modelo respectivamente. Y mediante la técnica MF-DFA, se determina si la serie conformada por los errores es multifractal o monofractal y sus características. Después, si la serie  $P_{f_e}$  resultaba multifractal, se aplicaba la técnica numérica CM para extraer el patrón multifractal de los errores de pronóstico primarios, obteniendo modelos híbridos (ARIMA-CM, AP-CM y Persistencia-CM), a los cuales se les calcula las métricas de rendimiento y finalmente dichas métricas se compararon con las métricas obtenidas de los modelos simples (ARIMA, AP y Persistencia) para determinar qué modelo tiene mejor desempeño. Si la serie  $P_{f_e}$  es monofractal, se modela con

las técnicas Box counting,  $R/S$  statistic y Power spectrum, utilizando el coeficiente  $H$ , de la misma manera que Cadenas et al. [136] y se define qué tipo de ruido se presenta, clasificándolo de la misma manera que Ihlen et al. [60].

### 5.1. Análisis cualitativo del comportamiento de los modelos generados

En la Fig. 5.1, se grafica el comportamiento de 50 puntos generados por los modelos desarrollados y los datos reales correspondientes a La Mata (cada 10 min) y La Venta (cada hora). La finalidad de esta representación es comparar cada modelo simple, con el modelo híbrido correspondiente, en el cual se integra la cascada multiplicativa.

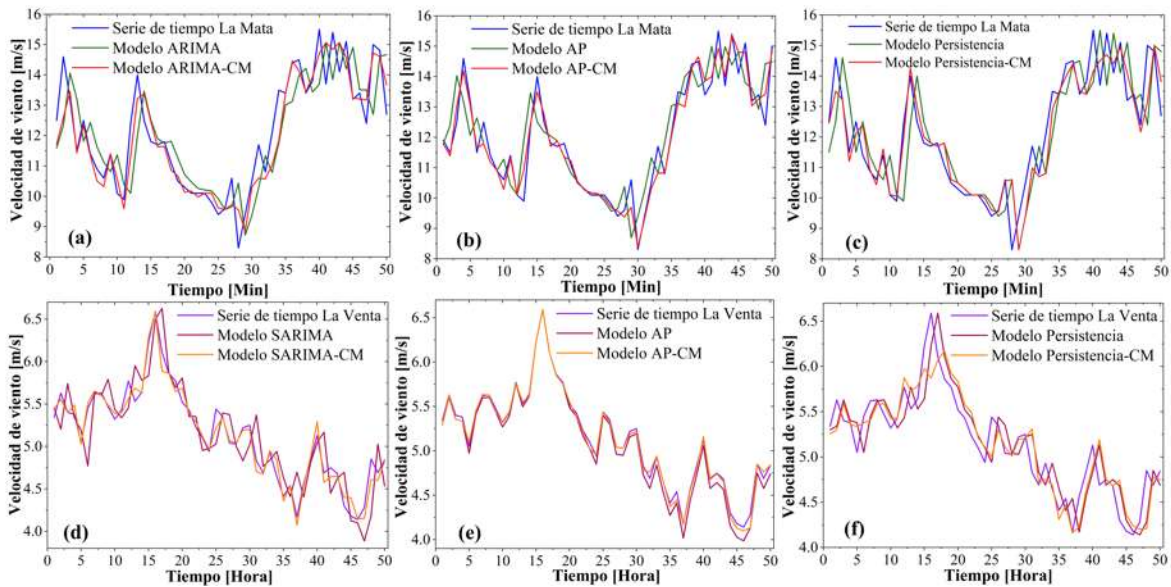


Figura 5.1: Comportamiento de los modelos generados.

En las Figs. 5.1 (a), 5.1 (b) y 5.1 (c) correspondientes a La Mata, se aprecia el mejor ajuste de los modelos híbridos para la mayoría de los puntos representados. Lo anterior se confirma con el cálculo de las medidas de error de pronóstico. Un comportamiento similar se refleja en los modelos de las Figs. 5.1 (d) y 5.1 (e), correspondientes a La Venta. No ocurre el mismo comportamiento en la Fig. 5.1 (f), en donde el modelo de Persistencia comienza con un mejor desempeño que el modelo híbrido Persistencia-CM, sin embargo, para la segunda mitad de los datos representados el modelo híbrido se vuelve competitivo. Para este caso en particular, fue necesario conocer la medida cuantitativa de ajuste para conocer que el modelo Persistencia-CM era mejor.

Se analizaron los errores de predicción primarios mediante la técnica MF-DFA, para detectar las características multifractales de las series de los errores. Posteriormente, se aplicó la técnica numérica CM extrayendo los patrones multifractales, a los 3 determinados. Además de calcular las métricas de rendimiento, las cuales permitieron realizar comparaciones y así determinar cuál modelo mejora la predicción, además de disminuir el error de pronóstico secundario  $P_{fs}$ . Finalmente, se encontró que

los modelos híbridos mejoraron la predicción de la velocidad de viento, alcanzando desde 0.331 % a 5.285 % de ganancia y un índice de comparación entre los modelos desde un 0.995  $m/s$  hasta un 0.999  $m/s$ .

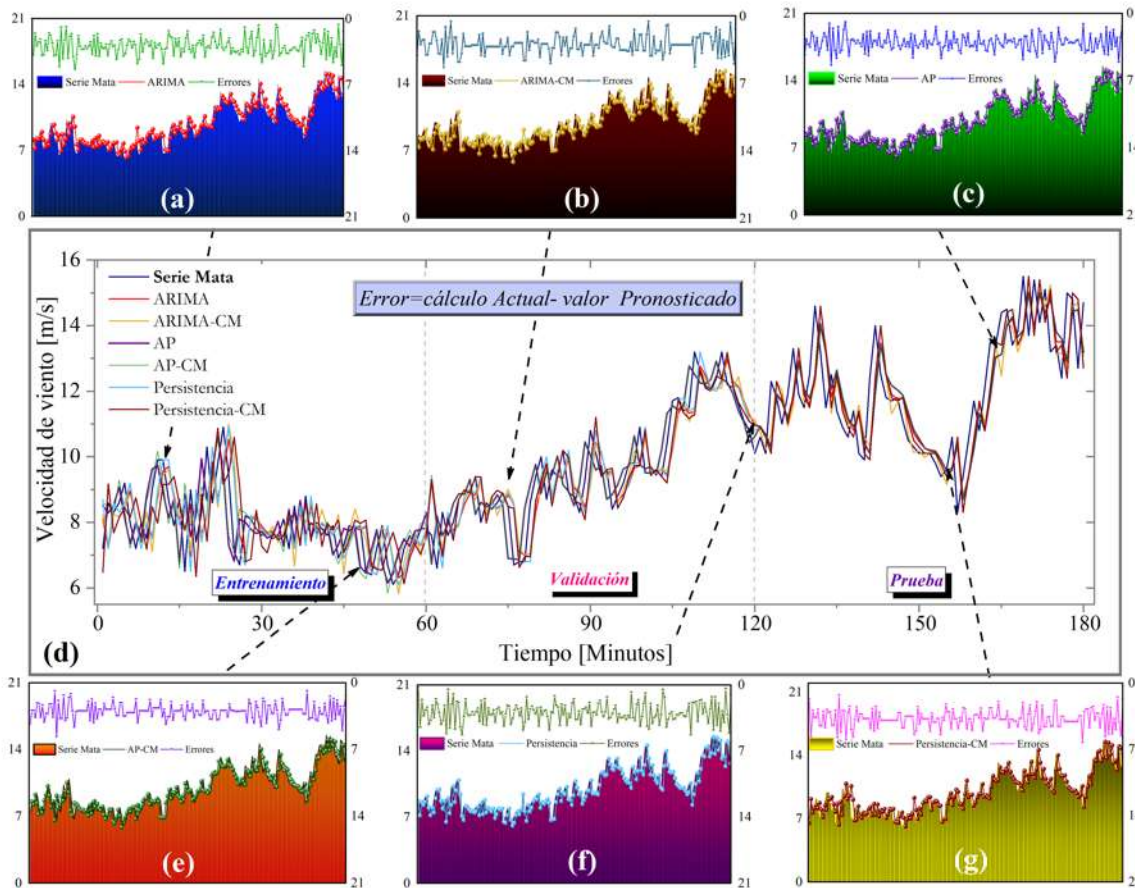


Figura 5.2: Resultados de los modelos simples, modelos híbridos y errores de pronóstico para la serie La Mata.

En la Fig. 5.2 y 5.3 incisos (a) y (b) se pueden apreciar las comparaciones de los modelos híbridos, modelos simples y series reales. Los incisos (a), (c) y (f) muestran el comportamiento de los modelos simples, los incisos (b), (e) y (g) los modelos híbridos para ambas series, resultado de utilizar la técnica numérica CM, la cual permitió la extracción del patrón multifractal a los errores  $P_{fe}$  y, con ello, mejores ajustes de los modelos respecto de las series reales, además de la disminución de los errores secundarios. Cabe mencionar que sólo se muestran 180 datos para los comportamientos del entrenamiento (0-60 datos), validación (61-120 datos) y prueba (121-180 datos), como una representativa de las series completas.

Notándose, el mejor ajuste se observó en los incisos (c) y (e) modelos AP y AP-CM en ambas series. Finalmente, mencionar que el modelo híbrido (incisos (g)) para la persistencia en ambas series, visualmente muestra fluctuaciones considerables, estadísticamente se corroboró que el modelo híbrido sigue siendo mejor.

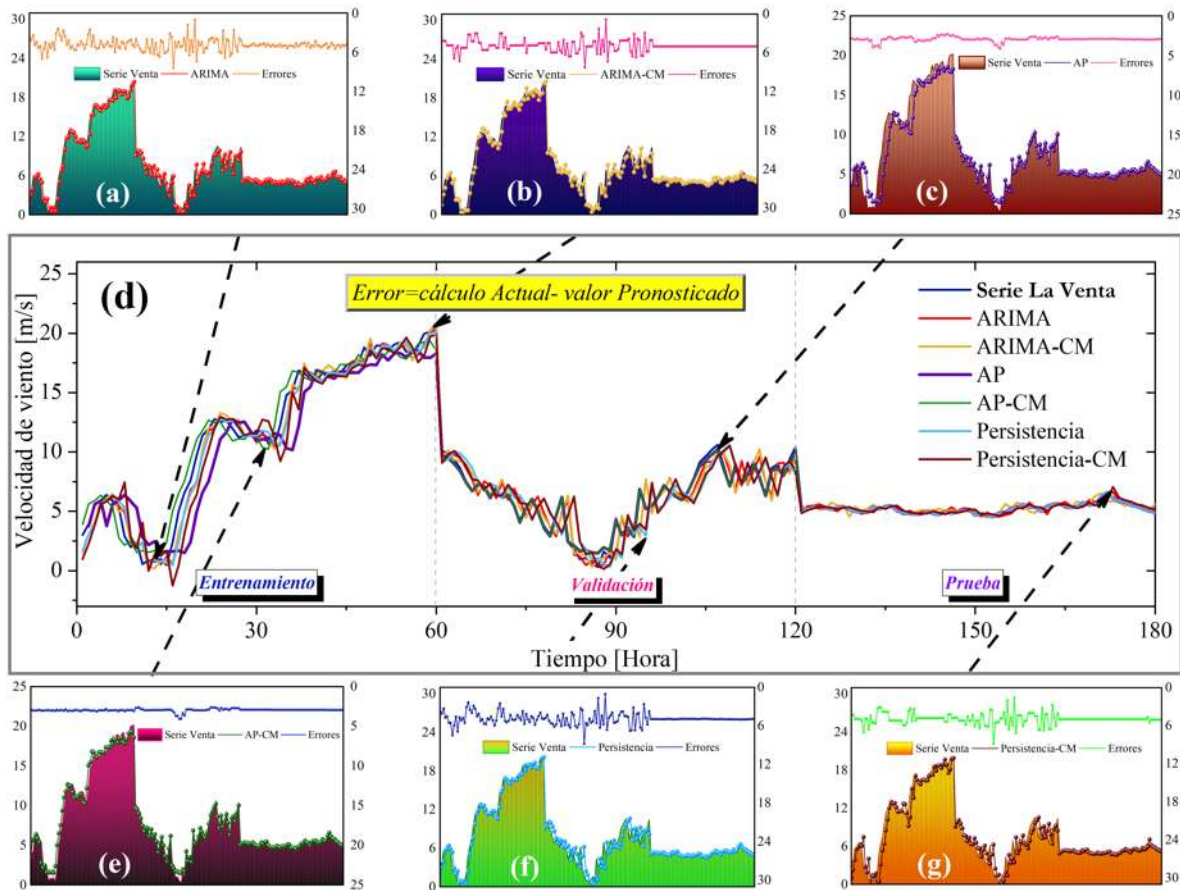


Figura 5.3: Resultados de los modelos simples, modelos híbridos y errores de pronóstico para la serie La Venta.

## 5.2. Características multifractales de los errores de los modelos

A continuación, se detallan las características multifractales de los errores primarios  $P_{fe}$  de los modelos simples (ARIMA, AP y Persistencia), respectivamente. Del estudio realizado, se logró determinar que los errores de las series de La Mata y La Venta fueron multifractales, es decir, conservaban un patrón, el cual se podría extraer mediante la técnica CM.

De la Fig. 5.4 a 5.7, revelan los resultados de aplicar la técnica MF-DFA a los errores primarios para los modelos ARIMA, AP y Persistencia para la serie La Mata.

La Fig. 5.4 muestra la caminata aleatoria de los errores primarios para los modelos del conjunto entrenamiento de serie La Mata. Así, la figura muestra el comportamiento para cada uno de los diferentes modelos utilizados, donde muestra una dependencia del dato anterior.

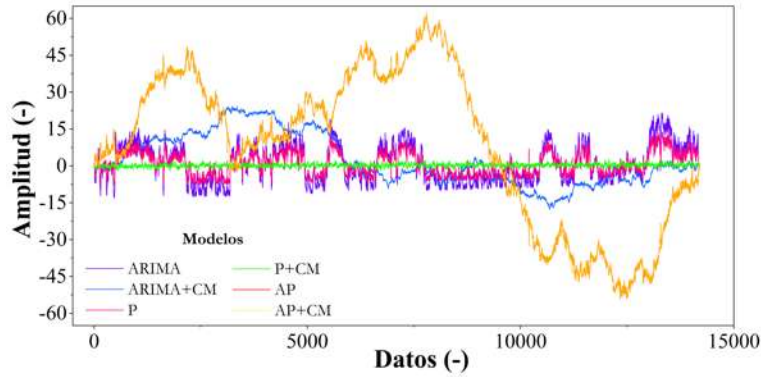


Figura 5.4: Caminata aleatoria para los errores primarios en cada uno de los modelos.

La Fig. 5.5 muestra las funciones de fluctuación, están ajustadas, en el que las pendientes  $H_q$  de las líneas de regresión son  $q$  dependientes de series de tiempo multifractal, presentando líneas rectas en escalas log-log.

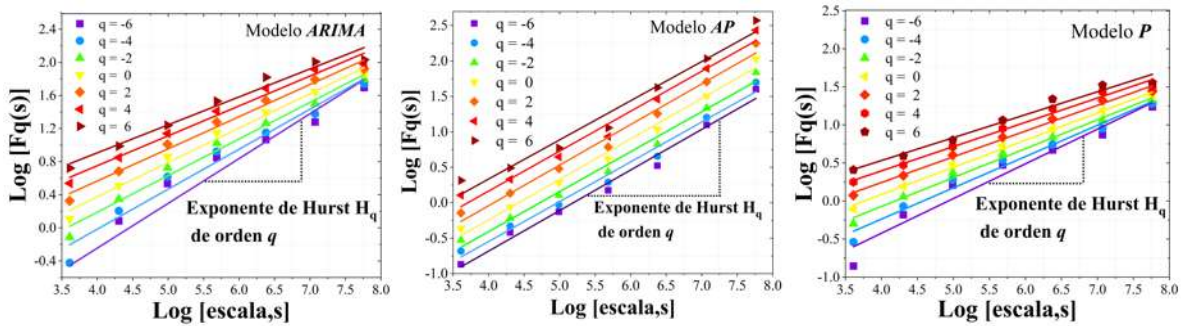


Figura 5.5: Función de fluctuación  $F_q(s)$  para cada uno de los modelos.

La Fig. 5.6 indica la dependencia del exponente generalizado de Hurst con  $q$ . Observándose que las funciones están disminuyendo para los modelos simple, lo que demuestra la existencia de patrones de multifractalidad en el tiempo Según [173]. Aunque con mayor decrecimiento para el modelo ARIMA y Persistencia, dejando notar mayor multifractalidad para ambos modelos con respecto al modelo AP.

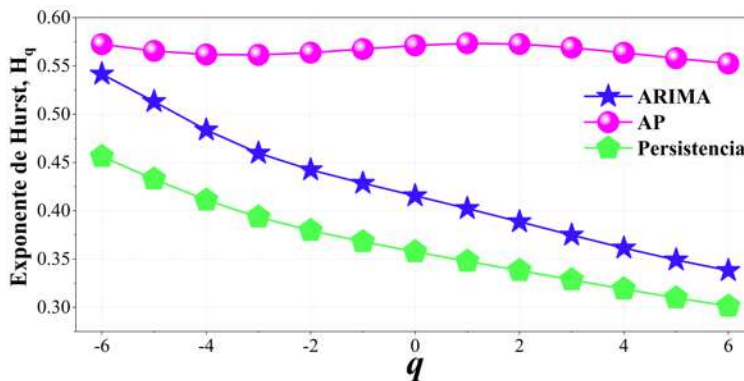


Figura 5.6: Exponentes generalizados de Hurst de los tres modelos simples.

Finalmente, la Fig. 5.7 muestra los espectros multifractales de los modelos, exhiben una sola cresta, que generalmente describe la serie multifractal como lo mencionan Balkissoon et al. [173]. El ancho del espectro multifractal de los modelos ARIMA y AP es mayor que el del modelo de Persistencia, lo que indica que a mayor ancho, existen mayores fluctuaciones, mientras a menor ancho hay menos fluctuaciones a lo largo de la serie. Además, como Cabrera menciona que cuando aparecen espectros que son simétricos, se mide como  $\alpha_{max} - \alpha_{min}$ . Por tanto, un factor importante para los cambios de escala múltiple en la volatilidad de la velocidad del viento son la velocidad de vientos fuertes y persistentes.

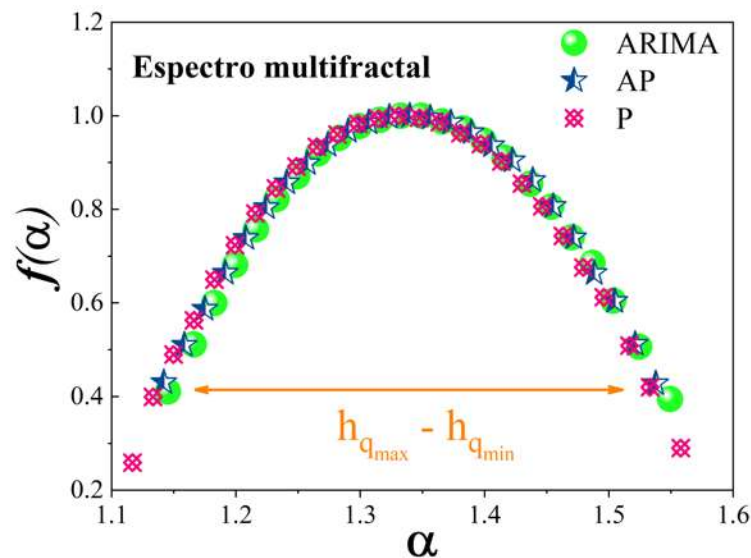


Figura 5.7: Espectro multifractal  $f(\alpha)$  para cada uno de los modelos.

Análogo para la serie La Venta, obteniéndose para cada uno de los errores primarios de modelos la presencia de la multifractalidad. De la Fig. 5.8 a 5.11, revelan los resultados de aplicar la técnica MF-DFA para los modelos SARIMA, AP y Persistencia para la serie La Venta.

La Fig. 5.8 muestra la caminata aleatoria para los errores primarios de los modelos del conjunto entrenamiento de serie La Venta. Así, esta figura muestra un comportamiento totalmente distinto entre los diferentes modelos utilizados, y, lo único en común es la dependencia del dato anterior.

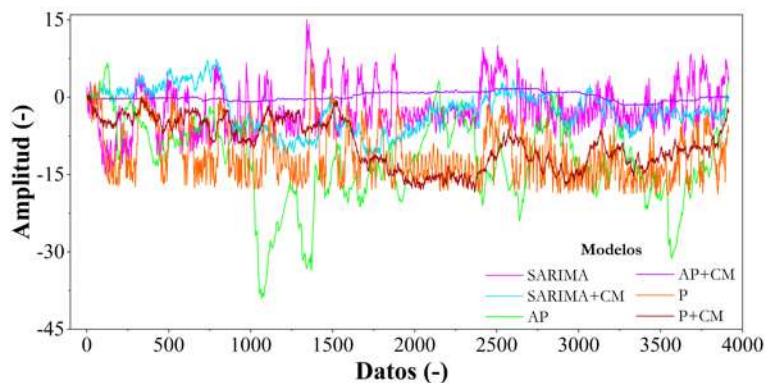


Figura 5.8: Caminata aleatoria para los errores primarios en cada uno de los modelos.

La Fig. 5.9 muestra las pendientes  $H_q$  de los incisos (c) y (d) son ligeramente mayores que en el inciso (b). Como lo establece Kantelhard [49] si existen entre dichas escalas regímenes lineales, con diferentes pendientes, es decir existen diferentes exponentes de Hurst.

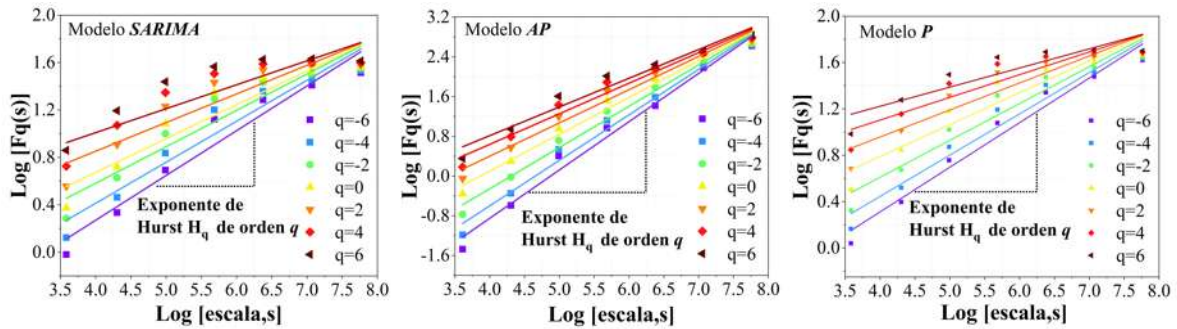


Figura 5.9: Muestran la función de fluctuación  $F_q(s)$  para cada uno de los modelos.

La Fig. 5.10 indica la dependencia del exponente generalizado de Hurst con  $q$ . Observándose que las funciones fueron disminuyendo para  $q$ 's positivos, es decir,  $H_q$  describe el comportamiento de escala de los segmentos con fluctuaciones altas, caso contrario para  $q$ 's negativos, donde el comportamiento son fluctuaciones pequeñas, como lo menciona Balkissoon et al. [173] en su estudio. Probando con ello, la existencia de patrones de multifractalidad en el tiempo. Mostrando mayor multifractalidad AP con respecto a los modelos SARIMA y Persistencia, indicando un decrecimiento menos pronunciado para ambos modelos.

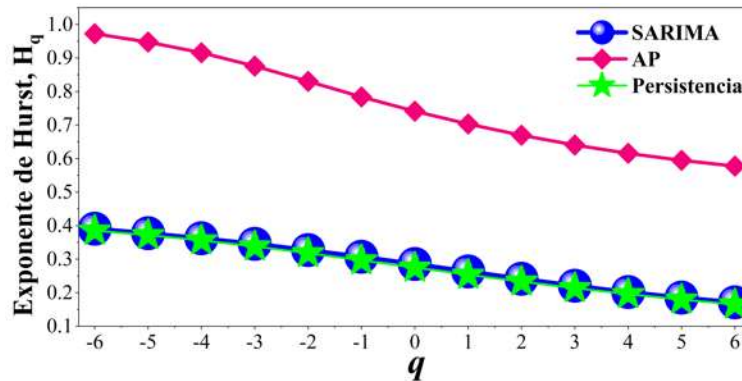


Figura 5.10: Exponentes generalizados de Hurst de los tres modelos simples.

En la Fig. 5.11 el ancho del espectro multifractal de los incisos de los modelos SARIMA y Persistencia son menores que el del modelo AP, indicando que existen menos fluctuaciones a lo largo de la serie para los modelos SARIMA y Persistencia. Por consiguiente y de manera análoga a la serie de La Mata, la variedad de modelos de series de tiempo de viento confirma que la velocidad de viento fuertes y persistentes es un factor importante para los cambios de escala múltiple en la volatilidad de la velocidad del viento.

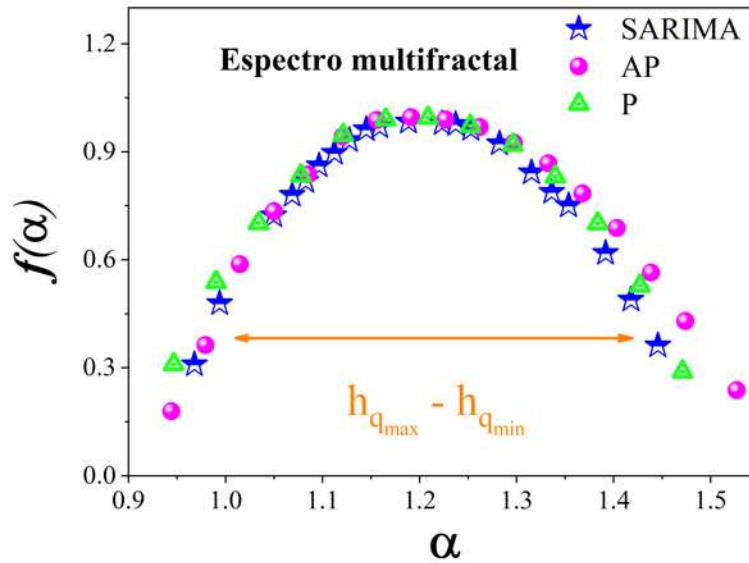


Figura 5.11: Espectro multifractal  $f(\alpha)$  para cada uno de los modelos.

### 5.3. Extracción del patrón multifractal en los errores de pronóstico de los modelos híbridos

Para conocer el patrón multifractal se graficaron los patrones obtenidos después de analizar los errores primarios mediante la técnica CM, los cuales se encuentran inmersos en la Ec. (4.16).

La Fig. 5.12 describe los patrones obtenidos en las series La Mata (Fig. 5.12 (a)) y La Venta (Fig. 5.12 (b)) en las diferentes técnicas utilizadas. Es decir, representan los diferentes  $H_q$  característico de la multifractalidad, y su apreciación para los errores de estos modelos, ya que se agrupan en paquetes. En la Fig. 5.12 es posible apreciar espigas de diferentes magnitudes, sin embargo, en el inciso (a) el patrón multifractal para el modelo Persistencia es posible identificarlo con mayor facilidad, por los grupos de tamaños iguales, así como en el inciso (b) el patrón multifractal para ARIMA y Persistencia.

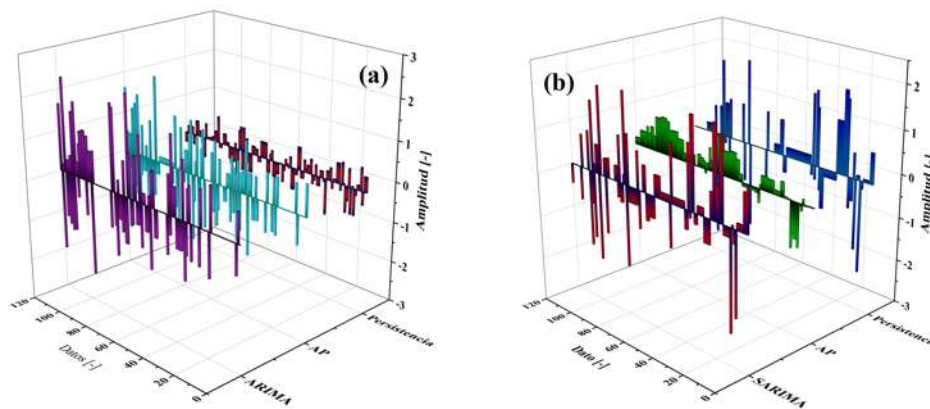


Figura 5.12: (a) Patrones multifractales extraídos  $P_{fe}$  en los diferentes modelos simples mediante la técnica de la cascada multiplicativa para (a) la serie La Mata y (b) la serie La Venta.

## 5.4. Comparación de las métricas de rendimiento entre los modelos simples y modelos híbridos

Para analizar la potencia de una técnica de predicción, es necesario revisar el error de pronóstico. Las Tablas 5.1 y 5.2 muestran los resultados obtenidos de las métricas de rendimiento para las series de tiempo La Mata y La Venta, las cuales fueron divididas en tres conjuntos (entrenamiento, validación y prueba), para poder comparar las técnicas utilizadas con la técnica de AP. En efecto, el 70 % de los datos que fue utilizado para el entrenamiento de los modelos ARIMA/SARIMA, AP y Persistencia, en particular para el caso de AP, además determina los parámetros que definen el modelo de red neuronal. Por otro lado, el 15 % de los datos para la validación de los modelos principalmente en el modelo AP sirve para vigilar el fenómeno de sobreajuste mediante validación cruzada. Por último, el 15 % restante de los datos, es utilizado para la prueba de los modelos, permitiendo así, probar que el modelo propuesto puede generalizar y pronosticar.

Tabla 5.1: Comparación estadística de los datos de entrenamiento, validación y prueba entre los modelos simples y los modelos híbridos para la serie de tiempo La Mata.

Datos	Modelos	MAE (m/s)	MSE (m/s)	RMSE (m/s)	MAPE (%)	IA -
Entrenamiento (70 %)	ARIMA	0.643	0.7666	0.875	10.370	0.998
	<b>ARIMA-CM</b>	<b>0.492</b>	<b>0.6020</b>	<b>0.776</b>	<b>7.221</b>	<b>0.999</b>
	AP	0.558	0.5860	0.765	10.297	0.999
	<b>AP -CM</b>	<b>0.407</b>	<b>0.4680</b>	<b>0.684</b>	<b>7.257</b>	<b>0.999</b>
	Persistencia	0.656	0.8080	0.899	9.988	0.999
	<b>Persistencia-CM</b>	<b>0.449</b>	<b>0.5840</b>	<b>0.764</b>	<b>6.580</b>	<b>1.000</b>
Validación (15 %)	ARIMA	0.618	0.698	0.836	15.400	0.995
	<b>ARIMA-CM</b>	<b>0.589</b>	<b>0.663</b>	<b>0.814</b>	<b>14.306</b>	<b>0.995</b>
	AP	0.557	0.770	0.759	16.277	0.999
	<b>AP -CM</b>	<b>0.406</b>	<b>0.452</b>	<b>0.673</b>	<b>11.307</b>	<b>0.999</b>
	Persistencia	0.620	0.708	0.842	14.604	0.998
	<b>Persistencia-CM</b>	<b>0.513</b>	<b>0.602</b>	<b>0.776</b>	<b>12.392</b>	<b>0.998</b>
Prueba (15 %)	ARIMA	0.729	1.040	1.020	8.537	0.999
	<b>ARIMA-CM</b>	<b>0.609</b>	<b>0.900</b>	<b>0.949</b>	<b>6.176</b>	<b>0.999</b>
	AP	0.615	0.739	0.860	9.285	0.999
	<b>AP -CM</b>	<b>0.550</b>	<b>0.681</b>	<b>0.825</b>	<b>8.075</b>	<b>0.999</b>
	Persistencia	0.751	1.101	1.049	8.281	0.999
	<b>Persistencia-CM</b>	<b>0.669</b>	<b>1.001</b>	<b>1.000</b>	<b>6.646</b>	<b>0.999</b>

A continuación con la Fig. 5.13 se verificó que los resultados obtenidos de la Tabla 5.1 fueron satisfactorios. Por ejemplo, los errores MAE para los modelos híbridos tienen mejor rendimiento que los modelos simples, esto a razón de cuanto más cercano sea el valor MAE a cero, mejor es el rendimiento del modelo, asimismo en RMSE los modelos híbridos son inferiores a los modelos simples. En relación a los tres conjuntos, también MAPE indica que los modelos híbridos alcanzan mejor exactitud, apreciándose en la validación.

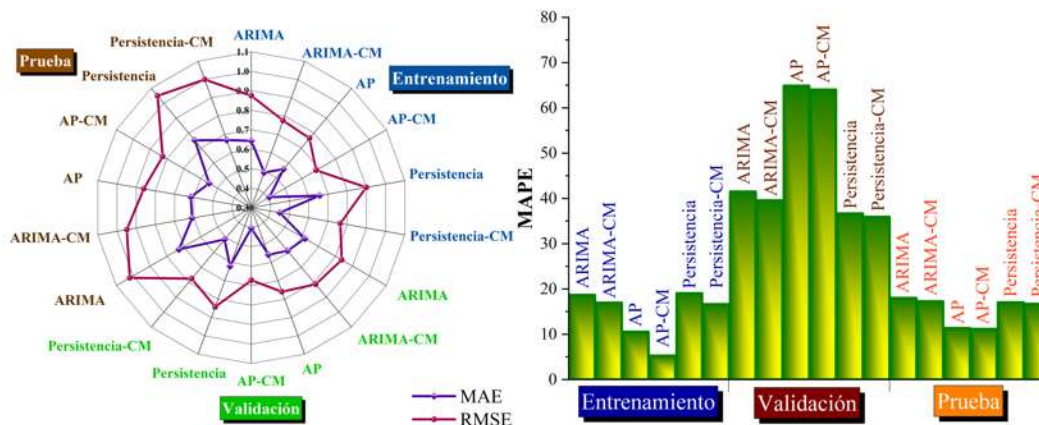


Figura 5.13: Resultados del pronóstico de los diferentes modelos y las métricas de pronóstico de MAE, RMSE y MAPE para cada modelo de la serie La Mata.

También se analizaron las métricas para la serie La Venta y los mejores modelos fueron los modelos híbridos de manera particular AP-CM alcanzando un IA de 0.999. La precisión de predicción para el 70 % de los datos de entrenamiento fue hasta un 16.96 %. Para la validación, se obtuvo que RMSE es menor en modelos híbridos reflejando con ello la desviación entre los datos observados y los originales. Además, en MSE se alcanzó una reducción de 0.214 % a 0.243 % en modelo híbridos. Finalmente, la prueba mostró también que la reducción de los errores fue alcanzada en los modelos híbridos.

Tabla 5.2: Comparación estadística de los datos de entrenamiento, validación y prueba entre los modelos simples y los modelos híbridos para la serie de tiempo La Venta.

Datos	Modelos	MAE	MSE	RMSE	MAPE	IA
		(m/s)	(m/s)	(m/s)	(%)	(-)
Entrenamiento (70 %)	SARIMA	0.819	1.332	1.154	18.659	0.994
	<b>SARIMA-CM</b>	<b>0.725</b>	<b>1.194</b>	<b>1.093</b>	<b>16.962</b>	<b>0.995</b>
	AP	0.251	0.120	0.346	10.543	0.999
	<b>AP -CM</b>	<b>0.243</b>	<b>0.114</b>	<b>0.337</b>	<b>5.258</b>	<b>0.999</b>
	Persistencia	0.922	1.378	1.174	19.057	0.998
	<b>Persistencia-CM</b>	<b>0.649</b>	<b>1.096</b>	<b>1.046</b>	<b>16.631</b>	<b>0.998</b>
Validación (15 %)	SARIMA	0.881	1.491	1.221	41.498	0.996
	<b>SARIMA-CM</b>	<b>0.789</b>	<b>1.353</b>	<b>1.163</b>	<b>39.551</b>	<b>0.996</b>
	AP	0.218	0.110	0.331	64.897	0.999
	<b>AP -CM</b>	<b>0.216</b>	<b>0.108</b>	<b>0.328</b>	<b>64.034</b>	<b>0.999</b>
	Persistencia	0.922	1.571	1.253	36.672	0.995
	<b>Persistencia-CM</b>	<b>0.871</b>	<b>1.475</b>	<b>1.215</b>	<b>35.905</b>	<b>0.996</b>
Prueba (15 %)	SARIMA	0.809	1.285	1.133	17.996	0.998
	<b>SARIMA-CM</b>	<b>0.769</b>	<b>1.234</b>	<b>1.111</b>	<b>17.245</b>	<b>0.998</b>
	AP	0.220	0.090	0.301	11.355	0.999
	<b>AP -CM</b>	<b>0.214</b>	<b>0.086</b>	<b>0.293</b>	<b>11.163</b>	<b>0.999</b>
	Persistencia	0.804	1.243	1.115	17.047	0.998
	<b>Persistencia-CM</b>	<b>0.772</b>	<b>1.202</b>	<b>1.096</b>	<b>16.716</b>	<b>0.998</b>

## 5.4 Comparación de las métricas de rendimiento entre los modelos simples y modelos híbridos

De manera análoga, la Fig. 5.14 es la representación gráfica de los resultados obtenidos en la Tabla 5.2, donde se puede corroborar, que los errores MAE para los modelos híbridos tienen un mejor rendimiento, puesto que cuanto más cercano sea el valor de MAE a cero, mejor el rendimiento respecto a los modelos simples, además la métrica RMSE refleja la desviación estándar, denotando también que los errores de los modelos híbridos son inferiores a los modelos simples. Por otra parte, la métrica MAPE indica el desempeño del pronóstico del modelo, asimismo en la Fig. 5.14, es claro que los modelos híbridos tienen mejor desempeño en comparación a los modelos simples para los tres conjuntos.

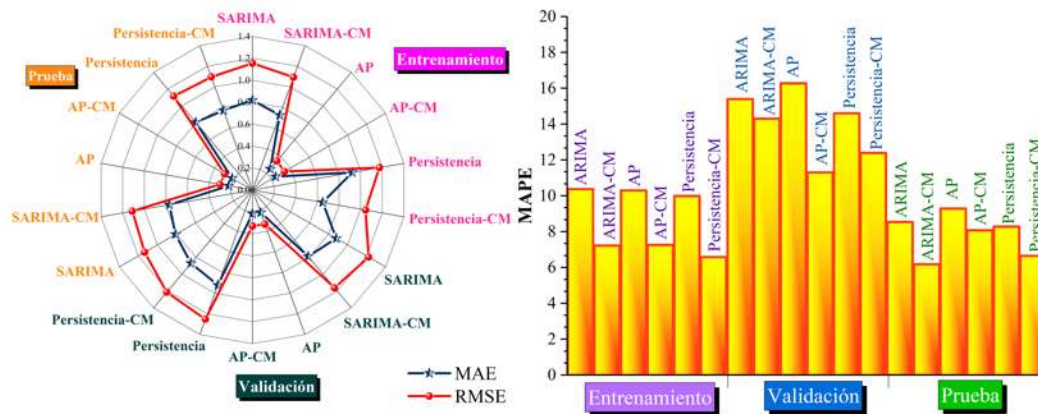


Figura 5.14: Resultados del pronóstico de los diferentes modelos y las métricas de pronóstico de MAE, RMSE y MAPE para cada modelo de la serie La Venta.

Este trabajo de investigación tuvo como objetivo la extracción de un patrón multifractal a los errores de pronóstico de los diferentes modelos ARIMA, Aprendizaje Profundo y Persistencia mediante la técnica de la Cascada Multiplicativa, cuya finalidad fue mejorar la predicción un paso adelante con modelos híbridos ARIMA-CM, Aprendizaje Profundo-CM y Persistencia-CM. Para demostrar esto, primero se realiza un análisis con la técnica MF-DFA a las series de tiempo de viento. Partiendo de los datos de dos series temporales; la primera serie, medida en el sitio denominado La Mata espaciada cada 10 minutos, y la segunda en La Venta, espaciada cada hora, ambos lugares en Oaxaca al sur de México. Las observaciones fueron las siguientes:

1. Los modelos ARIMA y AP fueron mejores que el modelo de la Persistencia, siendo este último considerado como el modelo más simple en cuestiones de predicciones.
2. Las estructuras de los errores de pronóstico para las series de tiempo en La Mata y La Venta, Oaxaca, México, son multifractales ya que, entre las escalas de la función fluctuación  $F_q(s)$  se encuentran diferencias de órdenes de magnitud según lo establece Kantelhard [49]. Y entre dichas escalas existen regímenes lineales, con diferentes pendientes, es decir diferentes exponentes de Hurst. Además, los exponentes generalizados de Hurst, para cada uno de los períodos, muestran una curva suave con cambios de pendiente, característica que definen la multifractalidad en ambas series. Asimismo, las curvas características de los espectros multifractales, en este caso asimétricas con truncamiento en los momentos negativos.
3. En ambas series de tiempo los errores de pronóstico generados por los modelos de predicción ARIMA, AP y la Persistencia fue posible extraer dicho patrón con la técnica de Cascada Multiplicativa.
4. Por otro lado, se considera que los errores de los modelos simples (ARIMA, AP y Persistencia) principalmente de ARIMA son ruido blanco, sin embargo, desde el enfoque fractal/multifractal se ha demostrado la existencia de patrones multifractales cuando se utiliza la técnica MF-DFA.
5. Se comparó el desempeño de los modelos individuales respecto a los modelos híbridos, encontrando que el ajuste de los modelos híbridos superan a los modelos individuales para ambas series. En el análisis cuantitativo se utilizaron diferentes métricas de error, siendo los modelos híbridos los más competitivos. Por ejemplo, en el error porcentual absoluto medio, las diferencias de rendimiento entre los modelos híbridos y los convencionales para el conjunto prueba fueron: 2.364, 1.210 y 1.635 puntos porcentuales para la serie temporal de diez minutos. En cambio, estas diferencias son de 0.751, 0.192 y 0.331 puntos porcentuales para la serie temporal horaria, demostrando que es conveniente el análisis de los errores de pronóstico generados por los modelos convencionales, con la finalidad de mejorar los pronósticos que ayuden en la planeación y despacho de la energía.
6. Finalmente, es importante comentar que la técnica de cascada multiplicativa, puede aplicarse a cualquier modelo de predicción cuya descomposición incluya errores de pronóstico equiespaciados.

## 6.1. Trabajos Futuros

Como continuación de este trabajo de tesis y como en cualquier otro proyecto de investigación, existen diversas líneas de investigación que quedan abiertas y en las que es posible continuar trabajando. Durante el desarrollo de esta tesis han surgido algunas líneas futuras que se han dejado abiertas y que se esperan atacar en un futuro; algunas de ellas, están más directamente relacionadas con este trabajo de tesis y son el resultado de cuestiones que han ido surgiendo durante la realización de la misma. Otras, son líneas más generales que, sin embargo, no son objeto de esta tesis; estas líneas pueden servir para retomarlas posteriormente o como opción a trabajos futuros para otros investigadores.

A continuación se presentan algunos trabajos futuros que pueden desarrollarse como resultado de esta investigación o que, por exceder el alcance de esta tesis, no han podido ser tratados con la suficiente profundidad. Además, se sugieren algunos desarrollos específicos para apoyar y mejorar el modelo y metodología propuestos. Entre los posibles trabajos futuros se destacan:

- Implementar el modelo y metodología propuesta tanto en diferentes dependencias y sectores de energía eólica, solar, económica, etc., lo que permitirá comprobar su eficacia, y además adecuar mejorar tanto el modelo como la metodología y herramientas propuestas.
- Implementar una interface completa y amigable para realizar el modelo y el pronóstico de cualquier serie de tiempo de viento que el usuario requiera.

## 6.2. Expectativas de posibles desarrollos

Se ha establecido una línea de investigación relativa al uso de la fractalidad de series de tiempo de viento que se demuestra con los artículos publicados (ver Anexo C) y la investigación en proceso y continuar desarrollando esa línea.

### A

- **Atributos Multifractales.**- Son muy utilizados en la práctica debido que permiten resumir la información multifractal de una señal en pocos coeficientes, a saber,  $\zeta(q)$ ,  $D(q)$ ,  $h(q)$  y  $c_m$ .

### C

- **Caminata Aleatoria o paseo aleatorio.**- Abreviado en inglés como RW (Random Walks), es una formalización matemática de la trayectoria, que describe una ruta que consiste en una sucesión de pasos aleatorios en algún espacio matemático. Los paseos aleatorios explican los comportamientos observados de muchos procesos en campos como la computación, la física, la ecología o la economía y, por lo tanto, sirven como modelo fundamental para la actividad estocástica registrada.
- **Coefficiente de Hurst ( $H$ ).**- Es un estadístico con media cero, expresado en términos de desviación estándar, cuyo análisis es de tipo no paramétrico. Fue estudiado como elemento para distinguir series fractales.

### D

- **Dimensión Fractal.**- Es un exponente que da cuenta de cuán completamente parece llenar un fractal el espacio conforme se amplía el primero hacia escalas más y más finas. En sentido genérico es el número que sirve para cuantificar el grado de irregularidad y fragmentación de un conjunto geométrico o de un objeto natural.

### E

- **Espectro Multifractal o Espectro de singularidad.**- Es una función utilizada en análisis Multifractal para describir la dimensión fractal de un subconjunto de puntos de una función perteneciente a un grupo de

puntos que tienen el mismo exponente de Hölder. Intuitivamente, el espectro de singularidad da un valor de cuán “fractal” es un conjunto de puntos en una función. Más formalmente, el espectro de singularidad de una función, se define como:  $f(\alpha)$ . La grafica de  $f(\alpha)$  frente a  $\alpha$  tiene un aspecto típico en forma de parábola cóncava cuyo máximo corresponde a  $q = 0$ . El ancho representa las desviaciones respecto a la estructura fractal media, dada por el exponente de escala del análisis monofractal, DFA, y generalmente, se mide como  $\alpha_{max} - \alpha_{min}$ . No obstante, cuando aparecen espectros que no son simétricos el ancho se calcula como el doble de la distancia que existe entre el  $\alpha$  correspondiente a la cola más corta de la parábola y el  $\alpha$  correspondiente al pico máximo, donde  $f(\alpha) = 1$ . En estos casos, para  $q < 0$  el espectro tendrá la cola izquierda más larga que la derecha, lo que implica que la serie presenta un comportamiento multifractal dominado por las fluctuaciones a largo plazo. Para  $q > 0$  ocurrirá lo contrario y el comportamiento multifractal estará denominado por las fluctuaciones a corto plazo. Una serie monofractal, donde  $q = 2$ , tendrá un espectro de ancho prácticamente nulo y centrado en  $\alpha = H_q = 0.5$

- **Exponente Generalizado de Hurst  $H(q)$ .**- Describe el comportamiento de escala de la función de fluctuación para diferentes órdenes  $q$ . Así, para  $q > 0$ , se tienen valores bajos de  $H_q$  que caracterizan segmentos de fluctuaciones largas, mientras que para  $q < 0$ ,  $H_q$  presenta valores elevado que describen segmentos con pequeñas fluctuaciones.

## F

- **Función de fluctuación  $F_q(s)$ .**- Se evalúa no sólo para la varianza como se hace en el DFA (Detrended Fluctuation Analysis, sus siglas en inglés),  $q = 2$ , sino que lo hace para estadísticos de orden tanto inferior como superior, es decir,  $q$  sólo toma valores reales y debe tomar tanto valores positivos como negativos con el fin de evaluar la función de fluctuaciones para periodos temporales y cortos.
- **Función de escalamiento.**- Brindan una descripción global de los valores que toma el exponente de Hölder. El exponente de escalamiento

también así conocido, debido a que depende del orden  $q$ , entonces, el comportamiento de invarianza a la escala de  $X$ , íntimamente relacionado con su regularidad local. De hecho, si se considera el efecto de variar el orden  $q$ , se observa que valores positivos y altos de  $q$  resaltarán los coeficientes  $T_X(a, x_k)$  de mayor valor absoluto, correspondientes a las regiones más irregulares de  $X$ . Por el contrario, los valores negativos de  $q$  resaltarán aquellos coeficientes más cercanos a 0, caracterizando las regiones más suaves y regulares de  $X$

- **Fractal.-** El fractal es un conjunto matemático que posee la propiedad de autosimilaridad. En otras palabras, es un objeto en el que el todo es exacta o aproximadamente similar a una parte de sí mismo: es decir, el todo tiene la misma forma que una o varias de sus partes.

## M

- **Monofractal.-** Se caracterizan por solo tener un exponentes de Hurt, mientras que las series multifractales se caracterizan con múltiples exponentes de Hurts y una dependencia no lineal de  $\tau(q)$  con  $q$ .
- **Multifractal.-** Es una generalización de un sistema fractal en el que un solo exponente (la dimensión fractal) no es suficiente para describir su dinámica; en cambio, se necesita un espectro continuo de exponentes (el llamado espectro de singularidad). Es decir, la suma de fractalidades de procesos que se pueden separar.

## P

- **Pronóstico.-** Son un método que se utiliza ampliamente en el análisis de las series de tiempo para predecir una variable de respuesta, como ganancias mensuales, comportamiento de acciones o cifras de desempleo, para un período de tiempo determinado. Los pronósticos se basan en patrones de datos existentes.

### B.1. Procesos ARIMA

Los modelos generales son donde las series de tiempo no son estacionarias, por lo que no es posible aplicar procesos AR, MA o ARMA estacionarios directamente. Una forma posible de manejar series no estacionarias es aplicar diferenciaciones para hacerlas estacionarias. Las primeras diferencias, a saber,  $(y_t - y_{t-1}) = (1 - B)y_t$ , pueden diferenciarse para dar segundas diferencias, y así sucesivamente. Las  $d$ -ésimas diferencias se pueden escribir como  $(1 - B)^d y_t$ . Si la serie de datos original se diferencia  $d$  veces antes de ajustar un proceso ARMA( $p, q$ ), entonces se dice que el modelo para la serie original no diferenciada es un proceso ARIMA( $p, d, q$ ) donde la letra  $I$  en el acrónimo significa integrado y la  $d$  denota el número de diferencias tomadas. Matemáticamente se generaliza para dar:

$$\phi(B)(1 - B)^d y_t = \theta(B)Z_t \quad (\text{B.1})$$

El operador AR combinado ahora es  $\phi(B)(1 - B)^d$ . Si reemplazamos el operador  $B$  en esta expresión con una variable  $y$ , se puede ver de inmediato que la función  $\phi(y)(1 - y)^d$  tiene  $d$  raíces en el círculo unitario (como  $(1 - y) = 0$  cuando  $x = 1$ ), lo que indica que el proceso no es estacionario, por lo que, por supuesto, es necesario diferenciar. Teniendo en cuenta que cuando  $\phi(B)$  y  $\theta(B)$  son ambos iguales a la unidad (de modo que  $p$  y  $q$  son ambos cero) y  $d$  es igual a uno, entonces el modelo se reduce a un Modelo ARIMA(0, 1, 0), dado por:

$$y_t - y_{t-1} = Z_t \quad (\text{B.2})$$

Obviamente, esto es lo mismo que el modelo de caminata aleatoria. Con los modelos ARIMA, existe un problema adicional al elegir el orden requerido de diferenciación (es decir, ¿cuál es  $d$ ?). Hay algunos procedimientos formales disponibles, incluida la prueba de la presencia de una raíz unitaria [131, 162],

pero muchos analistas simplemente diferencian las series hasta que el correlograma se reduce a cero “bastante rápido”. La diferenciación de primer orden suele ser adecuada para las series no estacionales, aunque en ocasiones se necesita la diferenciación de segundo orden. Una vez que la serie se ha vuelto estacionaria, se puede ajustar un modelo ARMA a los datos diferenciados de la forma habitual.

## B.2. Procesos SARIMA

Si la serie es *estacional*, con  $s$  periodos de tiempo por año, entonces se puede obtener un modelo ARIMA estacional (abreviado SARIMA). Sea  $B^s$  el operador tal que  $B^s y_t = y_{t-s}$ . Por lo tanto, la diferenciación estacional puede escribirse como  $(y_t - y_{t-s}) = (1 - B^s)y_t$ . Un término autorregresivo estacional, por ejemplo, es aquel en el que  $y_t$  depende linealmente de  $y_{t-s}$ . Un modelo SARIMA con términos de orden no estacionales ( $p, d, q$ ) y términos de orden estacionales ( $P, D, Q$ ) se abrevia modelo  $SARIMA(p, d, q)(P, D, Q)_s$  y puede escribirse

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta(B)\Theta(B^s)z_T \quad (\text{B.3})$$

donde  $\Phi, \Theta$  denota polinomios en  $B^s$  de orden  $P, Q$ , respectivamente.

Al ajustar modelos SARIMA, el analista primero debe elegir valores para los dos órdenes de diferenciación, tanto estacional ( $D$ ) como no estacional ( $d$ ), a fin de hacer que la serie sea estacionaria y eliminar (la mayor parte) de la estacionalidad. Luego se ajusta un modelo tipo ARMA a la serie diferenciada con la complicación adicional de que puede haber términos AR y MA en retrasos que son un múltiplo de la duración de la temporada  $s$  [131].

## C.0.1. Formalismo multifractal

Un formalismo multifractal está asociado con una elección particular de una cantidad multirresolución  $T_X(a, x)$ . En términos generales, se puede definir  $T_X(a, x)$  como el resultado de una comparación de  $X$  con un patrón de referencia, con soporte compacto, dilatado con un factor de escala  $a > 0$  y trasladado a una posición  $x$ .

El requisito principal que debe cumplir una cantidad multirresolución para que se pueda basar un formalismo multifractal en ella es que su decaimiento local con las escalas reproduzca al exponente de Hölder:

$$|T_X(A, X_0)| \leq K a^{h_X(x_0)}, a0^+, K > 0. \quad (\text{C.1})$$

Es decir, el comportamiento local de la magnitud de  $T_X(a, x)$ , a medida que la escala  $a$  se hace más finas, debe estar caracterizado por una ley de potencias gobernada por el exponente de Hölder. El primer paso del formalismo multifractal consiste en estudiar el comportamiento de los momentos de  $|T_X(a, x)|$ , llamados en este contexto *funciones de estructura*  $S_T(q, a)$ , definidas por:

$$S_T(q, a) = \frac{1}{n_a} \sum_{k=1}^{n_a} |T_X(a, x_k)|^q \quad (\text{C.2})$$

en donde las muestras de  $X : X(x_k)$ ,  $k \in \mathbb{N}$ .  $n_a$  denota la cantidad de coeficientes  $T_X(a, x_0)$  disponibles en la escala  $a$ . Si  $X$  muestra propiedades de invarianza a la escala, entonces las funciones de estructura siguen un decaimiento en ley de potencias:

$$S_T(q, a) \sim a^{\zeta_T(q)}, a0^+. \quad (\text{C.3})$$

El exponente  $\zeta_T(q)$  que caracteriza esta ley de potencia se conoce como exponente de escalamiento. Debido a que depende del orden  $q$ , también se lo llama función de escalamiento. Puede ser calculado como:

$$\zeta_T(q) = \liminf_{a \rightarrow 0^+} \frac{\log(S_T(q, a))}{\log(a)} \quad (\text{C.4})$$

En la práctica,  $\zeta_T(q)$  puede ser estimado mediante una regresión lineal de  $\log(S_T(q, a))$  en función de  $\log(a)$ . El exponente de escalamiento caracteriza, entonces, el comportamiento de invarianza a la escala de  $X$ , íntimamente relacionado con su regularidad local. De hecho, si consideramos el efecto de variar el orden  $q$ , a partir de la (C.2) se nota que valores positivos y altos de  $q$  resaltarán los coeficientes  $T_X(a, x_k)$  de mayor valor absoluto, correspondientes a las regiones más irregulares de  $X$ . Por el contrario, los valores negativos de  $q$  resaltarán aquellos coeficientes más cercanos a 0, caracterizando las regiones más suaves y regulares de  $X$ . Resulta claro de este análisis que para la caracterización completa de la regularidad de la señal mediante la función de escalamiento es necesario considerar todos los valores de  $q$ , tanto positivos como negativos, y tanto enteros como fraccionarios. A partir del análisis anterior se puede notar que tanto el espectro multifractal como la función de escalamiento brindan una descripción global de los valores que toma el exponente de Hölder [122].

### C.0.2. log-cumulantes

Desde un punto de vista práctico, las funciones  $\zeta$  y  $L$  pueden resultar difíciles de manipular. Por tal razón, en ciertas situaciones resulta provechoso obtener una aproximación polinomial de  $\zeta(q)$  alrededor de  $q = 0$ :

$$\zeta(q) = \sum_{m=1}^{\infty} c_m \frac{q^m}{m!}, \quad (\text{C.5})$$

y así reemplazar la determinación de la función  $\zeta$  completa por la estimación de un conjunto de coeficientes  $\{c_m\}_{m \in \mathbb{N}}$  que sinteticen la información importante de la función. Los coeficientes  $c_m$  pueden ser estimados directamente a partir de los cumulantes de la cantidad multirresolución utilizada. Los cumulantes  $C_m$  de una variable aleatoria  $Y$  son los coeficientes de la expansión polinomial alrededor del origen de su segunda función característica:

$$\log \mathbb{E}(e^{qY}) = \sum_{m=1}^{\infty} C_m \frac{q^m}{m!} \quad (\text{C.6})$$

Las funciones de estructura  $S_L(q, j)$  pueden ser interpretadas como estimadores muestrales de los momentos poblacionales  $\mathbb{E}\{L_X(j, \cdot)^q\}$ . Por lo tanto,

$$\mathbb{E}\{L_X(j, \cdot)^q\} = K_q 2^{j\zeta(q)} \quad (\text{C.7})$$

o, equivalentemente,

$$\log \mathbb{E}\{L_X(j, \cdot)^q\} = \log K_q + \zeta(q) \log(2^j) \quad (\text{C.8})$$

Por otro lado,

$$\log \mathbb{E}\{L_X(j, \cdot)^q\} = \log \mathbb{E}\{e^{q \log L_X(j, \cdot)}\}, \quad (\text{C.9})$$

es decir,  $\log \mathbb{E}\{L_X(j, \cdot)^q\}$  puede ser interpretado como la segunda función característica de la variable aleatoria  $\log L_X(j, \cdot)$ . Por lo tanto, admite una expansión polinomial de la forma:

$$\log \mathbb{E}\{L_X(j, \cdot)^q\} = \log \mathbb{E}\{e^{q \log L_X(j, \cdot)}\} = \sum_{m=1}^{\infty} C_m(j) \frac{q^m}{m!}. \quad (\text{C.10})$$

Comparando las Ecs. C.8 y C.10, vemos que

$$\log K_q + \zeta(q) \log(2^j) = \sum_{m=1}^{\infty} C_m(j) \frac{q^m}{m!}. \quad (\text{C.11})$$

Para que esta relación se cumpla,  $C_m(j)$  debe satisfacer:

$$C_m(j) = c_{0,m} + c_m \log(2^j). \quad (\text{C.12})$$

Efectivamente, de esta forma

$$\sum_{m=1}^{\infty} C_m(j) \frac{q^m}{m!} = \sum_{m=1}^{\infty} c_{0,m} \frac{q^m}{m!} + \left( \sum_{m=1}^{\infty} c_m \frac{q^m}{m!} \right) \log(2^j), \quad (\text{C.13})$$

de donde se reconoce que

$$\log K_q = \sum_{m=1}^{\infty} c_{0,m} \frac{q^m}{m!} \quad (\text{C.14})$$

$$\zeta(q) = \sum_{m=1}^{\infty} c_m \frac{q^m}{m!} \quad (\text{C.15})$$

Es decir, los coeficientes de la expansión polinomial de  $\zeta(q)$  alrededor de 0 se pueden obtener a partir de los cumulantes de la variable aleatoria  $\log L_X(j, \cdot)$  mediante una regresión lineal, de  $C_m(j)$  en función de  $\log(2^j)$ . Estos coeficientes  $c_m$  se conocen como *log-cumulantes* (en inglés, *log-cumulants*). Debido a las propiedades de los cumulantes, se tiene que si  $c_{m'} = 0$ , entonces  $c_m = 0, \forall m > m'$ . Además, la concavidad de  $\zeta(q)$  implica que  $c_2 \leq 0$ .

**Estimación** Los cumulantes  $C_m(j)$  pueden ser estimados a partir de los estimadores muestrales tradicionales [122]. Además, se propuso un método bayesiano para la estimación de  $c_2$  en ciertos tipos de procesos multifractales, y se demostró que tiene un buen desempeño para tamaños de muestra pequeños.

**Aplicación** Como se mencionó, el uso de los coeficientes  $c_m$  en lugar de las funciones  $\zeta(q)$  o  $L(h)$  es ventajoso desde el punto de vista práctico, en particular para la realización de tareas de clasificación o análisis, ya que concentran la información sobre el comportamiento multifractal de una función en unos pocos coeficientes. Por ejemplo, la discriminación entre una función monofractal y una multifractal puede hacerse en forma sencilla mediante un test con la hipótesis nula  $H_0 : c_2 = 0$ .

### C.0.3. Procesos Monofractales y Multifractales

Otra forma de clasificar los procesos es según la forma que adopta el espectro multifractal.

#### **Definición 1. Proceso Monofractal**

*Un proceso  $X$  es monofractal si su exponente de Hölder es constante en todo punto:  $h_X(t) \equiv H$ . Por lo tanto, su espectro de singularidades se reduce a un único punto:*

$$D(h) = \begin{cases} 1 & \text{si } h = H \\ -\infty & \text{en otro caso} \end{cases} \quad (\text{C.16})$$

#### **Definición 2. Proceso Multifractal**

*Un proceso  $X$  es multifractal si el soporte de su espectro de singularidades no se reduce a un único punto:  $\text{supp}(D) = [h_{\min}, h_{\max}]$ , con  $h_{\max} > h_{\min}$ .*

En los procesos monofractales, el exponente de Hölder es una función constante:  $h_X(t) \equiv H$ . Por el contrario, en los procesos multifractales la función  $h_X(t)$  es habitualmente una función altamente oscilatoria, en donde el exponente de Hölder cambia drásticamente de punto a punto.

La estimación práctica del espectro multifractal presenta dificultades. Para sortearlas, se recurre a un formalismo multifractal, el cual permite estimar dicho espectro a partir de cantidades globales fácilmente calculables e íntimamente ligadas con las propiedades de invarianza a la escala [122].

#### C.0.4. Invarianza a la escala

El primer paradigma utilizado para modelar funciones o procesos con invarianza a la escala fue el de los modelos auto-similares. De hecho, numerosos procesos H-sssi, muestran la siguiente forma de invarianza a la escala [122]:

$$\mathbb{E}[|\mathcal{L}_X(a, \cdot)|^q] \sim |a|^{qH}, \quad (\text{C.17})$$

es decir, los momentos de los incrementos del proceso decaen con las escalas según una ley de potencias. Además, existe una relación lineal entre el decaimiento en los distintos órdenes  $q$ , gobernada por el parámetro de Hurst  $H$ . Este tipo de comportamiento está íntimamente relacionado con la naturaleza aditiva de estos procesos. El hecho de que el comportamiento de los diferentes momentos esté gobernado por un único parámetro es a menudo restrictivo en la práctica. Un primer nivel de generalización puede lograrse al permitir que los exponentes de las leyes de potencia sigan una relación no lineal con el orden  $q$ . Esta elección corresponde al paradigma de los procesos multifractales. En este caso se tiene:

$$\mathbb{E}[|\mathcal{L}_X(a, \cdot)|^q] \sim |a|^{\zeta(q)}, \quad (\text{C.18})$$

en donde se puede ver que el comportamiento auto-similar es un caso particular de este paradigma que corresponde a la elección  $\zeta(q) = qH$ . La dependencia no lineal entre los decaimientos de los momentos está relacionada con la naturaleza multiplicativa involucrada en la construcción de estos procesos. A pesar del mayor nivel de generalidad que proveen, los procesos multifractales siguen manifestando una invarianza a la escala en forma de leyes de potencia. Se logra un segundo nivel de generalización al permitir que este decaimiento

tome otras formas. Esta situación corresponde al paradigma de las cascadas infinitamente divisibles, en donde se tiene:

$$\mathbb{E} [|\mathcal{L}_X(a, \cdot)|^q] \sim \exp(\eta(a)H(q)), \quad (\text{C.19})$$

Se puede ver que el paradigma de los procesos multifractales es un caso particular, que corresponde a la elección  $\zeta(a) = \log(a)$ . Las cascadas infinitamente divisibles fueron introducidas originalmente en el estudio de los fenómenos de turbulencia. Sin embargo, este tipo de procesos no es habitualmente utilizado en la práctica debido a la dificultad de desarrollar un algoritmo para su síntesis numérica.

### C.0.5. Relación entre las clasificaciones

El caso paradigmático de un proceso H-sssi es el movimiento browniano fraccionario. Este proceso es además monofractal y tiene una función de escalamiento lineal  $\zeta(q) = qH$ . Por lo tanto, un único parámetro  $H$  controla las propiedades de escalamiento, regularidad local y autosimilaridad del proceso. Debido a que el movimiento browniano fraccionario es considerado el proceso de referencia para modelar la invarianza a la escala, habitualmente se identifica en forma heurística que las tres propiedades son equivalentes: H-sssi, monofractalidad y función de escalamiento lineal. Además, su función de escalamiento es lineal a tramos. Habitualmente, la distinción entre un proceso multifractal y uno monofractal se hace en términos del segundo log-cumulante  $c_2$ . El caso  $c_2 = 0$  habitualmente se consideró una característica de procesos monofractales y con una función de escalamiento lineal. Sin embargo, como  $c_2$  sólo mide el comportamiento de  $\zeta(q)$  en una vecindad de  $q = 0$ , este coeficiente no puede medir desviaciones de la monofractalidad en las que la función de escalamiento es lineal a tramos. Por lo tanto, las conclusiones que se extraigan a partir del estudio de  $c_2$  deberían realizarse con cautela. Además, estas reflexiones resaltan la importancia de realizar el análisis multifractal con un rango amplio de valores de  $q$ . En general, se pueden realizar las siguientes afirmaciones sobre un proceso  $X$  con propiedades de escalamiento [122]:

1. Si es monofractal entonces su función de escalamiento es lineal para todo  $q$ . Esto es una consecuencia de la definición de  $\zeta$  como la transformación de Legendre de  $D$ .

2. Si la función de escalamiento es no lineal, entonces el proceso es multifractal.
3. Si  $X$  es H-sssi, su función de escalamiento es lineal por tramos y necesariamente se comporta como  $\zeta(q) = qH$  en una vecindad de  $q = 0$ .

### C.0.6. Atributos Multifractales

*Espectro Multifractal  $D(h)$*  Dicha técnica provee una descripción global de la distribución de las singularidades presentes en la señal, es una herramienta idónea para el análisis y la caracterización de señales en las que la información de interés está codificada en el comportamiento irregular [122].

#### Definición 3. Conjunto iso-Hölder

Los conjuntos iso-Hölder son los conjuntos de nivel de la función  $h_X$ :

$$E_X(h) = \{x \in \mathbb{R} : h_X(x) = h\}. \quad (\text{C.20})$$

En señales multifractales, los conjuntos  $E_X$  típicamente poseen una estructura compleja, que no puede ser caracterizada completamente mediante las dimensiones euclídeas. Por esta razón, su caracterización requiere del uso de dimensiones fractales. En particular, para la definición del espectro de singularidades se utiliza la dimensión de Hausdorff  $Dim_H$ , definida a continuación.

#### Definición 4. Medida de Hausdorff

Sea  $S$  un conjunto en un espacio métrico. Sea  $U_\epsilon(S)$  un recubrimiento de  $S$  mediante conjuntos  $u_n$  de diámetro  $|u_n| \leq \epsilon$ . Finalmente, sea  $\Upsilon_\epsilon$  la colección de todos los recubrimientos  $U_\epsilon(S)$  posibles. La medida de Hausdorff de  $S$  se define como:

$$M^{\delta}_H(S) = \liminf_{\epsilon \rightarrow 0^+} \inf_{U_\epsilon(S) \in \Upsilon_\epsilon} \sum_{u_n \in U_\epsilon(S)} |u_n|^\delta. \quad (\text{C.21})$$

Se puede demostrar que existe un valor crítico  $\delta_c$  tal que  $M^{\delta}_H(S) = 0$  si  $M^{\delta}_H(S) = \infty$  si  $\delta > \delta_c$ . La dimensión de Hausdorff de  $S$  se define como dicho valor  $\delta_c$ :

#### Definición 5. Dimensión de Hausdorff

La dimensión de Hausdorff del conjunto métrico  $S$  se define como:

$$Dim_H(S) = \inf(\delta \geq 0 : M^\delta(S) = 0) = \sup(\delta \geq 0 : M^\delta(S) = \infty). \quad (C.22)$$

Además, por convención  $Dim_H = -\infty$ .

**Definición 6. Espectro multifractal**

El espectro multifractal  $D(h)$  de una función  $X$  se define como la dimensión de Hausdorff de sus conjuntos iso-Hölder  $E_X(h)$ :

$$D(h) = Dim_H(E_X(h)) \quad (C.23)$$

Los conjuntos  $E_X(h)$  consisten en los subconjuntos del dominio de  $X$  en los que  $h_X$  toma un valor  $h$  determinado. Por lo tanto, brindan una descomposición del soporte de  $X$  según los exponentes de Hölder de sus singularidades. Por otro lado, el espectro multifractal informa el “tamaño” de dichos conjuntos. El concepto de tamaño está formalizado mediante la dimensión de Hausdorff. En forma poco rigurosa, podemos decir que  $D(h)$  informa sobre la proporción de puntos del dominio de  $X$  en donde la regularidad local es  $h$ . El caso extremo en que  $D(h) = 0$  corresponde a la situación en que las irregularidades de valor  $h$  están “aisladas”. El otro caso extremo, en donde  $D(h) = 1$ , ocurre cuando el exponente de Hölder  $h_X$  es constante en un intervalo.

# D

## Capítulo

# Anexo: Muestra de datos utilizados

Serie La Venta, Modelos simples y Modelos híbridos correspondientes.

No	Serie La Venta	Modelo ARIMA	Modelo ARIMA-CM	Modelo DL	Modelo DL-CM	Modelo Persistencia	Modelo Persistencia-CM
1	5.34	5.45	5.44	5.29	5.29	5.30	5.26
2	5.63	5.20	5.56	5.61	5.63	5.34	5.30
3	5.40	5.74	5.40	5.36	5.36	5.63	5.59
4	5.38	5.34	5.49	5.33	5.33	5.40	5.36
5	5.05	5.19	5.02	4.97	5.10	5.38	5.34
6	5.45	4.77	5.51	5.41	5.44	5.05	5.38
7	5.62	5.64	5.65	5.59	5.63	5.45	5.41
8	5.63	5.60	5.60	5.61	5.63	5.62	5.58
9	5.47	5.79	5.49	5.43	5.43	5.63	5.59
10	5.32	5.44	5.39	5.27	5.33	5.47	5.43
11	5.43	5.34	5.40	5.39	5.44	5.32	5.43
12	5.77	5.45	5.56	5.76	5.76	5.43	5.88
13	5.53	5.95	5.69	5.50	5.55	5.77	5.73
14	5.64	5.78	5.63	5.62	5.62	5.53	5.78
15	6.23	5.83	6.29	6.23	6.23	5.64	5.98
16	6.59	6.50	6.58	6.59	6.59	6.23	5.87
17	6.11	6.63	5.88	6.11	6.11	6.59	6.07
18	5.86	5.87	5.86	5.85	5.85	6.11	6.17
19	5.77	5.69	5.64	5.76	5.76	5.86	5.92
20	5.52	5.81	5.69	5.49	5.55	5.77	5.83
21	5.43	5.35	5.39	5.39	5.39	5.52	5.58
22	5.23	5.35	5.29	5.17	5.22	5.43	5.49
23	5.10	4.95	5.12	5.03	5.03	5.23	5.21
24	4.94	4.98	4.97	4.85	4.95	5.10	5.11
25	5.44	5.03	5.22	5.40	5.44	4.94	5.00
26	5.35	5.39	5.34	5.30	5.31	5.44	5.30
27	5.04	5.37	5.07	4.96	5.04	5.35	5.11
28	5.03	5.00	5.04	4.95	5.03	5.04	5.01
29	5.22	4.83	5.19	5.16	5.19	5.03	5.19
30	5.25	5.02	5.20	5.19	5.22	5.22	5.21
31	4.83	5.37	4.70	4.73	4.78	5.25	5.31
32	4.69	4.77	4.67	4.58	4.76	4.83	4.71
33	4.93	4.84	4.95	4.84	4.94	4.69	4.75
34	4.65	4.94	4.84	4.53	4.65	4.93	4.69
35	4.41	4.64	4.36	4.27	4.37	4.65	4.31
36	4.54	4.41	4.53	4.41	4.44	4.41	4.47
37	4.17	4.70	4.07	4.02	4.17	4.54	4.16
38	4.56	4.40	4.53	4.43	4.54	4.17	4.23
39	4.83	4.80	4.87	4.73	4.80	4.56	4.62
40	5.13	5.04	5.29	5.06	5.16	4.83	4.89
41	4.69	5.17	4.58	4.58	4.68	5.13	5.19
42	4.75	4.45	4.64	4.64	4.74	4.69	4.75
43	4.68	4.62	4.65	4.57	4.65	4.75	4.68
44	4.30	4.70	4.40	4.15	4.25	4.68	4.74
45	4.18	4.12	4.40	4.03	4.13	4.30	4.36
46	4.14	4.10	4.15	3.99	4.10	4.18	4.24
47	4.28	3.89	4.15	4.13	4.13	4.14	4.20
48	4.85	4.18	4.61	4.75	4.85	4.28	4.21
49	4.69	5.03	4.60	4.58	4.76	4.85	4.78
50	4.84	4.53	4.83	4.74	4.84	4.69	4.75

## Serie La Mata, Modelos simples y Modelos híbridos correspondientes.

No	Serie La Mata	Modelo ARIMA	Modelo ARIMA-CM	Modelo OL	Modelo OL-CM	Modelo Persistencia	Modelo Persistencia-CM
1	12.5	11.59	11.71	11.78	11.78	11.50	12.45
2	14.6	12.33	12.59	12.39	11.39	12.50	13.50
3	13.2	14.06	13.44	14.03	13.03	14.60	13.22
4	11.5	13.21	11.43	13.19	14.19	13.20	11.20
5	12.5	11.82	12.32	12.07	13.07	11.50	12.12
6	11.4	12.44	11.34	12.64	11.64	12.50	12.41
7	10.9	11.63	10.51	11.79	11.79	11.40	10.95
8	10.6	11.13	10.32	11.19	11.19	10.90	10.43
9	11.4	10.81	11.39	10.91	10.91	10.60	11.60
10	10.1	11.36	10.58	11.28	10.28	11.40	10.11
11	9.9	10.39	9.59	10.42	11.32	10.10	10.06
12	12.4	10.10	11.34	10.11	10.11	9.90	11.92
13	14	11.99	13.23	11.92	10.92	12.40	14.26
14	12.5	13.47	13.38	13.46	12.46	14.00	13.00
15	11.8	12.51	12.45	12.50	13.50	12.50	12.00
16	11.7	11.90	11.63	12.18	12.76	11.80	11.80
17	11.8	11.76	11.64	12.04	11.70	11.70	11.70
18	11.1	11.81	10.86	11.88	11.88	11.80	11.80
19	10.5	11.27	10.67	11.36	11.36	11.10	10.61
20	10.3	10.72	10.13	10.82	11.23	10.50	10.50
21	10.1	10.48	10.18	10.55	10.55	10.30	10.30
22	10.1	10.26	9.98	10.27	10.27	10.10	10.10
23	10.1	10.21	10.10	10.18	10.18	10.10	10.10
24	9.8	10.18	10.10	10.14	10.14	10.10	10.00
25	9.4	9.92	9.62	9.91	10.00	9.80	9.60
26	9.6	9.56	9.56	9.57	9.69	9.40	9.54
27	10.6	9.66	9.72	9.65	9.56	9.60	10.56
28	8.3	10.43	9.56	10.38	9.38	10.60	10.60
29	9.4	8.72	8.84	8.69	9.69	8.30	8.30
30	10.5	9.37	10.30	9.39	8.39	9.40	9.40
31	11.7	10.28	10.60	10.25	9.32	10.50	10.99
32	10.8	11.33	10.57	11.33	10.33	11.70	10.70
33	12.1	10.78	11.03	10.82	10.82	10.80	10.80
34	13.5	11.77	11.84	11.82	10.82	12.10	12.91
35	13.4	13.01	13.40	13.04	12.04	13.50	13.50
36	14.4	13.13	14.46	13.12	13.12	13.40	14.00
37	14.5	13.98	14.13	14.00	13.00	14.40	14.40
38	13.4	14.22	13.44	14.27	14.27	14.50	13.50
39	13.8	13.44	13.93	13.66	14.66	13.40	13.40
40	15.5	13.69	14.71	13.86	13.86	13.80	14.16
41	13.7	15.07	15.07	15.00	14.00	15.50	14.50
42	15.4	13.85	14.85	13.93	14.93	13.70	14.70
43	14.1	15.07	15.07	14.99	13.99	15.40	14.40
44	15.1	14.21	14.57	14.38	15.38	14.10	14.95
45	13.2	14.91	13.24	14.81	14.81	15.10	14.10
46	13.4	13.51	13.21	13.80	14.80	13.20	13.20
47	12.4	13.50	13.17	13.60	13.03	13.40	12.16
48	15	12.69	14.72	12.93	13.26	12.40	13.15
49	14.8	14.61	14.61	14.43	13.43	15.00	15.00
50	12.7	14.67	13.67	14.48	14.85	14.80	13.82

# E

## Capítulo

# Anexo: Publicaciones y Aportes

En esta sección se agregan las publicaciones y divulgaciones del presente trabajo de tesis.

---

Nombre del trabajo	<b>Wind speed forecasting using a hybrid model considering the turbulence of the airflow</b>
Revista	Renewable Energy
Editorial	ELSEVIER
Factor de impacto	8.634 <i>Q1</i>
Autores	Alma Rosa Méndez-Gordillo, Rafael Campos-Amezcu, Erasmo Cadenas
Año de publicación	26 Junio 2022
Doi	<a href="https://doi.org/10.1016/j.renene.2022.06.139">https://doi.org/10.1016/j.renene.2022.06.139</a>

---

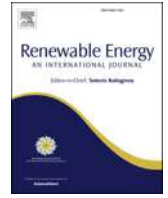
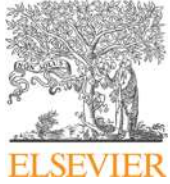
Nombre del trabajo	<b>Wind speed forecasting by the extraction of the multifractal patterns of time series through the multiplicative cascade technique</b>
Revista	Chaos, Solitons and Fractals
Editorial	ELSEVIER
Factor de impacto	9.922 <i>Q1</i>
Autores	Alma Rosa Méndez-Gordillo, Erasmo Cadenas
Año de publicación	13 Diciembre 2020
Doi	<a href="https://doi.org/10.1016/j.chaos.2020.110592">https://doi.org/10.1016/j.chaos.2020.110592</a>

---

---

Nombre del trabajo	<b>Wind speed variability study based on the Hurst coefficient and fractal dimensional analysis technique</b>
Revista	Energy Science & Engineering
Editorial	Wiley Online Library
Factor de impacto	4.17 <i>Q2</i>
Autores	Erasmus Cadenas <sup>1</sup> , Rafael Campos-Amezcu <sup>2</sup> , Wilfrido Rivera <sup>3</sup> , Marco Antonio Espinosa-Medina <sup>1</sup> , Alma Rosa Méndez-Gordillo <sup>1</sup> , Eduardo Rangel <sup>1</sup> , Jorge Tena <sup>1</sup>
Año de publicación	31 Enero 2019
Doi	<a href="https://doi.org/10.1002/ese3.277">https://doi.org/10.1002/ese3.277</a>

---



# Wind speed forecasting using a hybrid model considering the turbulence of the airflow

Alma Rosa Méndez-Gordillo <sup>a</sup>, Rafael Campos-Amezcuca <sup>b, \*</sup>, Erasmo Cadenas <sup>a</sup>

<sup>a</sup> Posgrado de la Facultad de Ingeniería Mecánica, Universidad Michoacana de San Nicolás de Hidalgo, Gral. Francisco J. Múgica S/N, Col. Felicitas del Río, C.P. 58040, Morelia, Mexico

<sup>b</sup> Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, Interior Internado Palmira S/N, Col. Palmira, C.P. 62490, Cuernavaca, Mexico

## ARTICLE INFO

### Article history:

Received 20 October 2021

Received in revised form

23 May 2022

Accepted 26 June 2022

Available online 7 July 2022

### Keywords:

Persistence

Multiplicative cascade

Multifractality

Wind speed time series forecasting

Turbulent wind

## ABSTRACT

This work presents the conformation of a hybrid model aimed at forecasting the wind speed one step ahead. The model comprises the Multiplicative Cascade and Persistence. To couple both techniques, a transition point was used, in this case, 6.6 m/s, defined from the spectrum of wind speed near the ground. The Multiplicative Cascade was used when speeds were higher than the transition point, while Persistence was used when lower. Two time series measured in Tamaulipas and Veracruz, Mexico, were used to build the hybrid model. The performance of the individual models concerning the hybrid model was compared, finding that the adjustment of the hybrid model exceeds that of the Multiplicative Cascade and Persistence models for both series. Different error metrics were used in the quantitative analysis, where the hybrid model was the most competitive. For example, in Mean Absolute Percentage Error, the difference in performance between the hybrid model and Persistence is 2.72% points for the 10-min time series. In contrast, this difference is 27.89% points for the hourly time series. Finally, this work shows that for the cases analyzed, differentiating the effects of turbulent from non-turbulent flows and modeling with the proper techniques improves forecasting accuracy.

© 2022 Elsevier Ltd. All rights reserved.

## 1. Introduction

The growing demand for electrical energy, as well as the need to reduce dependence on fossil fuels to mitigate their environmental impact, have attracted worldwide attention to the use of renewable energy sources, positioning wind energy as one of the most promising renewable energy technologies [1]. At the end of 2020, the accumulated installed wind power capacity worldwide reached 743 GW [2]. While in Mexico, the installed capacity came to 6681 MW [3].

In recent years, wind speed prediction has been rising popularity thanks to the increasing demand for electricity generation using wind energy and the evolution of the competitiveness of this renewable energy [4]. However, according to Moreno and dos Santos Coelho [5], one of the great problems with the electric generation process from wind power is the wind speed intermittence since it can be affected by many complex factors. Wind speed

is considered one of the most complex meteorological parameters to be predicted. So, the accurate prediction of wind speed remains a research topic. Consequently, numerous wind speed prediction models have been developed. These models could be classified into many ways according to different standards. On the one hand, depending on the prediction time-horizon: long-term prediction models [6–8], medium-term prediction models [9–11], and short-term prediction models [12–15]. On the other hand, the wind speed prediction models can be classified into five categories based on the techniques employed by the models, namely: artificial intelligence models, spatial correlation models, conventional statistical models, physical models and hybrid models [16].

Many studies have shown that hybrid models can get to have better prediction performance than individual models Liu et al. [17] since hybrid models take advantage of different techniques, such as physical, statistical, or artificial intelligence; in order to generate a more accurate wind speed forecast [18]. Shi et al. [19] consider that it is potentially feasible to forecast time series using hybrid models, obtaining better performance than conventional forecast modeling. They also mention that a hybrid model generally includes a prediction model for the linear component and another to handle the

\* Corresponding author.

E-mail address: [rafael.ca@cenidet.tecnm.mx](mailto:rafael.ca@cenidet.tecnm.mx) (R. Campos-Amezcuca).



Contents lists available at ScienceDirect

# Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: [www.elsevier.com/locate/chaos](http://www.elsevier.com/locate/chaos)

## Wind speed forecasting by the extraction of the multifractal patterns of time series through the multiplicative cascade technique

Alma Rosa Méndez-Gordillo, Erasmo Cadenas\*

División de Estudios de Posgrado de la Facultad de Ingeniería Mecánica, Universidad Michoacana de San Nicolás de Hidalgo, Gral. Francisco J. Múgica S/N, Col. Felicitas del Río, Morelia 58040, México

### ARTICLE INFO

#### Article history:

Received 10 October 2020

Revised 8 December 2020

Accepted 13 December 2020

#### Keywords:

Multifractal wind speed time series  
 Multifractal detrended fluctuation analysis  
 ARIMA  
 Deep learning  
 Forecasting errors  
 Multiplicative cascade

### ABSTRACT

In this work, the forecasting errors produced by the Autoregressive Integrated Moving Average, Deep Learning, and Persistence models were analyzed when these models were applied to two time series for predicting the wind speed one-step-ahead. The analysis consisted of verified, with the aid of the Multifractal Detrended Fluctuation Analysis, whether there were multifractal patterns in the forecasting errors. Then, the multifractal patterns of the errors were extracted by a Multiplicative Cascade model. The sum of the patterns extracted from the models generated new hybrid models such as Autoregressive Integrated Moving Average-Multiplicative Cascade, Deep Learning-Multiplicative Cascade, and Persistence-Multiplicative Cascade. Those new models showed a diminution of the final forecasting error. The wind speed time series used for this study were measured at two wind farms of the State of Oaxaca, Mexico, namely La Mata and La Venta. The first time series has five years of ten-minutes averaged data, while the second one has six years and eleven months of hourly data. Both time series were provided by Mexico's Federal Electricity Commission. The typical performance metrics tools were used for the quantitative analysis of the forecasting results, where the simple models were compared against the hybrid models by using the Mean Squared Error, the Mean Absolute Error, the Root Mean Squared Error, the Mean Absolute Percentage Error, and the Inter-rater Agreement. The latter was done to find out the actual input in a one-step-ahead forecast. The obtained results showed a gain between 0.331% and 5.285% in the fit of hybrid models in the metrics, which demonstrates the convenience of analyzing the forecasting errors generated by the models aforementioned to improve forecasts and therefore help in planning and dispatching energy flow.

© 2020 Elsevier Ltd. All rights reserved.

### 1. Introduction

The continued growth of wind energy in the world, stimulated by the universal agreements on climate change of the United Nations, generates a positive precedent that encourages the world community to continue promoting efforts oriented to limit the increase in global warming.

At the end of 2019, the installed capacity of wind energy in the world exceeds 651 GW, an increase of 10% compared to 2018 [1]. In Mexico, the installed wind power capacity increased to 6,238 MW in 2019. Thanks to the abundant wind resource, the Isthmus of Tehuantepec region in the state of Oaxaca, Mexico, concentrates 62% of the total installed wind energy of the country, according to the Mexican Wind Energy Association [2].

Wind speed forecasting can be done through physical and stochastic methods. Physical methods are based on the characteristics of the terrain and the atmosphere, such as the roughness, orography, obstacles as well as the local wind profile; they have advantages in long-term prediction. In the stochastic models, the prediction is based on the analysis of data and is carried out by using time series, statistical techniques, and artificial intelligence techniques; they have advantages mainly in short-term prediction.

The problem of wind speed forecasting is not a new topic. It is a complicated task since there is a wide range of behaviors in the data to be modeled. Hodge et al. [3] mention that it is not possible to have control of the wind. So, the best solution is to be able to forecast wind production in a certain period, which will allow us to know the behavior that this one will have. For Slini et al. [4] when modeling wind time series, different behaviors are observed due to their intermittent and uncertain nature, so it is necessary to generate a model for each case. Therefore, time series analysis is among the main research and development tools in this field.

\* Corresponding author.

E-mail addresses: [alma.mendez@umich.mx](mailto:alma.mendez@umich.mx) (A.R. Méndez-Gordillo), [ecadenas@umich.mx](mailto:ecadenas@umich.mx) (E. Cadenas).

## RESEARCH ARTICLE

# Wind speed variability study based on the Hurst coefficient and fractal dimensional analysis

Erasmio Cadenas<sup>1</sup> | Rafael Campos-Amezcuca<sup>2</sup>  | Wilfrido Rivera<sup>3</sup>  |

Marco Antonio Espinosa-Medina<sup>1</sup> | Alma Rosa Méndez-Gordillo<sup>1</sup> | Eduardo Rangel<sup>1</sup> |

Jorge Tena<sup>1</sup>

<sup>1</sup>Facultad de Ingeniería Mecánica,  
Universidad Michoacana de San Nicolás de  
Hidalgo, Morelia, Mich, Mexico

<sup>2</sup>Tecnológico Nacional de México/Instituto  
Tecnológico de Pachuca, Pachuca, Hgo,  
Mexico

<sup>3</sup>Instituto de Energías Renovables de la  
Universidad Nacional Autónoma de México  
(UNAM), Temixco, Mor, Mexico

## Correspondence

Rafael Campos-Amezcuca, Tecnológico  
Nacional de México/Instituto Tecnológico  
de Pachuca, Pachuca, Hgo, Mexico.  
Email: rca@itpachuca.edu.mx

## Funding information

the Universidad Michoacana de San Nicolás  
de Hidalgo; the UNAM-DGAPA-PAPIIT  
Program, Grant/Award Number: IA107416

## Abstract

This paper presents a study of a wind speed time series from La Venta, Oaxaca, Mexico. The time series consists of anemometric measurements taken by the Federal Electricity Commission of Mexico throughout a little over 6 years. The study was conducted to calculate the Hurst correlation coefficient using: box counting, rescaled range, power spectrum, detrended fluctuation analysis, and multifractal detrended fluctuation analysis techniques. The main objective of this research is to know the correlation among wind speed data to obtain a better description of real conditions of the time series, which is not always available, and to define the structure of its behavior. In this way, more suitable wind speed prediction models can be achieved. Results obtained from techniques above were used to generate fractals time series for a typical month, using the Hurst coefficient and a self-affine trace generator, which produces fractals time series whose probability distribution is always normal. These time series were compared against time series generated by using random numbers with Gaussian behavior and the characteristics of a typical month. Fractals time series highlight in the qualitative part regarding the modeling of wind speed variability and the descriptive statistics (average, standard deviation, and coefficient of variation), which is similar to the real series. Discordance tests were applied to the datasets to detect deviated values, and so ensure the normal behavior of the samples. These tests showed the existence of different populations with normal behavior in the samples that had bimodal characteristics. By separating the samples, it was possible to apply the self-affine trace generator to each population found, to generate the fractal time series. An additional objective was to find the level of change in the structure of the original series concerning its statistical and fractal characteristics at different window widths of the time series (daily, monthly, seasonal, and annual) to identify either a specific tendency or dynamic behavior. The results showed a wind speed time series with a negative correlation (antipersistent), a high degree of scale invariance (homothetic), and a fractal dimension very close to 2, thus indicating that the time series is more irregular than a random process.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Energy Science & Engineering* published by the Society of Chemical Industry and John Wiley & Sons Ltd.

# Referencia Bibliográfica

- [1] Karing Ohlenforst and Global Wind Energy Council. Global wind report 2020. <http://www.gwec.net>, accessed 11 June 2021.
- [2] Asociación Mexicana de Energía Eólica. Asociación mexicana de energía eólica. <https://amdee.org>, accessed 10 February 2021.
- [3] Bri-Mathias Hodge, Austin Zeiler, Duncan Brooks, Gary Blau, Joseph Pekny, and Gintaras Reklatis. Improved wind power forecasting with arima models. In *Computer Aided Chemical Engineering*, volume 29, pages 1789–1793. Elsevier, 2011.
- [4] Th Slini, K Karatzas, and N Moussiopoulos. Statistical analysis of environmental data as the basis of forecasting: an air quality application. *Science of the total environment*, 288(3):227–237, 2002.
- [5] Zhenkun Liu, Ping Jiang, Lifang Zhang, and Xinsong Niu. A combined forecasting model for time series: Application to short-term wind speed forecasting. *Applied Energy*, 259:114137, 2020.
- [6] Sinvaldo Rodrigues Moreno and Leandro dos Santos Coelho. Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system. *Renewable energy*, 126:736–754, 2018.
- [7] Saeid Mehdizadeh, Ali Kozekalani Sales, and Mir Jafar Sadegh Safari. Estimating the short-term and long-term wind speeds: implementing hybrid models through coupling machine learning and linear time series models. *SN Applied Sciences*, 2:1–15, 2020.
- [8] Wenzhe Li, Xiaodong Jia, Xiang Li, Yinglu Wang, and Jay Lee. A markov model for short term wind speed prediction by integrating the wind acceleration information. *Renewable Energy*, 164:242–253, 2021.

- [9] Jianming Hu, Jianzhou Wang, and Guowei Zeng. A hybrid forecasting approach applied to wind speed time series. *Renewable Energy*, 60:185–194, 2013.
- [10] Jianzhou Wang, Shanshan Qin, Qingping Zhou, and Haiyan Jiang. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in xinjiang, china. *Renewable Energy*, 76:91–101, 2015.
- [11] Chao Chen and Hui Liu. Medium-term wind power forecasting based on multi-resolution multi-learner ensemble and adaptive model selection. *Energy Conversion and Management*, 206:112492, 2020.
- [12] Sinvaldo Rodrigues Moreno, Viviana Cocco Mariani, and Leandro dos Santos Coelho. Hybrid multi-stage decomposition with parametric model applied to wind speed forecasting in brazilian northeast. *Renewable Energy*, 164:1508–1526, 2021.
- [13] Yuxiu Hua, Zhifeng Zhao, Rongpeng Li, Xianfu Chen, Zhiming Liu, and Honggang Zhang. Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*, 57(6):114–119, 2019.
- [14] Chaoshun Li, Zhengguang Xiao, Xin Xia, Wen Zou, and Chu Zhang. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. *Applied Energy*, 215:131–144, 2018.
- [15] Yu Jiang, Zhe Song, and Andrew Kusiak. Very short-term wind speed forecasting with bayesian structural break model. *Renewable energy*, 50:637–647, 2013.
- [16] Erasmo Cadenas and Wilfrido Rivera. Short term wind speed forecasting in la venta, oaxaca, méxico, using artificial neural networks. *Renewable Energy*, 34(1):274–278, 2009.
- [17] Xiwei Mi, Hui Liu, and Yanfei Li. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. *Energy Conversion and Management*, 180:196–205, 2019.

- [18] Jianzhou Wang, Jianming Hu, Kailiang Ma, and Yixin Zhang. A self-adaptive hybrid approach for wind speed forecasting. *Renewable Energy*, 78:374–385, 2015.
- [19] Sarbjit Singh, Kulwinder Singh Parmar, Jatinder Kumar, and Sidhu Jitendra Singh Makkhan. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (arima) models in application to one month forecast the casualties cases of covid-19. *Chaos, Solitons & Fractals*, page 109866, 2020.
- [20] Juan Cormán Trujillo. *Comportamiento del gasto público de una entidad pública usando el modelo Box-Jenkins, 2009-2017*. PhD thesis, Universidad Nacional Agraria La Molina, 2020.
- [21] Peter J Brockwell and Richard A Davis. *Introduction to time series and forecasting*. springer, 2016.
- [22] Theodoros Koutroumanidis, Konstantinos Ioannou, and Garyfallos Arabatzis. Predicting fuelwood prices in greece with the use of arima models, artificial neural networks and a hybrid arima–ann model. *Energy Policy*, 37(9):3627–3634, 2009.
- [23] Tian Zhongda, Li Shujiang, Wang Yanhong, and Sha Yi. A prediction method based on wavelet transform and multiple models fusion for chaotic time series. *Chaos, Solitons & Fractals*, 98:158–172, 2017.
- [24] Damián Jorge Matich. *Redes neuronales: Conceptos básicos y aplicaciones*. Universidad Tecnológica Nacional, México, 41, 2001.
- [25] Fernando Berzal. *Redes neuronales & deep learning*. Independently published, 2018.
- [26] Salim Lahmiri and Stelios Bekiros. Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, 118:35–40, 2019.
- [27] Xabier Basogain Olabe. *Redes neuronales artificiales y sus aplicaciones*. *Publicaciones de la Escuela de Ingenieros*, 1998.
- [28] Anirban DasGupta. *Probability for statistics and machine learning: fundamentals and advanced topics*. Springer Science & Business Media, 2011.

- [29] Jianbo Gao, Yinhe Cao, Wen-wen Tung, and Jing Hu. *Multiscale analysis of complex time series: integration of chaos and random fractal theory, and beyond*. John Wiley & Sons, 2007.
- [30] Shaun Lovejoy and Daniel Schertzer. Scale, scaling and multifractals in geophysics: twenty years on. In *Nonlinear dynamics in geosciences*, pages 311–337. Springer, 2007.
- [31] Andrei Nikolaevich Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 434(1890):9–13, 1991.
- [32] Wen-wen Tung, Mitchell W Moncrieff, and Jian-Bo Gao. A systemic analysis of multiscale deep convective variability over the tropical pacific. *Journal of climate*, 17(14):2736–2751, 2004.
- [33] Anthony Davis, Alexander Marshak, Warren Wiscombe, and Robert Cahalan. Multifractal characterizations of nonstationarity and intermittency in geophysical fields: Observed, retrieved, or simulated. *Journal of Geophysical Research: Atmospheres*, 99(D4):8055–8072, 1994.
- [34] Anthony Davis, A Marshak, WARRENJ Wiscombe, and R Cahalan. Multifractal characterizations of intermittency in nonstationary geophysical signals and fields. *Current topics in nonstationary analysis*, pages 97–158, 1996.
- [35] Benoit B Mandelbrot. *Fractals and scaling in finance: Discontinuity, concentration, risk. Selecta volume E*. Springer Science & Business Media, 2013.
- [36] Thomas M Over and Vijay K Gupta. A space-time theory of mesoscale rainfall using random cascades. *Journal of Geophysical Research: Atmospheres*, 101(D21):26319–26331, 1996.
- [37] Marco Aurelio Alzate Monroy. *Aplicaciones y mediciones-Derivas de complejidad*. Editorial Universidad del Rosario, 2012.
- [38] Jeferson Wilian de Godoy Stênico and Lee Luan Ling. Modern network traffic modeling based on binomial multiplicative cascades. *The Journal of Supercomputing*, 71(5):1712–1735, 2015.

- [39] Xiyang IA Yang and Adrián Lozano-Durán. A multifractal model for the momentum transfer process in wall-bounded flows. *Journal of fluid mechanics*, 824, 2017.
- [40] Peter Molnar and Paolo Burlando. Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. *Atmospheric Research*, 77(1-4):137–151, 2005.
- [41] Eduardo Faleiro Usanos. *Estructura Multifractal y aplicaciones de las fluctuaciones en cascadas atmosféricas producidas por rayos cósmicos*. Universidad Complutense de Madrid, Servicio de Publicaciones, 2003.
- [42] Brendan Fox, Damian Flynn, Leslie Bryans, Nick Jenkins, David Milborrow, Mark O'Malley, Richard Watson, and Olimpo Anaya-Lara. *Wind power integration: connection and system operational aspects*, volume 50. Iet, 2007.
- [43] Kasım Koçak. A method for determination of wind speed persistence and its application. *Energy*, 27(10):967–973, 2002.
- [44] Kasım Koçak. Practical ways of evaluating wind speed persistence. *Energy*, 33(1):65–70, 2008.
- [45] Hui, Hong-Qi Tian, Chao Chen, and Yan-fei Li. A hybrid statistical method to predict wind speed and wind power. *Renewable energy*, 35(8):1857–1861, 2010.
- [46] Jing Shi, Jinmei Guo, and Songtao Zheng. Evaluation of hybrid forecasting approaches for wind speed and power generation time series. *Renewable and Sustainable Energy Reviews*, 16(5):3471–3480, 2012.
- [47] Laura Cabrera Brito. *Análisis fractal de procesos dinámicos en el océano*. PhD thesis, Universidad de Las Palmas de Gran Canaria, 2016.
- [48] Jonas Olsson, Janusz Niemczynowicz, and Ronny Berndtsson. Fractal analysis of high-resolution rainfall time series. *Journal of Geophysical Research: Atmospheres*, 98(D12):23265–23274, 1993.
- [49] Jan W Kantelhardt. Encyclopedia of complexity and systems science. *Fractal and Multifractal Time Series*, pages 3754–3779, 2009.

- [50] Patrice Abry, Paolo Goncalves, and Jacques Lévy Véhel. *Scaling, fractals and wavelets*. John Wiley & Sons, 2013.
- [51] N Kalamaras, K Philippopoulos, D Deligiorgi, CG Tzanis, and G Karvounis. Multifractal scaling properties of daily air temperature time series. *Chaos, Solitons & Fractals*, 98:38–43, 2017.
- [52] Mohamed Laib, Jean Golay, Luciano Telesca, and Mikhail Kanevski. Multifractal analysis of the time series of daily means of wind speed in complex regions. *Chaos, Solitons & Fractals*, 109:118–127, 2018.
- [53] Jan W Kantelhardt, Stephan A Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, and H Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1-4):87–114, 2002.
- [54] Caiping Xi, Shuning Zhang, Gang Xiong, and Huichang Zhao. A comparative study of two-dimensional multifractal detrended fluctuation analysis and two-dimensional multifractal detrended moving average algorithm to estimate the multifractal spectrum. *Physica A: Statistical Mechanics and its Applications*, 454:34–50, 2016.
- [55] David Harte. *Multifractals: theory and applications*. Chapman and Hall/CRC, 2001.
- [56] Piotr Baranowski, Magdalena Gos, Jaromir Krzyszcak, Krzysztof Siwek, Adam Kieliszek, and Przemysław Tkaczyk. Multifractality of meteorological time series for poland on the base of merra-2 data. *Chaos, Solitons & Fractals*, 127:318–333, 2019.
- [57] Tao Feng, Zuntao Fu, Xing Deng, and Jiangyu Mao. A brief description to different multi-fractal behaviors of daily wind speed records over china. *Physics Letters A*, 373(45):4134–4141, 2009.
- [58] Rajesh G Kavasseri and Radhakrishnan Nagarajan. A multifractal description of wind speed records. *Chaos, Solitons & Fractals*, 24(1):165–173, 2005.
- [59] Román Rodríguez Aguilar. El coeficiente de hurst y el parámetro  $\alpha$ -estable para el análisis de series financieras aplicación al mercado cambiario mexicano. *Contaduría y administración*, 59(1):149–173, 2014.

- [60] Espen Alexander Fürst EAFI Ihlen. Introduction to multifractal de-trended fluctuation analysis in matlab. *Frontiers in physiology*, 3:141, 2012.
- [61] R Rodríguez. Predicción de crisis financieras a través de la volatilidad implícita del precio de las opciones, asumiendo distribuciones estables (tesis doctorado). *México DF, México: Instituto Politécnico Nacional*, 2013.
- [62] Ruiz Julio, Gutierrez Doricela, Rico Molina, and Albarran Yaroslaf. Estimación del coeficiente de hurst y minería de datos en el precio del petróleo mexicano. *Revista Aristas: Investigación Básica y Aplicada*, 7(14), 2019.
- [63] Sergio Juárez Hernández and Gabriel León. Energía eólica en el istmo de tehuantepec: desarrollo, actores y oposición social. *Problemas del Desarrollo*, 45(178):139 – 162, 2014.
- [64] C Martín del Campo-Márquez, PF Nelson-Edelstein, and MÁ García-Vázquez. Wind power in mexico: Simulation of a wind farm and application of probabilistic safety analysis. *Ingeniería Investigación y Tecnología*, 10(4):343–352, 2009.
- [65] Mohamed Laib, Jean Golay, Luciano Telesca, and Mikhail Kanevski. Multifractal analysis of the time series of daily means of wind speed in complex regions. *Chaos, Solitons & Fractals*, 109:118 – 127, 2018.
- [66] Luis Alfredo Fernández Jiménez. *Modelos avanzados para la predicción a corto plazo de la producción eléctrica en parques eólicos*. PhD thesis, Universidad de La Rioja, 2007.
- [67] Kumar Molugaram, G Shanker Rao, Anil Shah, and Naresh Davergave. *Statistical techniques for transportation engineering*. Butterworth-Heinemann, 2017.
- [68] Martin H. Trauth. *Time-Series Analysis*, pages 107–159. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [69] Gonzalo Rios. Series de tiempo. Technical report, Universidad de Chile, 2008.

- [70] Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman. *Forecasting methods and applications*. John Wiley & Sons, 2008.
- [71] Isaac Van der Hoven. Power spectrum of horizontal wind speed in the frequency range from 0.0007 to 900 cycles per hour. *Journal of Atmospheric Sciences*, 14(2):160–164, 1957.
- [72] I Fernández and V Elena Parnás. Elementos para la simulación numérica de series temporales de velocidad de viento. *Revista ingeniería de construcción*, 32(2):83–90, 2017.
- [73] J. Guillermo Sierra. *Procesos Hurts y Movimiento Browniano Fraccional en Mercados Fractales- Edición Únicas*. PhD thesis, Instituto Tecnológico de Estudios Superiores de Monterrey Campus Ciudad de México, 2007.
- [74] Ian Stewart. *The nature of fractal geometry*. Springer, 2010.
- [75] R. Lopes and N. Betrouni. Fractal and multifractal analysis: A review. *Elsevier*, 13(4):634–649, 2009.
- [76] Tang Chengkun. From experiments in visualizing fractal theory to rethinking social networks as moistmedia. *Technoetic Arts*, 14(3):263–273, 2016.
- [77] Rolph Schwarzenberger. Fractal geometry: mathematical foundations and applications, by falconer kenneth. *The Mathematical Gazette*, 74(469):316–317, 1990.
- [78] Aniceto Murillo Mas. Geometría fractal: el diseño de la naturaleza. *Recursos didácticos en matemáticas: Descartes y Ultra Fractal*, page 24, 2008.
- [79] Rufino Iturriaga, Carina Jovanovich, et al. Los fractales y el diseño en las construcciones. *TRIM: revista de investigación multidisciplinar*, pages 5–19, 2012.
- [80] Benoit B Mandelbrot, Benoît B Mandelbrot, Benoît B Mandelbrot, and Benoît B Mandelbrot. *Les objets fractals: forme, hasard et dimension*, volume 17. Flammarion Paris, 1975.

- [81] Ariel Osvaldo Quezada Len et al. Fractales, más allá de 1d, 2d o 3d. *Revista Digital Universitaria*, 2005.
- [82] Benoit B Mandelbrot. *The fractal geometry of nature*, volume 1. WH freeman New York, 1982.
- [83] Gustavo Rubiano. El conjunto de mandelbrot. *Boletín de Matemáticas*, 3(1):25–36, 1996.
- [84] ML Sahari and I Djellit. Fractal newton basins. *Discrete Dynamics in Nature and Society*, 2006, 2006.
- [85] JS Thorp, SA Naqavi, and H-D Chiang. More load flow fractals. In *29th IEEE Conference on Decision and Control*, pages 3028–3030. IEEE, 1990.
- [86] John Ramírez. Introducción a las series de tiempo fractales. *Matemática*, 2(1), 2020.
- [87] Paredes Hernández Ulises. *Caracterización Fractal y Multifractal de Registros de Respiración usando Bombesina sobre el Complejo Pre-B<sup>0</sup>tzinger*. PhD thesis, Universidad Autónoma del Estado del Hidalgo Instituto de Ciencias Básicas e Ingeniería, 2019.
- [88] Gabriel Enrique Gajardo Cea. Estudio de las características multifractales en la correlación entre índices bursátiles latinoamericanos y el precio del crudo. Master's thesis, Universidad Técnica Federico Santa María, 2017.
- [89] Norman Gaither and Greg Frazier. *Administración de producción y operaciones*, volume 8. International Thomson, 2000.
- [90] Maimouna Diagne, Mathieu David, Philippe Lauret, John Boland, and Nicolas Schmutz. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27:65–76, 2013.
- [91] John Hanke and Arthur Reitsch. *Pronósticos en los negocios. Quinta Edición*. Prentice Hall Hispanoamericana, México, 1996.
- [92] Juan Inzunza. *Meteorología descriptiva*. Departamento de Geofísica, Universidad de Concepción, Concepción, 2012.

- [93] John E Hanke and Dean W Wichern. *Pronósticos en los negocios*. Pearson educación, 2006.
- [94] Yu-Wei Chang and Meng-Yuan Liao. A seasonal arima model of tourism forecasting: The case of taiwan. *Asia Pacific journal of Tourism research*, 15(2):215–221, 2010.
- [95] Bruce L Bowerman, Richard T O’Connell, Anne B Koehler, and Miguel Balderas Lozada. *Pronósticos, series de tiempo y regresión: un enfoque aplicado*, volume 4. Thomson, 2007.
- [96] Loren Pankratz. A new technique for the assessment and modification of feigned memory deficit. *Perceptual and Motor Skills*, 57(2):367–372, 1983.
- [97] Najib Bekkali. *Modelos ARIMA y Función de Transferencia para Previsiones Temporales*. PhD thesis, Universitat Politècnica de València, 2014.
- [98] Jordi Torres. *Python Deep Learning: Introducción práctica con Keras y TensorFlow 2*. Marcombo, 2020.
- [99] Teuvo Kohonen. An introduction to neural computing. *Neural networks*, 1(1):3–16, 1988.
- [100] Soumen Chakrabarti, Earl Cox, Eibe Frank, Ralf Hartmut Güting, Jia-wei Han, Xia Jiang, Micheline Kamber, Sam S Lightstone, Thomas P Nadeau, Richard E Neapolitan, et al. *Data mining: know it all*. Morgan Kaufmann, 2008.
- [101] José Ramón Hilera González, Víctor José Martínez Hernando, et al. *Redes neuronales artificiales: fundamentos, modelos y aplicaciones*. PhD thesis, Universidad Autónoma de Nayarit (UAN), México, 2000.
- [102] Carlos A Scolari. *Las leyes de la interfaz: diseño, ecología, evolución, tecnología*, volume 141. Editorial Gedisa, 2021.
- [103] Víctor Manuel Alcaraz. *Estructura y función del sistema nervioso*. UNAM, 2000.
- [104] Martín Javier Alarcón Mondéjar. *Algoritmos de decisión para priorizar las acciones de restauración en el mantenimiento de redes*. PhD thesis, Universidad Politécnica de Madrid, 2016.

- [105] Josep M Sopena, Pedro J Ramos, Joan López-Moliner, and Elizabeth Gilboy. Composicionalidad, cómputo de estructura y redes neuronales. *Estudios de Psicología*, 23(2):175–235, 2002.
- [106] Duane E Haines, PhD Faaas Faaa, and Gregory A Mihailoff. *Principios de neurociencia: aplicaciones básicas y clínicas*. Elsevier, 2019.
- [107] Vahid Mirjalili and Sebastian Raschka. *Python machine learning*. Marcombo, 2020.
- [108] Robert Hecht-Nielsen. Neurocomputing: picking the human brain. *IEEE spectrum*, 25(3):36–41, 1988.
- [109] AL Hodgkin and AF Huxley. This week’s citation classic. *J. Physiol*, 117:500–44, 1952.
- [110] Yoh-Han Pao. Neural net computing for pattern recognition. In *Handbook of Pattern Recognition and Computer Vision*, pages 125–162. World Scientific, 1993.
- [111] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323, 2011.
- [112] A Ronald Gallant and Halbert White. There exists a neural network that does not make avoidable mistakes. In *Proceedings of the Second Annual IEEE Conference on Neural Networks, San Diego, CA, I*, 1988.
- [113] A Requena, R Quintanilla, JM Bolarín, A Vázquez, A Bastida, and JZ LMT. *Nuevas tecnologías y contaminación de atmósferas*, 2015.
- [114] Raúl Pino Díez, Alberto Gómez Gómez, and Nicolás de Abajo Martínez. *Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva*. Universidad de Oviedo, 2001.
- [115] Andrés Sánchez Mangas. Análisis de componentes principales: versiones dispersas y robustas al ruido impulsivo. Master’s thesis, Universidad Carlos III de Madrid, 2012.

- [116] Raquel Flórez López and José Miguel Fernandez Fernandez. *Las redes neuronales artificiales*. Netbiblo, 2008.
- [117] Marcos Redondo Fonseca. *Simulación de redes neuronales como herramienta Big Data en el ámbito sanitario*. Lulu. com, 2016.
- [118] Docente Investigador Principal, Ing Morris William Díaz Saravia, Docente Co Investigador, and Ing José Manuel Trejo Peraza. *Diseño de prototipo de vehículo autónomo utilizando red neuronal*. PhD thesis, Consorcio de Bibliotecas Universitarias de El Salvador, 2019.
- [119] Ernesto Varela-Arregoces and Edwin Campbells. Redes neuronales artificiales: una revisión del estado del arte, aplicaciones y tendencias futuras. *Investigación y Desarrollo en TIC*, 2(1):18–27, 2011.
- [120] Alba Centeno Franco. *Deep learning*. PhD thesis, Universidad de Sevilla, 2019.
- [121] Juan José Montaña Moreno et al. *Redes neuronales artificiales aplicadas al análisis de datos*. PhD thesis, Universitat de les Illes Balears, 2002.
- [122] Roberto Fabio Leonarduzzi. *Análisis multifractal basado en coeficientes ondita líderes: formalismo multifractal basado en p-líderes y aplicación a señales biomédicas*. PhD thesis, Universidad Nacional del Litoral (UNL), 2014.
- [123] Lidia Soraya Contreras Ávila, Gabriel Armando Ospina García, Marco Aurelio Alzate Monroy, et al. Cascadas conservadoras aplicadas a la predicción de tráfico multifractal. *Ingeniería*, 11(1):62–67, 2006.
- [124] Benoit B Mandelbrot. Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of fluid Mechanics*, 62(2):331–358, 1974.
- [125] Patrice Abry, Herwig Wendt, Stéphane Jaffard, and Gustavo Didier. Multivariate scale-free temporal dynamics: From spectral (fourier) to fractal (wavelet) analysis. *Comptes Rendus Physique*, 20(5):489–501, 2019.
- [126] Rudolf Riedi. Multifractal scaling: General theory and approach by wavelets. *Scaling, Fractals and Wavelets*, pages 139–177, 2009.

- [127] Jens Feder. *Fractals*. Springer Science & Business Media, 2013.
- [128] Kenneth Falconer. *Fractal geometry: mathematical foundations and applications*. John Wiley & Sons, 2004.
- [129] Francisco Javier Caniego Monreal. Análisis multifractal de la distribución de la porosidad del suelo. Master's thesis, Universidad Politécnica de Madrid, 1997.
- [130] Peyton Z Peebles Jr. *Probability, random variables, and random signal principles*. McGraw Hill Book Company, 1987.
- [131] Chris Chatfield. *Time-series forecasting*. CRC press, 2000.
- [132] Albert-László Barabási and Tamás Vicsek. Multifractality of self-affine fractals. *Physical Review A*, 44(4):2730, 1991.
- [133] D Schertzer and S Lovejoy. Lecture notes: nonlinear variability in geophysics 3: scaling and multifractal processes in geophysics. In *NVAG3 Conference, Cargâese, Institut d'Etudes Scientifique de Cargâese, France*, page 292, 1993.
- [134] François Schmitt. *Turbulence développée et Multifractals universels en soufflerie et dans l'Atmosphère*. PhD thesis, Paris 6, 1993.
- [135] Jonthan E Nyquist and Michel Boufadel. Multifractal characterization of geologic noise for improved uxu detection and discrimination. Technical report, Temple Univ Philadelphia PA, 2008.
- [136] Erasmo Cadenas, Rafael Campos-Amezcuca, Wilfrido Rivera, Marco Antonio Espinosa-Medina, Alma Rosa Méndez-Gordillo, Eduardo Rangel, and Jorge Tena. Wind speed variability study based on the hurst coefficient and fractal dimensional analysis. *Energy Science & Engineering*, 7(2):361–378, 2019.
- [137] Dmitriy Piskarev. Calculating the hurst exponent, April 2017.
- [138] Sergio Cerutti and Carlo Marchesi. *Advanced methods of biomedical signal processing*, volume 27. John Wiley and Sons, 2011.
- [139] Oscar Yesid Quintero Delgado, Jonathan Ruiz Delgado, et al. Estimación del exponente de hurst y la dimensión fractal de una superficie

- topográfica a través de la extracción de perfiles. *UD y la geomática*, 5(5):84–91, 2011.
- [140] Michael J. Ostwald and Josephine Vaughan. *Introducing the Box-Counting Method*, pages 39–66. Springer International Publishing, Cham, 2016.
- [141] Carl Bovill and Carl Bovill. *Fractal geometry in architecture and design*. Springer, 1996.
- [142] Apuntes de Teoría de Señales. Introducción a la transformada wavelet. *Departamento de Señales y sistemas. Universidad de Navarra*, 2006.
- [143] E Gómez-Luna, D Silva, and G Aponte. Selección de una wavelet madre para el análisis frecuencial de señales eléctricas transitorias usando wpd. *Ingeniare. Revista chilena de ingeniería*, 21(2):262–270, 2013.
- [144] Jie Cao Guangxi Cao, Ling-Yun He. *Multifractal Detrended Analysis Method and Its Application in Financial Markets*. Springer, 2018.
- [145] Asha Chelani. Long-memory property in air pollutant concentrations. *Atmospheric Research*, 171:1–4, 2016.
- [146] Bernice E Rogowitz and R Voss. Shape perception and low-dimension fractal boundary contours. In *Human Vision and Electronic Imaging: Models, Methods, and Applications*, volume 1249, pages 387–394. International Society for Optics and Photonics, 1990.
- [147] Khaled Harrar and Mohamed Khider. Texture analysis using multifractal spectrum. *International Journal of Modeling and Optimization*, 4(4):336, 2014.
- [148] Esther Lubián Bermejo. Métodos wavelets para el análisis estadístico de series funcionales fractales. Technical report, Universidad de Granada, 2011.
- [149] C Yarlequé. *Análisis de campos de biomasa del altiplano usando wavelet y parámetros universales multifractales*. PhD thesis, Tesis de Licenciatura en Física, 2009.
- [150] Alejandro Salazar Couso. *Máster Oficial en Estadística Aplicada*. PhD thesis, Universidad de Granada, 2016.

- [151] Weipeng Gao and Youchan Zhu. A cloud computing fault detection method based on deep learning. *Journal of Computer and Communications*, 5(12):24–34, 2017.
- [152] Hongmin Li, Jianzhou Wang, Ranran Li, and Haiyan Lu. Novel analysis–forecast system based on multi-objective optimization for air quality index. *Journal of cleaner production*, 208:1365–1383, 2019.
- [153] Ping Jiang, Hufang Yang, and Xuejiao Ma. Coal production and consumption analysis, and forecasting of related carbon emission: evidence from china. *Carbon Management*, 10(2):189–208, 2019.
- [154] David R Stauffer and Nelson L Seaman. Use of four-dimensional data assimilation in a limited-area mesoscale model. part i: Experiments with synoptic-scale data. *Monthly Weather Review*, 118(6):1250–1277, 1990.
- [155] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.
- [156] James Douglas Hamilton. *Time series analysis*. Princeton university press, 2020.
- [157] Ruey S Tsay. *Analysis of financial time series*, volume 543. John wiley & sons, 2005.
- [158] David Stephen Geoffrey Pollock, Richard C Green, and Truong Nguyen. *Handbook of time series analysis, signal processing, and dynamics*. Elsevier, 1999.
- [159] Jianzhou Wang, Ying Wang, Zhiwu Li, Hongmin Li, and Hufang Yang. A combined framework based on data preprocessing, neural networks and multi-tracker optimizer for wind speed prediction. *Sustainable Energy Technologies and Assessments*, 40:100757, 2020.
- [160] Zaccheus O Olaofe. A 5-day wind speed & power forecasts using a layer recurrent neural network (lrnn). *Sustainable Energy Technologies and Assessments*, 6:1–24, 2014.

- [161] Rob J Hyndman and Anne B Koehler. Effect of question formats on item endorsement rates in web surveys. *International Journal of Forecasting*, 22(4):679–688, 2006.
- [162] Douglas C Montgomery, Cheryl L Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.
- [163] Hongmin Li, Jianzhou Wang, Haiyan Lu, and Zhenhai Guo. Research and application of a combined model based on variable weight for short term wind speed forecasting. *Renewable Energy*, 116:669–684, 2018.
- [164] Li Liu, Qianru Wang, Jianzhou Wang, and Ming Liu. A rolling grey model optimized by particle swarm optimization in economic prediction. *Computational Intelligence*, 32(3):391–419, 2016.
- [165] Ma Lei, Luan Shiyang, Jiang Chuanwen, Liu Hongling, and Zhang Yan. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews*, 13(4):915–920, 2009.
- [166] Guillermo Santamaría-Bonfil, A Reyes-Ballesteros, and C Gershenson. Wind speed forecasting for wind farms: A method based on support vector regression. *Renewable Energy*, 85:790–809, 2016.
- [167] Søren Bisgaard and Murat Kulahci. *Time series analysis and forecasting by example*. John Wiley & Sons, 2011.
- [168] Erasmo Cadenas and Wilfrido Rivera. Wind speed forecasting in the south coast of Oaxaca, Mexico. *Renewable Energy*, 32(12):2116–2128, 2007.
- [169] David Quintero Plaza. *Posproceso estadístico*, chapter 14, pages 193–207. Agencia Estatal de Meteorología (AEMET), 2018.
- [170] Pierre Chainais, Rudolf Riedi, and Patrice Abry. On non-scale-invariant infinitely divisible cascades. *IEEE transactions on Information Theory*, 51(3):1063–1083, 2005.
- [171] Carlos E Puente. Lecciones a partir de la complejidad la hipotenusa, el camino de la paz. *Complexus*, 1(3):32–53, 2005.

- 
- [172] Charles Meneveau and Katepalli R Sreenivasan. The multifractal spectrum of the dissipation field in turbulent flows. *Nuclear Physics B- Proceedings Supplements*, 2:49–76, 1987.
- [173] Sarah Balkissoon, Neil Fox, and Anthony Lupo. Fractal characteristics of tall tower wind speeds in missouri. *Renewable Energy*, 154:1346–1356, 2020.