

UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICA



TRABAJO DE GRADO

GENERACIÓN DE MODELOS ESTADÍSTICOS UTILIZANDO REDES NEURONALES
Y SERIES DE TIEMPO PARA EL PRONÓSTICO DE LOS NIVELES DEL RESERVORIO
DE LA PRESA HIDROELÉCTRICA CERRÓN GRANDE DE EL SALVADOR

PARA OPTAR AL GRADO DE

LICENCIADO(A) EN ESTADÍSTICA

PRESENTADO POR

LILIAN MARICELA MARTÍNEZ MANCÍA

FRANCISCO JAVIER SANTOS MARTÍNEZ

DOCENTE ASESOR

LICENCIADO JAIME ISAAC PEÑA MEJÍA

OCTUBRE, 2019

SANTA ANA, EL SALVADOR, CENTROAMÉRICA

UNIVERSIDAD DE EL SALVADOR

AUTORIDADES



M.Sc. ROGER ARMANDO ARIAS ALVARADO

RECTOR

DR. MANUEL DE JESÚS JOYA ÁBREGO

VICERRECTOR ACADÉMICO

ING. NELSON BERNABÉ GRANADOS ALVARADO

VICERRECTOR ADMINISTRATIVO

LICDO. CRISTÓBAL HERNÁN RÍOS BENÍTEZ

SECRETARIO GENERAL

M.Sc. CLAUDIA MARÍA MELGAR DE ZAMBRANA
DEFENSORA DE LOS DERECHOS UNIVERSITARIOS

LICDO. RAFAEL HUMBERTO PEÑA MARÍN

FISCAL GENERAL

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

AUTORIDADES



DR. RAÚL ERNESTO AZCÚNAGA LÓPEZ

DECANO

M.Ed. ROBERTO CARLOS SIGÜENZA CAMPOS

VICEDECANO

M.Sc. DAVID ALFONSO MATA ALDANA

SECRETARIO

LICDA. ROSARIO CÁCERES AGUILAR

JEFA DEL DEPARTAMENTO DE MATEMÁTICA

*Dedicado a mi familia,
y a cada una de las personas
que estuvieron presentes apoyando
en el desarrollo del trabajo de
grado, y durante el trayecto
de la carrera universitaria.*

Agradecimientos

Los éxitos que más se valoran son los logrados con el esfuerzo propio. Gracias por estar presente no solo en esta etapa de mi vida sino en todo momento, ofreciendo lo mejor y permitiendo sonreír ante los logros que son el resultado de tu ayuda. Que, al caer y ponerme a prueba, haces que mejore como ser humano y crezca de diversas formas. Este trabajo de grado ha sido una gran bendición en todo sentido y por eso te lo agradezco profundamente Dios mío.

Gracias a mis padres por ser los principales promotores de mis sueños, por creer en mí y confiar en cada uno de mis expectativas, porque se preocupaban por el avance y el desarrollo del proyecto. Gracias por la compañía en las noches agotadoras de estudio, por cada consejo y palabras que me guiaron durante mi trayecto académico y de mi vida, siendo motivadoras para seguir adelante.

El desarrollo del proyecto de graduación no resultó en un proceso fácil, pero con el apoyo de cada uno de mis amigos que estuvieron presentes durante todo o la mayor parte de la realización de éste, se pudo convertir en una experiencia amena y gratificante al observar los resultados. Gracias a mi asesor, docentes y a todos aquellos que con mucho respeto y decencia realizaron aportes para el proyecto de grado, además un agradecimiento para las personas que tuvieron un momento para leer dicho proyecto.

Gracias a la vida por este triunfo y a todos por el apoyo, por creer y confiar siempre.

Índice

Resumen	xiv
Introducción	xvi
Capítulo 1. Planteamiento del Problema	18
1.1. Descripción del Problema	18
1.2. Justificación	20
1.3. Objetivos	21
1.3.1. Objetivo General.....	21
1.3.2. Objetivos Específicos	21
Capítulo 2. Fundamento Teórico: Redes Neuronales Artificiales.....	22
2.1. Introducción a la Minería de Datos.....	22
2.2. Redes Neuronales	25
2.3. Redes Neuronales Artificiales.....	29
2.3.1. Historia de las Redes Neuronales Artificiales	30
2.3.2. Las Redes Neuronales Artificiales (RNA)	33
2.3.3. Funciones de activación.....	37
2.3.4. Arquitectura de una red neuronal	42
2.3.6. Tipos de Redes Neuronales Artificiales	51
2.3.7. Entrenamiento de redes neuronales artificiales	57
2.3.8. El Sobreaprendizaje o sobreajuste en una Red Neuronal	66
2.3.9. Fase de Validación o Test de una Red Neuronal.....	68
2.3.10. Aplicaciones de las redes neuronales artificiales	70
Capítulo 3. Fundamento Teórico: Series de Tiempo.....	73
3.1. Componentes de una serie de tiempo	75
3.2. Clasificación de una serie de tiempo	76
3.3. Frecuencias de las series de tiempo	77
3.4. Procesos Clásicos.....	77
3.5. Proceso estocástico	79
3.5.1. Características de un proceso estocástico.....	80
3.5.2. Procesos estocásticos estacionarios	82

3.5.3.	Función de autocovarianzas y de autocorrelación	83
3.5.4	Estimación de los momentos	86
3.5.5.	Ruido Blanco	87
3.6.	Modelos ARMA	88
3.6.1.	Modelos autorregresivos	88
3.6.2.	Modelo de medias móviles.....	92
3.6.3.	Modelo Autorregresivos y de Medias Móviles (ARMA).....	95
3.7.	Estacionariedad.....	98
3.7.1.	Identificación de Estacionariedad.....	100
3.8.	Raíz Unitaria	105
3.8.1.	Pruebas de Raíz Unitaria	106
3.9.	Modelo Autorregresivo Integrado de Media Móvil ARIMA (p, d, q).....	108
3.10.	Modelos SARIMA.....	109
3.10.1.	Modelo Multiplicativo SARIMA	111
3.10.2.	Pruebas de Raíz Unitaria Estacional.....	113
3.11.	Validación de los modelos Univariantes	116
3.11.1.	Medición del error de pronóstico.....	116
3.11.2.	Criterios de selección del modelo.....	120
Capítulo 4. Marco Metodológico.....		122
4.1.	Tipo de Investigación.....	122
4.2.	Población y Muestra	122
4.3.	Recolección de los datos.....	123
4.4.	Procesamiento de los datos	123
4.4.1.	Diseño del Modelo de Redes Neuronales.....	123
4.4.2.	Diseño del Modelo de Series de tiempo	126
4.4.3.	Contraste de los modelos de pronóstico	128
Capítulo 5. Resultados.....		129
5.1.	Resultados Generales	129
5.2.	Análisis de la Red Neuronal	132
5.2.1.	Red Neuronal con Función Sigmoide.....	133
5.2.2.	Red Neuronal con Función Tangente Hiperbólica	135
5.2.3.	Resultados de la Red Neuronal.....	138
5.2.4.	Pronóstico con Redes Neuronales	144

5.3.	Análisis de la Serie de Tiempo	147
5.3.1.	Análisis de la serie original	147
5.3.2.	Identificación del modelo	151
5.3.3.	Estimación del modelo	157
5.3.4.	Diagnosís del modelo	165
2.3.5.	Pronóstico con el Modelo de Series de Tiempo	172
5.4.	Comparación de modelos.....	175
	Conclusiones y Recomendaciones	178
	Conclusiones.....	178
	Recomendaciones	180
	Bibliografía.....	181
	Anexos.....	184
	Anexo 1. Datos originales de los niveles del reservorio.....	185
	Anexo 2. Nivel del reservorio generados por la red neuronal con la función Tangente Hiperbólica	186
	Anexo 3. Nivel del reservorio generados por la red neuronal con la función Sigmoide.....	186
	Anexo 4. Arquitectura de la Red Neuronal	187
	Anexo 5. Estimaciones de los parámetros generados por el modelo de la red neuronal.....	188
	Anexo 6. Comportamiento de los residuos del modelo de red neuronal	190

Índice de figuras

Figura 1: Representación de una neurona biológica	27
Figura 2: Nodo de una red neuronal.....	33
Figura 3: Interconexión entre una neurona presináptica y una neurona postsináptica.....	35
Figura 4: Modelo de neurona estándar	36
Figura 5: Función de Activación Lineal.....	38
Figura 6: Función de activación Escalón	39
Figura 7: Función de activación Tangente hiperbólica	40
Figura 8: Función de activación Sigmoidea.....	41
Figura 9: Función de activación Gaussiana.....	41
Figura 10: Esquema de una red neuronal monocapa.....	43
Figura 11: Esquema de una red neuronal multicapa	44
Figura 12: Esquema de una red neuronal recurrente.....	44
Figura 13: Esquema de una red neuronal totalmente conectada	45
Figura 14: Influencia de la neurona N_j en la neurona N_i	48
Figura 15: Estructura de un Perceptrón.....	51
Figura 16: El Perceptrón de Rosenblatt.....	52
Figura 17: Estructura de una red multicapa o perceptrón multicapa	55
Figura 18: Neurona oculta conectada con la neurona de la capa de salida	59
Figura 19: Modelo con sobregeneralización (izquierda), ajustado (centro) y sobreajustado (derecha).....	67
Figura 20: Componentes de una serie de tiempo	76
Figura 21: Regiones de contraste del estadístico d de Durbin-Watson.....	104
Figura 22: Niveles mensuales del reservorio de la presa Cerrón Grande de El Salvador desde el 2000 al 2018	130
Figura 23: Diagrama de caja-bigote de los niveles mensuales del reservorio del Cerrón Grande desde el 2000 al 2018.....	131
Figura 24: Estructura de la Red Neuronal para el nivel del reservorio del Cerrón Grande desde el 2000 al 2018	132

Figura 25: Correlación de los valores reales y pronósticos con el modelo de red neuronal del nivel del reservorio	138
Figura 26: Serie Pronosticada del nivel del reservorio utilizando el modelo de Redes Neuronales con la Función Sigmoide.....	139
Figura 27: Correlación de los valores reales y pronósticos con el modelo de red neuronal del nivel del reservorio	140
Figura 28: Serie Pronosticada utilizando el modelo de Redes Neuronales.....	140
Figura 29: Datos reales vs. pronósticos con red neuronal sigmoideal de los niveles del reservorio de la presa Cerrón Grande	142
Figura 30: Datos reales vs. pronósticos con red neuronal hiperbólica de los niveles del reservorio de la presa Cerrón Grande	143
Figura 31: Datos reales vs. pronósticos de la red neuronal del nivel del reservorio de la Presa Cerrón Grande	144
Figura 32: Valores reales y valores pronósticos del nivel del reservorio generados con Red neuronal desde el año 2000 hasta el 2021	145
Figura 33: Gráfico Temporal de los niveles del reservorio de la presa Cerrón Grande de El Salvador	147
Figura 34: Gráfico estacional de los niveles del reservorio de la presa Cerrón Grande de El Salvador	148
Figura 35: Gráfico estacional Polar de los niveles del reservorio de la presa Cerrón Grande de El Salvador	149
Figura 36: Gráfico estacional mensual de los niveles del reservorio de la presa Cerrón Grande de El Salvador.....	150
Figura 37: Gráfico estacional mensual por año de los niveles del reservorio de la presa Cerrón Grande de El Salvador	150
Figura 38: Componentes de la serie de los niveles del reservorio de la presa Cerrón Grande de El Salvador	152
Figura 39: Funciones de autocorrelaciones de los daos del nivel del reservorio del Cerrón Grande de El Salvador	152
Figura 40: Serie diferenciada estacional (D=1) de los niveles del reservorio de la presa Cerrón Grande	155

Figura 41: Funciones de autocorrelación de la serie diferenciada estacionalmente del nivel del reservorio del Cerrón Grande	155
Figura 42: Invertibilidad y estacionariedad de los modelos seleccionados	164
Figura 43: Residuos del modelo ARIMA(1,0,2)(1,1,1), de la serie nivel del reservorio	165
Figura 44: Normalidad de los residuos del modelo ARIMA(1,0,2)(1,1,1)	166
Figura 45: Funciones de autocorrelación del modelo ARIMA(1,0,2)(1,1,1)	166
Figura 46: Residuos del modelo ARIMA(1,0,2)(1,1,1) de la serie nivel del reservorio	168
Figura 47: Funciones de Autocorrelación del modelo ARIMA(1,0,2)(1,1,1) con análisis de intervención	170
Figura 48: Pronósticos del nivel del reservorio del Cerrón Grande generados con el modelo de series temporales.....	173
Figura 49: Comparación de los niveles del reservorio con los niveles creados con los modelos de pronóstico	177
Figura 50: Arquitectura de la Red Neuronal	187
Figura 51: Comportamiento de los residuos del modelo de red neuronal.....	190

Índice de tablas

Tabla 1: Estadísticos descriptivos de los niveles del reservorio de la presa Cerrón Grande de El Salvador	129
Tabla 2: Resumen del modelo 1 utilizando red neuronal artificial con función de activación sigmoide	133
Tabla 3: Errores del modelo 1 utilizando red neuronal artificial con función de activación sigmoide	134
Tabla 4: Resumen del modelo 2 utilizando red neuronal artificial con función de activación sigmoide	134
Tabla 5: Errores del modelo 2 utilizando red neuronal artificial con función de activación sigmoide	135
Tabla 6: Resumen del modelo 3 utilizando red neuronal artificial con función de activación tangente hiperbólica.....	136
Tabla 7: Errores del modelo 3 utilizando red neuronal artificial con función de activación tangente hiperbólica.....	136
Tabla 8: Resumen del modelo 4 utilizando red neuronal artificial con función de activación tangente hiperbólica.....	137
Tabla 9: Errores del modelo 4 utilizando red neuronal artificial con función de activación tangente hiperbólica.....	137
Tabla 10: Valores reales del nivel del reservorio y valores pronosticados con los modelos de redes neuronales identificados.....	141
Tabla 11: Precisión de pronóstico de los modelos de redes neuronales identificados	143
Tabla 12: Valores reales vs. valores pronosticados con Redes neuronales del nivel del reservorio de la presa Cerrón Grande	145
Tabla 13: Pronóstico con el modelo de Red Neuronal del nivel del reservorio de la presa Cerrón Grande desde el año 2019 hasta el 2021	146
Tabla 14: Test de raíces unitarias de Dickey-Fuller	153
Tabla 15: Estadísticos descriptivos de la serie diferenciada estacional de los niveles del reservorio Cerrón Grande de El Salvador.....	154
Tabla 16: Parámetros del modelo ARIMA (2,0,0)(0,1,1)	158

Tabla 17: Parámetros del modelo ARIMA (2,0,0)(1,1,1)	159
Tabla 18: Parámetros del modelo ARIMA (2,0,1)(0,1,1)	159
Tabla 19: Parámetros del modelo ARIMA (1,0,2)(1,1,1)	160
Tabla 20: Parámetros del modelo ARIMA(2,0,2)(0,1,1)	161
Tabla 21: Parámetros del modelo ARIMA (2,0,2)(1,1,0)	161
Tabla 22: Expresiones de los modelos ARIMA identificados	162
Tabla 23: Criterios de Estacionariedad e invertibilidad de los parámetros de los modelos ARIMA identificados	163
Tabla 24: Análisis autocorrelativos de los modelos propuestos.....	164
Tabla 25: Ljung-Box Test	167
Tabla 26: Modelo de la serie temporal	169
Tabla 27: Valores atípicos del modelo ARIMA(1,0,2)(1,1,1) de la serie temporal	169
Tabla 28: Resumen del modelo ARIMA(1,0,2)(1,1,1) sin intervención.....	171
Tabla 29: Resumen del modelo ARIMA(1,0,2)(1,1,1) con intervención.....	171
Tabla 30: Valores reales vs. valores pronosticados de los niveles del reservorio con series temporales.....	172
Tabla 31: Valores pronosticados con el modelo de series de tiempo del nivel del reservorio del 2019 al 2021	174
Tabla 32: Comparación de los valores pronosticados de los modelos de pronóstico con los valores reales	175
Tabla 33: Resumen de los errores de los modelos de redes neuronales y series temporales .	176
Tabla 34: Datos originales de los niveles del reservorio de la presa Hidroeléctrica Cerrón Grande	185
Tabla 35: Valores del nivel del reservorio generados por la red neuronal con la función de activación Tangente Hiperbólica	186
Tabla 36: Valores del nivel del reservorio generados por la red neuronal con la función de activación Sigmoide	186
Tabla 37: Estimaciones de los parámetros generados por el modelo de la red neuronal artificial.....	188

Resumen

El modelado de los niveles de los reservorios resulta ser importante para la operación del mismo en la presa, de igual forma funciona como apoyo para el diseño de estructuras hidráulicas y su continuo mantenimiento, además de poder hacer la determinación de la contaminación que se produce en el reservorio y la seguridad de la represa. Es por ello que el presente trabajo de investigación de grado tiene como objetivo el formular los modelos estadísticos con series temporales y redes neuronales artificiales para pronosticar los niveles del reservorio de la presa hidroeléctrica Cerrón Grande de El Salvador.

El trabajo de grado consta de cinco capítulos, el cual se basa a partir de los datos del nivel del reservorio producidos desde el año 2000 hasta el 2018, para realizar pronósticos del embalse para la presa Hidroeléctrica Cerrón Grande ubicada en el cauce del río Lempa en el territorio de El Salvador proporcionados por la Comisión Ejecutiva Hidroeléctrica del Río Lempa (CEL), y es en el primer capítulo donde se presenta el problema de investigación y los objetivos que se han propuesto cumplir con esta investigación.

En los capítulos dos y tres se encuentra el desarrollo teórico de las dos técnicas que se emplean para la creación de modelos de pronóstico, definiendo aspectos que ayudan a comprender el funcionamiento de las redes neuronales (capítulo dos), iniciando desde la introducción de las redes neuronales artificiales, su arquitectura y tipos de redes que se pueden trabajar, además, de los algoritmos necesarios para que una red realice el proceso de aprendizaje para una clasificación o predicción. Posteriormente, en el capítulo tres, primero se estudia los conceptos principales de la técnica y la clasificación de modelos de series de tiempo. En el capítulo cuatro se desarrolla la metodología de la investigación, de igual forma se presenta la metodología de trabajo de las redes neuronales y la metodología necesaria para modelar series de tiempo.

En el quinto capítulo se desarrolla la implementación en la base de datos los procedimientos de las redes neuronales y la serie de tiempo. En la primera sección se muestran los resultados generales, para continuar con los resultados de la construcción del modelo, la arquitectura de la red y los pronósticos generados por el modelo de redes neuronales artificiales. Seguidamente se trata los resultados de las series temporales en primer lugar se hace la identificación el modelo, después se estudia la estimación de parámetros en el modelo, posterior a esto se describe el proceso de diagnóstico y finalmente se realiza el pronóstico, el cual es de suma importancia pues es uno de los principales objetivos de la construcción de modelos de series de tiempo. Finalmente, en la última parte del capítulo se incluyen la comparación de los modelos de pronóstico de la serie de tiempo y de redes neuronales contrastándolo con el error cuadrático medio, para determinar el mejor modelo de pronóstico entre las dos técnicas.

Introducción

Actualmente la sociedad vive en constante desarrollo, donde el avance tecnológico relacionado a las computadoras ha provocado una revolución digital, permitiendo mejoras en el estilo de vida de las personas. Por medio de las computadoras se ha podido tener con diversas posibilidades de captura, obtención, almacenamiento y la disposición de cantidades enormes de datos. A partir de esto, la implementación de minería de datos¹ se vuelve necesaria para el procesamiento y gestión de los datos, pero en algunas ocasiones las técnicas tradicionales de estadística o herramientas de gestión convencionales no están preparadas para asumir la capacidad de trabajar con información en abundancia.

Sin embargo, el poder combinar las herramientas convencionales con las que ofrece la minería de datos proporciona nuevas oportunidades para explorar en el mundo que ofrece el Big data². Dentro del data mining se encuentran herramientas que permiten hacer uso de prácticas estadísticas, para realizar análisis a los conjuntos de datos encontrando modelos que representen la información, mediante el uso de algoritmos de búsqueda próximos a la Inteligencia Artificial³, siendo una de ellas las Redes Neuronales, en donde el conocimiento se incorpora mediante el aprendizaje a partir de ejemplos y el reconocimiento de patrones donde se adquiere, almacena y utiliza el conocimiento experimental.

Es por ello que implementar redes neuronales se vuelve llamativo debido a que en el ámbito educativo su aplicación es mínima y escasa, se observa más la utilización de series de tiempo, las cuales trabajan mediante secuencias de observaciones medidos en determinados momentos del tiempo, ordenados cronológicamente y espaciados entre sí de manera uniforme dado que los datos que se utilizan en ellas usualmente son dependientes entre sí, las serie de

¹ La minería de datos o data mining, es el conjunto de técnicas y tecnologías que permiten explorar grandes bases datos con el objetivo de encontrar patrones que aporten información en la toma de futuras decisiones.

² Big data (en español, grandes datos o grandes volúmenes de datos.) es un término evolutivo que describe cualquier cantidad voluminosa de datos estructurados, semiestructurados y no estructurados que tienen el potencial de ser extraídos para obtener información.

³ Es un campo de la informática que estudia y desarrolla sistemas capaces de realizar tareas que normalmente se atribuyen a la inteligencia humana.

tiempo tienen como principal objetivo el dar un análisis para hacer pronóstico (Üneş, Demirci, Kişi, & Özgür, 2015). Debido a esto, se espera brindar una visión más amplia de la forma de crear pronósticos de fenómenos con la implementación de las redes neuronales.

Este estudio es de carácter predictorio, el cual busca abrir el camino hacia la implementación de técnicas que traten información en cantidades grandes, partiendo desde el plantear la problemática y el fundamento teórico que se desarrolla en cada uno de los apartados, siguiendo una metodología específica para hacer el uso de las Redes Neuronales Artificiales y de las Series Temporales en software estadísticos, que permitirán estudiar cada una de las variables que influyen los niveles del reservorio de la presa Cerrón Grande, creando análisis que lleven a realizar modelos de predicción con las técnicas de Redes Neuronales o modelos de series de tiempo. Posterior a ello, presentar conclusiones que ayuden a generar soluciones a las problemáticas que estén relacionadas con el nivel de dicho lugar, y de otro modo presentar técnicas que ayuden a solucionar problemas que no sean de un comportamiento lineal.

Para ello se busca modelar resultados de los datos proporcionados por los encargados de la presa hidroeléctrica Cerrón Grande de El Salvador, con la finalidad de poder conocer datos a futuro sobre el nivel del agua del reservorio de la presa. Comparando los resultados obtenidos en cada una de las fases del trabajo y seleccionando el mejor modelo que brinde los pronósticos más acertados, confiables y seguros con el fin de pronosticar resultados que ayuden a la mejor toma de decisiones del reservorio de la presa. Para ello se contará con el apoyo del Ministerio de Medio Ambiente y Recursos Naturales (MARN) y la Comisión Ejecutiva Hidroeléctrica del Río Lempa (CEL), que se encuentran monitoreando fenómenos que se relacionan con el comportamiento del nivel del embalse de dicha presa.

Capítulo 1. Planteamiento del Problema

1.1. Descripción del Problema

El Salvador es un país en vías de desarrollo y en su territorio cuenta con recurso del agua que forma parte fundamental para el consumo, el uso cotidiano, para usos tradicionales agrícolas y la generación de fuentes de energía eléctrica para abastecer las ciudades grandes y pequeñas, incluyendo también al uso industrial que se da actualmente debido a la modernización nacional e internacional, es por ello que el agua se convierte en parte esencial dentro de la sociedad pues es demandada en gran escala, siendo importante conocer la capacidad y el nivel del agua que alcanzan los reservorios naturales (lagos, lagunas) o los artificiales.

Dado esto se dice que: “Un reservorio, embalse o represamiento se utiliza para definir la laguna artificial que se forma cuando se construye una presa en la boquilla de un vaso respectivo. Es un almacenamiento de agua que no necesariamente se habrá de ocupar el 100% de la extensión del vaso, porque de ello depende del balance hídrico⁴ correspondiente, en estricta concordancia con la capacidad que tenga” (Ministerio de Agricultura y Riego, 2017); es por ello que El Salvador cuenta con reservorios naturales y artificiales que ayudan a minimizar el desabastecimiento de la época seca, aprovechando la época lluviosa.

Los pronósticos de los niveles de agua en los embalses o reservorios en El Salvador no se han contemplado en los estudios estadísticos o de carácter hidrológicos, puesto que la cantidad de datos que se podrían generar en sus constantes mediciones torna complejo el manejo de dicha información con lo cual se dificulta identificar modelos que permitan prever posibles comportamientos de dichos niveles, y poder maximizar las operaciones de las presas hidroeléctricas que funcionan a través de los embalses y evitar posibles inundaciones en las

⁴ El balance global es aquel que se realiza sobre el conjunto de toda la cuenca incluyendo tanto el agua que escurre superficialmente como el agua que queda almacenada en el terreno.

zonas que rodean a estas estructuras, además de ver cuál es el comportamiento que tienen dicho nivel en las diversas épocas en el territorio salvadoreño.

Sin embargo, no se encuentran estudios que se asemejen al estudio de los niveles de dichos embalses, tanto así que las técnicas de series temporales con las que se cuentan no han sido aplicadas para modelizar el comportamiento que se presenta en el tiempo de los niveles de agua en los reservorios, que se pueden abordar con modelos de autorregresión (AR), media móvil autoregresiva (ARMA) y el modelo autorregresivo integrado de media móvil (ARIMA). Además, se ha dejado a un lado la implementación de técnicas de tratamiento de datos grandes manejados desde inteligencias artificiales, en las cuales las redes neuronales artificiales (ANN) han sido aceptadas como una herramienta apropiada para modelar fenómenos complejos no lineales en hidrología y sistemas de recursos hídricos, lo que lleva a la expansión de sus aplicaciones.

Estas dos técnicas permitieron el desarrollo de estrategias para pronosticar los niveles de reservorio en diferentes periodos de tiempo. Aunque es posible identificar modelos sofisticados que tengan en cuenta parámetros hidrológicos y meteorológicos, como precipitación, régimen, humedad y temperatura, es preferible realizar un modelo que simule variaciones de nivel basadas en registros de niveles pasados, y que simule dichas variaciones y se encuentren disponible para fines de investigación y también para fines prácticos. Dado esto surge la necesidad de crear modelos que ayuden al pronóstico de los niveles del reservorio de la Presa Hidroeléctrica El Cerrón Grande de El Salvador. Con la finalidad de ayudar a contestar a ¿cuál es el comportamiento de los datos pronosticados con los modelos estadísticos realizados con Redes Neuronales Artificiales y las Series Temporales? y ¿cuál modelo pronostica mejor el nivel del reservorio en El Cerrón Grande?

1.2. Justificación

La acumulación de agua producida por la construcción de una presa en un río, lago o laguna se le conoce como reservorio o embalse, cuya función es cerrar parcial o totalmente el cauce de un río. Los reservorios se construyen para regular el caudal de un río almacenar el agua proveniente de lagunas o lagos durante los períodos húmedos para utilizarlos durante los períodos más secos; también es utilizado para la generación de energía eléctrica, riego de cultivos, el abastecimiento de agua potable, uso industrial, contener y atenuar los caudales extremos de las crecidas (función de regulación de crecientes), además de crear espacios de esparcimiento y deportes acuáticos.

Conocer el nivel de agua con el que debe contar un reservorio es vital para que la presa realice la función para la que fue creada, se debe garantizar el nivel de agua adecuado a lo largo de un año, considerando como actuar ante un problema que pueda surgir en la temporada seca (verano) pues puede generar un nivel bajo de agua en el reservorio, al igual que la temporada de lluvias que puede crear un exceso de agua en la presa.

Los problemas que pueden generar las temporadas de sequía o el exceso de lluvia en el país, afectan a más de una persona y los encargados de las presas hidroeléctricas año con año lidian con dichos problemas, debido a que pueden surgir preguntas como: ¿En qué momento se debe abrir el vertedero para liberar agua del reservorio? ¿Qué cantidad de agua se debe liberar? ¿Cuánto tiempo se tardará el reservorio en volver a su nivel si se libera agua en temporada seca? entre otras interrogantes.

Por lo cual, se realiza un estudio para la Presa hidroeléctrica del Cerrón Grande, en donde se enfoca en los pronósticos de los niveles del agua del embalse de la presa, dado que no existen los suficientes antecedentes de creación de modelos en el estudio de los reservorios de presas hidroeléctrica del país. Para esto, surge la necesidad de realizar análisis de modelos estadísticos ARIMA, incorporando Redes Neuronales debido a que es capaz de predecir un modelo basado en la forma de trabajo del cerebro humano, conocido como procesos neuronales, para con ello poder resolver problemas desarrollado con patrones, reconocimiento de formas, clasificación, predicción de control.

1.3. Objetivos

1.3.1. Objetivo General

Formular modelos estadísticos con series temporales y redes neuronales artificiales para pronosticar los niveles del reservorio de la presa hidroeléctrica Cerrón Grande de El Salvador.

1.3.2. Objetivos Específicos

- Evaluar un modelo con redes neuronales artificiales para obtener pronósticos por periodos en los niveles del reservorio de la presa hidroeléctrica Cerrón Grande de El Salvador.
- Identificar modelos de autoregresión (AR), de media móvil autoregresiva (ARMA) y el Modelo promedio móvil progresivo autorregresivo estacional (SARIMA) para observar y pronosticar el nivel del agua de los reservorios de la Presa Hidroeléctrica Cerrón Grande de El Salvador.
- Contrastar los modelos de pronóstico generado por las series temporales, con el modelo de pronóstico generado por las Redes Neuronales Artificiales (ANN), en el nivel del agua de los reservorios de la Presa Hidroeléctrica El Cerrón Grande de El Salvador.

Capítulo 2. Fundamento Teórico: Redes Neuronales Artificiales

2.1. Introducción a la Minería de Datos

La revolución digital despuntada en las últimas décadas, donde las computadoras y los softwares han hecho posible que la información pueda ser digitalizada para su fácil captura, cumpliendo con el proceso de almacenar, distribuir y el proceso de transmitir de manera rápida y eficiente los datos. Con el importante progreso de las tecnologías informáticas y la expansión de su uso en diferentes aspectos de la vida, se continúa recogiendo y almacenando en bases de datos gran cantidad de información. Sin embargo, el tratar un enorme volumen de datos es un reto grande y complejo, es por ello que la minería de datos (MD) es un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada. La tecnología de Internet actual y su creciente demanda necesita el desarrollo de técnicas de minería de datos más avanzadas para interpretar la información y el conocimiento de los datos distribuidos por todo el mundo. (Mitra, 2003)

La Minería de Datos es un paso particular en el proceso del tratamiento de la información, consistiendo en la aplicación de algoritmos específicos para extraer patrones (modelos) de los datos, en donde se pueden apreciar la conformación de dos pasos que consisten en: la elección del modelo y por otro el ajuste final de éste a los datos, tomando uso y ventaja de varias áreas como la Estadística, la Inteligencia Artificial, la Computación Gráfica, las Bases de Datos y el Procesamiento Masivo, principalmente usando como materia prima las bases de datos. Pudiendo tener una definición podemos mencionar que la minería de dato es: “Un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos”. (Fayyad, 1996)

El concepto de la minería de datos está relacionado con conceptos de aprendizaje automático y de estadística. En general, la estadística es la que tradicionalmente se ha encargado

de extraer la información de los datos, apoyándose de metodologías de investigación y de procesos que se encuentran cimentados en técnicas procedentes de procesos matemáticos. Posteriormente cuando se empezó a usar la tecnología, el emplear los ordenadores como apoyo fundamental para facilitar la tarea, trajo los procesos Machine Learning⁵ que se basan en algoritmo o patrones basados en informaciones previas. (Benítez, Escudero, & Kanaan, 2009)

En ocasiones el tratamiento por parte de la estadística (principalmente estadística no paramétrica) se auxilia de la minería de datos, pues el tratamiento de los casos o fenómenos se diferencian de los clásicos en la cantidad de datos que se analizan, puesto que son mayores y como consecuencia las técnicas estadísticas clásicas no pueden ser aplicadas. La Minería de datos permite trabajar con grandes cantidades de observaciones sin ningún inconveniente de igual forma trata una gran cantidad de variables presentes en las bases de datos actuales. Las predicciones estadísticas de Minería de datos deben ser inspeccionadas constantemente de manera que comprendan y verifiquen lo que se ha producido y eso implica encontrar un término medio entre la capacidad explicativa del modelo (claridad) y su poder de predicción. En general, conforme aumenta el poder de predicción del modelo baja su capacidad de interpretar el fenómeno objeto de estudio. Mientras más sencilla sea la forma del modelo, más fácil será su comprensión, pero tendrá menor capacidad para tener en cuenta dependencias sutiles o demasiado variadas (no lineales).

La minería de datos surge ante el problema cada vez más apremiante de extraer información automáticamente a partir de masas de textos, surgiendo técnicas que se podrían proponer para los análisis dividiéndose dependiendo del objetivo que se persiga, teniendo las técnicas supervisadas y no supervisadas⁶. Pero dentro de éstas existen herramientas que se utilizan de manera frecuente en la minería de datos, dentro de las cuales se puede mencionar:

⁵ Aprendizaje Automático.

⁶ En la técnica supervisada hay una variable que debe ser explicada por las otras. Y en las no supervisadas no hay una variable preferente que debe ser explicada por las otras.

Redes neuronales

Es una herramienta de análisis estadístico que permite la construcción de un modelo de comportamiento a partir de una determinada cantidad de ejemplos (constituidos por una determinada cantidad de variables descriptivas de dicho comportamiento). La red neuronal completamente ignorante al principio, efectúa un aprendizaje partiendo de los ejemplos para luego transformarse a través de modificaciones sucesivas en un modelo susceptible de rendir cuenta del comportamiento observado en función de las variables descriptivas. Se ha popularizado porque actualmente es cuando se ha llegado a un cierto poder de cálculo necesario para su puesta en aplicación y los componentes informáticos se han convertido en herramientas potentes a la hora de realizar el procesamiento de la información.

Árboles de decisión

Son modelos que tienen estructuras de forma de árbol que representan conjuntos de decisiones generando reglas para la clasificación de un conjunto de datos. Existen métodos específicos de árboles de decisión los cuales incluyen Árboles de Clasificación y Regresión (CART: Classification And Regression Tree). Los árboles de decisión, son representaciones gráficas de la lógica de las probabilidades aplicada a las alternativas de decisión. En este tipo de técnicas el tronco del árbol representa el punto de partida de la decisión, las ramas de éste comienzan con la probabilidad del primer acontecimiento. La probabilidad de cada acontecimiento produce dos o más efectos posibles, algunos de los cuales conducen a otros acontecimientos de probabilidad y a puntos de decisión sub-consecuentes. Los valores en los que se cifran las ramas del árbol, provienen de un análisis muy cuidadoso que se basa en el establecimiento de un criterio para la toma de decisión.

Algoritmos Genéticos

El algoritmo genético permite obtener soluciones a problemas que no tiene ningún método de resolución descrito de forma precisa, o cuya solución exacta si se conoce es demasiado complicada para ser calculada en un tiempo aceptable. En el caso particular de cuando se encuentran restricciones múltiples y complejas e incluso contradictorias que deben ser satisfechas simultáneamente, por ejemplo, casos como: formar equipos de trabajo, planificar rondas de entregas, implantar puntos de venta de manera óptima o construir modelos estadísticos.

Criterios para la toma de decisión (clasificación)

Es posible utilizar diversos criterios para la toma de decisiones se puede bien referir a la toma de una decisión de clasificación así que los criterios presentados a continuación podrán ser utilizados para resolver problemas de dicha índole. Entre distintos criterios se pueden destacar los siguientes:

- Criterio de Laplace
- Criterio de Wald
- Criterio de Hurwicz
- Criterio de Savage
- Criterio de Bayes

Dada la importancia de implementar estas técnicas, se tiene la intención de aplicar las redes neuronales, y es por ello que se trata la temática de Redes Neuronales.

2.2. Redes Neuronales

Una red neuronal artificial es un modelo matemático inspirado en el comportamiento biológico de las neuronas y en la estructura del cerebro, inclusive también puede ser vista como un sistema inteligente que lleva a cabo tareas de manera distinta a como lo hacen las computadoras actuales. Aunque los ordenadores de la actualidad son eficientes y rápidas, en el procesamiento de la información existen tareas muy complejas como el reconocimiento y

clasificación de patrones que demandan demasiado tiempo y esfuerzo aún en las computadoras más potentes de la actualidad, pero que el cerebro humano es más apto para resolverlas, muchas veces sin aparente esfuerzo.

El cerebro puede considerarse un sistema altamente complejo cuya unidad básica es la neurona, que está masivamente distribuida con conexiones entre ellas (se calcula que hay aproximadamente 10 billones de neuronas en la corteza cerebral y 60 trillones de conexiones neuronales). Fue Ramón y Cajal (1888) quién descubrió la estructura celular (neurona) del sistema nervioso y quien se encargó de defender la teoría la cual consistía en que: “las neuronas se interconectan entre sí de forma paralela, y no formando un circuito cerrado como el sistema sanguíneo” (Ramón & Agüeras, 1992). Una neurona consta de un cuerpo celular (soma⁷) de entre 10 y 80 μ m, del que surge un denso árbol de ramificaciones (dendritas⁸) y una fibra tubular (axón⁹.) de entre 100 μ m y 1 metro.

Una neurona cerebral puede recibir unas 10,000 señales de entradas y enviar a su vez su salida a varios cientos de neuronas. La conexión entre neuronas se llama sinapsis, pero no es una conexión física, sino que hay unos 2 μ m de separación, además que son conexiones unidireccionales en la que la transmisión de la información se hace de forma eléctrica en el interior de la neurona y de forma química entre neuronas; gracias a unas sustancias específicas llamadas neurotransmisores. Es de realzar que no todas las neuronas son iguales y existen muchos tipos diferentes según el número de ramificaciones de sus dendritas, la longitud del axón y otros detalles estructurales. Sin embargo, todas ellas operan con los mismos principios básicos.

⁷ Es el elemento central y más importante de la célula. La función consiste en controlar la parte que rige el funcionamiento y mantiene con vida a la neurona.

⁸ Las dendritas son prolongaciones nerviosas de las neuronas, que se encargarán de recibir los estímulos además de la alimentación celular y son las encargadas del aprendizaje y la memoria, al igual que las enfermedades degenerativas.

⁹ La función de los axones es el de hacer que las señales eléctricas que viajan a través de las neuronas vayan a parar a otro lugar del organismo, es una especie de conducto por el que pasan impulsos nerviosos a toda velocidad.

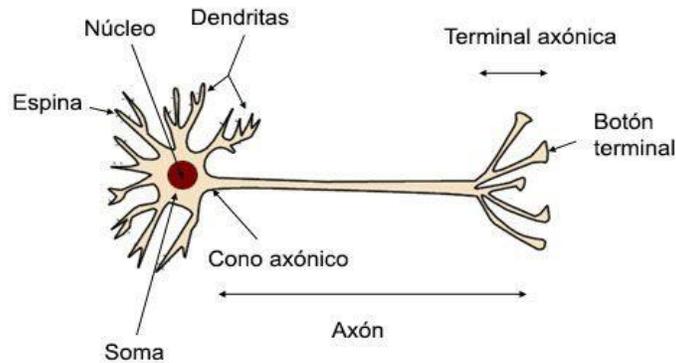


Figura 1: Representación de una neurona biológica

La función principal de las neuronas es la transmisión de los impulsos nerviosos, éstos viajan por toda la neurona comenzando por las dendritas hasta llegar a las terminaciones del axón, donde pasan a otra neurona por medio de la conexión sináptica. La manera en que responde ante los estímulos del mundo exterior y el aprendizaje del mismo está directamente relacionado con las conexiones neuronales del cerebro, y las redes neuronales artificiales son un intento de emular este hecho.

A partir de la estructura del sistema nervioso humano, las redes neuronales artificiales imitan hasta cierto punto la estructura física y el trabajo que puede realizar el cerebro. Dicho órgano es el encargado de manejar la información procedente de los estímulos, señales e impulsos eléctricos recibidos por parte de las neuronas que se encuentran distribuidas en las diferentes partes del cuerpo, presentando diversas cualidades de procesamiento distribuido en paralelo (Rumelhart & McClelland, 1992)¹⁰, y de adaptabilidad que un sistema de redes neuronales posee para comprender y reaccionar a los estímulos que se presentan constantemente al interactuar con el ambiente en donde los seres humanos se desarrollan.

¹⁰ El procesamiento distribuido en paralelo (también conocido como PDP) es una teoría sobre el aprendizaje que proponen una explicación al fenómeno de la adquisición de conocimiento basado en los procesos neuronales que tienen lugar en el cerebro.

Un sistema de redes neuronales artificiales resulta ser intrínsecamente¹¹ paralelo, al igual como el funcionamiento del sistema nervioso, pues dicha red artificial está formada por unidades elementales de procesamiento llamadas neuronas y a su vez cada neurona realiza un tipo de procesamiento simple, que ayuda a que el proceso de aprendizaje sea completado y llevado a cabo de manera normal y efectiva.

La información de las redes neuronales es distribuida debido a que no se almacena localmente en determinadas partes de la misma red, teniendo la ventaja que se encuentra presente por todo este sistema nervioso, para ser más identificables la información se aloja en la sinapsis entre cada una de las neuronas. A partir de esto, en la computación se comporta de manera distribuida, cuando la red neuronal artificial está en proceso de aprendizaje intervienen cada uno de los elementos que la componen, los cuales se distribuyen por la arquitectura de red que se encuentre funcionando, gracias a esta característica se brinda a la red la capacidad de ser tolerante a fallos, es decir, si se pierde una neurona o parte de las neuronas no se pierde en su totalidad la información.

Además, la red neuronal artificial posee un grado de adaptabilidad que está sustentada en la capacidad de aprender y la de generalizar. Cuando la red se encuentra en el proceso de aprendizaje se refiere a la capacidad de recoger la información de las experiencias pasadas y utilizarlas de forma eficiente para actuar ante situaciones presentadas en un futuro. A partir de ello se relaciona con la generalización para lograr abstraer información útil poniendo de lado los casos particulares, de esta manera es como una red neuronal es capaz de asumir el reto de responder casos desconocidos o fenómenos particulares. (Olabe, 2012)

¹¹ Intrínseco se refiere a una característica propia de un objeto que se expresa por sí misma y no depende de las circunstancias.

2.3. Redes Neuronales Artificiales

Las redes neuronales artificiales permiten resolver problemas relacionados con la clasificación y predicción, una de sus principales ventajas es que son capaces de lidiar con problemas de alta dimensionalidad y encontrar soluciones basadas en hiperplanos no lineales. Por el contrario, dos de sus principales problemas o inconvenientes son: que el conocimiento está oculto en las redes, y en determinados casos puede ser complejo explicitar este conocimiento de forma clara, seguidamente donde la preparación de los datos de entrada y salida no es una tarea trivial. Es recomendable asegurarse que los datos de entrada se encuentren en el intervalo $[0,1]$, lo cual implica unos procesos previos de transformación de los datos. Las redes neuronales se utilizan muy a menudo para problemas de clasificación relacionados con imágenes, así como otros problemas en los cuales tenemos datos de muy alta dimensionalidad. (Roig, Alfonso, Quiles, & Roma, 2017)

Es necesario tener en cuenta que existen ventajas de las redes neuronales artificiales frente a otros sistemas de procesamiento de información¹² de las cuales se pueden mencionar algunas de las siguientes:

- Las redes neuronales artificiales pueden sintetizar algoritmos a través de un proceso de aprendizaje.
- Para utilizar la tecnología neuronal no es necesario conocer los detalles matemáticos.
- Sólo se requiere estar familiarizado con los datos del trabajo.
- La solución de problemas no lineales es uno de los fuertes de las redes neuronales.
- Las redes neuronales son robustas¹³, pueden fallar algunos elementos de procesamiento, pero la red continúa trabajando; esto es contrario a lo que sucede en programación tradicional.

¹² Es la capacidad de un sistema para efectuar cálculos de acuerdo con una secuencia de operaciones preestablecida, dichos cálculos pueden efectuarse con datos introducidos recientemente en el sistema o bien con datos que están almacenados. Dicha característica de estos sistemas permite la transformación de datos que puede ser utilizada para la toma de decisiones.

¹³ Dada la cantidad de datos que maneja, la red neuronal debe ser capaz de enfrentar errores o llegue a minimizarlos al momento que se encuentre en ejecución.

Sin embargo, en todo tipo de técnicas existen desventajas, y ese no es la excepción con las redes neuronales, las cuales podrían ser:

- Las redes neuronales artificiales se deben entrenar¹⁴ para cada problema. Además, es necesario realizar múltiples pruebas para determinar la arquitectura adecuada. El entrenamiento es largo y puede consumir varias horas de la computadora (CPU).
- Debido a que las redes se entrenan en lugar de programarlas, éstas necesitan muchos datos.
- Las redes neuronales artificiales representan un aspecto complejo para un observador externo que desee realizar cambios. Para añadir nuevo conocimiento es necesario cambiar las iteraciones entre muchas unidades para que su efecto unificado sintetice este conocimiento. Para un problema de tamaño considerable es imposible hacer esto manualmente, por lo tanto, una red con representación distribuida debe emplear algún esquema de aprendizaje.

2.3.1. Historia de las Redes Neuronales Artificiales

- **Alan Turing (1936)**

Fue el primero en estudiar el cerebro como una forma de ver el mundo de la computación. Sin embargo, los primeros teóricos que concibieron los fundamentos de la computación neuronal fueron Warren McCulloch, un neurofisiólogo, y Walter Pitts, un matemático, quienes, en 1943, lanzaron una teoría acerca de la forma de trabajar de las neuronas. Ellos modelaron una red neuronal simple mediante circuitos eléctricos.

- **Donald Hebb (1949)**

Fue el primero en explicar los procesos del aprendizaje (que es el elemento básico de la inteligencia humana) desde un punto de vista psicológico, desarrollando una regla de como el aprendizaje ocurría. Aún hoy, este es el fundamento de la mayoría de las funciones de aprendizaje que pueden hallarse en una red neuronal. Su idea fue que el aprendizaje ocurría

¹⁴ El proceso de entrenamiento consiste en la aplicación secuencial de diferentes conjuntos o vectores de entrada para que se ajusten los pesos de las interconexiones según un procedimiento predeterminado.

cuando ciertos cambios en una neurona eran activados. También intentó encontrar semejanzas entre el aprendizaje y la actividad nerviosa. Los trabajos de Hebb formaron las bases de la Teoría de las Redes Neuronales.

- **Karl Lashley (1950)**

En sus series de ensayos, encontró que la información no era almacenada en forma centralizada en el cerebro, sino que era distribuida encima de él.

- **Congreso de Dartmouth (1956)**

Este Congreso frecuentemente se menciona para indicar el nacimiento de la inteligencia artificial.

- **Frank Rosenblatt (1957)**

Comenzó el desarrollo del perceptrón. Esta es la red neuronal más antigua; utilizándose hoy en día para aplicación como identificador de patrones. Este modelo era capaz de generalizar, es decir, después de haber aprendido una serie de patrones podía reconocer otros similares, aunque no se le hubiesen presentado en el entrenamiento. Sin embargo, tenía una serie de limitaciones, por ejemplo, su incapacidad para resolver el problema de la función OR-exclusiva y, en general, era incapaz de clasificar clases no separables linealmente.

- **Frank Rosenblatt (1959)**

Principios de Neurodinámica: En este libro confirmó que, bajo ciertas condiciones, el aprendizaje del perceptrón convergía hacia un estado finito (Teorema de Convergencia del perceptrón).

- **Bernard Widroff/Marcian Hoff (1960)**

Desarrollaron el modelo Adaline (Adaptative Linear Elements). Esta fue la primera red neuronal aplicada a un problema real (filtros adaptativos para eliminar ecos en las líneas telefónicas) que se ha utilizado comercialmente durante varias décadas.

- **Karl Steinbeck: Die Lernmatrix (1961)**
Red neuronal para simples realizaciones técnicas (memoria asociativa).

- **Marvin Minsky/Seymour Papert (1969)**
En este año casi se produjo la “muerte abrupta” de las Redes Neuronales; puesto que Minsky y Papert probaron (matemáticamente) que el perceptrón no era capaz de resolver problemas relativamente fáciles, tales como el aprendizaje de una función no-lineal. Esto demostró que el perceptrón era muy débil, dado que las funciones no-lineales son extensamente empleadas en computación y en los problemas del mundo real.

- **Paul Werbos (1974)**
Desarrolló la idea básica del algoritmo de aprendizaje de propagación hacia atrás (backpropagation); cuyo significado quedó definitivamente aclarado en 1985.

- **Stephen Grossberg (1977)**
Teoría de Resonancia Adaptada (TRA). La Teoría de Resonancia Adaptada es una arquitectura de red que se diferencia de todas las demás previamente inventadas. La misma simula otras habilidades del cerebro: memoria a largo y corto plazo.

- **John Hopfield (1985)**
Provocó el renacimiento de las redes neuronales con su libro: “Computación neuronal de decisiones en problemas de optimización”.

- **David Rumelhart/G. Hinton (1986)**
Redescubrieron el algoritmo de aprendizaje de propagación hacia atrás (backpropagation).

A partir de 1986, el panorama fue alentador con respecto a las investigaciones y el desarrollo de las redes neuronales (Matich, 2001). Actualmente el uso se refleja no solo en investigaciones o trabajos, sino también se ha extendido en software domésticos (se puede encontrar modelos de redes neuronales en juegos, programas de reconocimiento de voz, juegos de ordenadores, programas o tutoriales, por mencionar algunos ejemplos) y así deja de estar restringida exclusivamente al mundo de la investigación o el ámbito empresarial.

2.3.2. Las Redes Neuronales Artificiales (RNA)

Las Redes Neuronales Artificiales se definen como sistemas de mapeos no lineales cuya estructura se basa en principios observados en los sistemas nerviosos de humanos y animales, que constan de procesadores simples ligados por conexiones con pesos. Las unidades de procesamiento de información se denominan neuronas y cada unidad recibe entradas de otros nodos y genera una salida simple escalar que depende de la información local disponible, guardada internamente o que llega a través de las conexiones con pesos. Esto permite realizar muchas funciones complejas dependiendo de las conexiones.

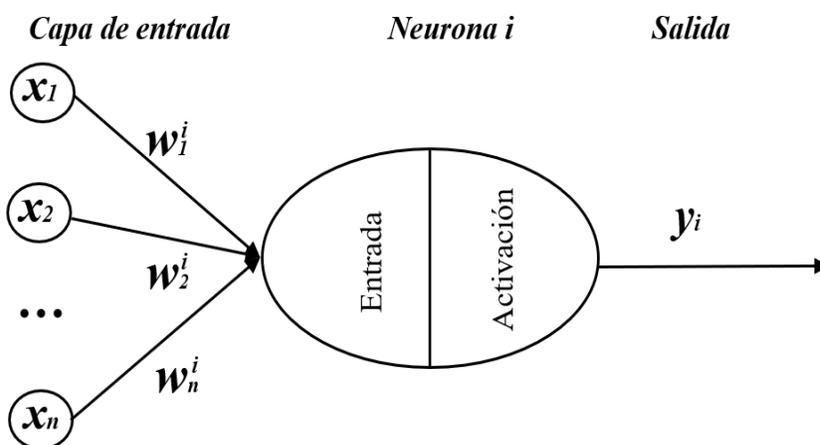


Figura 2: Nodo de una red neuronal

Las neuronas artificiales simples fueron introducidas por McCulloch y Pitts en 1943. Una red neuronal se caracteriza por los siguientes elementos:

- Un conjunto de unidades de procesamiento o neuronas.
- Un estado de activación para cada unidad, equivalente a la salida de la unidad.
- Conexiones entre las unidades, generalmente definidas por un peso que determina el efecto de una señal de entrada en la unidad.
- Una regla de propagación, que determina la entrada efectiva de una unidad a partir de las entradas externas.
- Una función de activación que actualiza el nuevo nivel de activación basándose en la entrada efectiva y la activación anterior.

- Una entrada que corresponde a un término determinado como vías para cada unidad.
- Un método para reunir la información, correspondiente a la regla del aprendizaje.
- Un ambiente en el que el sistema operar, con señales de entrada o señales de error.

Una neurona es un procesador elemental tal que a partir de un vector de entrada procedente del exterior o de otras neuronas, proporciona una única respuesta o salida y en muchas redes las unidades de proceso tienen respuesta de la forma:

$$y = f\left(\sum_{ij} w_{ij} x_{ij}\right)$$

Una red puede tener una estructura arbitraria, pero las capas que contienen estas estructuras están definidas de acuerdo con su ubicación en la topología de la red neuronal. Las entradas son aplicadas en la primera capa y las salidas se consideran la última capa, dentro de éstas dos capas se encuentran las capas internas las cuales se denominan capas ocultas. Las variables de entrada x no se consideran como capa porque no realizan procesamiento¹⁵. La función más habitual es de tipo lineal, que se basa en una suma ponderada de las entradas con los pesos sinápticos de las entradas conectadas más un bias θ ¹⁶:

$$\sum_{ij} w_{ij} X_{ij} + \theta$$

donde los pesos sinápticos (w) en una red neuronal representa la intensidad de interacción entre cada neurona presináptica j y la neurona postsináptica i , apoyándose de las reglas de propagación las cuales proporciona el valor del potencial postsináptico, $h(t)$, de la neurona en función de sus pesos (w) y entradas (x). Y si se considera a las entradas como funciones del tiempo, la expresión anterior se convierte en:

$$\sum_{ij} w_{ij}(t) X_{ij}(t) + \theta(t)$$

¹⁵ Las entradas pueden ser binarias (digitales) o continuas (analógicas) dependiendo del modelo de aplicación.

¹⁶ Se refiere a una neurona siempre activa, y se sitúa en la capa anterior a la neurona.

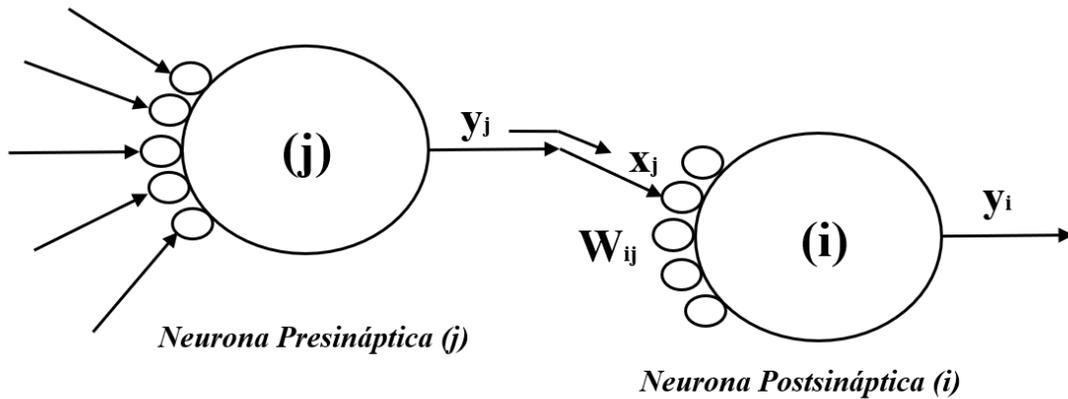


Figura 3: Interconexión entre una neurona presináptica y una neurona postsináptica

Cada neurona combina los valores de entrada, aplicando sobre ellos una función de entrada o combinación, la función f proporciona el estado de activación actual de la neurona en función de su estado anterior y del potencial postsináptico actual, dicha función puede ser sigmoideal¹⁷, tangente hiperbólica¹⁸, escalón¹⁹, entre otras, las cuales facilitan las aproximaciones que se requieran hacer empleando redes neuronales artificiales. Cada unidad de proceso tiene una tarea simple: recibe la entrada de otras unidades o de fuentes externas y procesa la información para obtener una salida que se propaga a otras unidades.

El valor resultante es procesado por la función de activación quien modula el valor de las entradas para generar el valor de salida Y_i . Este valor, generalmente se propaga a las conexiones de la neurona i con otras neuronas o bien es empleado como valor de salida de la red (Berzal, 2018). Frecuentemente la función de salida es simplemente la identidad, de modo que el estado de activación de la neurona se considera como la propia salida. Considerando que la regla de propagación es la suma ponderada y que la función de salida es la identidad, la neurona estándar consiste en:

¹⁷ Es una función continua no lineal la cual posee las unidades de proceso de entrada y salida están comprendidas entre 0 y 1.

¹⁸ La función hiperbólica ofrece flexibilidad y un amplio rango de resultados que ofrece, pues comprende desde -1 hasta 1, tomando en cuenta los posibles resultados negativos.

¹⁹ La función escalón se utiliza cuando las salidas de la red son binarias.

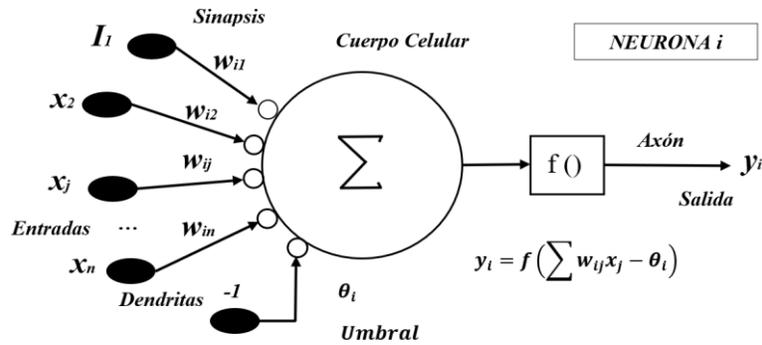


Figura 4: Modelo de neurona estándar

Cada conexión de entrada x_j tiene un peso determinado w_i^j que refleja su importancia o estado. El objetivo de la función de entrada es combinar las distintas entradas con su peso y agregar los valores obtenidos de todas las conexiones de entrada para obtener un único valor. Entonces, para un conjunto de n conexiones de entrada en el que cada una tiene un peso w_i^j , las funciones más utilizadas para combinar los valores de entrada son las siguientes:

- **La función suma ponderada:**

$$z(x) = \sum_{j=1}^n x_j w_j^i$$

- **La función máximo:** $z(x) = \max(x_1 w_1^i, \dots, x_n w_n^i)$
- **La función mínimo:** $z(x) = \min(x_1 w_1^i, \dots, x_n w_n^i)$
- **La función lógica AND (\wedge) o OR (\vee)** Aplicable solo en el caso de entradas binarias:

$$z(x) = (x_1 w_1^i \wedge \dots \wedge x_n w_n^i)$$

$$z(x) = (x_1 w_1^i \vee \dots \vee x_n w_n^i)$$

La utilización de una función u otra está relacionada con el problema y los datos concretos con los que se trabaja. Sin embargo, generalmente la suma ponderada²⁰ suele ser la más utilizada. (Roig, Alfonso, Quiles, & Roma, 2017)

²⁰ Es la función que representa la sumatoria de cada una de las entradas con los pesos que se ajustan de manera automática a medida que la red neuronal realiza el proceso de aprendizaje.

2.3.3. Funciones de activación

Una neurona biológica puede estar activa (excitada) o inactiva (no excitada); es decir, que tiene un estímulo que recibe dicha neurona supera un cierto umbral entonces la neurona emite un impulso, en caso contrario permanece en reposo. Las neuronas artificiales también tienen diferentes estados de activación, algunas de ellas solamente dos, pero otras pueden tomar cualquier valor dentro de un conjunto determinado. De esta idea se implementa en la práctica calculando la suma ponderada de las entradas que recibe la neurona, y filtrándola posteriormente con una función de activación para obtener la salida:

$$Y_i = f\left(\sum_{j=1}^n w_{ij} x_j - \theta_i\right) = f\left(\sum_{j=1}^n w_{ij} x_j\right)$$

donde:

$f(x)$ es la función activación y

θ_i es el umbral de activación²¹ de la neurona.

La función activación calcula el estado de actividad de una neurona; transformando la entrada global (menos el umbral, θ_i) en un valor (estado) de activación, cuyo rango normalmente va de (0 a 1) o de (-1 a 1). Esto es así, porque una neurona puede estar totalmente inactiva (0 o -1) o activa (1). (Cruz, 2010)

El umbral de activación se puede incluir en el sumatorio considerando una nueva neurona auxiliar $x_0 = -1$ conectada a y_i con un peso $w_{i0} = \theta_i$. Por tanto, la salida de una neurona y_i se obtiene simplemente transformando la suma ponderada de las entradas que recibe usando la función de activación. Las funciones de activación más comúnmente utilizadas se detallan a continuación:

²¹ Las neuronas tienen un umbral de activación que es la cantidad de “inputs” (entradas) necesario para activar la neurona, de tal modo que, si la suma es igual o mayor que el umbral de la neurona, ésta se activa y permite una salida, caso contrario no se activará la neurona.

Función Lineal

Los valores de salida obtenidos por medio de esta función de activación serán: $f(x) = x$, cuando el argumento de $f(x)$ esté comprendido dentro del rango $\left(\frac{-1}{a}, \frac{1}{a}\right)$. Por encima o por debajo de esta zona se fija la salida en 1 o -1 respectivamente. Cuando $a = 1$ (siendo que la misma afecta la pendiente de la gráfica), la salida es igual a la entrada. Teniendo a $f(x) = x$ y $a > 0$. La expresión analítica de esta función es la siguiente:

$$f(x) = \begin{cases} -1 & x \leq \frac{-1}{a} \\ ax & \frac{-1}{a} < x < \frac{1}{a} \\ 1 & x \geq \frac{1}{a} \end{cases}$$

La función lineal permite generar combinaciones lineales de las entradas. En su forma más básica se puede ver a continuación:

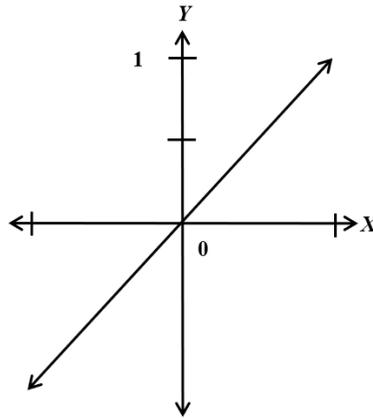


Figura 5: Función de Activación Lineal

Función Escalón

La función de activación escalón se asocia a neuronas binarias en las cuales, cuando la suma de las entradas es mayor o igual que el umbral de la neurona la activación es 1; si es menor la activación es 0 (o -1). La salida binaria se puede establecer en los valores $\{-1,1\}$, pero también se utilizan a menudo los valores $\{0,1\}$.

Analíticamente su comportamiento es el siguiente, seguido de su representación gráfica.

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad f(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

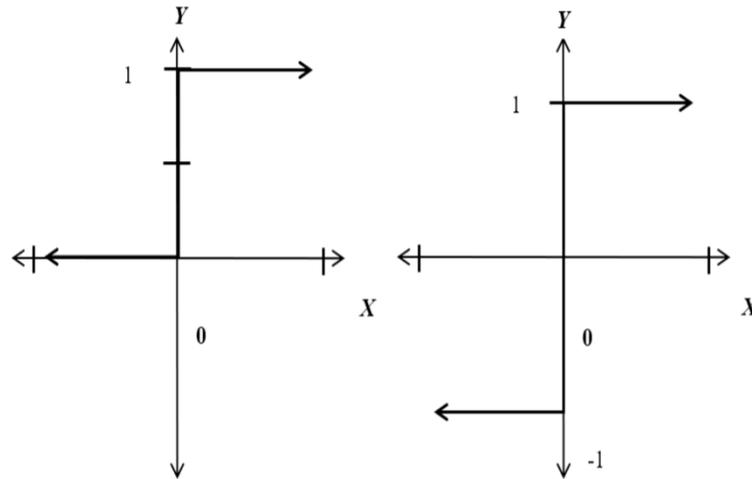


Figura 6: Función de activación Escalón

Función Tangente hiperbólica

La función de activación tangente hiperbólica se emplea en los casos que presentan variaciones suaves de valores positivos y negativos de la señal a clasificar. Es una de las funciones más empleadas en entrenamientos supervisados, como en el caso del entrenamiento de retropropagación del error.

Los valores de salida de la función tangente hiperbólica están comprendidos dentro de un rango que va de -1 a 1. Al modificar el valor de x se ve afectada la pendiente de la función de activación. A partir de ello, la expresión analítica de esta función es la siguiente:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Gráficamente la función está representada así:

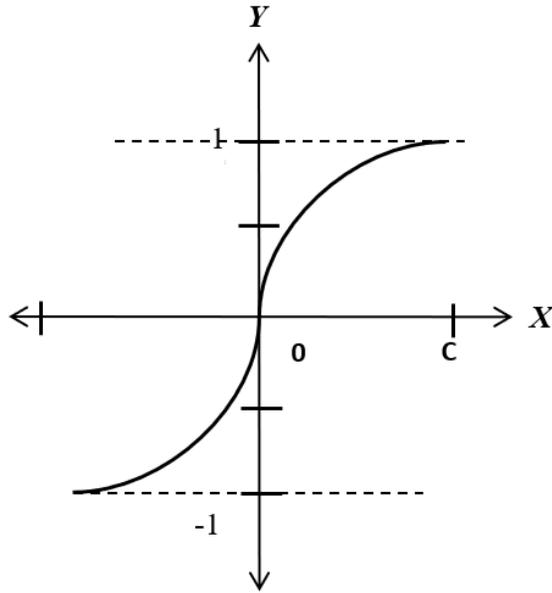


Figura 7: Función de activación Tangente hiperbólica

Función Sigmoidea

Con la función sigmoideal el valor dado por la función es cercano a uno de los valores asíntóticos (ya sea el valor de 0 o 1). Esto hace que, en la mayoría de los casos, el valor de salida está comprendido en la zona alta o baja del sigmoide. De hecho, cuando la pendiente es elevada esta función tiende a la función escalón.

Sin embargo, la importancia de la función sigmoideal es que su derivada siempre es positiva y cercana a cero para los valores grandes positivos o negativos; además, toma su valor máximo cuando $x = 0$ en el intervalo de 0 a 1. Esto hace que se puedan utilizar reglas de aprendizaje definidas para las funciones escalón, con la ventaja respecto a esta función es que la derivada está definida en todo el intervalo.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Gráficamente la función sigmoidea tiene el comportamiento siguiente:

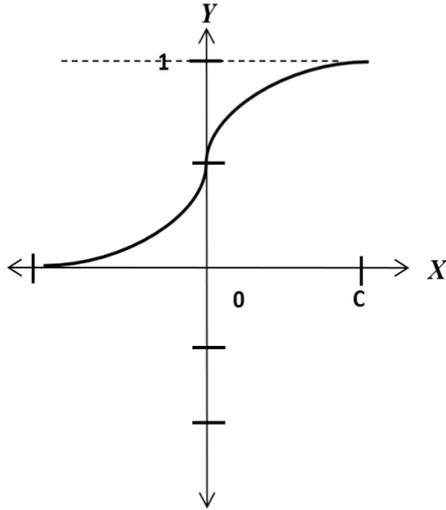


Figura 8: Función de activación Sigmoidea

Función de Gauss

Los mapeos ocultos algunas veces pueden realizarse con un solo nivel de neuronas mediante el uso de funciones de activación tipo Gauss, en lugar de funciones tipo sigmoidea. Donde función está centrada en cero, es decir con $\mu = 0$ y con desviación estándar σ .

Con una representación gráfica y analítica de la siguiente forma:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

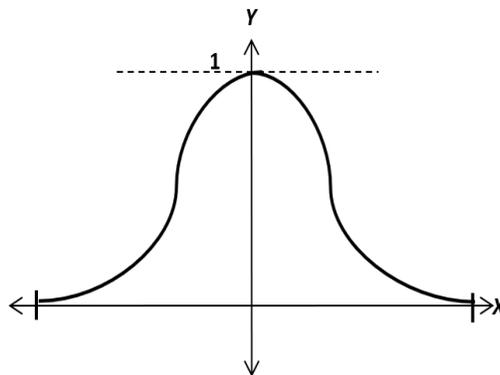


Figura 9: Función de activación Gaussiana

2.3.4. Arquitectura de una red neuronal

La arquitectura o topología de una red neuronal se define como la organización de las neuronas en distintas capas, así como los parámetros que afectan a la configuración de las neuronas, tales como las funciones de entrada o activación. Cada arquitectura o topología puede ser válida para algunos tipos de problemas, presentando diferentes niveles de calidad de los resultados y también diferentes niveles de coste computacional. En analogía a la estructura del cerebro de un ser vivo, los modelos computacionales de redes neuronales están compuestos por un gran número de unidades simples de proceso (neuronas) conectadas entre sí con base a una topología definida.

La arquitectura empleada para modelar una red neuronal define la configuración para cada uno de los elementos básicos, en el paradigma de las redes neuronales artificiales está caracterizado por como los algoritmos de aprendizajes determinan el cambio de los pesos provenientes de las funciones de activación. Debido a esto los elementos de una red neuronal dan paso a modelos conexionistas que se pueden clasificar de diferentes formas según el criterio usado: Según el número de capas, el tipo de conexión y el grado de conexión.

2.3.4.1. Según Número de Capas

Redes neuronales monocapa

La red neuronal más simple está formada por una única neurona, conectada a todas las entradas disponibles y con una única salida, es de las arquitecturas más simple de redes neuronales pues se tiene una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan diferentes cálculos.

Estas redes presentan la capa de entrada, por una capa oculta de procesamiento: formada por un conjunto variable de neuronas y finalmente la capa de salida con una o más neuronas. Es de mencionar que una red monocapa con múltiples salidas es del tipo de redes capaces de clasificar patrones de entrada más complejos o realizar predicciones sobre dominios de dimensionalidad más alta.

El número de neuronas de la capa oculta está relacionado con la capacidad de procesamiento y clasificación, pero un número demasiado grande de neuronas en esta capa aumenta el riesgo de sobre especialización en el proceso de entrenamiento. Su representación es la siguiente:

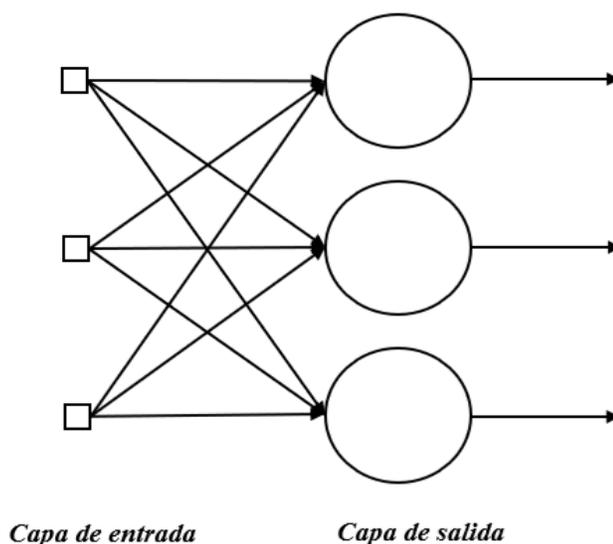


Figura 10: Esquema de una red neuronal monocapa

Redes neuronales multicapas

Es una generalización de las redes monocapa existiendo un conjunto de capas intermedias entre las de entrada y salida (capas ocultas), en donde se puede añadir un número indeterminado de este tipo de capas. Cuando se añaden neuronas y capas a una red se aumenta generalmente su poder de predicción, la calidad de dicha predicción y la capacidad de separación. Aunque también se aumenta su tendencia a la sobre especialización, y se aumenta el coste computacional y temporal de entrenamiento. Este tipo de red puede estar total o parcialmente conectada.

El número de neuronas de la capa de salida depende, en gran medida, del problema concreto que se está tratando y de la codificación empleada para resolver dicha situación. Y cuya representación de la arquitectura de una red neuronal multicapa se muestra a continuación:

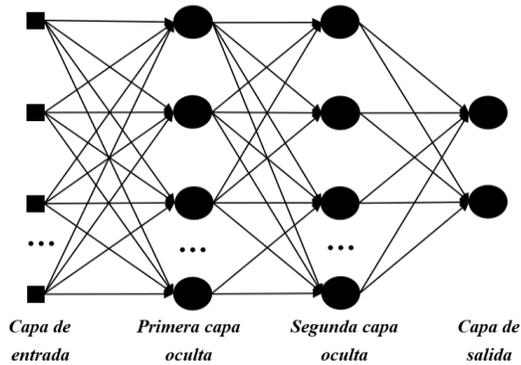


Figura 11: Esquema de una red neuronal multicapa

2.3.4.2. Según el tipo de conexiones

Redes neuronales no recurrentes

En esta red la propagación de las señales se produce en un sentido solamente, no existiendo la posibilidad de retroalimentarse, es por ello que estas estructuras no tienen memoria.

Redes neuronales recurrentes

Esta red viene caracterizada por la existencia de lazos de retroalimentación que pueden ser conectadas entre neuronas de diferentes capas, neuronas de la misma capa o más sencillamente entre una misma neurona. Esta estructura recurrente la hace especialmente adecuada para estudiar la dinámica de sistemas no lineales. La siguiente figura representa el esquema de una red recurrente:

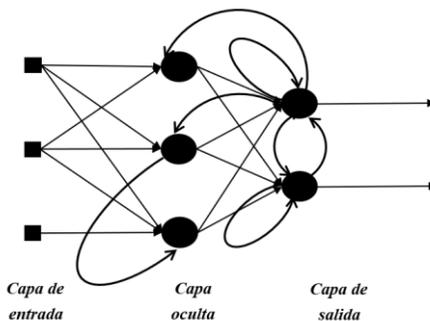


Figura 12: Esquema de una red neuronal recurrente

2.3.4.3. Según el grado de conexión

Redes neuronales parcialmente conectadas

En este caso no se da la conexión total entre neuronas de diferentes capas. (Serrano, Soria, & Martín, 2010)

Redes neuronales totalmente conectadas

Para este caso todas las redes neuronales una capa se encuentran conectadas con las de la capa siguiente (redes no recurrentes) o con las de la anterior (redes recurrentes).

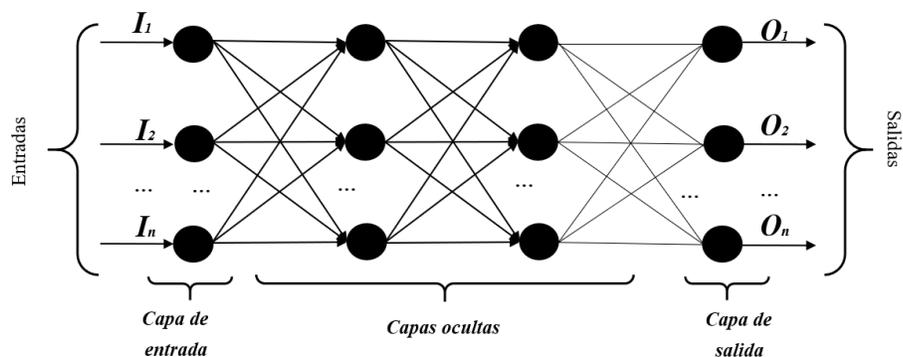


Figura 13: Esquema de una red neuronal totalmente conectada

Aprendizaje de una red neuronal

Una de las principales características de las redes neuronales artificiales es su capacidad de aprendizaje. El entrenamiento de las redes neuronales artificiales muestra algunos paralelismos con el desarrollo intelectual de los seres humanos, no obstante, aun cuando parece que se ha conseguido entender el proceso de aprendizaje conviene ser moderado porque el aprendizaje de las redes neuronales artificiales está limitado.

El objetivo del entrenamiento de una red neuronal artificial es conseguir una aplicación determinada, para que un conjunto de entradas produzca el conjunto de salidas deseadas o mínimamente consistentes. El proceso de entrenamiento consiste en la aplicación secuencial de diferentes conjuntos o vectores de entrada para que se ajusten los pesos de las interconexiones según un procedimiento predeterminado. Durante la sesión de entrenamiento los pesos

convergen gradualmente hacia los valores que hacen que cada entrada produzca el vector de salida deseado, esto gracias a los algoritmos de entrenamiento o los procedimientos de ajuste de los valores de las conexiones de las redes neuronales artificiales.

Otro criterio que se puede utilizar para diferenciar las reglas de aprendizaje se basa en considerar si la red puede aprender durante su funcionamiento habitual o si el aprendizaje supone la desconexión de la red; es decir, su inhabilitación hasta que el proceso termine. En el primer caso, se trataría de un aprendizaje online, mientras que el segundo es lo que se conoce como offline.

Cuando el aprendizaje es offline se distingue entre una fase de aprendizaje o entrenamiento y una fase de operación o funcionamiento, existiendo un conjunto de datos de entrenamiento (data train) y un conjunto de datos de test o prueba, que serán utilizados en la correspondiente fase; además, los pesos de las conexiones permanecen fijos después que termina la etapa de entrenamiento de la red. Debido precisamente a su carácter estático, estos sistemas no presentan problemas de estabilidad en su funcionamiento.

Una generalización de la fórmula o regla para decir los cambios en los pesos es la siguiente: *Peso Nuevo = Peso Viejo + Cambio de Peso*. Matemáticamente esto es:

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij}(t)$$

donde t hace referencia a la etapa de aprendizaje, $w_{ij}(t + 1)$ al peso nuevo y $w_{ij}(t)$ al peso viejo. Dichos aprendizajes se pueden clasificar en dos grupos: Supervisado y No Supervisado.

2.3.5.1. Entrenamiento Supervisado

El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. Estos algoritmos requieren el emparejamiento de cada vector de entrada con su correspondiente vector de salida.

El entrenamiento consiste en presentar un vector de entrada a la red, calcular la salida de la red, compararla con la salida deseada, y el error o diferencia resultante se utiliza para realimentar la red y cambiar los pesos de acuerdo con un algoritmo que tiende a minimizar el error. Las parejas de vectores del conjunto de entrenamiento se aplican secuencialmente y de forma cíclica. Se calcula el error y el ajuste de los pesos por cada pareja hasta que el error para el conjunto de entrenamiento entero sea un valor pequeño y aceptable.

En este tipo de aprendizaje se suelen considerar tres formas de llevarlo a cabo, que dan lugar a los siguientes aprendizajes supervisados: El aprendizaje por corrección de error, por refuerzo y estocástico.

Aprendizaje por corrección de error

Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos a la salida de la red, es decir, en función del error cometido en la salida.

Un ejemplo de este tipo de algoritmos lo constituye la regla de aprendizaje del Perceptrón, utilizada en el entrenamiento de la red del mismo nombre que desarrolló Rosenblatt en 1958. Esta es una regla muy simple, para cada neurona en la capa de salida se le calcula la desviación a la salida objetivo como el error δ . El cual luego se utiliza para cambiar los pesos sobre la conexión de la neurona precedente. El cambio de los pesos por medio de la regla de aprendizaje del Perceptrón se realiza según la siguiente regla:

$$\Delta w_{ij} = \sigma(out_j)(a_{qi} - out_i)$$

donde: a_{qi} es la salida deseada/objetivo de la neurona de salida N_i , $\Delta_i = (a_{qi} - out_i)$ la desviación objetivo de la neurona N_i y σ el aprendizaje.

La salida de la neurona $N_j(out_j)$ se utiliza, porque este valor influye en la entrada global y, por ende, en la activación y luego en la salida de la neurona N_i . Esto es semejante a un “efecto en cadena”.

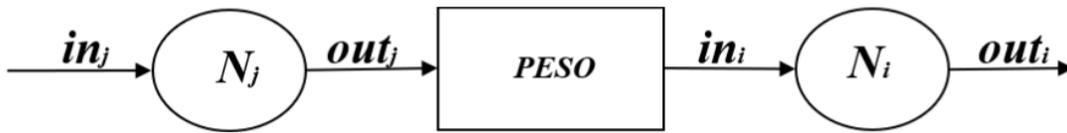


Figura 14: Influencia de la neurona N_j en la neurona N_i

Otro algoritmo muy conocido y que pertenece a esta clasificación es la regla de aprendizaje Delta o regla del mínimo error cuadrado (LMS Error: Least Mean Squared Error), que también utiliza la desviación a la salida objetivo, pero toma en consideración a todas las neuronas predecesoras que tiene la neurona de salida. Esto permite cuantificar el error global cometido en cualquier momento durante el proceso de entrenamiento de la red, lo cual es importante pues cuanto más información se tenga sobre el error cometido, más rápido se puede aprender, luego el error calculado (Δ) es igualmente repartido entre las conexiones de las neuronas predecesoras. La regla de aprendizaje de propagación hacia atrás o de backpropagation, también conocido como regla LMS multicapa es una generalización de la regla de aprendizaje Delta. Esta es la primera regla de aprendizaje que permitió realizar cambios sobre los pesos en las conexiones de la capa oculta.

Aprendizaje por refuerzo

Se trata de un aprendizaje supervisado más lento que el anterior, que se basa en la idea de no disponer de un ejemplo completo del comportamiento deseado; es decir, de no indicar durante el entrenamiento exactamente la salida que se desea que proporcione la red ante una determinada entrada.

En el aprendizaje por refuerzo la función del supervisor se reduce a indicar mediante una señal de refuerzo si la salida obtenida en la red se ajusta a la deseada (éxito = +1 o fracaso = -1), y en función de ello se ajustan los pesos basándose en un mecanismo de probabilidades. Se podría decir que en este tipo de aprendizaje la función del supervisor se asemeja más a la de un crítico (que opina sobre la respuesta de la red) que a la de un maestro (que indica a la red la respuesta concreta que debe generar), como ocurría en el caso de supervisión por corrección del error.

Aprendizaje estocástico

Consiste en realizar cambios aleatorios en los valores de los pesos de las conexiones de la red y evaluar su efecto a partir del objetivo deseado y de distribuciones de probabilidad. Además, en el aprendizaje estocástico se suele hacer una analogía en términos termodinámicos²², asociando a la red neuronal con un sólido físico que tiene cierto estado energético. En el caso de la red, la energía de la misma representaría el grado de estabilidad de la red, de tal forma que el estado de mínima energía correspondería a una situación en la que los pesos de las conexiones consiguen que su funcionamiento sea el que más se ajusta al objetivo deseado.

El aprendizaje consistiría en realizar un cambio aleatorio de los valores de los pesos y determinar la energía de la red (habitualmente la función energía es una función de Lyapunov²³). Si la energía es menor después del cambio, es decir si el comportamiento de la red se acerca al deseado se acepta el cambio; si, por el contrario, la energía no es menor se aceptaría el cambio en función de una determinada y preestablecida distribución de probabilidades.

2.3.5.2. Entrenamiento No Supervisado

Las redes que utilizan el aprendizaje no supervisado (también conocido como autosupervisado) no requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta.

Estas redes deben encontrar las características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presenten en su entrada. Existen varias posibilidades en cuanto a la interpretación de la salida de estas redes, que dependen de su estructura y del algoritmo de aprendizaje empleado.

²² Se refiere a las transferencias de energía entre determinados sistemas y en la predicción y sistematización de cómo ocurrirá la transferencia de energía.

²³ Es una función pseudo-energética definida positivamente y en las que el valor que toma a lo largo de la trayectoria de un sistema termodinámico disminuye a lo largo del tiempo, hasta alcanzar un mínimo en los equilibrios estables de un sistema.

En algunos casos, la salida representa el grado de familiaridad o similitud entre la información que se le está presentando en la entrada y las informaciones que se le han mostrado hasta entonces (en el pasado). En otro caso, podría realizar una clusterización (clustering) o establecimiento de categorías, indicando la red a la salida a qué categoría pertenece la información presentada a la entrada, siendo la propia red quien debe encontrar las categorías apropiadas a partir de las correlaciones entre las informaciones presentadas.

En cuanto a los algoritmos de aprendizaje no supervisado, en general se suelen considerar dos tipos, que dan lugar a los siguientes aprendizajes: El aprendizaje hebbiano junto al aprendizaje competitivo y comparativo.

Aprendizaje hebbiano

Esta regla de aprendizaje pretende medir la familiaridad o extraer características de los datos de entrada. El fundamento es una suposición bastante simple: si dos neuronas N_i y N_j toman el mismo estado simultáneamente (ambas activas o ambas inactivas), el peso de la conexión entre ambas se incrementa.

Las entradas y salidas permitidas a la neurona son: $-1,1$ o $0,1$ (neuronas binarias). Esto puede explicarse porque la regla de aprendizaje de Hebb se originó a partir de la neurona biológica clásica, que solamente puede tener dos estados: activa o inactiva.

Aprendizaje competitivo y comparativo

Se orienta a la clusterización o clasificación de los datos de entrada. Como característica principal del aprendizaje competitivo se puede decir que, si un patrón nuevo se determina que pertenece a una clase reconocida previamente, entonces la inclusión de este nuevo patrón a esta clase matizará la representación de la misma. Si el patrón de entrada determina que no pertenece a ninguna de las clases reconocidas anteriormente, entonces la estructura y los pesos de la red neuronal serán ajustados para reconocer la nueva clase.

2.3.6. Tipos de Redes Neuronales Artificiales

Los tipos de redes neuronales que destacan son el perceptrón simple y el perceptrón multicapa, tal y como se detalla a continuación.

2.3.6.1 Perceptrón Simple

El perceptrón simple fue introducido por Rosenblatt (1962) y es un modelo unidireccional compuesto por dos capas de neuronas, una de entrada y otra de salida. La operación en un perceptrón simple que consta de n neuronas de entrada y m neuronas de salida. La arquitectura del Perceptrón, llamada mapeo de patrones (pattern-mapping), aprende a clasificar modelos mediante un aprendizaje supervisado. Los modelos que clasifica suelen ser generalmente vectores con valores binarios (0,1) y las categorías de la clasificación se expresan mediante vectores binarios. Se puede expresar como:

$$y = f \left(\sum_{j=1}^n w_{ij} x_j - \theta_i \right) \quad \text{con } i = 1, \dots, m$$

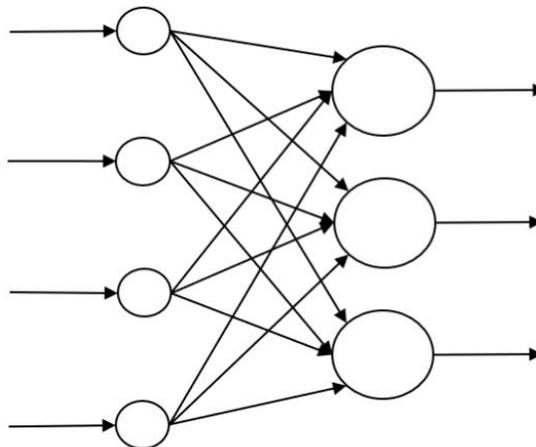


Figura 15: Estructura de un Perceptrón

El Perceptrón presenta dos capas de unidades procesadoras (PE) y sólo una de ellas presenta la capacidad de adaptar o modificar los pesos de las conexiones. La arquitectura del perceptrón admite capas adicionales, pero éstas no disponen la capacidad de modificar sus propias conexiones.

El perceptrón simple puede utilizarse como clasificador, radicando su importancia histórica en su carácter de dispositivo entrenable, esto gracias al algoritmo de aprendizaje del modelo introducido por Rosenblatt (1962) permite determinar automáticamente los pesos sinápticos que clasifican un conjunto de patrones a partir de un conjunto de ejemplos etiquetados.

El algoritmo de aprendizaje del perceptrón simple pertenece al grupo de los algoritmos que se fundamentan en la corrección de errores. Los algoritmos de este tipo ajustan los pesos de manera proporcional a la diferencia existente entre la salida actual de la red neuronal y la salida deseada, con el objetivo de minimizar el error actual de la red. Dicho proceso de aprendizaje es iterativo.

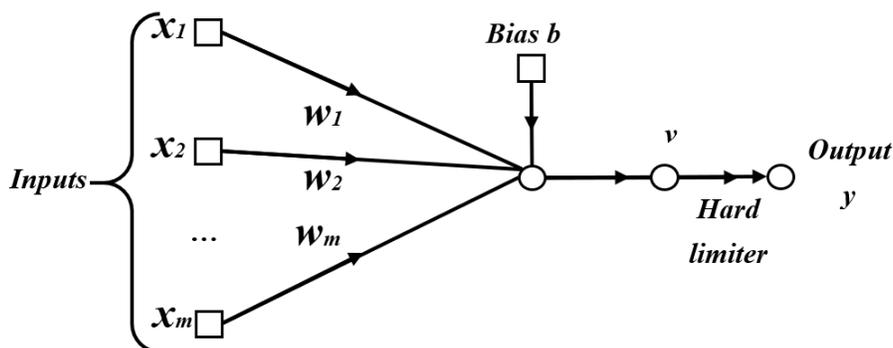


Figura 16: El Perceptrón de Rosenblatt

2.3.6.2 El Adaline: La regla de aprendizaje Delta

Otro modelo clásico de redes neuronales es el Adaline o neurona con adaptación lineal. Es similar al perceptrón simple, pero utiliza como función de activación la función identidad, $f(x) = x$, en lugar de la función signo. La salida del Adaline es simplemente una función lineal de las entradas ponderadas con los pesos sinápticos y ahora la salida de la red es continua en lugar de binaria 0,1 y ésta sigue siendo:

$$y = f \left(\sum_{j=1}^{n+1} w_j x_j \right)$$

La regla de aprendizaje supervisado que sigue es la regla Delta, con la cual se trata de determinar los pesos sinápticos de manera que se minimice la función de error cuadrático siguiente:

$$E = \frac{1}{2} \sum_{k=1}^p (z(k) - y(k))^2 = \frac{1}{2} \sum_{k=1}^p \left(z(k) - f \left(\sum_{j=1}^{n+1} w_j x_j(k) \right) \right)^2$$

Este proceso se realiza de forma iterativa, donde se va presentando los patrones una a uno y se van modificando los parámetros de la red mediante la regla del descenso de gradiente, es decir, en la misma dirección, pero en sentido opuesto al gradiente. Y consiste en realizar un cambio en cada peso proporcional a la derivada del error respecto del peso:

$$\begin{cases} w_j(k+1) = w_j(k) + \Delta w_j(k), & j = 1, \dots, n+1 \\ \text{siendo} & \Delta w_j(k) = -\eta \frac{\partial E}{\partial w_j(k)} \end{cases}$$

Es decir:

$$w_j(k+1) = w_j(k) - \eta \frac{\partial E}{\partial w_j(k)}$$

Por consiguiente, la regla de aprendizaje general es la siguiente:

$$w_j(k+1) = w_j(k) + \eta [z(k) - y(k)] f'(h) x_j(k)^{24}, \quad j = 1, \dots, n+1$$

Como en este caso f es la función identidad, tenemos la siguiente regla de aprendizaje para el Adaline:

$$w_j(k+1) = w_j(k) + \eta [z(k) - y(k)] x_j(k), \quad j = 1, \dots, n+1$$

El parámetro η controla la longitud de pasos que se deben de dar en la dirección opuesta del gradiente. Conforme mayor sea η , mayor será la cantidad por la que se modificarán los pesos

²⁴ Donde:

η es el coeficiente de aprendizaje.

$f(h)$ es la función de activación.

$f'(h)$ la primera derivada de la función de activación.

sinápticos. Dicho parámetro debe ser un valor pequeño para evitar dar pasos demasiado largos; es decir, que no lleven a soluciones peores de las que se tienen pues el método del gradiente solamente garantiza el decrecimiento de la función de error si se desplaza en la dirección opuesta del gradiente, pero en un entorno suficientemente pequeño.

En el proceso de entrenamiento se introduce un patrón en cada iteración, pues se dice que se ha realizado un aprendizaje en línea. También se puede introducir los p patrones directamente y comparar las salidas de la red con las salidas deseadas, pasando entonces a actualizar los pesos sinápticos, en cuyo caso se dice que el aprendizaje es por lotes.

La modificación de los pesos se hace tomando como función de error el error cuadrático medio, dividiendo por p el error total para interpretarla como error cuadrático medio por patrón, es decir:

$$E = \frac{1}{2p} \sum_{k=1}^p (z(k) - y(k))^2 = \frac{1}{2p} \sum_{k=1}^p \left(z(k) - f \left(\sum_{j=1}^{n+1} w_{j(k)} x_j \right) \right)^2$$

Y así la regla de aprendizaje es:

$$w_j(k + 1) = w_j(k) + \eta \frac{1}{p} [z(k) - y(k)] x_j(k)$$

La principal diferencia entre el Perceptrón simple y el Adaline es la forma en que se usa la salida de la red en la regla de aprendizaje. El primero solo tiene en cuenta si se ha equivocado o no, mientras que el Adaline considera cuánto se ha equivocado. (Munt, 2018)

2.3.6.3. Perceptrón Multicapa

Es una red dentro del aprendizaje supervisado y que en la mayoría de aplicaciones utiliza o se ha utilizado, la cual es formado por una capa de entrada, al menos de una capa oculta y una capa de salida. Los nodos de la capa de entrada alimentan la red hacia delante, no existen las conexiones laterales y en donde cada nodo en las capas ocultas y de salida recibe de forma progresiva una entrada de los nodos de las capas previas y calcula un valor de salida para la siguiente capa.

Las capas ocultas dan una mayor flexibilidad al modelo resultante, que puede considerarse un modelo no paramétrico de regresión no lineal. Sin embargo, el problema de este tipo de redes es su diseño, pues a priori no existe ninguna regla orientativa sobre el número de capas óptimo para un problema dado o el número de neuronas en cada capa. Posterior de la elección del diseño para la red neuronal es necesario utilizar un algoritmo de aprendizaje para ajustar los parámetros (pesos) a los datos. (Gutiérrez, Cofiño, & Sordo, 2004)

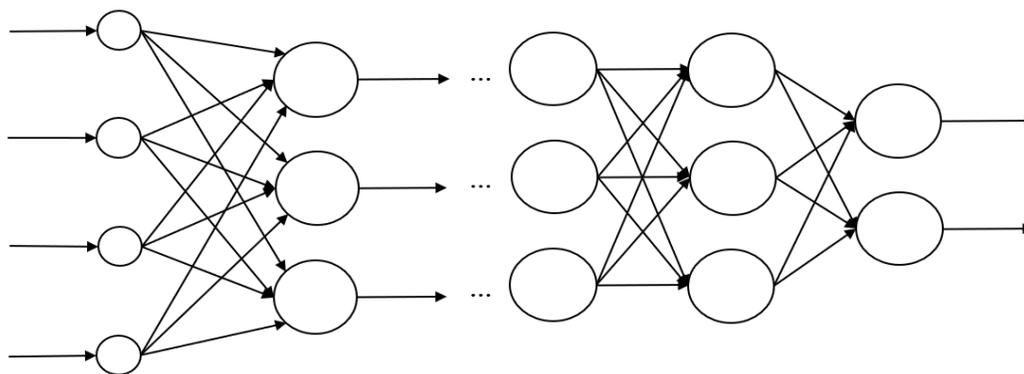


Figura 17: Estructura de una red multicapa o perceptrón multicapa

El perceptrón multicapa es capaz de actuar como un aproximador universal de funciones: una red backpropagation conteniendo al menos una capa oculta con suficientes unidades no lineales puede aproximar cualquier tipo de función o relación continua entre un grupo de variables de entrada y salida. Esta propiedad convierte a las redes perceptrón multicapa en herramientas de propósito general, flexibles y no lineales, contando con la capacidad de dotar a la red de generalización de tal modo que el sistema obtiene una salida correcta para un conjunto de datos de entrada que no se había usado antes.

El método de aprendizaje más popular para redes multicapa es conocido como retro-propagación (backpropagation) y está basado en minimizar la función de error cuadrático total usando el método del descenso de gradiente de forma apropiada, dado que los pesos de las dos capas dependen unos de otros.

Proceso de aprendizaje de un Perceptrón Multicapa

El proceso de entrenamiento de una red es:

- Se inician los pesos con los valores aleatorios próximos a 0.
- Se presenta un patrón n de entrenamiento $(X(n), S(n))$, se propaga hacia la salida y se obtiene la respuesta de la red $Y(n)$.
- Se evalúa el error que conecte la red para cada neurona.
- Se aplica la regla delta generalizada para modificar los pesos de la red, empezando por la capa de salida hasta llegar a la capa de entrada.
- Se repite el proceso hasta alcanzar un mínimo de error.

Ventajas e inconvenientes

La principal ventaja de la Backpropagation es su capacidad genérica de mapeo de patrones, puesto que la red es capaz de aprender una gran variedad de estas relaciones y no requiere un conocimiento matemático de la función que relaciona los patrones de la entrada y los patrones de salida. La Backpropagation sólo necesita ejemplos de mapeo para aprender, además la flexibilidad de esta red se ve aumentada con la posibilidad de elegir el número de capas, interconexiones, unidades procesadoras, constante de aprendizaje y representación de datos, como resultado de estas características este tipo de red es capaz de participar con éxito en una amplia gama de aplicaciones.

El mayor inconveniente es el tiempo de convergencia y las aplicaciones reales pueden llegar a tener miles de ejemplos en el conjunto de entrenamiento y ello requiere días de tiempo de cálculo, además la Backpropagation es susceptible de fallar en el entrenamiento, es decir, la red puede que nunca llegue a converger.

Existe una variedad de técnicas desarrolladas para disminuir el tiempo de convergencia y evitar los mínimos locales. Entre ellas destacan cambiar la red, cambiar el conjunto de entrenamiento y añadir ruido aleatorio a los pesos. (Cano, 2012)

2.3.7 Entrenamiento de redes neuronales artificiales

En el contexto de las redes neuronales el aprendizaje puede ser visto como el proceso de ajuste de los parámetros libres de la red (Yao, 1995). Partiendo de un conjunto de pesos sinápticos aleatorio, el proceso de aprendizaje busca un conjunto de pesos que permitan a la red desarrollar correctamente una determinada tarea. El proceso de aprendizaje es un proceso iterativo, en el cual se va refinando la solución hasta alcanzar un nivel de operación suficientemente bueno.

La mayoría de los métodos de entrenamiento utilizados en las redes neuronales con conexión hacia delante consisten en proponer una función de error que mida el rendimiento actual de la red en función de los pesos sinápticos. El objetivo del método de entrenamiento es encontrar el conjunto de pesos sinápticos que minimizan (o maximizan) la función. El método de optimización proporciona una regla de actualización de los pesos que en función de los patrones de entrada modifica iterativamente los pesos hasta alcanzar el punto óptimo de la red neuronal.

2.3.7.1. Métodos de gradiente descendente

El método de entrenamiento más utilizado es el método del gradiente descendente. Este método define una función $E(W)$ que proporciona el error que comete la red en función del conjunto de pesos sinápticos W . El objetivo del aprendizaje será encontrar la configuración de pesos que corresponda al mínimo global de la función de error, aunque en muchos casos es suficiente encontrar un mínimo local lo suficientemente bueno (Chinrungrueng, 1993).

El principio general del método es el siguiente: dado un conjunto de pesos $W(0)$ para el instante de tiempo $t = 0$, se calcula la dirección de máxima variación del error. La dirección de máximo crecimiento de la función $E(W)$ en $W(0)$ viene dado por el gradiente $\nabla E(W)$. Luego, se actualizan los pesos siguiendo el sentido contrario al indicado por el gradiente $\nabla E(W)$, dirección que indica el sentido de máximo decrecimiento (Falco, Cioppa, Natale, & Tarantino, 1997). De este modo se va produciendo un descenso por la superficie de error hasta alcanzar un mínimo local.

$$W(t + 1) = W(t) - \alpha \nabla E(W)$$

donde α indica el tamaño del paso tomado en cada iteración, pudiendo ser diferente para cada peso e idealmente debería ser infinitesimal. El tamaño del paso es un factor importante a la hora de diseñar un método de estas características. Si se toma un paso muy chico el proceso de entrenamiento resulta muy lento, mientras que si el tamaño del paso es muy grande se producen oscilaciones en torno al punto mínimo.

2.3.7.2. Algoritmo de aprendizaje backpropagation

La red Perceptrón multicapa utiliza un algoritmo de aprendizaje supervisado por corrección del error con aprendizaje offline conocido como propagación hacia atrás de errores o retropropagación (Backpropagation) o también conocido como Regla delta generalizada. Dicho algoritmo consiste en retropropagar la señal desde la capa de salida hasta la capa de la entrada, optimizando los valores de los pesos a través de un proceso que se basa en la minimización de la función de coste²⁵. El algoritmo se divide en dos fases:

- **Propagación hacia adelante:** Las señales se propagan desde la capa de entrada hasta la capa de salida, generalizando la salida y el error cometido por la red; al comparar la salida obtenida con la esperada que se le facilita a la red durante el entrenamiento.
- **Propagación hacia atrás:** De acuerdo a los errores cometidos en la capa de salida Backpropagation se encarga de corregir el valor de los pesos entre las conexiones de las neuronas mediante la retropropagación del error desde la capa de salida hacia la capa de entrada a través de las capas ocultas.

²⁵ Se denomina función de coste al valor absoluto del error.

Una neurona de la capa oculta se conecta con otra de la capa de salida, donde x_1, \dots, x_n son las entradas a la capa oculta, x_0 es la entrada que dependiendo si es 1 será el bías²⁶, por lo contrario, si es -1 será el umbral²⁷. Los valores de w_{m1}, \dots, w_{mn} son los pesos que conectan las entradas con la neurona oculta m y w_{m0} es el peso sináptico correspondiente a x_0 :

$$v_m(t) = \sum_{i=0}^n (w_{mi})(x_i(t))$$

En la retropropagación los primeros pesos que se actualizan son los de la capa de salida y si existen más capas ocultas el proceso es el mismo, se atraviesan todas las capas hasta llegar a la entrada.

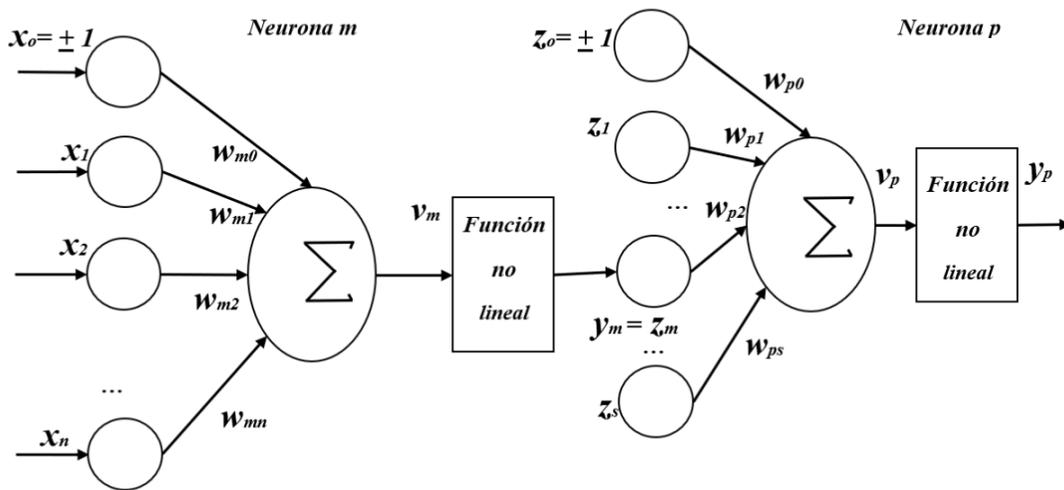


Figura 18: Neurona oculta conectada con la neurona de la capa de salida

²⁶ El algoritmo de backpropagation en cualquiera de sus capas ocultas se utiliza el bías que presentan un nivel de activación de valor 1. Su objetivo es mejorar la convergencia de la red y ofrecer un nuevo efecto umbral sobre la unidad que opera.

²⁷ El umbral representa la mínima entrada total ponderada necesaria para provocar la activación de la neurona.

Deducción del algoritmo Backpropagation

El algoritmo propone una actualización iterativa de los pesos de la siguiente manera:

$$W(t + 1) = W(t) + \nabla W(t) \quad (1)$$

si se toma una variación proporcional al gradiente de una función de error $E(W)$ ²⁸ se tiene que:

$$W(t + 1) = W(t) + \alpha \nabla W[E(W)] \quad (2)^{29}$$

El primer paso de este algoritmo consiste en propagar hacia delante un patrón de entrada X_p y obtener la salida de la red Y_p y la salida de la neurona i viene dada según su estado de activación. Si consideramos la función de salida identidad se tiene que:

$$y_i(t) = F_i(a_i(t)) = a_i(t) \quad (3)$$

Siendo:

$$a_i(t) = f_i(h_i(t)) \quad (4)$$

La regla de propagación más simple y utilizada consiste en realizar una suma de las entradas ponderadas con sus pesos sinápticos correspondientes.

$$h_j(t) = \sum_j w_{ij} (x_j(t)) \quad (5)$$

Se compara la salida obtenida Y_p con la salida deseada D_p ³⁰, obteniéndose un error que viene dado por:

$$e_p = \frac{1}{2} \sum_{k=1}^M (d_{pk} - y_{pk})^2 \quad (6)$$

donde k es el índice de neurona para las neuronas de la última capa y M el total de neuronas de la misma (de la última capa).

²⁸ Función de los errores del peso.

²⁹ Se sustituye el tamaño del paso en cada iteración (α) multiplicado por la función de los errores del peso ($\alpha (E(W))$) en lugar conjunto de pesos de en un instante t ($W(t)$).

³⁰ Salida deseada de la neurona p para un k -ésimo patrón de entrenamiento.

El error total de la red está dado por:

$$e = \frac{\sum_{p=1}^P e_p}{P} \quad (7)$$

siendo p el índice de ejemplo y P el número total de ejemplos.

De acuerdo a la ecuación (2) la variación de los pesos sinápticos será proporcional al gradiente de la función de error:

$$\Delta w_{ij} = -\alpha \frac{\partial e_p}{\partial w_{ji}} \quad (8)$$

Si se aplica la regla de la cadena a la ecuación (8), se obtiene que:

$$\frac{\partial e_p}{\partial w_{ji}} = \frac{\partial e_p}{\partial h_j} \frac{\partial h_j}{\partial w_{ji}} \quad (9)$$

La ecuación (9) expresa la derivada del error en función de dos derivadas. La derivada del error respecto al potencial resultante h_j indica como varia el error al variar la entrada de la neurona j , mientras que la derivada con respecto al peso sináptico W_{ji} indica como varia la entrada de la neurona j al variar el peso de la conexión que va desde la neurona i hasta la neurona j . El segundo término de la ecuación (9) se puede expresar a partir de la ecuación (5) de la siguiente manera:

$$\frac{\partial h_j}{\partial w_{ji}} = \frac{\partial \sum_i w_{ji} y_{pi}}{\partial w_{ji}} = y_{pi} \quad (10)$$

si se escribe al primer término de la ecuación (9) como:

$$\frac{\partial e_p}{\partial h_j} = -\delta_{pj} \quad (11)$$

Luego se tiene que:

$$\frac{\partial e_p}{\partial w_{ji}} = -\delta_{pj} y_{pi} \quad (12)$$

y por lo tanto la ecuación (8) queda expresada de la siguiente manera:

$$\Delta w_{ij} = -\alpha \delta_{pj} y_{pj} \quad (13)$$

Para calcular el valor de delta se vuelve a aplicar la regla de la cadena:

$$\delta_{pj} = -\frac{\partial e_p}{\partial h_j} = -\left(\frac{\partial e_p}{\partial y_{pj}} \frac{\partial y_{pj}}{\partial h_j}\right) \quad (14)$$

El cálculo del segundo término de la ecuación anterior es simple si se observa las ecuaciones (3) y (4):

$$\frac{\partial y_{pj}}{\partial h_j} = \frac{\partial f_j(h_j)}{\partial h_j} = f'_j(h_j) \quad (15)$$

sin embargo, para el cálculo del primer término de la ecuación anterior es necesario distinguir entre dos casos diferentes.

- **La neurona j es una neurona de salida**

En este caso se puede obtener el segundo término a partir de la ecuación (6) con el subíndice j es igual al subíndice k :

$$\frac{\partial e_p}{\partial y_{pj}} = \frac{\partial \frac{1}{2} \sum_{j=1}^M (d_{pj} - y_{pj})^2}{\partial y_{pj}} = -(d_{pj} - y_{pj}) \quad (16)$$

Así, la variación de los pesos de una conexión que va hacia la capa externa de la red se calcula como:

$$\Delta w_{ji} = \alpha (d_{pj} - y_{pj}) f'_j(h_j) y_{pi} \quad (17)$$

- **La neurona j es una neurona oculta**

En este caso es necesario aplicar nuevamente la regla de la cadena:

$$\frac{\partial e_p}{\partial y_{pj}} = \sum_k \left(\frac{\partial e_p}{\partial h_k} \frac{\partial h_k}{\partial y_{pj}} \right) \quad (18)$$

Donde k es el subíndice de las neuronas que pertenecen a la próxima capa. La ecuación anterior se puede reescribir:

$$\frac{\partial e_p}{\partial y_{pj}} = \sum_k \left(\frac{\partial e_p}{\partial h_k} \frac{\partial (\sum_j w_{kj} y_{pj})}{\partial y_{pj}} \right) = \sum_k \left(\frac{\partial e_p}{\partial h_k} w_{kj} \right) \quad (19)$$

por la ecuación (11) se tiene:

$$\frac{\partial e_p}{\partial y_{pi}} = \sum_k -\delta_{pj} w_{kj} = -\sum_k \delta_{pj} w_{kj} \quad (20)$$

Así, la variación de los pesos de una conexión que va desde una capa hacia otra capa de la red que no sea la externa se calcula como:

$$\Delta w_{ij} = \alpha \sum_k (\delta_{pk} w_{kj}) f'_j(h_j) y_{pj} \quad (21)$$

En la implementación del algoritmo, se toma una amplitud de paso que viene dado por la tasa de aprendizaje (α). A mayor tasa de aprendizaje el proceso será más rápido. Sin embargo, si la tasa de aprendizaje es muy alta puede dar lugar a oscilaciones en torno a un mínimo local. Es posible disminuir el impacto de dichas oscilaciones mediante la adición de un momento (β), quedando una expresión que se puede expresar de la siguiente manera:

$$\Delta w_{ji}(t + 1) = \alpha \delta_{pj} y_{pj} + \beta \Delta w_{ji}(t) \quad (22)$$

De esta manera el momento β determina el efecto en el instante $t + 1$ del cambio de los pesos realizado en el instante t . Con esto se consigue la convergencia de la red en menor número de iteraciones, pues la modificación de los pesos en los instantes t y $t + 1$ es en la misma dirección, entonces el descenso por la superficie de error en $t + 1$ es mayor. En cambio, si la modificación en los pesos en los instantes t y $t + 1$ se produce en direcciones opuestas, el paso que se da en $t + 1$ es más pequeño, lo que es adecuado, significando que se ha pasado por un mínimo.

Resumiendo, el algoritmo backpropagation queda expresado de la siguiente manera:

$$w_{ji}(t + 1) = w_{ji}(t) + [\alpha \delta_{pj} y_{pj} + \beta \Delta w_{ji}(t)] \quad \text{si:}$$

$$\delta_{pj} = (d_{pj} - y_{pj}) f'_j(h_j) \quad \text{si } j \text{ es una neurona de salida a}$$

$$\delta_{pj} = \left(\sum_k \delta_{pk} w_{kj} \right) f'_j(h_j) \quad \text{si } j \text{ es una neurona oculta}$$

Modos de entrenamiento

Durante la aplicación del algoritmo backpropagation, el aprendizaje se produce mediante la presentación sucesiva de un set de entrenamiento. Cada presentación completa al perceptrón multicapa del set de entrenamiento se denomina epoch. Así, el proceso de aprendizaje se repite epoch tras epoch hasta que los pesos sinápticos se estabilizan y la performance de la red converge a un valor aceptable. La forma en que se actualizan los pesos sinápticos da lugar a dos modos de entrenamientos distintos, cada uno con sus ventajas y desventajas.

- **Modo Secuencial**

La actualización de los pesos sinápticos se produce tras la presentación de cada ejemplo de entrenamiento (Yao, 1993), conocido también como: modo por patrón. Si un set de entrenamientos posee N ejemplos, este modo tiene como resultado N correcciones de pesos sinápticos durante cada epoch.

- **Modo Batch**

En este modo de entrenamiento la actualización de los pesos sinápticos se produce una única vez, tras la presentación de todo el set de entrenamiento. Para cada epoch se calcula el error cuadrático medio (9) producido por la red. La variación de los pesos sinápticos para un set de entrenamiento se puede calcular como:

$$\Delta w_{ji} = -\alpha \frac{\partial e}{\partial w_{ji}} = -\frac{\alpha}{N} \sum_{n=1}^N e_j \frac{\partial e_j}{\partial w_{ji}} \quad (23)$$

Si los patrones de entrenamiento se presentan a la red de manera aleatoria, el modo de entrenamiento secuencial convierte la búsqueda en el espacio de pesos en estocástica por naturaleza y disminuye la probabilidad de que el algoritmo backpropagation quede atrapado en un mínimo local. Sin embargo, la naturaleza estocástica del modo de entrenamiento secuencial dificulta el establecimiento de condiciones teóricas para la convergencia del algoritmo. Por su parte, el uso del modo de entrenamiento batch provee una estimación precisa del vector gradiente, garantizando de esta manera la convergencia hacia un mínimo local.

Aceleración del aprendizaje

El algoritmo backpropagation ha sido utilizado con éxito en gran cantidad de aplicaciones, pero el éxito y la velocidad de convergencia de este mecanismo de entrenamiento tienen un alto grado de dependencia de la configuración del mismo. Por ello se han realizado una serie de métodos que permiten mejorar significativamente la performance del algoritmo.

- **Modo de actualización:** La actualización secuencial es más rápida computacionalmente y demanda menos recursos que la actualización batch. Esto es especialmente cierto cuando el set de datos es grande y altamente redundante (Chinrungrueng, 1993).
- **Set de datos:** La calidad del set de datos es un factor muy importante a tener en cuenta. Cada ejemplo presentado a la red debe cumplir con las siguientes dos premisas, la de maximizar el error de entrenamiento y maximizar la información.
- **Presentación de los ejemplos:** El modo en que se presentan los ejemplos es otro factor importante a tener en cuenta. La aleatorización del orden en que se presentan los ejemplos en los distintos epochs evita que los resultados se vean distorsionados por el orden de los ejemplos.
- **Función de activación:** El uso de una función de activación adecuada puede acelerar notoriamente el tiempo de entrenamiento. El uso de funciones antisimétricas en general produce mejores tiempos que el uso de funciones no simétricas.

- **Valores objetivos:** La selección de los valores objetivos debe hacerse de acuerdo a la función de activación seleccionada. Una técnica que permite acelerar el tiempo de aprendizaje es desplazar el valor objetivo del valor máximo de la función. El algoritmo backpropagation tiende a saturar las neuronas ocultas cuando el valor objetivo es igual al máximo de la función de activación. Cuando esto sucede se produce el efecto de parálisis, lo que produce un aumento en el tiempo total de entrenamiento. Desplazando el valor objetivo un offset del valor máximo de la función se reduce el riesgo de que las neuronas ocultas se saturen.
- **Normalización de las entradas:** Si bien el algoritmo backpropagation no exige que los valores de entrada a la red se encuentren normalizados, esta es una buena técnica para acelerar los tiempos de entrenamiento (Bishop & Jordan, 1996). La normalización de las entradas debe realizarse de manera tal que el valor medio de la misma se encuentre cercano a cero.
- **Preprocesamiento de los ejemplos:** Se aplica en aquellos casos en que un atributo toma un conjunto discreto de valores. Si un atributo sólo puede tomar N valores diferentes, la entrada de la red puede subdividirse en N entradas, representando a una clase. Cada una de estas entradas puede tomar dos valores, verdadero o falso. Esta técnica puede ayudar a mejorar los tiempos de entrenamiento de la red neuronal.

2.3.8. El Sobreaprendizaje o sobreajuste en una Red Neuronal

Un posible problema del proceso de entrenamiento es el sobreaprendizaje o pérdida de generalización, puesto que al momento de estimar un modelo neuronal puede ocurrir que el error alcanzado durante la etapa de entrenamiento es muy pequeño, pero frente a nuevos datos (etapa de validación) la red neuronal arroja soluciones con un alto error. Es decir, dado un conjunto de datos es posible que la red neuronal reproduzca muy bien el comportamiento de dichos datos durante el entrenamiento, pero no es capaz de generalizar frente a una nueva situación con datos nuevos.

Este problema se acentúa en el caso de que los datos tengan ruido o errores y la generalización no es garantizada cuando el error durante el entrenamiento se reduce a cero.

Aparte de que exista sobreajuste (Overfitting) también puede ocurrir una sobregeneralización (underfitting) el cuál sucede cuando existe un exceso de generalización del modelo e ignora todas o la mayoría de las muestras del entrenamiento.

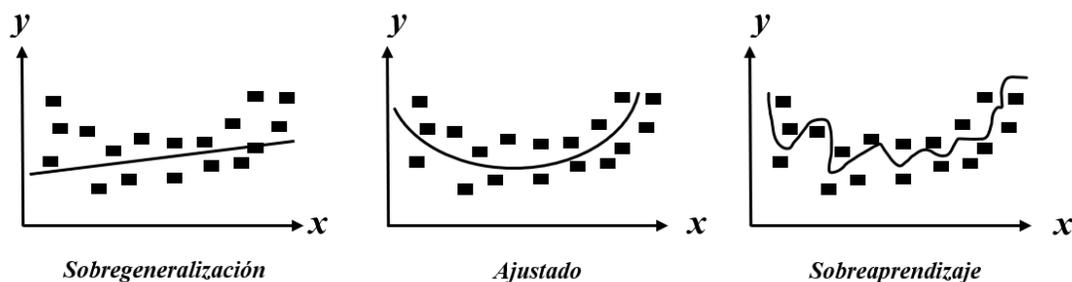


Figura 19: Modelo con sobregeneralización (izquierda), ajustado (centro) y sobreajustado (derecha)

Una señal inequívoca de que un modelo sobreaprende es que aún dotado de la capacidad suficiente su error sobre el conjunto de entrenamiento es mucho menor que su error sobre el conjunto de prueba. Debido a esta situación existen diferentes modos de evitar el sobreaprendizaje:

- **Obtener más datos**

Si al momento de reajustar el modelo se disponen de la capacidad de cálculo necesaria para entrenar la red utilizando un conjunto de datos de entrenamiento más grande.

- **Ajustar la capacidad del modelo**

Ajustar adecuadamente los parámetros de la red también puede servir para que la red tenga la capacidad adecuada y a la vez suficiente para identificar las regularidades relevantes en los datos de entrenamiento, pero no demasiada para que también se ajuste a las espurias. Pues se supone que las espurias³¹ serán más débiles que las auténticas, ya que de otro modo el aprendizaje resultaría por completo imposible.

³¹ Es una situación donde dos o más variables de medidas se creen estadísticamente relacionadas, pero no tienen relación de causalidad entre ellas.

2.3.9. Fase de Validación o Test de una Red Neuronal

Tras la fase de entrenamiento viene la fase de ejecución, en la que se pide a la red que responda a estímulos diferentes a los presentados durante la fase de entrenamiento. Gracias a esta fase la red deberá ser capaz de generalizar y dar respuestas correctas ante patrones de estímulos nuevos, una vez terminado el aprendizaje la red puede generalizar; es decir, ante entradas similares a las del entrenamiento será capaz de producir salidas correctas. Y es de tener en cuenta que es muy difícil conseguir la capacidad de generalización de una red sin utilizar grandes cantidades de datos y que estos sean muy variados.

Para operar con una red entrenada el proceso es el mismo que cuando se realizaba el entrenamiento; por lo tanto, se le sigue suministrando información de entrada a la red, solo que en este proceso no se realiza ningún ajuste en los pesos sinápticos. La red reconoce o evalúa y da una respuesta a partir de los pesos establecidos durante el entrenamiento. Para conseguir el mejor rendimiento de generalización, los datos usados para el entrenamiento deben cubrir un rango de hechos suficientemente amplio, cuando aumenta el tamaño y variedad de los ensayos disminuye la necesidad de que los datos de entrada durante la fase de trabajo normal se parezcan mucho a los patrones del ensayo; es decir, la red generaliza de mejor manera. Si los datos se diferencian demasiado de todos los patrones del ensayo la red tiene dificultades para encontrar la respuesta correcta.

Debido a la gran cantidad de información que se le puede proporcionar a una red, se ha de buscar un criterio de selección para crear el ensayo. Durante los ensayos debe de existir suficientes hechos, es decir, parejas “patrones de estímulos - respuesta correcta”. Además, los hechos del ensayo deben cubrir ampliamente la totalidad de las características a los que la red debe de enfrentarse.

Para una red evaluadora es importante mostrarle tanto los patrones de entrada que llevan a evaluaciones positivas, como los patrones de entrada que llevan a evaluaciones negativas, debido a esto el entrenamiento de la red debe incluir situaciones que se evalúen negativamente, pues de lo contrario la red simplemente es un proceso de aprendizaje y reconoce todos los patrones como correctos.

Además, no se puede incluir en los ensayos una colección exageradamente grande de hechos, pues es necesario seleccionar aquellos hechos que reflejen claramente cada uno de los patrones a reconocer y las situaciones extremas de evaluación en una red evaluadora, lo ideal es preparar una colección amplia de hechos de entrenamiento que cubran todos los problemas a los que se pueda tener que enfrentar la red.

Una vez que ha terminado el proceso de aprendizaje y los pesos de la red neuronal han sido calculados, es importante comprobar la calidad del modelo resultante. Es preciso asegurarse de que nuestra red funciona correctamente, ya que podría no hacerlo por diferentes motivos. Los más usuales son los siguientes:

- Entrenamiento insuficiente.
- Conjunto de datos no representativo.
- Sobreentrenamiento.

Dentro de las técnicas de validación, uno de los principales objetivos que se persigue es encontrar cuál es exactamente el punto en el que debemos parar el entrenamiento.

Generalmente, se suelen seguir los siguientes criterios:

- Al llegar a evaluar un número determinado de ejemplos.
- Cuando el error alcanza un mínimo.
- Si la reducción del error es inferior al estipulado convenientemente.

Método de validación simple

Este es el método conocido como “hold-out”. Su metodología divide el conjunto de datos en 3 grupos, que deberán estar claramente diferenciados y escogidos a conciencia:

- Datos para el entrenamiento: serán los datos con los que se entrenará la red, y a través de los que se realizará el aprendizaje.
- Datos para la validación: serán los datos que servirán como comprobación de que nuestra red neuronal funciona adecuadamente y en la manera que se espera.

- Datos para el test: serán un conjunto reducido de datos escogidos al azar, que darán finalmente un resultado del trabajo de la red.

Método de validación cruzada

Es más conocido como “cross-validation”. Este método de validación busca eliminar la dependencia del azar a la hora de seleccionar los conjuntos de datos para la validación. Así, en cada serie de conjuntos de datos, se seleccionará uno de ellos para formar parte de la validación. Produce mayor fiabilidad en la validación, aunque consume un tiempo algo superior.

Otros métodos de validación

Simplemente, mencionar otros métodos usuales en la validación de redes neuronales:

- Bootstrapping.
- Stacked generalization.

En el caso del aprendizaje supervisado, una medida de la calidad puede darse en términos de los errores entre los valores de salida deseados y los obtenidos por la red neuronal. Algunas medidas estándar del error son:

- La suma de los cuadrados de los errores.
- La raíz cuadrada del error cuadrático medio.
- El error máximo.

2.3.10. Aplicaciones de las redes neuronales artificiales

- **Clasificación:** puede identificar datos y clasificar en clases. A diferencia de las técnicas tradicionales puede generar para un problema concreto estructuras específicas de identificación de características.

- **Asociación:** la red funciona como un proceso de recuperación de un dato a partir de una entrada relacionada con el dato almacenado.
- **Agrupamiento:** cuando no se produce una clasificación de los datos de entrada, es la red la que genera esta clasificación a medida que se va produciendo la señal.
- **Generalización:** las redes pueden detectar las relaciones subyacentes entre diferentes elementos, y por tanto, pueden extrapolar las relaciones numéricas presentes en los datos disponibles en situaciones similares.
- **Optimización** cuando es necesario elegir la opción óptima entre diferentes alternativas, las redes proporcionan un procedimiento relativamente rápido para generar una solución sub-óptima que es suficientemente satisfactoria.
- **Predicción y control:** la flexibilidad de las redes neuronales las hace adecuadas para la predicción ya que el sistema va aprendiendo y puede seguir de manera versátil estos cambios.

Las redes neuronales artificiales en la actualidad se están utilizando principalmente en una doble dirección: **predicción y clasificación**. (Torres, 1997)

2.3.10.1. Aplicación de las Redes Neuronales a los Pronósticos

El verdadero poder de las redes neuronales recae en su habilidad para capturar relaciones no lineales inherentes en los datos, mientras que los modelos lineales describen una relación lineal entre observaciones actuales y posteriores, las redes neuronales describen una relación no lineal entre las dos. Debido a este motivo son de gran utilidad en el área de las predicciones.

A partir de esto, se considera que el perceptrón multicapa es el tipo de red neuronal artificial más utilizado y con los mejores resultados para este tipo de aplicación (predicciones) (Dorffner, 2008). Para diseñar un modelo de una red neuronal artificial con aplicaciones a predicciones, se propone el siguiente procedimiento:

- Selección de variables.
- Recolección de datos.
- Preprocesamiento de datos.
- Conjunto de entrenamiento, prueba y validación.
- Paradigmas de redes neuronales: Elección del número de capas ocultas, número de neuronas ocultas, número de neuronas de salida y seleccionar la función de transferencia.
- Criterio de evaluación.
- Entrenamiento de red neuronal: Seleccionar el número de iteraciones de entrenamiento y la tasa de aprendizaje e impulso.
- Aplicación.

El procedimiento empieza identificando las variables de entrada, para luego proceder a la recopilación de datos asegurando la consistencia de los mismos disminuyendo el ruido inherente en ellos (puede ser aplicando estandarización), luego se definen los conjuntos de entrenamiento, prueba y validación en base a los datos obtenidos. A partir de este momento es tiempo de comenzar a tomar en cuenta los paradigmas de diseño de la red neuronal artificial como el número de capas, el número de neuronas en cada capa, el número de salida y la función de transferencia. Una vez diseñada la red se debe de seleccionar los criterios de evaluación de los resultados que arroja la misma, inmediatamente después se debe de planificar como va a ser la fase de aprendizaje de la red, teniendo en cuenta factores tales como el número de interacciones y la tasa de aprendizaje y luego se implementa el modelo haciendo uso de un software adecuado. (Kaastra & Boyd, 1996)

Capítulo 3. Fundamento Teórico: Series de Tiempo

Una serie temporal es una secuencia ordenada de observaciones cada una de las cuales está asociada a un momento de tiempo e implica como objetivo el analizar la evolución de una variable a través de él. La diferencia esencial entre las series temporales y los análisis no temporales (Estadística descriptiva, Diseño de experimentos o Regresión) es que en los análisis previos no importa el orden en que estaban tomadas las observaciones y éste se podía variar sin problemas.

La mayoría de los métodos estadísticos elementales suponen que las observaciones individuales que forman un conjunto de datos son realizaciones de variables aleatorias mutuamente independientes. El supuesto de independencia mutua se justifica por la atención prestada a diversos aspectos del experimento, incluyendo la extracción aleatoria de la muestra de una población más grande, la asignación aleatoria del tratamiento a cada unidad experimental, por mencionar algunos aspectos. Además, en este tipo de datos (tomamos una muestra aleatoria simple de una población más grande) el orden de las observaciones no tiene mayor importancia.

Sin embargo, en el caso de las series temporales se debe de tener en cuenta que el orden es fundamental, pues se posee un conjunto de datos ordenado, y por otra parte el supuesto de independencia no se sostiene debido a que las observaciones son dependientes entre sí y la naturaleza de su dependencia es de interés en sí misma y el análisis debe llevarse a cabo teniendo en cuenta el orden temporal de las observaciones. Los métodos estadísticos basados en la independencia de las observaciones no son válidos para el análisis de series temporales porque las observaciones en un instante de tiempo dependen de los valores de la serie en el pasado ya que dichos análisis utilizan datos históricos como base para estimar resultados futuros.

El proceso de modelización de series de tiempo está siendo entendido como el aprendizaje de una serie de pasos para identificar sólo aquellos elementos salientes de un fenómeno y para evaluar, según distintas reglas, el modelo escogido que permitirá percibir de otra manera la nueva descripción de los fenómenos o como el proceso de razonamiento integrado que “hace uso de un modelaje analógico y visual y de experimentos pensados en la creación y transformación de las representaciones de un problema”.

El conjunto de técnicas de estudio de series de observaciones dependientes ordenadas en el tiempo se denomina análisis de series temporales, el cual utiliza un instrumento de análisis que suele ser un modelo que permite reproducir el comportamiento de la variable de interés. Dichos Modelos de Series Temporales pueden ser: Univariantes o Multivariantes³².

Los objetivos importantes al analizar una serie temporal se pueden encontrar:

- **Descripción**

Se describe las características de la serie, en términos de sus componentes de interés, puesto que cuando se estudia una serie temporal lo primero que se realiza es graficar los datos y considerar las medidas descriptivas básicas, considerando si los datos presentan forma creciente o decreciente (tendencia), si existe influencia de ciertos periodos de cualquier unidad de tiempo (estacionalidad) y si aparecen outliers (observaciones extrañas o discordantes).

- **Predicción**

Cuando se observan los valores de una serie, se pretende normalmente no sólo explicar el pasado, sino también predecir el futuro. Y si se pueden predecir exactamente los valores, se dice que las series son determinísticas. Por otro caso si el futuro sólo se puede determinar de modo parcial por las observaciones pasadas y no se pueden determinar exactamente, se considera que los futuros valores tienen una distribución de probabilidad que está condicionada a los valores pasados, siendo series estocásticas. (Moreno, 2018)

³² Un modelo univariante es cuando sólo se analiza una serie temporal en función de su propio pasado y el modelo multivariante analiza varias series temporales a la vez.

3.1. Componentes de una serie de tiempo

El estudio descriptivo de series temporales se basa en la idea de descomponer la variación de una serie en varias componentes básicas. Este enfoque no siempre resulta ser el más adecuado, pero es interesante cuando en la serie se observa cierta tendencia o cierta periodicidad. Hay que resaltar que esta descomposición no es en general única. Las componentes o fuentes de variación que se consideran habitualmente son las siguientes:

- **Tendencia (T):** Es la componente general de la serie y puede considerarse como el movimiento global de la serie en relación al nivel medio, o el cambio a largo plazo de la media. Se identifica con un movimiento suave de la serie a largo plazo, además, suele obtenerse o describirse mediante ajustes a una función matemática, por medias móviles o aislamiento exponencial.
- **Estacionalidad (E):** Movimiento periódico que se producen en un lapso corto y conocido, es decir variaciones inferiores a un año y reconocibles todos los años.
- **Ciclo (C):** caracterizado por oscilaciones alrededor de la tendencia con una larga duración, y sus factores no son claros.
- **Aleatorio (A):** son movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas, siendo prácticamente impredecible. Este comportamiento representa todos los tipos de movimientos de una serie de tiempo que no son tendencia, variaciones estacionales ni fluctuaciones cíclicas (Hurtado C. , 2008).

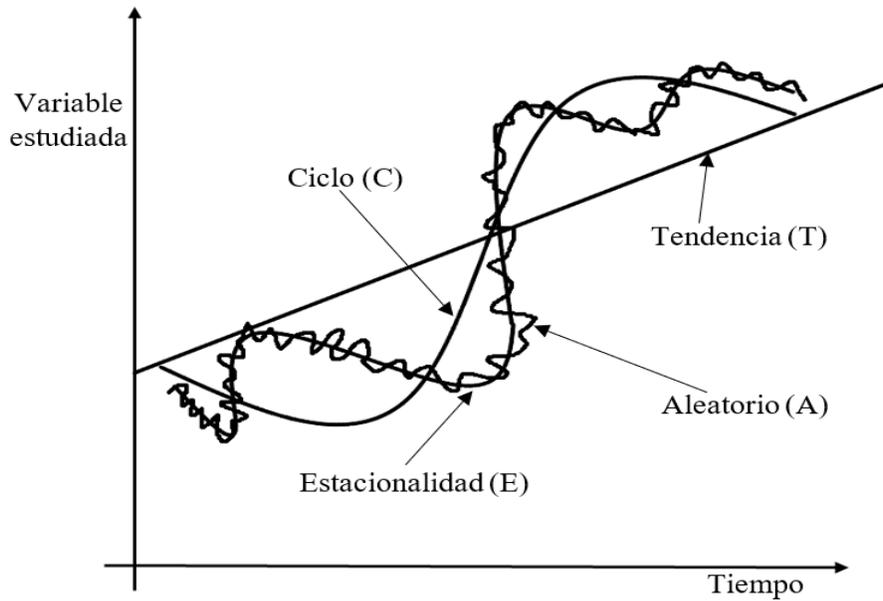


Figura 20: Componentes de una serie de tiempo

3.2 Clasificación de una serie de tiempo

Las series temporales se pueden clasificar en:

- **Estacionarias:**

Una serie es estacionaria cuando es estable 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.

- **No 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.

3.3. Frecuencias de las series de tiempo

Con las series de tiempo se pueden realizar estudios en diferentes lapsos del tiempo, esto es muy importante, pues se puede optar por diversos modelos para poder realizar el mejor análisis a los datos con los que se cuenta. La periodicidad puede ser:

- Anual: se toma un dato cada año.
- Trimestral: se toma un dato cada tres meses
- Mensual: se toma un dato cada mes.
- Semanal: se toma un dato cada semana.
- Diaria: se toma un dato cada día.

3.4. Procesos Clásicos

El análisis más clásico de las series temporales se basa en la idea de que los valores que toma la variable de observación son la consecuencia de las componentes anteriores (tendencia, estacionalidad, ciclo y aleatoriedad), aunque no siempre aparecen todas. Luego este enfoque descriptivo determinista consiste en encontrar componentes que se correspondan a una tendencia a largo plazo, un comportamiento estacional y una parte aleatoria. Para ello se consideran dos modelos habituales para relacionar los valores de la serie con los componentes:

Modelo aditivo: Es un modelo donde cada componente contribuye al comportamiento de la variable de interés en forma aditiva (unidades):

$$Y_t = T_t + E_t + A_t + C_t$$

Modelo multiplicativo: Este modelo se caracteriza porque cada componente contribuye al comportamiento de la variable de interés en forma multiplicativa:

$$Y_t = T_t \cdot E_t \cdot A_t \cdot C_t$$

Donde T_t es la tendencia, E_t es la componente estacional que constituyen la señal o parte determinística y A_t es la parte aleatoria. Para conocer qué tipo se adapta mejor a la serie, se pueden seguir los siguientes procedimientos:

De manera visual:

- La tendencia y la estacionalidad se mantienen relativamente constantes (modelo aditivo).
- La tendencia y la estacionalidad varían creciendo o decreciendo (modelo multiplicativo).

De forma matemática:

- Calcular las series diferencia y cociente.
- Calcular el coeficiente de variación para cada serie: coeficiente de variación de la diferencia (CVD) y coeficiente de variación del cociente (CVC).
- Comparar sendos coeficientes (Si $CVC < CVD$ es un modelo multiplicativo, pero si $CVC > CVD$ es un modelo aditivo.).

Una vez que se conoce la forma en que se relacionan los componentes, se descompone la serie estimando los componentes de tendencia y estacionalidad.

Componente de tendencia: La forma de la tendencia se puede modelar mediante los siguientes métodos:

- Modelos lineales.
- Modelos polinómicos.
- Filtrado (Medias móviles) (no se recomienda ya que no predice, solo describe).
- Diferenciación (diferencias regulares).

Es importante reseñar que como la finalidad del análisis no es solo describir la serie sino predecir, hay que tener en cuenta cómo continúa la tendencia estimada para valores futuros.

Componente de estacionalidad:

Para estimar el efecto estacional se pueden obtener los índices o coeficientes de estacionalidad que representan el valor promedio para cada elemento de la estación³³. Además, la serie original se puede desestacionalizar también mediante una diferenciación estacional de la serie.

- Medias móviles centradas.
- Diferenciación (diferencias estacionales)

En este punto se tiene una descomposición de la serie en componentes que separan tendencia, estacionalidad, ciclo y ruido. Estos componentes obtenidos de la serie la describen, pero no la predicen.

Las predicciones de valores futuros se consiguen usando las componentes T_t y E_t con valores de tiempo $t + 1, t + 2, \dots$. Para ello se realiza un pronóstico futuro de la tendencia, y se le añade la predicción de la estacionalidad (índice de estacionalidad) correspondiente a cada periodo (la componente irregular o aleatorio no es predecible y por lo tanto no se considera) (Villavicencio, 2010).

3.5. Proceso estocástico

Es una sucesión de variables aleatorias Y_t ordenadas, pudiendo tomar t cualquier valor entre $-\infty$ y ∞ . El subíndice t no tiene en principio, ninguna interpretación a priori, pero en los procesos estocásticos, en el contexto del análisis de series temporales, este subíndice representará el paso del tiempo.

³³ Si es anual, para cada mes, si es trimestral cada trimestre, esto es, el periodo de la serie y la suma de los índices estacionales debe de ser igual al número de periodos.

Cada una de las variables Y_t que configuran un proceso estocástico tendrán su propia función de distribución con sus correspondientes momentos. Así mismo, cada par de esas variables tendrán su correspondiente función de distribución conjunta y sus funciones de distribución marginales. Esto mismo ocurrirá, no para cada par de variables, sino para los conjuntos más amplios. De esta forma, para caracterizar un proceso estocástico es necesario especificar las funciones de distribución conjunta de cualquier conjunto de variables:

$$(Y_{t_1}, Y_{t_2}, Y_{t_3}, \dots, Y_{t_m})$$

cualesquiera que fueran los valores de (t_1, t_2, \dots, t_m) y cualquiera que fuera el valor de m :

$$\begin{aligned} y_1, y_2, y_3 & \quad (t_1 = 1 \quad y \quad m = 3) \\ y_3, y_4, y_5, y_6 & \quad (t_1 = 3 \quad y \quad m = 4) \end{aligned}$$

Habitualmente, conocer esas funciones de distribución resulta complejo de forma que, para caracterizar un proceso estocástico, basta con especificar la media y la varianza para cada y_t y la covarianza para variables referidas a distintos valores de t :

$$\begin{aligned} E[Y_t] &= \mu_t \\ \sigma_t^2 &= Var(y_t) = E[y_t - \mu_t]^2 \\ \gamma_t &= cov(Y_y, Y_s) = E[(y_t - \mu_t)(y_s - \mu_s)] \end{aligned}$$

3.5.1. Características de un proceso estocástico

Un proceso estocástico se puede caracterizar bien por su función de distribución o por sus momentos.

Función de distribución

Para conocer la función de distribución de un proceso estocástico es necesario conocer las funciones de distribución univariantes de cada una de las variables aleatorias del proceso: $F[Y_{t_i}], \forall t_i$ y las funciones bivariantes correspondientes a todo par de variables aleatorias del proceso: $F[Y_{t_i}, Y_{t_j}], \forall (t_i, t_j)$, del mismo modo para cualquier análisis multivariante.

La función de distribución de un proceso estocástico incluye todas las funciones de distribución para cualquier subconjunto finito de variables aleatorias del proceso: $F[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}], \forall (t_1, t_2, \dots, t_n)$, siendo n finito.

Momentos del proceso estocástico

Como suele ser muy complejo determinar las características de un proceso estocástico a través de su función de distribución se suele recurrir a caracterizarlo a través de los dos primeros momentos.

El primer momento de un proceso estocástico viene dado por el conjunto de las medias de todas las variables aleatorias del proceso:

$$E(Y_t) = \mu_t < \infty, \quad t = 0, \pm 1, \pm 2, \dots,$$

El segundo momento centrado del proceso viene dado por el conjunto de las varianzas de todas las variables aleatorias del proceso y por las covarianzas entre todo par de variables aleatorias:

$$V(Y_t) = E[Y_t - \mu_t]^2 = \sigma_t^2, < \infty, \quad t = 0, \pm 1, \pm 2, \dots,$$

$$cov(Y_t Y_s) = E[Y_t - \mu_t][Y_s - \mu_s] = \rho_{t,s}, \forall t, s (t \neq s)$$

Si la distribución del proceso es normal y se conocen sus dos primeros momentos (medias, varianzas y covarianzas), el proceso está perfectamente caracterizado y se conoce su función de distribución (Arce, 2001).

3.5.2. Procesos estocásticos estacionarios

El objetivo del análisis de series temporales es utilizar la teoría de procesos estocásticos y de esta manera determinar qué proceso estocástico ha sido capaz de generar bajo estudio con el fin de explicar el comportamiento de la serie de tiempo y predecir a través de éste. Para obtener métodos de predicción consistentes, no se puede utilizar cualquier tipo de proceso estocástico, y es necesario que la estructura probabilista del mismo sea estable en el tiempo.

El concepto de estacionariedad se puede caracterizar bien en términos de la función de distribución o de los momentos del proceso. En el primer caso, se habla de estacionariedad en sentido estricto y en el segundo de estacionariedad de segundo orden o en covarianza.

Estacionariedad estricta

Un proceso estocástico Y_t es estacionario en sentido estricto si y solo si:

$$F[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}] = F[Y_{t_1+k}, F[Y_{t_2+k}, \dots, F[Y_{t_n+k}]] \quad \forall (t_1, t_2, \dots, t_n) \text{ y } k$$

es decir, si la función de distribución de cualquier conjunto finito de n variables aleatorias del proceso no se altera si se desplaza k periodos en el tiempo.

Estacionariedad en covarianza

Un proceso estocástico Y_t es estacionario en covarianza si y solo si:

- a) Es estacionario en media, es decir, todas las variables aleatorias del proceso tienen la misma media y es finita: $E(Y_t) = \mu < \infty, \forall t$
- b) Todas las variables aleatorias tienen la misma varianza y es finita, es decir, la dispersión en torno a la media constante a lo largo del tiempo es la misma para todas las variables del proceso.

$$V(Y_t) = E[Y_t - \mu]^2 = \sigma_Y^2 < \infty, \forall t$$

- c) Las autocovarianzas solo dependen del número de periodos de separación entre las variables y no del tiempo, es decir, la covarianza lineal entre dos variables aleatorias del proceso que disten k periodos de tiempo es la misma que existe entre cualesquiera otras dos variables que estén separadas también k periodos, independientemente del momento concreto de tiempo al que estén referidas

$$cov(Y_t Y_s) = E[Y_t - \mu][Y_s - \mu] = \gamma_{|t-s|} = \gamma_k < \infty, \forall k$$

Por lo tanto, un proceso estocástico es estacionario en covarianza si y solo si:

1) $E(Y_t) = \mu < \infty$

$$2) Cov(Y_t Y_s) = \begin{cases} V(Y_t) = \sigma_Y^2 < \infty & t = s \\ \sigma_{|t-s|} = \gamma < \infty & t \neq s \end{cases}$$

Un proceso estocástico estacionario en covarianza está caracterizado si se conoce:

$$\mu, V(Y_t) \text{ y } \gamma_k, k \pm 1, \pm 2, \dots,$$

Si un proceso estocástico es estacionario en covarianza y su distribución es Normal, es estacionario en sentido estricto.

3.5.3. Función de autocovarianzas y de autocorrelación

En principio, si se considera el proceso estocástico teórico Y_t que comienza en algún momento del pasado lejano y acaba en un futuro indeterminado se pueden calcular un número indefinido de autocovarianzas, por lo que conviene definir una función que las agrupe a todas.

3.5.3.1. Función de autocovarianzas

La función de autocovarianzas de un proceso estocástico estacionario es una función del k (número de periodos de separación entre las variables) que recoge el conjunto de las autocovarianzas del proceso y se denota por:

$$\gamma_k, \quad k = 0, 1, 2, 3, \dots,$$

Características de la función de autocovarianzas:

- Incluye la varianza del proceso para $k = 0$:

$$\gamma_0 = E[Y_t - \mu][Y_t - \mu] = V(Y_t)$$

- Es una función simétrica:

$$\gamma_k = E[Y_t - \mu][Y_{t+k} - \mu] = E[Y_t - \mu][Y_{t-k} - \mu] = \gamma_{-k}$$

La función de autocovarianzas de un proceso estocástico recoge toda la información sobre la estructura dinámica lineal del mismo. Pero depende de las unidades de medida de la variable, por lo que, en general, se suele utilizar la función de autocorrelación.

3.5.3.2. Función de autocorrelación

El coeficiente de autocorrelación de orden k de un proceso estocástico estacionario mide el grado de asociación lineal existente entre dos variables aleatorias X e Y del proceso separadas k periodos:

$$\rho_k = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\gamma_k}{\sqrt{\gamma_0\gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

Por ser un coeficiente de correlación, no depende de unidades y $|\rho_k| \leq 1, \forall k$.

La función de autocorrelación de un proceso estocástico estacionario es una función de k que recoge el conjunto de los coeficientes de autocorrelación del proceso y se denota por ρ_k , $k = 0,1,2,3, \dots$. La función de autocorrelación se suele representar gráficamente por medio de un gráfico de barras denominado correlograma. Las características de la función de autocorrelación de un proceso estocástico estacionario están determinadas por:

- El coeficiente de autocorrelación de orden 0 es, por definición, 1. Por eso, a menudo, no se le incluye explícitamente en la función de autocorrelación.

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

- Es una función simétrica: $\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_{-k}}{\gamma_0} = \rho_{-k}$. Por ello, en el correlograma se representa la función de autocorrelación solamente para los valores positivos del retardo k .
- La función de autocorrelación de un proceso estocástico estacionario tiende a cero rápidamente cuando k tiende a ∞ .

La función de autocorrelación es el principal instrumento utilizado para recoger la estructura dinámica lineal del modelo mediante los gráficos de los correlogramas correspondientes a diferentes series temporales. Los correlogramas que decrecen rápidamente hacia cero cuando aumenta k , ya sea de forma exponencialmente o truncándose el decrecimiento, a partir de esto correspondiente a series no estacionarias. Por el contrario, los coeficientes de autocorrelación de correlogramas que decrecen lentamente, de forma lineal corresponden a una serie estacionaria.

Función de Autocorrelación Parcial (FAP)

Es la función que proporciona la relación directa que existe entre observaciones separadas por k retardos. Esta es una información muy valiosa sobre la estructura de la serie, ya que elimina el problema que presentaba la función de autocorrelación simple de que si z_1 influye en z_2 y z_2 influye en z_3 , entonces z_1 influye en z_3 . Es de notar que en la función de

autocorrelación parcial el primer retardo es significativo, mientras que ninguno de los demás lo es. Esto implica que la serie presenta relación directa entre una observación y la siguiente, pero no existe ninguna otra relación directa.

Además, la autocorrelación parcial existe entre los valores de una serie en diferentes instantes después de que se hayan eliminado los efectos de los instantes intermedios y se encuentra definida:

$$\rho_{kk} = \text{Corr}[z_t, z_{t+k} | z_{t+1}, z_{t+2}, z_{t+3}, \dots, z_{t+(k-2)}, z_{t+(k-1)}]$$

También se puede representar por medio de una regresión:

$$z_{t+k} = \phi_{k_1} z_{t+(k-1)} + \phi_{k_2} z_{t+(k-2)} + \phi_{k_3} z_{t+(k-3)}, \dots, \phi_{k_{k-1}} z_{t-1} + \phi_{k_k} z_t = \varepsilon_{t+k}$$

ya que cada coeficiente equivale a ρ_{kk}

$$\rho_{kk} = \frac{\text{cov}(Y_t - \hat{Y}_t, Y_{t-k} - \hat{Y}_{t-k})}{\sqrt{(Y_t - \hat{Y}_t)} \sqrt{Y_{t-k} - \hat{Y}_{t-k}}}$$

donde:

ρ_{kk} : Es el coeficiente de autocorrelación parcial.

\hat{Y}_{t-k} : valor estimado en el periodo $t - k$.

Y_t : es la observación en el periodo t .

3.5.4 Estimación de los momentos

Los momentos poblacionales de un proceso estocástico estacionario se estiman a través de los correspondientes momentos muestrales.

a) **Media:**

$$\hat{\mu} = \bar{Y} = \frac{1}{T} \sum_{t=1}^T Y_t$$

b) **Varianza:**

$$\hat{\gamma}_0 = \hat{V}(Y) = \frac{1}{T} \sum_{t=1}^T (Y_t - \hat{Y})^2$$

c) **Función de Autocovarianzas:**

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (Y_t - \hat{Y})(Y_{t+k} - \hat{Y}), \quad k = 1, 2, \dots, K$$

d) **Función de Autocorrelación:** $\hat{\rho} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}; \quad k = 1, 2, \dots, K$

Aunque para el proceso estocástico teórico se cuenta con un número indefinido de autocovarianzas y coeficientes de autocorrelación, cuando se dispone de una serie temporal finita de tamaño T , como máximo se pueden estimar $T - 1$ coeficientes de autocorrelación, pero se recomienda un máximo de $T/3$ número de observaciones. Esto es debido a que cuanto mayor sea k menos información hay para estimar ρ_k y la calidad de la estimación es menor.

La estimación de los momentos poblacionales de un proceso estocástico proporciona otra razón por la que es necesario imponer la restricción de estacionariedad de un proceso. Si el proceso que ha generado la serie (Y_1, Y_2, \dots, Y_T) no fuera estacionario, sería preciso estimar T medias diferentes (μ_t , T varianzas diferentes (σ_t^2), y un gran número de autocorrelaciones, lo que no es posible contando con T observaciones solamente.

3.5.5. Ruido Blanco

El proceso estocástico más sencillo es el denominado Ruido Blanco que es una secuencia de variables aleatorias de media cero, varianza constante y covarianzas nulas. Se denota habitualmente por $a_t, t = 0, \pm 1, \pm 2, \dots$:

$$E(a_t) = 0, \forall t; V(a_t) = \sigma^2, \forall t; \quad Cov(a_t a_s) = 0, \forall t \neq s$$

Así, un proceso ruido blanco $a_t \sim RB(0, \sigma^2)$, es estacionario si la varianza σ^2 es finita con función de autocovarianzas (FACV):

$$\gamma_k = \sigma^2, k = 0 \quad y \quad \gamma_k = 0, k > 0$$

y función de autocorrelación (FAC):

$$\gamma_k = 1, k = 0 \quad y \quad \gamma_k = 0, k > 0$$

3.6. Modelos ARMA

3.6.1. Modelos autorregresivos

Se define un modelo como autorregresivo si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados más un término de error (Llamas & Ponce, 2011).

Un modelo autorregresivo de orden p , o abreviadamente AR(p), se define como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + u_t$$

pudiéndose escribir de forma abreviada como:

$$\phi_p(L)Y_t = \delta + u_t$$

donde $\phi_p(L)$ es lo que se conoce como operador polinomial de retardos, siendo:

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

Estos modelos, deben cumplir la condición de estacionariedad, y una condición necesaria, aunque no suficiente, para que un modelo autorregresivo de cualquier orden sea estacionario es la siguiente:

$$\phi_1 + \phi_2 + \dots + \phi_p < 1$$

Al cumplir el supuesto de estacionariedad, la esperanza matemática viene dada por:

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$$

Su varianza:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \dots + \phi_p\gamma_p + \sigma_u^2$$

Función de autocovarianzas

$$\gamma_j = \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \dots + \phi_p\gamma_{j-p} \quad \text{para } j > 0$$

Considerando γ_0 y γ_j para $j = 1, 2, 3, \dots, p$ se obtiene el sistema de $p + 1$ ecuaciones:

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \dots + \phi_p\gamma_p + \sigma_u^2$$

$$\gamma_1 = \phi_1\gamma_1 + \phi_2\gamma_2 + \dots + \phi_p\gamma_{p-1}$$

$$\gamma_2 = \phi_1\gamma_0 + \phi_2\gamma_1 + \dots + \phi_p\gamma_{p-2}$$

.....

$$\gamma_p = \phi_1\gamma_{p-1} + \phi_2\gamma_{p-2} + \dots + \phi_p\gamma_0$$

Si se conocen los valores de los parámetros $\phi_1, \phi_2, \dots, \phi_p$ y σ_u^2 , pueden calcularse las varianzas y autocovarianzas de orden no superior al del proceso. A partir de dicho orden, es decir, para $j > p$, las autocovarianzas se obtendrán de forma recursiva según las ecuaciones siguientes:

$$\gamma_{p+1} = \phi_1\gamma_p + \phi_2\gamma_{p-1} + \dots + \phi_p\gamma_1$$

$$\gamma_{p+2} = \phi_1\gamma_{p+1} + \phi_2\gamma_p + \dots + \phi_p\gamma_2$$

$$\gamma_{p+3} = \phi_1\gamma_{p+2} + \phi_2\gamma_{p+1} + \dots + \phi_p\gamma_3$$

Función de autocorrelación: se Divide por la función de autocovarianzas, γ_0 , se tiene:

$$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \dots + \phi_p\rho_{j-p} \quad \text{para } j > 0$$

Las ecuaciones que se obtienen directamente de la función de autocorrelación, serán:

$$\rho_1 = \phi_1 + \phi_2\rho_2 + \dots + \phi_p\rho_{p-1}$$

$$\rho_2 = \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2}$$

.....

$$\rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \dots + \phi_p$$

Resolviendo dicho sistema para $\phi_1, \phi_2, \dots, \phi_p$ y expresado en forma de matrices se tiene:

$$\begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_p \end{bmatrix}^{-1} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{bmatrix}$$

Y se tiene la función de Autocorrelaciones Parciales:

$$\phi_{ij} = \begin{cases} \rho_1 & \text{para } j = 1 \\ \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \frac{\phi_2[(1 - \phi_2)^2 - \phi_1^2]}{(1 - \phi_2)^2 - \phi_1^2} & \text{para } j = 2 \\ \dots & \dots \dots \\ \frac{\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-2} & \rho_1 \\ \rho_1 & 1 & \dots & \rho_{p-3} & \rho_2 \\ \dots & \dots & \dots & & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_1 & 1 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-2} & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-3} & \rho_{p-2} \\ \dots & \dots & \dots & & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & \rho_1 & 1 \end{bmatrix}} & \text{para } j = p \\ 0 & \text{para } j > p \end{cases}$$

Propiedades de los modelos AR(p):

- **Estacionariedad:** La condición necesaria y suficiente de estacionariedad en un modelo $AR(p)$ es que las raíces de su polinomio autorregresivo $\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p = 0$ estén fuera del círculo unidad del plano complejo (sean en modulo estrictamente mayores que la unidad).
- **Invertibilidad:** Invertir un modelo AR consiste en transformarlo en su modelo MA equivalente. Si se tiene el modelo $AR(1)$ estacionario se puede transformar en un modelo de medias móviles de orden infinito $MA(\infty)$. La condición de invertibilidad en los modelos autorregresivos de un número finito de términos se cumple siempre de forma automática.

- **Correlograma simple:** El correlograma simple de un modelo $AR(p)$ está formado por infinitos coeficientes de correlación no nulos, decreciendo según una mezcla de exponenciales y sinusoidales amortiguadas.
- **Correlograma parcial:** El correlograma parcial de un modelo $AR(p)$ tiene no nulos los p primeros coeficientes de correlación ρ'_k (desde $k = 1$ hasta $k = p$), siendo nulos a partir de $k = p + 1$.

3.6.2. Modelo de medias móviles

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Estos modelos se denotan normalmente con las siglas MA , seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis (Hernández, 2007). Así, un modelo con q términos de error $MA(q)$ se representaría por la siguiente expresión:

$$Y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \dots + \theta_q u_{t-q}$$

Abreviando mediante la utilización del polinomio de retardos (como en el caso de los modelos AR):

$$Y_t = \theta_q(L)u_t + \mu$$

donde $\theta_q(L)$ es lo que se conoce como operador polinomial de retardos, definido como:

$$\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

Siendo L el operado de retardo, tal que el valor de su exponente indica el orden de retardo de la variable a la que multiplica, donde la esperanza matemática del modelo de medias móviles, $MA(q)$ viene dada por:

$$E(y_t) = \mu$$

La varianza del modelo $MA(q)$:

$$\gamma_0 = \sigma_u^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)$$

La función de autocovarianzas:

$$\gamma_j = \begin{cases} (-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q)\sigma_u^2 & \text{para } j = 1, 2, 3, \dots, q \\ 0 & \text{para } j > q \end{cases}$$

Y la función de autocorrelación:

$$\rho_j = \begin{cases} \frac{-\theta_j + \theta_1\theta_{j+1} + \dots + \theta_{q-j}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{para } j = 1, 2, 3, \dots, q \\ 0 & \text{para } j > q \end{cases}$$

Finalmente, la función de autocorrelación parcial:

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{33} = \frac{\rho_1^3 - \rho_1\rho_2(2 - \rho_2) + \rho_3(1 - \rho_1^2)}{1 - \rho_2^2 - 2\rho_1^2(1 - \rho_2)}$$

.....

$$\phi_{jj} = \frac{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{j-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{j-3} & \rho_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \cdots & \rho_1 & 1 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{j-2} & \rho_{j-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{j-3} & \rho_{j-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho_{j-1} & \rho_{j-2} & \rho_{j-3} & \cdots & \rho_1 & 1 \end{bmatrix}}$$

Propiedades del modelo MA (q):

- **Estacionariedad:** Los procesos de medias móviles de orden q son siempre estacionarios, al cumplir tres condiciones de la estacionariedad en sentido amplio.
- **Invertibilidad:** Un modelo $AR(\infty)$ siempre y cuando las raíces de la ecuación $\theta(L)$ se ubiquen fuera del círculo unidad, donde $\theta(L) = 0$ y éste se define como:

$$\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q$$

- **El correlograma simple:** El correlograma simple de un proceso $MA(q)$ tiene los q primeros coeficientes de correlación ρ'_k desde $k = 1$ hasta $k = q$ no nulos, siendo el resto nulos. Es decir, las autocovarianzas se anulan cuando el desfase temporal considerado es superior al orden del modelo. Consecuentemente, se dice que la memoria de un proceso de medias móviles es igual al orden q del proceso.
- **Correlograma parcial:** El correlograma parcial de un proceso $MA(q)$ consta de infinitos coeficientes de correlación no nulos, decreciendo según una mezcla de exponenciales y sinusoidales amortiguadas. Es decir, las autocorrelaciones parciales no

se anulan, presentado un comportamiento amortiguado hacia cero (caso de raíces reales) o sinusoidal también tendiendo a cero (en caso de que las raíces sean complejas).

A partir de las características de los modelos $AR(p)$ y $MA(q)$ se aprecia cierta simetría que se denomina dualidad, la cual es consecuencia del hecho de que un mismo proceso podemos representarlo como autorregresivo o como de medias móviles según nuestra conveniencia, dado que un $AR(p)$ puede representarse como un $MA(\infty)$ y un $MA(q)$ como un $AR(\infty)$. La ventaja de usar en cada caso la representación adecuada reside en tener un modelo con pocos parámetros.

3.6.3. Modelo Autorregresivos y de Medias Móviles (ARMA)

Es un modelo mixto que tiene parte autorregresiva y parte de medias móviles ($ARMA(p, q)$) y éste se representa como:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \delta + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \dots - \theta_q u_{t-q}$$

Y el cual, al igual que en los casos anteriores, puede expresarse simplifícadamente, usando los polinomios de retardo como sigue:

$$\phi(L)Y_t = \delta + \theta(L)u_t$$

Dado que este modelo contiene una parte AR y una parte MA, estos deberán cumplir las condiciones de estacionariedad como de invertibilidad. Es decir, que el modelo es estacionario si su parte AR lo es y se dirá que es invertible si su parte MA lo es.

Este modelo es estacionario e invertible si tanto las raíces de:

$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ como las de $\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ caigan fuera del círculo unidad.

Además, una condición necesaria, aunque no suficiente, para que un modelo mixto autorregresivo-medias móviles de cualquier orden sea estacionario e invertible es la siguiente:

$$\theta_1 + \theta_2 + \dots + \theta_q < 1$$

$$\phi_1 + \phi_2 + \dots + \phi_p < 1$$

Adicionalmente a la estacionariedad e invertibilidad supondremos que las p raíces de $\phi_p(L) = 0$ y las q raíces de $\theta_q(L) = 0$ no se cancelan, esto es, no hay raíces comunes. Bajo estas condiciones, la esperanza matemática y la varianza del modelo ARMA (p, q) se definen así:

Esperanza matemática: $\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots - \phi_p}$

Varianza:

$$\gamma_0 = \frac{\sigma_u^2(1 + \theta_1^2 + \dots + \theta_q^2 - 2\phi_1\theta_1 - 2\phi_2\theta_2 - \dots - 2\phi_h\theta_h)}{1 - \phi_1^2 - \phi_2^2 - \dots - \phi_p^2}$$

Función de autocovarianzas:

$$\gamma_j = \begin{cases} \phi_1\gamma_0 + \phi_2\gamma_1 + \dots + \phi_p\gamma_{p-1} - \theta_1\sigma_u^2 - \theta_2E(\tilde{y}_{t-1}u_{t-2}) - \dots - \theta_qE(\tilde{y}_{t-1}u_{t-q}) \\ \text{para } j = 1 \\ \phi_1\gamma_1 + \phi_2\gamma_0 + \dots + \phi_p\gamma_{p-2} - \theta_2\sigma_u^2 - \theta_3E(\tilde{y}_{t-1}u_{t-3}) - \dots - \theta_qE(\tilde{y}_{t-2}u_{t-q}) \\ \text{para } j = 2 \\ \dots\dots\dots \\ \phi_1\gamma_{q-1} + \phi_2\gamma_{q-2} + \dots + \phi_p\gamma_{p-q} - \theta_q\sigma_u^2 \\ \text{para } j = q \\ \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \dots + \phi_p\gamma_{p-j} \\ \text{para } j > q \end{cases}$$

Función de autocorrelaciones

$$\begin{aligned}
 & \rho_j \\
 = & \left\{ \begin{array}{l}
 \phi_1 + \phi_2\rho_1 + \dots + \phi_p\rho_{p-1} + \frac{1}{\gamma_0} [-\theta_1\sigma_u^2 - \theta_2E(\tilde{Y}_{t-1}u_{t-2}) - \dots - \theta_qE(\tilde{Y}_{t-1}u_{t-q})] \\
 \text{para } j = 1 \\
 \\
 \phi_1\rho_1 + \phi_2 + \dots + \phi_p\rho_{p-2} + \frac{1}{\gamma_0} [-\theta_2\sigma_u^2 - \theta_3E(\tilde{Y}_{t-2}u_{t-3}) - \dots - \theta_qE(\tilde{Y}_{t-2}u_{t-q})] \\
 \text{para } j = 2 \\
 \\
 \dots\dots\dots \\
 \phi_1\rho_{q-1} + \phi_2\rho_{p-2} + \dots + \phi_p\rho_{p-q} + \frac{1}{\gamma_0}\theta_q\sigma_u^2 \\
 \text{para } j = q \\
 \\
 \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \dots + \phi_p\rho_{p-j} \\
 \text{para } j > q
 \end{array} \right.
 \end{aligned}$$

Propiedades del modelo ARMA (p, q)

- **Estacionariedad:** Un proceso $ARMA(p, q)$ es estacionario si y solo si las raíces de su polinomio autorregresivo ($\phi_p(L) = 0$) están todas fuera del círculo unidad del plano complejo.
- **Invertibilidad:** Un proceso $ARMA(p, q)$ es invertible si y solo si las raíces de su polinomio de medias móviles ($\theta_q(L) = 0$) están todas fuera del círculo unidad del plano complejo.
- **Función de Autocorrelación:** La función de autocorrelación no se anula, si bien tiende a cero a medida que aumentan los desfases temporales considerados. Consecuentemente, la memoria de este modelo es infinita. La presencia del término de medias móviles en el modelo $ARMA(p, q)$ afecta la determinación de las q primeras autocorrelaciones ($\rho_1, \rho_2, \dots, \rho_q$) estando el resto de las autocorrelaciones afectadas únicamente por la parte autorregresiva. Concretamente, el análisis de la función de autocorrelación muestra que,

una vez superado el orden de la parte de medias móviles, las autocorrelaciones del modelo $ARMA(p, q)$ se comportan como las de un modelo $AR(p)$.

- **Correlograma Simple:** La representación gráfica de autocorrelación, presentará q autocorrelaciones, hasta el retardo q , en los que influye tanto la parte autorregresiva como el de las medias móviles del modelo y , a continuación, su forma coincide con los del modelo $AR(p)$, es decir, un comportamiento amortiguado hacia cero sin llegar a anularse, si las raíces son reales, o comportamiento sinusoidal, pero también amortiguado hacia cero, si las raíces son complejas.
- **Correlograma Parcial:** La función de autocorrelación parcial no se anula, como consecuencia de que el modelo $ARMA(p, q)$ contiene el modelo $MA(q)$ como un caso especial. La representación gráfica de la función de autocorrelación parcial tendrá también un comportamiento amortiguado hacia cero. Los Modelos $ARMA$ permiten una máxima flexibilidad de representación con muy pocos parámetros, y rara vez los órdenes p y q sobrepasan el valor 2.

3.7. Estacionariedad

Generalmente se asocia a comprobar la estacionariedad en la media de la serie temporal, de tal forma que esta se encuentre cercana al valor de cero, lo cual indica que la serie de tiempo no presenta cambios en su estructura, pero para comprobar la variabilidad de las observaciones de la serie de tiempo, se debe comprobar la estacionariedad la cual debe permanecer constante a lo largo del tiempo que se estudiaron las observaciones de dicha serie.

Si dentro de una serie existe una perturbación transitoria sobre una variable estacionaria esta puede tener efectos puramente transitorios; pueden durar varios períodos, pero sus efectos terminan desapareciendo, es decir, no deambula durante períodos largos de tiempo a un mismo lado de su media muestral, sino que cruza frecuentemente dicho nivel medio. Por el contrario, una perturbación de carácter transitorio sobre una variable no estacionaria tiene efectos permanentes. La función de autocorrelación de una variable no estacionaria converge a cero

muy lentamente, y su serie temporal muestra claramente largos períodos de tiempo en que deambula sin cruzar su nivel medio.

Si se tiene que la función de distribución de probabilidad de una variable no varía con respecto a un desplazamiento en el tiempo significando que la distribución conjunta de m observaciones $y_{t_1}, y_{t_2}, \dots, y_{t_m}$ es similar a la de $y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k}$ para todo m y k retardos. Con esto implica una restricción con cada distribución de y_t es independiente del tiempo y se define la estacionariedad débil. Una serie es débilmente estacionaria si su media es constante e independiente del tiempo, su varianza es finita y constante, y el valor de su covarianza entre dos periodos no depende del tiempo sino de las distancias entre ellos. Por tanto, una serie presenta estacionariedad débil si reúne las siguientes condiciones:

Media: $E(y_t) = \mu$, constante \forall_t

Varianza: $Var(y_t) = E(y_t - \mu)^2 = \sigma^2 = \gamma_0$, constante \forall_t

Covarianza: $Cov(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)] = \gamma_k$, siendo $t \neq t + k$

Un proceso que es usado en el análisis de series de tiempo, es el denominado caminata aleatoria (random walk), que se manifiesta de dos formas, una que contempla rumbo y otro que no. Se puede expresar en términos de la siguiente especificación:

$$y_t = y_{t-1} + \varepsilon_t$$

donde ε_t reúne las características de ser ruido blanco y la variable t_y es una serie cuyo valor es igual al inmediatamente anterior más un choque aleatorio. Si se retrocede dos periodos, se puede expresar $y_{t-1} = y_{t-2} + \varepsilon_{t-1}$ y $y_{t-2} = y_{t-3} + \varepsilon_{t-2}$ para un y dos retardos respectivamente, y sustituyendo progresivamente se tiene:

$$y_t = y_{t-1} + \varepsilon_t = \varepsilon_t + \varepsilon_{t-1} + y_{t-2} = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + y_{t-3} = \dots =$$

Debido a esto, el proceso no tiene un pasado finito pero siempre se inicia desde el valor de y_0 como primer parámetro, teniendo en cuenta esto la expresión anterior se puede escribir como: $y_t = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + y_0$ teniendo $y_t = \sum \varepsilon_j + y_0$ para $j = 1, 2, \dots, t$.

La media se puede expresar como:

$$\mu = e(y_t) = (\sum \varepsilon_j + y_0) = y_0 \sum E(\varepsilon_j) = y_0$$

y la varianza como:

$$var(y_t) = E(y_t - y_0)^2 = E(\sum \varepsilon_j)^2 = \sum E(\varepsilon_j)^2 = t\sigma^2$$

La media de y_t es igual al valor inicial que es constante, pero la varianza no lo es, ya que crece de manera indefinida en la medida que t aumenta, lo cual viola una condición de estacionariedad.

3.7.1. Identificación de Estacionariedad

3.7.1.1. Funciones de autocorrelación simple y autocorrelación parcial

Para diagnosticar si una serie es estacionaria como primera herramienta, se hace el análisis de la función de autocorrelación simple de la serie de tiempo y_t . Este instrumento permite determinar el grado de autocorrelación entre las observaciones individuales que conforman la muestra total de la serie y_1, y_2, \dots, y_T considerando un desfase temporal igual a k . Su representación gráfica se ilustra a través de un correlograma que está conformado por todos los coeficientes muestrales. En particular, para un desfase de k periodos la autocorrelación se define a través de la siguiente expresión:

$$\rho_k = \frac{E[(Y_t - \mu)(y_{y+k} - \mu)]}{\sqrt{E(y_t - \mu)^2 E(y_{t+k} - \mu)^2}} = \frac{Cov(y_t, y_{t+k})}{\sigma_{y_t} \sigma_{y+k}}$$

donde ρ_k es el coeficiente de autocorrelación, μ la media de la serie analizada y el subíndice k representa el orden del rezago. Deduciendo que para una serie estacionaria son iguales las varianzas σ_{y_t} y σ_{y+k} la expresión quedaría de la siguiente forma:

$$\rho_k = \frac{E[(Y_t - \mu)(y_{y+k} - \mu)]}{\sigma_y^2} = \frac{Cov(y_t, y_{t+k})}{\sigma_y^2} = \frac{\gamma_k}{\gamma_0}$$

Siendo γ_k la covarianza al rezago k y γ_0 la varianza de y_t . En el caso particular de que la longitud del rezago fuera $k = 1$, se tendría:

$$\rho_0 = \frac{E(y_t - \mu_y)}{\sigma_y^2} = 1$$

Para inferir sobre los coeficientes poblacionales, es necesario estimar los coeficientes muestrales, los cuales oscilan entre -1 y $+1$.

Entonces, la función de autocorrelación se expresa para el conjunto de la muestra cómo:

$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$, donde:

$$\hat{\gamma}_k = \frac{\sum(y_t - \bar{y})(y_{t+k} - \bar{y})}{T}$$

$$\hat{\gamma}_0 = \frac{\sum(y_t - \bar{y})^2}{t}$$

Es por ello que la función muestral está dada por:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

donde \bar{y} es la media de la población.

Para obtener el correlograma de la función de autocorrelación simple, es recomendable que se cuente con un tamaño de muestra lo suficientemente grande como para evaluar la autocorrelación de orden superior. Es importante destacar que no se podrían efectuar las estimaciones si $k > T - 1$. En la práctica es común que se estimen los coeficientes considerando $k = \frac{n}{4}$

3.7.1.2. Estadísticos de Box – Pierce y Ljung - Box

Para el análisis y determinar si una serie temporal es estacionaria, consiste en verificar si un valor particular de $\hat{\rho}_k$ es estadísticamente significativo. En este caso la hipótesis a contrastar es si $\rho_k = 0$ para una k particular o bien si $\rho_k = 0$ para $k > 0$. Por ello en Bartlett

plantea que si una serie de tiempo se caracteriza por ser ruido blanco (serie puramente aleatoria), entonces los coeficientes de autocorrelación muestral para $k > 0$ están distribuidos aproximadamente de acuerdo con una distribución normal con media 0 y $\sigma = \frac{1}{\sqrt{T}}$, donde T es el número de observaciones y $\rho_k \sim N(0, \frac{1}{T})$.

El estadístico de Box y Pierce evalúa la hipótesis de que todos los coeficientes de autocorrelación son iguales a cero, $h_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$ o bien de que $\rho_k = 0$ para $k > 0$.

El estadístico de prueba está definido por:

$$Q = T \sum_{k=1}^k \rho_k^2$$

donde T es el tamaño de la muestra y k la longitud del rezago. Éste es un estadístico que está distribuido como una χ^2 con k grados de libertad, por lo que $Q(k) \sim \chi(k)$.

Si la Q calculada excede el valor teórico al nivel de significancia seleccionado, se puede rechazar la hipótesis de que todos los ρ_k son iguales a cero, por lo que al menos algunos de ellos deben ser diferentes de cero.

En el análisis de la estacionariedad también es frecuente utilizar el estadístico sugerido por Ljung y Box y que en realidad es una variante de la prueba anterior. La hipótesis que se contrasta es $h_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$ y el estadístico está definido por la siguiente expresión:

$$LB = T(T + 2) = \sum_{k=1}^2 \left(\frac{\rho_k^2}{n - k} \right) \sim \chi^2(k)$$

que al igual que el estadístico Q , sigue una distribución χ^2 con k grados de libertad.

3.7.1.3. Contraste d de Durbin-Watson

El contraste desarrollado por Durbin y Watson permite verificar la hipótesis de no autocorrelación frente a la alternativa de autocorrelación de primer orden bajo un esquema autorregresivo AR (1), teniendo como hipótesis:

$H_0: r = 0$ No existe autocorrelación AR (1)

$H_1: 0 < |r| < 1$ Existe autocorrelación AR (1)

Teniendo el estadístico de prueba:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

A partir de este estadístico se puede interpretar que:

- Si hay autocorrelación positiva las diferencias entre residuos que distan un periodo es muy pequeña por lo que el valor del estadístico d será próximo a cero.
- Si hay autocorrelación negativa los residuos serán prácticamente iguales, pero de signo contrario, su diferencia será por tanto grande y el estadístico será más próximo al límite superior que, como se verá, se establece en cuatro.
- Si no hay autocorrelación, la relación entre los residuos será intermedia y, por tanto, el valor del estadístico experimental también alcanzará un valor intermedio.

Para establecer los límites de variación del estadístico d la fórmula anterior se puede desarrollar obteniéndose una expresión en función del coeficiente de autocorrelación muestral de primer orden para los residuos \hat{r} , el estadístico experimental se puede expresar:

$$d \approx 2(1 - \hat{r})$$

Teniendo en cuenta los límites de variación del coeficiente de correlación empírico $-1 \leq \hat{r} \leq 1$, se puede deducir el rango de variación del estadístico de Durbin-Watson y el signo de la autocorrelación:

$\hat{r} = -1 \rightarrow d \approx 4$ se considera que existe autocorrelación negativa.

$\hat{r} = 0 \rightarrow d \approx 2$ indica ausencia de autocorrelación.

$\hat{r} = 1 \rightarrow d \approx 0$ se puede admitir que existe autocorrelación positiva.

De esta manera se determina que el estadístico experimental tomará valores entre 0 y 4 de tal modo que cuánto más próximo a cero (a cuatro) sea el valor del estadístico d mayor es la evidencia de autocorrelación positiva (negativa). Si el valor del estadístico experimental d es dos, entonces la correlación muestral será nula y por tanto no se detectará un problema de autocorrelación entre las perturbaciones.

Dichos valores (0, 2 y 4) son límites extremos que deben matizarse estableciendo regiones más amplias en las que pueda considerarse si existe o no autocorrelación y, en caso de detectarse, si ésta es positiva o negativa, por tanto, no existe un valor crítico único que permita establecer una regla de decisión. Para solucionar esta dificultad Durbin y Watson encontraron valores para los límites superior (d_u) e inferior (d_L) que permiten tomar decisiones acerca de la presencia o ausencia de autocorrelación.

Éstos señalan el límite superior (d_u) para considerar autocorrelación positiva³⁴ y el límite inferior (d_L) para no rechazar la hipótesis nula y suponer que las covarianzas de las perturbaciones del modelo son nulas y, por tanto, no están autocorrelacionadas.

Si el valor del estadístico d es superior a dos se puede contrastar la hipótesis nula de no autocorrelación frente a la alternativa de autocorrelación negativa. El análisis es similar, pero considerando el valor máximo de 4 como límite para la autocorrelación negativa por tanto los límites anteriores se establecen en los puntos $4 - d_u$ y $4 - d_L$.

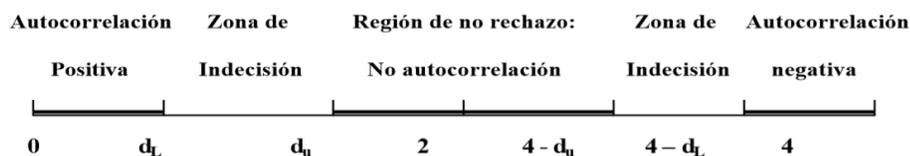


Figura 21: Regiones de contraste del estadístico d de Durbin-Watson

³⁴ Esto es, para valores del estadístico experimental superiores a este límite no se rechaza la hipótesis de ausencia de autocorrelación.

Dando paso a:

$0 < d < d_L$ se rechaza H_0 , existe entonces autocorrelación positiva con un esquema AR (1)

$4 - d_L < d < 4$ se rechaza H_0 , existe autocorrelación negativa con un esquema AR (1)

$d_u < d < 4 - d_u$ no se rechaza H_0 , no existe autocorrelación;

$d_L < d < d_u$ el contraste no es concluyente.

$4 - d_u < d < 4 - d_L$ el contraste no es concluyente.

Estos límites dependen del tamaño de la muestra (n) y del número de regresores del modelo (k). Las tablas originales sirven para muestras entre 15 y 100 observaciones y un máximo de 5 regresores. Años más tarde, Savin y White publicaron unas tablas más completas que incluyen tamaños de muestra superiores: $5 < n < 200$ y hasta 20 regresores.

3.8. Raíz Unitaria

Un proceso estocástico lineal tiene una raíz unitaria si el valor de la raíz de la ecuación característica del proceso es igual a 1, por lo tanto, tal proceso es no estacionario. Si las demás raíces de la ecuación característica se encuentran dentro del círculo unitario, es decir, tienen un valor absoluto menor a uno, entonces la primera diferencia del proceso es estacionaria. Así, se dice que una serie I (1) tiene una raíz unitaria y que una serie I(d) tiene d raíces unitarias.

Las pruebas de raíz unitaria que se desarrollan dependen del modelo básico que genera la serie, lo importante es contrastar la hipótesis nula (H_0), la cual propone que la serie tiene una raíz unitaria ($\alpha = 1$). La prueba de hipótesis se hace con el valor de t estadística, si el resultado es positivo o está por debajo del valor crítico, se acepta H_0 . Este resultado se confirma un poco más intuitivamente cuando se observa que la probabilidad de la prueba es mayor al 95% de confianza, lo cual se advierte en el valor de probabilidad, que debe ser mayor a 0.05. En tal caso se sabe de la existencia de al menos una raíz unitaria, la obtención de una probabilidad mayor a 0.05 sería prueba de que la serie tiene dos raíces unitarias y que sería necesario sacar otra diferencia para tenerla estacionaria.

3.8.1. Pruebas de Raíz Unitaria

Dentro de las pruebas formales que existen para determinar la presencia de raíz unitaria en un proceso estocástico, mencionamos las sugeridas por Dickey y Fuller (1979), la Dickey Fuller Aumentada que se probó en Said y Dickey (1984), la prueba Phillips-Perron (1988).

3.8.1.1. Prueba Dickey-Fuller (DF)

La prueba básica DF se aplica generalmente a las siguientes regresiones:

$$\begin{aligned}\Delta y_t &= \alpha y_{t-1} + u_t & \Delta y_t &= \phi_1 + \alpha y_{t-1} + u_t \\ \Delta y_t &= \phi_1 + \phi_2 t + \alpha y_{t-1} + u_t\end{aligned}$$

Donde Δ es el operador de la primera diferencia y $\alpha = (\phi - 1)$. La primera ecuación representa una caminata aleatoria simple, la segunda una caminata aleatoria con variación ϕ_1 y la tercera añade un componente de tendencia t . En cualquiera de las tres especificaciones si la hipótesis nula $H_0: \alpha = 0$ no se rechaza y se concluye que la serie objeto de estudio presenta una raíz unitaria.

Cabe destacar que uno de los supuestos fundamentales de la prueba DF es que el término de error es un proceso de ruido blanco. En este sentido es que se recomienda analizar el estadístico Durban- Watson para verificar si se cumple con el supuesto de no correlación serial. En caso de que las observaciones u_t estén autocorrelacionadas, se recomienda desarrollar la prueba Dickey-Fuller-Aumentada (DFA) que corrige por esta deficiencia.

3.8.1.2. Prueba Dickey-Fuller Aumentada (DFA)

Esta prueba consiste en estimar la siguiente regresión:

$$\Delta y_t = \phi_1 + \phi_2 t + \alpha y_{t-1} + \sum_{i=1}^p \delta_i \Delta_{t-i} + \varepsilon_i$$

La característica de esta técnica es que incorpora como regresores adicionales las diferencias rezagadas de la variable dependiente y se puede aplicar a cualquiera de los tres modelos en la prueba Dickey-Fuller. El propósito de incluir los rezagos de las diferencias es precisamente controlar por el problema de autocorrelación en el término de error, del cual se alertó en la especificación DF. En este caso, se evalúa la hipótesis $H_0: \alpha = 0$ contra $H_0: \alpha < 1$. Es importante señalar que se debe especificar el número de rezagos del término Δ_{t-i} .

Si la serie original presenta tendencia, es recomendable incluir en la regresión el término independiente y de tendencia lineal, si la serie no la refleja y tiene un valor medio distinto de cero, es conveniente incluir el término constante. Si por el contrario la serie fluctúa en torno al valor medio cero, no es imprescindible que se incorpore ningún regresor adicional en la ecuación (Mogni, 2013).

3.8.1.3. Prueba Phillips-Perron (PP)

La prueba PP se basa en un método no-paramétrico para controlar la presencia de correlación serial de orden elevado en una serie de un proceso AR(p) para $p > 1$. En particular, a diferencia de la prueba DFA que resuelve el problema por medio de incluir términos de rezago de la primera diferencia de la variable dependiente, la PP incluye un factor de corrección estimando un estadístico de prueba diferente al utilizado en la prueba DFA. Si se considera el proceso autorregresivo:

$$y_t = \mu + \rho y_{t-1} + u_t$$

Una prueba en el espíritu de la DFA utiliza un estadístico de prueba como $T(\hat{\rho} - 1)$ mientras que la prueba PP emplea el estadístico $Z \hat{\rho}_\mu = T(\hat{\rho} - 1) - CF$, donde CF es un factor de corrección. De tal forma que el estadístico modificado incorpora el posible problema de autocorrelación. Al igual que en las dos pruebas anteriores, aquí se debe determinar la prudencia de incluir una constante o tendencia determinística.

3.8.1.4. Prueba KPSS

A diferencia de las pruebas DF, DFA, y PP que toman como hipótesis nula la no estacionariedad de las series, esto es $H_0: \alpha = 0$. la prueba KPSS considera como hipótesis nula la estacionariedad de la serie, así que la hipótesis nula se refiere a $H_0: \alpha < 1$.

3.9. Modelo Autorregresivo Integrado de Media Móvil ARIMA (p, d, q)

Los modelos de series de tiempo se basan en el supuesto de estacionariedad, esto es, la media y la varianza para una serie de tiempo son constantes en el tiempo y la covarianza es invariante en el tiempo. Algunas series de tiempo y en especial las series económicas no son estacionarias, porque pueden ir cambiando de nivel en el tiempo o sencillamente la varianza no es constante en el tiempo, a este tipo de proceso se les considera procesos integrados, se debe diferenciar una serie de tiempo d veces para hacerla estacionaria y luego aplicarla a esta serie diferenciada un modelo ARMA(p, q) se dice que la serie original es ARIMA(p, d, q), es decir, una serie de tiempo autoregresiva integrada de media móvil. Donde p denota el número de términos autorregresivos, del número de veces que la serie debe ser diferenciada para hacerla estacionaria y q el número de términos de la media móvil invertible. Su expresión algebraica es:

$$X_t^d = c + \phi_1 X_{t-1}^d + \dots + \theta X_{t-p}^d + \theta_1 \varepsilon_{t-1}^d + \theta_2 \varepsilon_{t-2}^d + \dots + \theta_q \varepsilon_{t-q}^d + \varepsilon_t^d$$

Donde X_t^d es la serie de las diferencias de orden d , ε es un proceso de ruido blanco, θ y ϕ son los parámetros del modelo.

La construcción de los modelos ARIMA (p, d, q) es un proceso iterativo en el cual se pueden distinguir cuatro etapas:

1. **Identificación:** Utilizando los datos ordenados cronológicamente se pretende sugerir un modelo ARIMA (p, d, q). El objetivo es determinar los valores p, d, q que sean apropiados para reproducir la serie de tiempo. En esta etapa es posible identificar más de un modelo candidato que pueda describir la serie.
2. **Estimación:** Considerando el modelo apropiado para la serie de tiempo se realiza inferencia sobre los parámetros.
3. **Validación:** Se realizan contraste de diagnóstico de los residuos, para validar si el modelo seleccionado se ajusta a los datos, si no es así, se debe de escoger el próximo modelo candidato y repetir los pasos anteriores.
4. **Predicción:** Una vez seleccionado el mejor modelo candidato ARIMA (p, d, q) se pueden hacer pronósticos en términos probabilísticos de los valores futuros.

Cabe destacar que el proceso por ser autoregresivo, tiene tendencia a la repetición simultánea de los parámetros que inciden en la construcción del modelo, y además otorga una mejor aproximación en los resultados cuando este ha sido escogido como el mejor modelo de predicción, por tanto, su pronóstico para valores futuros será más acertado y confiable para cualquier serie de tiempo que cumpla con las cuatro etapas.

3.10. Modelos SARIMA

Los modelos ARIMA (p, d, q) son no estacionarios, también se cuenta con los modelos de series de tiempo estacionales autorregresivos integrados de media móvil ARIMA(p,d,q)(P,D,Q) (o SARIMA(p,d,q)(P,D,Q)³⁵). Aunque la metodología para construir estos procesos es igual, su utilidad mayormente es cuando la serie de tiempo observada tiene intervalos de observación menores a un año debido a que es más frecuente que éstas tengan variaciones o patrones sistemáticos cada cierto periodo. Esta estacionalidad debe ser captadas e

³⁵ Autorregresive Integrated Movile Average.

incluida dentro de la estructura del modelo a construir. Este tipo de procesos tiene las siguientes características:

- Contiene una componente $ARIMA(p,d,q)$ que modela la dependencia regular, que es la dependencia asociada a observaciones consecutivas.
- Contiene una componente $ARIMA(P,D,Q)$ que modela la dependencia estacional, que está asociada a observaciones separadas por periodos.

Si la serie Y_t tiene una componente con período s es posible eliminarla diferenciando con un rezago de orden s , es decir transformando Y_t a:

$$W_t = (1 - L^s)^D Y_t = \Delta_s^D Y_t, \quad D = 0,1,2$$

Y se busca una estructura $ARMA$ para W_t

En principio la parte estacional se puede modelizar de la misma forma que la parte regular. La identificación se realiza estudiando la ACF y la PACF en los retardos estacionales. Cuando una serie tiene parte regular y parte estacional, la FAS presenta una interrelación entre ellas. Básicamente la ACF y PACF de los modelos estacionales presentan las siguientes características:

- **ACF:** en los primeros rezagos se observa únicamente la parte regular, mientras que la parte estacional se observa en los retardos del orden del periodo. También se puede observar la interacción entre ambos, mediante una componente regular en los retardos cercanos a los estacionales.
- **PACF:** los primeros rezagos se observa únicamente la parte regular, mientras que la parte estacional se observa en los retardos del orden del periodo y se puede observar la interacción entre ambos, mediante una componente regular en los retardos cercanos a los estacionales (positiva o negativa a la derecha y positiva a la izquierda).

3.10.1. Modelo Multiplicativo SARIMA

Sin embargo, al encontrar la estructura ARMA y existe la estacionalidad, pueden existir estructuras *ARIMA intra* período y entre los s períodos. Las estructuras *ARIMA intra* se refieren a modelos de la forma:

$$\varphi_p(L)\Delta^d Y_t = \theta(L)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2)^{36}$$

En otro caso, los *ARIMA entre* se refiere a los modelos $ARIMA(P, D, Q)_s$ de la forma:

$$\Phi_P(L^s)\Delta_s^D Y_t = \Theta_Q(L^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

Y si se define el modelo $ARIMA(p, d, q)(P, D, Q)_s$ como:

$$\varphi_p(L)\Phi_P(L^s)\Delta^d \Delta_s^D Y_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t$$

Donde $X_t = \Delta^d \Delta_s^D Y_t = (1 - L)^d (1 - L^s)^D Y_t$ es la transformación (filtro lineal) que elimina la tendencia y la estacionalidad dejando una estructura $ARMA(p + P_s, q + Q_s)$. Teniendo la idea es que el modelo:

$$\varphi_p(L)\Phi_P(L^s)X_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2)$$

es similar o equivalente al modelo:

$$Y_t = a + bt + \sum_{j=1}^{s-1} \delta_j I_j(t) + X_t$$

En ambos la operación inicial de calcular la serie diferenciada X_t es similar a eliminar una tendencia y estacionalidad aleatorias en y determinísticas en el modelo anterior, y luego ajustar un modelo ARMA a los residuos estructurales.

Si se supone que se tiene un modelo $SARIMA(1,1,0)(1,1,0)_{12}$ en donde se tiene $p = 1, d = 1, q = 0, P = 1, D = 1, Q = 0, s = 12$.

³⁶ RB se refiere al Ruido Blanco de la Serie,

sustituyendo se tiene:

$$(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})(1 - L)(1 - L^{12})(Y_t - \mu) = \varepsilon_t$$

$$(1 - \varphi_1 L)(1 - \varphi_{12} L^{12})X_t = \varepsilon_t$$

con

$$X_t = (1 - L)(1 - L^{12})Y_t = Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

Cuando se desarrollan los polinomios autoregresivos del modelo se obtiene:

$$(1 - \varphi_1 L - \varphi_{12} L^{12} + \varphi_1 \varphi_{12} L^{12})X_t = \varepsilon_t$$

$$X_t = \varphi_1 X_{t-1} - \varphi_{12} X_{t-12} + \varphi_1 \varphi_{12} X_{t-13} + \varepsilon_t$$

Este modelo se puede considerar como un AR (13) con restricciones sobre los coeficientes. El parámetro $\varphi_1 \varphi_{12}$ se estima como un solo valor φ_{13} y puede ser no significativo. En el caso que sea no significativo el modelo se dice sin interacción, es decir:

$$(1 - \varphi_1 L - \varphi_{12} L^{12})X_t = \varepsilon_t$$

Si se define como un modelo multiplicativo de la forma $SARIMA(0,1,1)(0,1,1)_{12}$ es un modelo utilizado con frecuencia dado por:

$$(1 - L)(1 - L^{12})Y_t = (1 + \theta_1)(1 + \theta_{12} L^{12})\varepsilon_t, \quad |\theta_1| < 1, |\theta_{12}| < 1$$

Al utilizar $W_t = (1 - L)(1 - L^{12})Y_t$ y examinar la función de autocovarianza de W_t se observa que $W_t = (1 + \theta_1 L)(1 + \theta_{12} L^{12})\varepsilon_t$ es un $MA(13)$. Se puede comprobar que la función de W_t tiene valores diferentes de cero sólo en los rezagos 1,11,12,13.

La función de autocovarianza de W_t es:

$$\gamma_0 = (1 + \theta_1^2)(1 + \theta_{12}^2)\sigma^2$$

$$\gamma_1 = \theta_1(1 + \theta_{12}^2)\sigma^2$$

$$\gamma_{11} = \theta_1\theta_{12}^2\sigma^2$$

$$\gamma_{12} = \theta_{12}(1 + \theta_1^2)\sigma^2$$

$$\gamma_{11} = \theta_1\theta_{12}\sigma^2$$

3.10.2. Pruebas de Raíz Unitaria Estacional

A partir de la definición de los modelos SARIMA hay dos posibles modelos para una serie con componente estacional:

1. Modelos con componente determinística, es decir, el modelo de descomposición.
2. Modelos no estacionarios integrados estacionalmente, es decir, $SARIMA(p, d, q)(P, D, Q)_s$, con $d \neq 0$ y $D \neq 0$.

Sin embargo, un modelo de la forma $SARIMA(p, 0, q)(P, 0, Q)_s$, equivalente a un $ARMA(p + P, q + Q)$, podría ser tanto estacionario como mostrar características estacionales. Por tanto, un tercer modelo es el anterior, que se denomina un modelo “estacionario estacional”.

El modelo con componente determinística no es estacionario, pero si se elimina la componente de tendencia, el proceso filtrado resulta la suma de un proceso determinístico periódico y otro estacionario en covarianza, concretamente $Y_t - \hat{T}_t = \hat{S}_t + \hat{\varepsilon}_t$. Este modelo también se denomina “estacionario estacional”.

Se necesitan técnicas para discriminar cuál modelo aplicar. Se pueden implementar criterios como la validación cruzada y la comparación del $MAPE$ ³⁷ para escoger el mejor modelo para pronosticar, en la teoría de series de tiempo se ha dado mucho desarrollo en las pruebas de

³⁷ Error Porcentual Absoluto Medio (MAPE o Mean Absolute Percentage Error).

raíces unitarias, como las pruebas para detectar raíces unitarias estacionales utilizando pruebas de como las de Hylleberg, indicada por HEGY, y la prueba Canova-Hansen, indicada por CH.

La prueba HEGY

La prueba HEGY es una generalización de la prueba aumentada de Dickey-Fuller. Permite decidir cuáles de las frecuencias estacionales corresponden a raíces estacionales significativas. En caso de detectarse se concluye que la serie en cuestión está integrada estacionalmente.

Los estadísticos para detectar las raíces unitarias estacionales son tipo t-Student y F, correspondientes a los estadísticos para pruebas de hipótesis sobre los parámetros de la regresión. Se asume que la serie Y_t está generada por un proceso $AR(p)$, es decir, se cumple:

$$\varphi(L)Y_t = \varepsilon$$

para ε_t ruido blanco.

Se asume que están dados p números complejos $\theta_j \in \mathbb{C}, j = 1, \dots, p$ en los cuales $\varphi(\theta_j)$ es un número finito. Estos p números son las p raíces estacionales. El valor p es el período indicado anteriormente por s . Adicionalmente se definen las cantidades auxiliares:

$$\begin{aligned} \delta_k(z) &= 1 - \frac{z}{\theta_k} \\ \Delta(z) &= \prod_{k=1}^p \delta_k(z), \\ \lambda_k &= \frac{\varphi(\theta_k)}{\prod_{j \neq k} \delta_j(\theta_k)} \end{aligned} \quad (3.1)$$

Entonces se cumple por una identidad (debida a Lagrange), que:

$$\varphi(z) = \sum_{k=1}^p \frac{\lambda_k \Delta(z)(1 + \delta_k(z))}{\delta_k(z)} + \Delta(z)\varphi^*(z) \quad (3.2)$$

donde $\varphi^*(z)$ es un polinomio posiblemente infinito o racional, cumpliendo:

$$\varphi(\theta_k) = 0 \Leftrightarrow \lambda_k = 0$$

por lo tanto, para determinar si θ_k es raíz estacional se debe comprobar $\lambda_k = 0$. Pero las λ_k se pueden asimilar a los coeficientes en una regresión lineal múltiple que se define a partir de la identidad (3.2). Y, por tanto, probar $\lambda_k = 0$ se puede hacer, en principio, mediante una prueba t-Student de significación de parámetros o con una prueba F.

Prueba Canova-Hansen

Los modelos de Descomposición con errores ARMA y los SARIMA se consideran, inicialmente, equivalentes. Pero es difícil saber a priori cuál posibilidad [modelo] produce la mejor descripción de la serie. Algunas series muestran cambios en los patrones estacionales.

La prueba CH se basa en esta idea. La inestabilidad estructural de la componente estacional se toma como una característica equivalente a la existencia de una raíz unitaria estacional y opuesta a estacionario estacional.

El modelo que asume la prueba es un modelo sin tendencia y con componente estacional, descrita mediante funciones trigonométricas, con el primer rezago y como variable exógena es:

$$Y_t = bY_{t-1} + \sum_{j=1}^{s/2} \gamma_j \cos\left(\frac{2\pi jt}{s}\right) + \gamma_j^* \sin\left(\frac{2\pi jt}{s}\right)$$

donde se asume que s es par, $t = 1, 2, \dots$

Cuando $j = s/2$ entonces $\sin\left(\frac{2\pi jt}{s}\right) = \sin(\pi t) = 0$. Luego, el número de parámetros en el modelo anterior es $s - 1$, el mismo del modelo con variables indicadoras. Los coeficientes se colocan en vectores:

$$\gamma = (\gamma_1, \dots, \gamma_{s/2})', \gamma^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)', \gamma^* = (\gamma_1^*, \dots, \gamma_{s/2-1}^*)'$$

La hipótesis nula en la prueba CH es que los coeficientes γ, γ^* no cambian con t , en cuanto varían con t según un modelo de marcha aleatoria:

$$\gamma_t = \gamma_{t-1} + u_t, \gamma_t^* = \gamma_{t-1}^* + v_t$$

donde u_t, v_t son vectores iid³⁸, independientes de media cero.

Esta hipótesis alterna es la forma en la que se establece la presencia de raíces unitarias estacionales, equivalente a inestabilidad estructural en la componente estacional, las cuales se pueden escribir como sigue:

$$H_0: \gamma_t \equiv \gamma, \gamma_t^* \equiv \gamma^*$$

$$H_a: \gamma_t = \gamma_{t-1}^* + u_t, \gamma_t^* = \gamma_{t-1} + v_t$$

3.11. Validación de los modelos Univariantes

3.11.1. Medición del error de pronóstico

En las series de tiempo, se suele utilizar notación matemática para referirse a cada periodo específico, la letra Y se usa para representar una variable de serie de tiempo, a menos que haya más de una variable. El periodo asociado con una observación se identifica como un subíndice. De modo que Y_t se refiere al valor de la serie de tiempo en el periodo t . De igual forma se distinguirá un valor real de la serie (Y) y un valor pronostico \hat{Y} , por lo que \hat{Y}_t será el valor pronóstico de Y_t .

Existen diversos métodos cuya finalidad es resumir los errores generados por una técnica específica de pronósticos. La mayoría de estas medidas son el promedio de alguna función de la diferencia entre su valor real y su valor pronosticado. Estas diferencias se denominan residuos, y se expresan como:

$$e_t = \hat{Y} - \hat{Y}_t$$

³⁸ Independiente e idénticamente distribuido (iid).

Un método para evaluar una técnica de pronósticos usa la suma de los errores absolutos. El MAD ³⁹ mide la exactitud del pronóstico, promediando las magnitudes de los errores del pronóstico (los valores absolutos de los errores). MAD está en las mismas unidades que la serie original, y proporciona un tamaño promedio de los “errores” sin importar la dirección.

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

donde:

Y_t es el valor real en el periodo t

\hat{Y}_t es el valor del pronóstico en el periodo t

n el número de periodos

El MSE ⁴⁰ es un método para evaluar una técnica de elaboración de pronósticos. Cada error o residuo se eleva al cuadrado; luego éstos se suman y se dividen entre el número de observaciones. Este enfoque sanciona errores grandes en la elaboración de pronósticos, ya que los errores están elevados al cuadrado, lo cual es importante porque una técnica que produce errores moderados quizá sea preferible a una que usualmente tenga pequeños errores, pero ocasionalmente produce errores extremadamente grandes.

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

La raíz cuadrada del MSE, o la raíz cuadrada del error cuadrado medio (RMSE), también se usa para evaluar los métodos de elaboración de pronósticos. Tanto la RMSE como la MSE sancionan los errores grandes, pero tienen las mismas unidades de la serie que se está pronosticando, de modo que su magnitud se interpreta con mayor facilidad. La RMSE se presenta a continuación.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

³⁹ La desviación media absoluta.

⁴⁰ El error cuadrático medio.

El error porcentual absoluto medio (MAPE) se calcula obteniendo el error absoluto de cada periodo, dividiendo éste entre el valor real observado en ese periodo y promediando estos errores porcentuales absolutos. El resultado final se multiplica después por 100 y se expresa como porcentaje. Este enfoque es útil cuando el error relativo al tamaño respectivo del valor de la serie de tiempo es importante, para la evaluación de la exactitud del pronóstico y resulta útil cuando los valores Y_t son grandes. Además no tiene unidades de medición y sirve para comparar la exactitud de la misma técnica o de otras técnicas en dos series completamente diferentes.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{|Y_t|}$$

donde Y_t es diferente de cero.

En ocasiones es necesario determinar si el método para pronosticar está sesgado (con pronósticos consistentemente altos o bajos). En estos casos, se usa el error porcentual medio (MPE), el cual se calcula obteniendo el error en cada periodo, dividiendo éste entre el valor real de ese periodo y luego promediando estos errores porcentuales. El resultado usualmente se multiplica por 100 y se expresa como un porcentaje. Si el enfoque del pronóstico no tiene sesgo, el MPE producirá un resultado que esté cercano a cero. Si el resultado es un porcentaje negativo grande, el método de elaboración del pronóstico está sobrestimando consistentemente. Si el resultado es un porcentaje positivo grande, el método de elaboración del pronóstico está subestimando consistentemente. El MPE está dado por:

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{(Y_t - \hat{Y}_t)}{Y_t}$$

La decisión para usar una técnica de elaboración de pronósticos específica se basa, en parte, en la determinación de si la técnica producirá errores en el pronóstico que se consideren lo suficientemente pequeños. En efecto, se espera que una buena técnica de elaboración de pronósticos produzca errores relativamente pequeños de manera consistente.

El error de escala absoluta de media (MASE) es una media de error para pronósticos de diferente escala y resulta la mejor medida disponible. Se expresa de la siguiente forma:

$$MASE = \frac{1}{n} \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{\left(\frac{1}{n-1} \sum_{i=2}^n |Y_i - \hat{Y}_{i-1}| \right)} = \text{Media}|q_t|$$

donde:

n es el número de muestras

Y_t es el valor actual

\hat{Y}_t es estimación del modelo t

La medida MASE es simétrica y resistente a los valores atípicos y la división por cero ocurre solamente en un caso trivial donde todos los valores en el registro de series de tiempo son iguales.

Al utilizar esta medida se debe tener en cuenta que si $MASE > 1$ implica que el pronóstico actual no es adecuado que un método de pronóstico de punto de referencia primitivo calculado como muestra. Caso contrario si $MASE < 1$ implica que el comportamiento del pronóstico actual es mejor que el método primitivo.

Y en todo caso se puede usar los valores MASE para comparar distintos métodos de pronóstico. Entre más bajo sea el valor del MASE, es decir, entre más bajo sea el error de pronóstico absoluto relativo, mayor es la eficiencia del método.

A parte de los errores anteriores, se puede mencionar el Error absoluto medio (MAE), el cual es el promedio de los valores absolutos de los errores calculados, definido por:

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

donde: n es el número de muestras, y_t es el valor actual y \hat{Y}_t es la estimación de Y_t

3.11.2. Criterios de selección del modelo

Es posible seleccionar los modelos ARIMA examinando una gráfica de la serie y al establecer una correspondencia entre los patrones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra con los patrones teóricos conocidos de los procesos ARIMA. Sin embargo, existe cierta subjetividad implicada en este procedimiento, y es posible que dos o más modelos iniciales sean congruentes con los patrones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Más aún, después de la estimación y verificación, ambos modelos pueden representar adecuadamente los datos. Si los modelos contienen el mismo número de parámetros, el modelo con menor error cuadrático medio (s^2), es por lo general el preferido. Si los modelos contienen diferentes números de parámetros, el principio de parsimonia conduce a la selección del modelo más sencillo. Sin embargo, el modelo con más parámetros puede tener un error cuadrático medio considerablemente menor.

Se ha desarrollado un método para la selección del modelo que considere tanto el ajuste del modelo como el número de parámetros. El criterio de información de Akaike o *AIC*, selecciona el mejor modelo de un grupo de modelos candidatos que minimiza.

$$AIC = \ln \hat{\sigma}^2 + \frac{2}{n}r$$

donde

\ln es el logaritmo natural

$\hat{\sigma}^2$ es la suma residual de cuadrados dividida entre el número de observaciones

n es el número de observaciones (residuos)

r es el número total de parámetros (incluyendo el término constante) en el modelo ARIMA

El criterio de información bayesiano que desarrolló Schwarz (1978), o *BIC*, selecciona el modelo que minimiza

$$BIC = \ln \hat{\sigma}^2 + \frac{\ln n}{n}r$$

El segundo término tanto en el criterio *AIC* como en el criterio *BIC* es un factor que “penaliza” la inclusión de parámetros adicionales en el modelo. Como el criterio *BIC* impone una penalización mayor para el número de parámetros que el criterio *AIC*, el uso del *BIC* mínimo para la selección del modelo tendrá como resultado un modelo cuyo número de parámetros no será mayor que el seleccionado por el *AIC*. A menudo los dos criterios producen el mismo resultado.

El *AIC* y el *BIC* deben considerarse como procedimientos adicionales para ayudar en la selección del modelo. No se deben usar como sustitutos de una revisión cuidadosa de las autocorrelaciones de la muestra y las autocorrelaciones parciales. (Díaz, 2011)

Capítulo 4. Marco Metodológico

En el proceso de investigación para crear modelos estadísticos utilizando Redes Neuronales y Series de Tiempo para el pronóstico de los niveles del reservorio de la presa Hidroeléctrica Cerrón Grande de El Salvador, se cuenta con una metodología que consta de las siguientes partes:

4.1. Tipo de Investigación

Para conocer los niveles que posee el reservorio la Presa del Cerrón Grande, mediante los modelos de predicción estadísticos de series de tiempo y de las redes neuronales artificiales, se define el tipo investigación utilizada es de carácter predictiva. La investigación predictiva es el que tiene como propósito prever o anticiparse a situaciones futuras observando eventos durante cierto tiempo, en donde se analiza y describe para lograr presagiar el comportamiento o la tendencia del evento estudiado.

Desde el punto de vista estadístico no se trata de poner a prueba hipótesis, sino de construir modelos predictivos; para ello se aplican técnicas específicas, como las ecuaciones estructurales, las series de tiempo y el análisis de supervivencia; así como la minería de datos.

Para construir un modelo predictivo se debe de definir claramente la variable endógena (variable a predecir) y las exógenas (variables predictivas), de modo que no se somete a contraste su pertinencia o su participación en la ocurrencia del evento adverso, sino se usa con fines predictivos. Pero, cuando se centra en predecir la ocurrencia de un evento en función al tiempo, para lo cual necesita información del pasado para predecir el futuro, esta predicción se hace con un determinado nivel de confianza que se expresará mediante un intervalo de confianza. (Hurtado J. , 2000)

4.2. Población y Muestra

- **Población**

La población utilizada para realizar pronósticos de los niveles del reservorio en la Presa Hidroeléctrica Cerrón Grande de El Salvador, son las mediciones del nivel en metros sobre el nivel del mar (m s. n. m.), agrupados en meses comprendidos desde enero del 2000 hasta diciembre del 2018, haciendo un total de 216 meses.

- **Muestra**

Debido al almacenaje de los datos son registrados y manipulados por la CEL, no es necesario establecer una muestra, por lo que es conveniente trabajar con la población de los datos, lo cual indica que se trabaja con una población muestral.

4.3. Recolección de los datos

Los datos utilizados son obtenidos por parte de la Comisión Ejecutiva Hidroeléctrica del Río Lempa (CEL), los cuales consiste en las mediciones mensuales de los niveles del reservorio en la Presa Hidroeléctrica Cerrón Grande de El Salvador. Las mediciones que presenta la base de datos corresponden a los últimos 18 años presentados de forma mensual que comprenden desde enero del año 2000 hasta el diciembre de 2018.

4.4. Procesamiento de los datos

Para procesar los datos proporcionados por la Comisión Ejecutiva Hidroeléctrica del Río Lempa (CEL), se utilizan dos técnicas: Las redes neuronales artificiales y las series de tiempo, con las cuales se busca encontrar los modelos de pronósticos para el nivel del reservorio de la Presa Cerrón Grande, para cada una de las técnicas se ocuparon metodologías para la construcción de dichos modelos.

4.4.1. Diseño del Modelo de Redes Neuronales

Para realizar la construcción del modelo de pronóstico con la técnica de redes neuronales, se toma en cuenta los siguientes criterios:

4.4.1.1. Selección de la Variable

Determinar y seleccionar qué variables de entrada son importantes en el pronóstico de los niveles del reservorio de la Presa, sin embargo, se diseña una red neuronal univariante por lo que se toma la variable nivel del embalse del Cerrón Grande, detallándose como sigue:

- **Variable a Predecir:** Nivel del reservorio de la presa Hidroeléctrica Cerrón Grande de El Salvador.
- **Tipo de Variable:** Cuantitativa.
- **Unidad:** Metros sobre el nivel del mar (m s. n. m.).

4.4.1.2. Conjunto de Entrenamiento y Prueba

- **Conjunto de Entrenamiento**

Es el conjunto de datos que ayuda a la red neuronal para aprender los patrones que presentan los datos (calcular la obtención de los pesos de la red) y le enseña a la red cómo sobrellevar diferentes características, ajustando los coeficientes de acuerdo con su probabilidad de minimizar los errores en sus resultados. En esta ocasión, se toma aproximadamente el 70% de los datos para incluirse en el conjunto de entrenamiento, lo que representa 160 registros de 228 disponibles.

- **Conjunto de Prueba**

Corresponde a los datos restantes del conjunto de entrenamiento. Este conjunto de datos se utiliza para evaluar la precisión de la red, el cual está compuesto por 68 registros de los datos que se poseen.

El conjunto de entrenamiento y de prueba fue mediante la selección aleatoria de una muestra de 70% y de 30%, pues dentro del machine learning se toma en cuenta la regla general de la división 2/3 para el training y 1/3 para el test, esto es debido a que si se introduce toda la cantidad de información con la que se cuenta, no habría manera de evaluar el error que está

cometiendo, el modelo pues no tendrían datos fiables contra los que contrastar, de esa forma se evita contar con un modelo sobreajustado⁴¹.

4.4.1.3. Selección de la Arquitectura de Redes Neuronales

Para desarrollar una red neuronal de predicción, en la mayoría de los casos se utilizan las redes con algoritmo de entrenamiento Back propagation.

Para crear una red neuronal se debe seleccionar:

- **Número de neuronas de entrada**

Estas corresponden a las mediciones registradas por la CEL en el intervalo de un año de forma diaria, lo que en la base de datos que equivale a 12 meses. En este caso la cantidad de neuronas de entrada son 12 datos.

- **Número de Capas Ocultas**

Las capas ocultas dan a la red la habilidad de generalizar, y en la práctica las redes neuronales con una o dos capas ocultas son las más utilizadas y han tenido un buen desempeño. Para la red utilizada en esta investigación solo se requiere de una capa oculta. Para este caso se selecciona dos capas ocultas, en donde tiene 9 y 6 neuronas respectivamente.

- **Número de Neuronas de Salida**

El número de neuronas de la capa de salida de la red será una, porque solo se intenta predecir el valor de la variable nivel para el siguiente periodo.

⁴¹ Un modelo va a estar sobreajustado cuando se desempeña bien con los datos de entrenamiento, pero su precisión es notablemente más baja con los datos de evaluación; esto se debe a que el modelo ha memorizado los datos que ha visto y no pudo generalizar las reglas para predecir los datos que no ha visto. De aquí también la importancia de siempre contar con dos conjuntos de datos distintos, uno para entrenar el modelo y otro para evaluar su precisión; pues si se utiliza el mismo conjunto de datos para las dos tareas, no habría forma de determinar cómo el modelo se comporta con datos que nunca ha visto.

- **Función de Transferencia**

Con esta función se busca prevenir a las salidas de alcanzar valores muy elevados, los cuales pueden paralizar la red y detener el entrenamiento de la misma. Como función de transferencia se usa la función sigmoideal y se usa la función tangente hiperbólica.

- **Topología de la red**

Se opta por el uso de dos capas ocultas, la cual es suficiente para asegurar la capacidad de generalización de la red. Por lo que la red tendría un total de 3 capas, la capa de entrada, la capa oculta y la capa de salida.

4.4.1.4. Criterios de parada de la red neuronal

El criterio de parada se aplica cuando los valores de los pesos reduzcan la función del error (diferencia cuadrática entre la salida deseada y la salida de la red neuronal), y este proceso se realiza por el método del gradiente descendente. La idea del método es realizar un cambio en los pesos proporcional a la derivada de la función de error respecto al peso para cada patrón.

4.4.2. Diseño del Modelo de Series de tiempo

Para poder desarrollar el modelo para la predicción, se hace uso de la metodología Box-Jenkins, que fue desarrollada en el año 1970 buscando realizar una pequeña revolución en el enfoque del análisis de series temporales, para poder explicar su estructura y su evolución en el futuro, a partir de eso sus procedimientos se utilizan ampliamente desde entonces en diferentes ramas de la ciencia, conociéndose como modelos ARIMA y también como modelos Box-Jenkins. A continuación, se detalla el procedimiento utilizado para realizar la predicción de acuerdo a esta metodología

4.4.2.1. Análisis inicial de la serie temporal

Al inicio se debe de plantear a búsqueda de características importantes dentro de la serie, de esta forma es necesario detectar si existe tendencia (creciente o decreciente), la existencia de

ciclos, presencia de valores atípicos. posterior a ello es conveniente plantear la construcción del comportamiento de la serie de tiempo, mediante un gráfico de secuencia, que es empleado para observar cómo es el comportamiento a través del tiempo.

Después de la construcción del gráfico, se busca la presencia de estacionalidad que se observa cuando ocurre un patrón sistemático que se repite periódicamente.

4.4.2.2. Desarrollo del modelo

Para desarrollar un modelo de series temporales se procede:

- Inicialmente se busca convertir las observaciones en una serie estacionaria, es decir, que los valores de la media, varianza, autocorrelaciones dependan del tiempo.
- Luego se procede a estudiar la presencia de irregularidades en la serie, determinando si en la serie si el modelo se debe formular con los datos originales o con series transformadas, verificando el componente regular (tendencia) y estacional, si existe tendencia se realiza una diferenciación regular, y una diferencia estacional, si existe estacionalidad en la serie, para identificar un posible modelo matemático, calculando a través de la función de autocorrelación simple y parcial, seleccionando el modelo que más se adecue a los datos.
- Posteriormente se determinan los parámetros de estimación por medio de la autocorrelación de la serie que permite identificar el nivel de ruido de ésta y consecuentemente determinar los parámetros de estimación.
- Para el proceso de estimación de los parámetros, se debe de revisar los residuales determinará la selección del modelo ARIMA adecuado.
- Finalmente, el proceso de predicción supone la aplicación sobre la serie de datos.

4.4.3. Contraste de los modelos de pronóstico

Para realizar el contraste de los modelos de pronóstico se hace uso de los errores MAPE (error porcentual absoluto medio) y el RMSE (La raíz del error cuadrático medio), obteniendo dichos errores para cada una de las técnicas de series de tiempos y redes neuronales artificiales, e identificando el menor error entre ambos para seleccionar el modelo que pronostica los niveles del reservorio de la presa Hidroeléctrica Cerrón Grande de El Salvador.

Capítulo 5. Resultados

Se presentan los resultados de los procedimientos para encontrar los modelos con las técnicas de series de tiempo y redes neuronales artificiales para generar pronósticos del nivel del reservorio, y realizar una comparación entre los modelos resultantes.

5.1. Resultados Generales

Es conveniente analizar la parte gráfica y los estadísticos descriptivos, con el propósito de descartar la existencia de los datos atípicos o también conocidos como outliers en el comportamiento de la serie de datos nivel del reservorio.

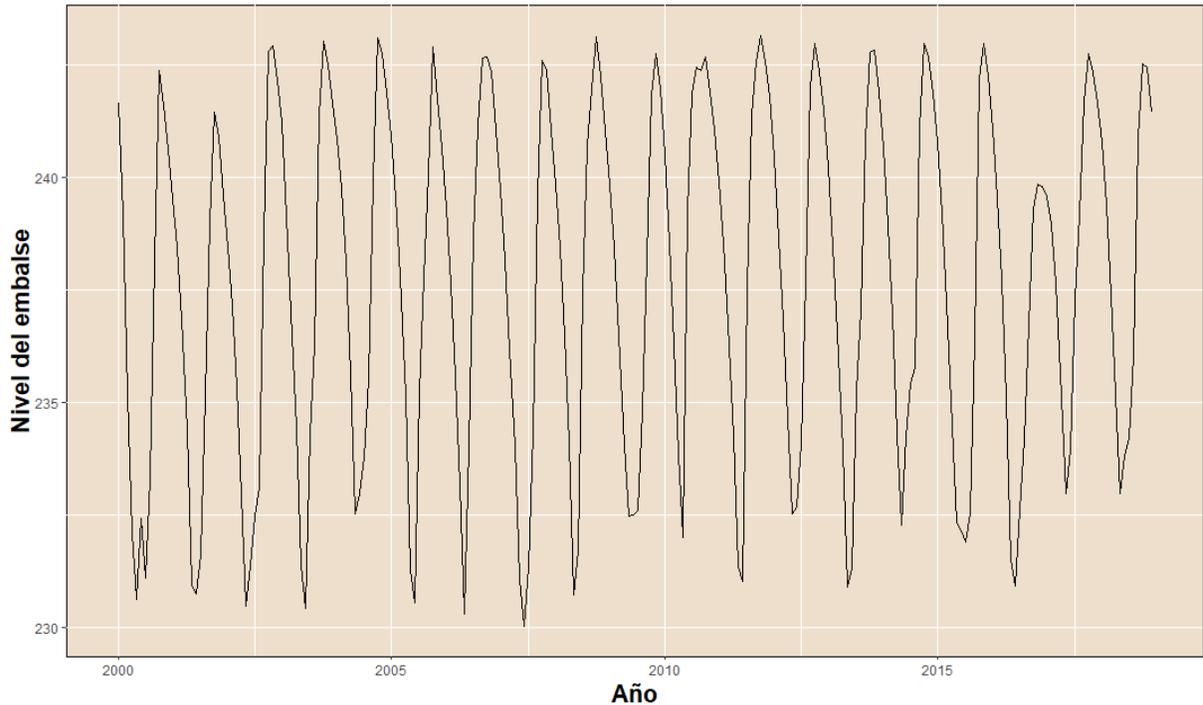
Tabla 1: *Estadísticos descriptivos de los niveles del reservorio de la presa Cerrón Grande de El Salvador*

Estadísticos	Valor
Número de datos	228
Media	237.6
Desviación Estándar	3.95
Valor Mínimo	230.0
Cuartil 1	234.2
Mediana	238.4
Cuartil 3	241.2
Valor Máximo	243.1

Fuente: Elaboración Propia

En los estadísticos principales presentados en la tabla 1, donde se observa que la serie de los datos oscila entre los 230.0 y 243.1 m s. n. m. del nivel del reservorio desde el año 2000 hasta el año 2018, con un promedio de 237.6 m s. n. m. durante dicho periodo con una variación de 3.95 metros con respecto al promedio del embalse.

Nivel del reservorio de la Presa Hidroeléctrica Cerrón Grande de El Salvador del año 2000 al 2018



Fuente de datos: Comisión Ejecutiva Hidroeléctrica del Río Lempa (CEL)

Figura 22: Niveles mensuales del reservorio de la presa Cerrón Grande de El Salvador desde el 2000 al 2018

Esto se observa de manera gráfica en la figura 22, donde evidencia que no existe presencia de datos atípicos, puesto que los datos se presentan patrones parecida en cada uno de los años, En el caso de las partes máximas poseen un comportamiento similar en cada año, únicamente en el 2016 el valor máximo que posee no se acerca al nivel máximo que tiene en los años anteriores, lo cual se puede asociar a escasez en las lluvias alrededor del río que alimenta el reservorio, además la variación se presentan son los valores bajos, debido a que presentan valores similares durante el periodo de tiempo estudiado de los niveles.

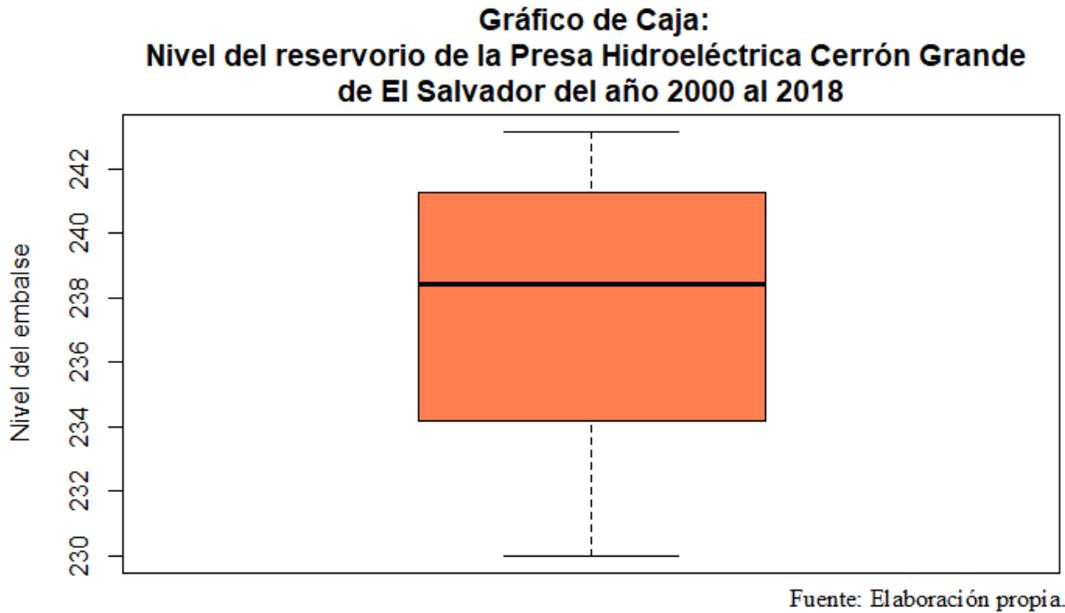


Figura 23: Diagrama de caja-bigote de los niveles mensuales del reservorio del Cerrón Grande desde el 2000 al 2018

Al observar la evolución de los datos no se aprecia la presencia de datos atípicos, esto se afirma con el diagrama de cajas-bigotes presentado en la figura 23, en donde no existen datos que se encuentren por fuera de los bigotes de la gráfica, además no presentan una dispersión alta, debido a que tiene una forma alargada. En cuanto a la concentración de los datos cuenta con un comportamiento asimétrico negativo, esto se debe a la existencia de mayor concentración de los valores del nivel del reservorio por debajo de la mediana de los datos.

Debido a la nula presencia de outliers significativos, se llega a la conclusión que todos los valores de la serie original, serán tomados en cuentas para el análisis y la creación de modelos de predicción creadas por medio de las técnicas: Series Temporales y Redes Neuronales.

5.2. Análisis de la Red Neuronal

La aplicación del método de redes neuronales artificiales para pronosticar valores futuros de la serie de los niveles del reservorio de la Presa Hidroeléctrica el Cerrón Grande de El Salvador, es necesario construir un tipo de red neuronal denominado perceptrón multicapa, el cual consta de una capa de entrada que con doce entradas, seguidamente de dos capas ocultas que contienen 9 y 6 neuronas respectivamente, y en la capa de salida se encuentra una única neurona que se refiere al nivel del reservorio. La red neuronal que se construyó es la que se presenta en la figura 22.

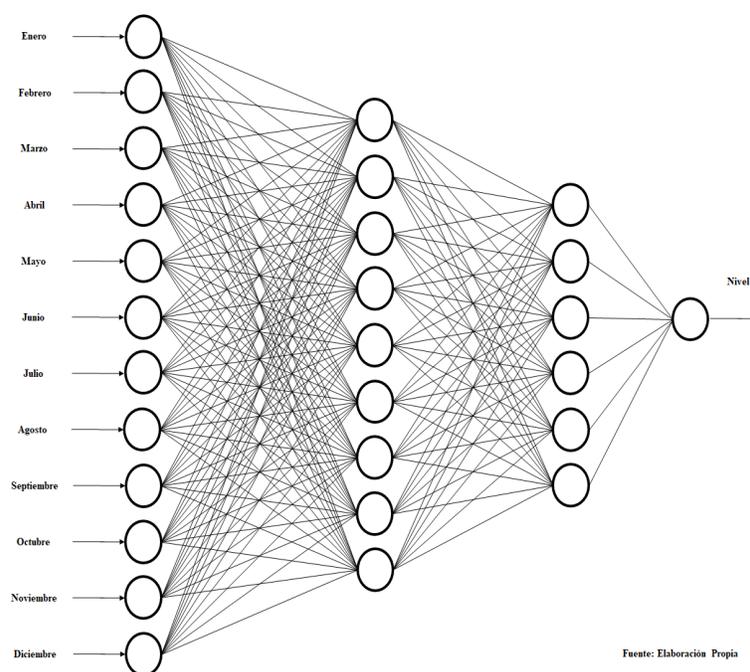


Figura 24: Estructura de la Red Neuronal para el nivel del reservorio del Cerrón Grande desde el 2000 al 2018

Para trabajar con esta estructura del total de los datos referidos al nivel del reservorio, se tomó el 70% de las observaciones para obtener el conjunto de entrenamiento y el 30% restante se toma para formar el conjunto de validación, utilizando el algoritmo de entrenamiento backpropagation. Dentro de las capas ocultas de la red las funciones de activación fueron la Sigmoide y la Tangente Hiperbólica, que son las encargadas de relacionar la suma ponderada de las unidades de una capa con los valores de unidades en la capa correcta. Además, teniendo en la capa de salida la función identidad como activadora.

5.2.1. Red Neuronal con Función Sigmoide

A partir de la estructura de la red neuronal utilizando la función de activación sigmoide, se presentan a continuación el comportamiento de dos modelos.

Modelo 1 de Red Neuronal

El resumen del primero modelo de red neuronal se muestra en la tabla 2, donde de los 228 datos del nivel del reservorio representan el 100%, se ha tomado 154 casos que son el 67.5% para formar el conjunto para realizar el proceso de entrenamiento de la red, y los 74 casos restantes que representa el 32.5% que conforma el conjunto de validación en el proceso.

Tabla 2: *Resumen del modelo 1 utilizando red neuronal artificial con función de activación sigmoide*

Resumen de procesamiento de casos para el modelo 1 de Red Neuronal			
		N	Porcentaje
Ejemplo	Entrenamiento	154	67.5%
	Pruebas	74	32.5%
Válido		228	100.0%
Excluido		0	
Total		228	

Fuente: Elaboración Propia

A partir de los valores de error obtenidos se puede evaluar el ajuste del modelo de red neuronal. En la tabla 3 muestra el error de la suma de cuadrados de los dos conjuntos, en donde el error posee un valor mayor en el conjunto de entrenamiento que el de validación. Sin embargo, teniendo el error relativo presentan un valor menor al 15% en los conjuntos del modelo.

Tabla 3: Errores del modelo 1 utilizando red neuronal artificial con función de activación sigmoide

Resumen del modelo 1 de red neuronal con función Sigmoide		
	Error de suma de cuadrados	6.929
	Error relativo	0.094
Entrenamiento	Regla de parada utilizada	1 paso(s) consecutivo(s) sin disminución del error
	<hr/>	
Pruebas	Error de suma de cuadrados	5.972
	Error relativo	0.130

Fuente: Elaboración Propia

Modelo 2 de Red Neuronal

El modelo 2 de la red neuronal, entrenada con la función sigmoide presenta en la tabla 4 que se toma un 69.3% de los datos del nivel del reservorio para realizar el entrenamiento, para luego evaluar la red en el conjunto de validación representada por el 30.7%.

Tabla 4: Resumen del modelo 2 utilizando red neuronal artificial con función de activación sigmoide

Resumen de procesamiento de casos para el modelo 2 de Red Neuronal			
		N	Porcentaje
Ejemplo	Entrenamiento	154	69.3%
	Pruebas	74	30.7%
Válido		228	100.0%
Excluido		0	
Total		228	

Fuente: Elaboración Propia

La tabla 5 muestra el error de la suma de cuadrados en la etapa de entrenamiento que posee el modelo 2 de la red neuronal, teniendo un valor de 7.097 con un error relativo de 0.09. El cual cambia en el conjunto de prueba, debido a que posee un error de 0.106 este valor no se aleja considerablemente ante el conjunto anterior.

Tabla 5: Errores del modelo 2 utilizando red neuronal artificial con función de activación sigmoide

Resumen del modelo 2 de red neuronal con función Sigmoide		
	Error de suma de cuadrados	7.097
	Error relativo	0.090
Entrenamiento	Regla de parada utilizada	1 paso(s) consecutivo(s) sin disminución del error
	Error de suma de cuadrados	4.321
Pruebas	Error relativo	0.106

Fuente: Elaboración Propia

Debido a que dicho modelo posee menor error que el error del modelo 1 (ver tabla 3), se propone al segundo modelo de redes neuronales artificiales trabajado con la función de activación sigmoide para realizar los pronósticos del nivel del reservorio del de la Presa Hidroeléctrica del Cerrón Grande de El Salvador.

5.2.2. Red Neuronal con Función Tangente Hiperbólica

Además, de los dos modelos entrenado con la función de activación sigmoide, también se presentan dos modelos utilizando la función de activación tangente hiperbólica que posee una función de salida identidad.

Modelo 3 de Red Neuronal

Se muestra el resumen del modelo de red neuronal, en la tabla 6 que presenta el porcentaje retomando los dos conjuntos que procesa la red neuronal. Haciendo énfasis que, en este caso, se toma el 65.8% de los datos del nivel para entrenar la red, además el restante para realizar el proceso de validación de la red. Es de recalcar que en dicho proceso todos los valores del nivel reservorio son válidos.

Tabla 6: Resumen del modelo 3 utilizando red neuronal artificial con función de activación tangente hiperbólica

Resumen de procesamiento de casos para el modelo 3 de Red Neuronal			
		N	Porcentaje
Ejemplo	Entrenamiento	150	65.8%
	Pruebas	78	34.2%
Válido		228	100.0%
Excluido		0	
Total		228	

Fuente: Elaboración Propia

El modelo número 3 presentado con redes neuronales, muestra errores parecidos a los modelos planteados anteriormente, teniendo en la parte de entrenamiento un error de 0.053, junto al de la suma de cuadrados con un valor de 3.958. A partir de ello se puede considerar tomar en cuenta que este modelo es candidato para realizar los pronósticos del nivel del embalse, esto se puede apreciar en la tabla 7.

Tabla 7: Errores del modelo 3 utilizando red neuronal artificial con función de activación tangente hiperbólica

Resumen del modelo 1 de red neuronal con función Sigmoide		
Entrenamiento	Error de suma de cuadrados	3.958
	Error relativo	0.053
	Regla de parada utilizada	1 paso(s) consecutivo(s) sin disminución del error
Pruebas	Error de suma de cuadrados	6.144
	Error relativo	0.143

Fuente: Elaboración Propia

Modelo 4 de Red Neuronal

Se presenta un cuarto modelo trabajado con la red neuronal, donde el conjunto de entrenamiento se encuentra representado por el 71.5% de los datos relacionados con el nivel del reservorio, tomando el resto como el conjunto de validación o de prueba, se puede observar en la tabla 8.

Tabla 8: Resumen del modelo 4 utilizando red neuronal artificial con función de activación tangente hiperbólica

Resumen de procesamiento de casos para el modelo 1 de Red Neuronal			
		N	Porcentaje
Ejemplo	Entrenamiento	163	71.5%
	Pruebas	65	28.5%
Válido		228	100.0%
Excluido		0	
Total		228	

Fuente: Elaboración Propia

Tabla 9: Errores del modelo 4 utilizando red neuronal artificial con función de activación tangente hiperbólica

Resumen del modelo 1 de red neuronal con función Sigmoide		
Entrenamiento	Error de suma de cuadrados	6.951
	Error relativo	0.091
	Regla de parada utilizada	1 paso(s) consecutivo(s) sin disminución del error
Pruebas	Error de suma de cuadrados	4.952
	Error relativo	0.132

Fuente: Elaboración Propia

En la tabla 9, se detecta una disminución considerable en el valor de los errores en los conjuntos de entrenamiento y de prueba con respecto a los planteados anteriormente, teniendo errores no mayores a 12%, además se presenta el error relativo en la fase de entrenamiento el cual es de 0.014 y en la fase de prueba presenta un valor de 0.112.

Para presentar los pronósticos generados por la red neuronal, se toman dos modelos con los menores errores relativos, los cuales se encuentran representados por el modelo 2 de red neuronal que se encuentra activado por la función sigmoide, y el modelo 4 con la función tangente hiperbólica como activadora, esto según lo muestra la tabla 5 y 9.

5.2.3. Resultados de la Red Neuronal

Se observan los resultados de los modelos de redes neuronales aceptados anteriormente, mostrando el comportamiento de los pronósticos en cada modelo, posteriormente se compara los resultados obtenidos.

Red Neuronal con Función Sigmoide

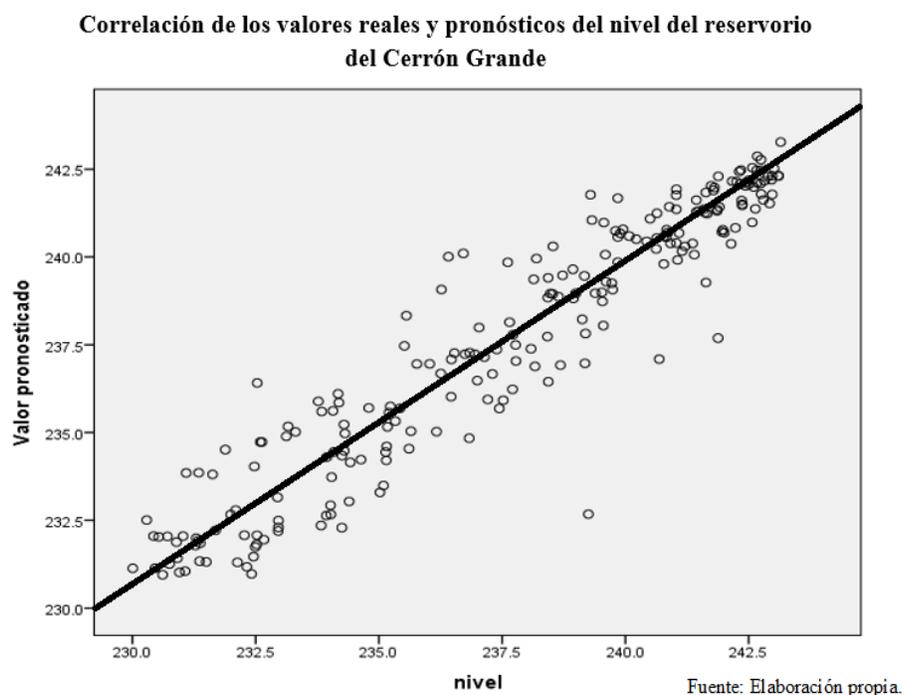


Figura 25: Correlación de los valores reales y pronósticos con el modelo de red neuronal del nivel del reservorio

La gráfica de correlaciones (figura 25) presenta relación alta entre los valores reales del nivel del reservorio con los valores pronosticados el modelo de Red Neuronal artificial con la función sigmoide, existiendo una baja cantidad de datos que no se ajustan a la recta de ajuste, existiendo una poca dispersión en relación a la información del reservorio.

Con el modelo de la red neuronal activado por la función Sigmoide ha realizado un aprendizaje cercano a los niveles del reservorio de la presa Hidroeléctrica Cerrón Grande proporcionados por la CEL, y se puede observar en la figura 26 el comportamiento de los datos entrenados con la red neuronal tomando el mismo periodo que los datos reales del nivel desde el año 2000 hasta el 2018.

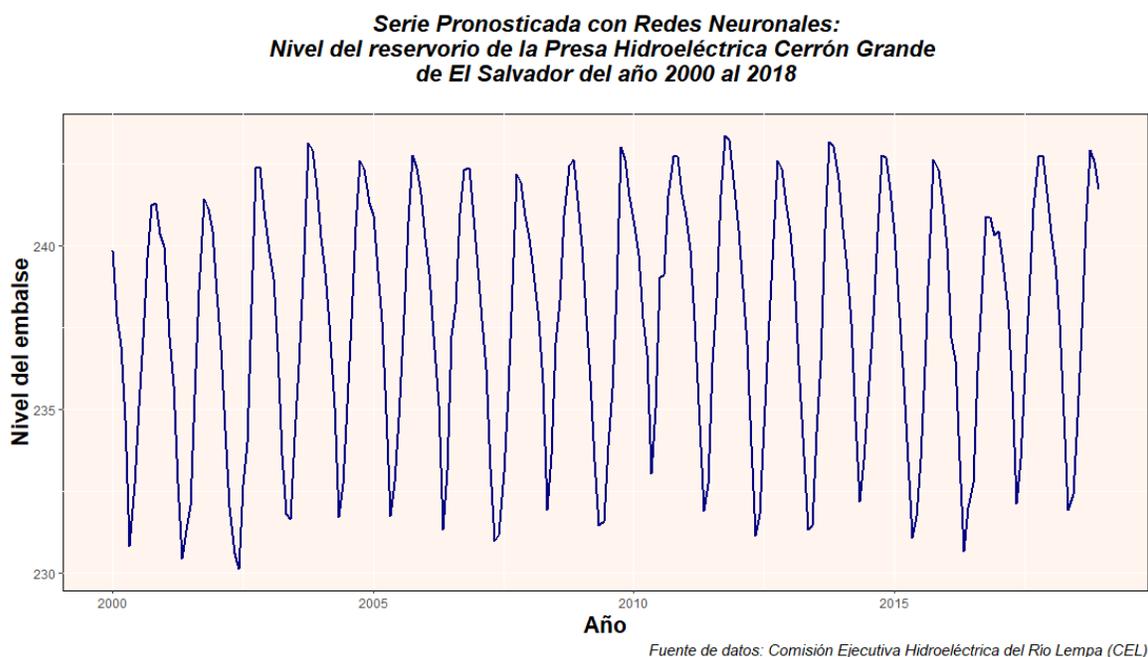


Figura 26: Serie Pronosticada del nivel del reservorio utilizando el modelo de Redes Neuronales con la Función Sigmoide

Red Neuronal con Función Tangente Hiperbólica

El modelo de red neuronal activado con la función tangente hiperbólica, en primera instancia se observa que existe una correlación fuerte entre el valor pronosticado con el valor real, acoplándose de mejor manera a la recta de mejor ajuste, esto se puede apreciar de mejor forma al observar la figura 27.

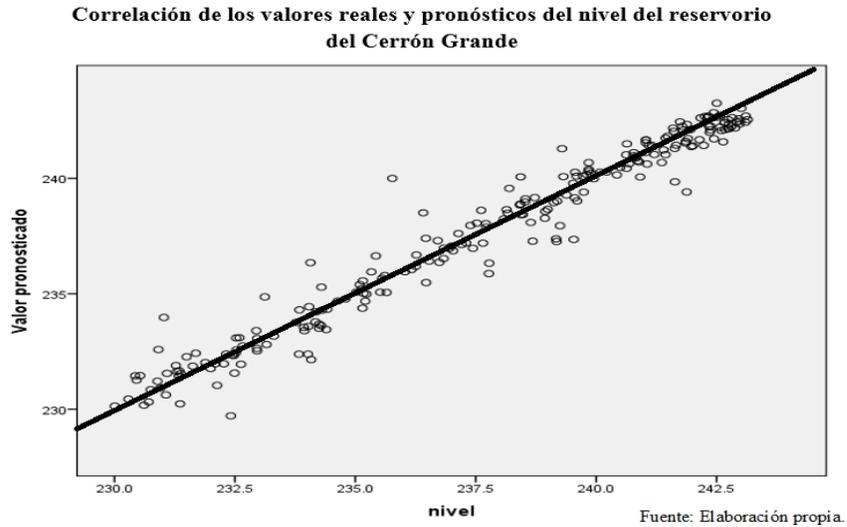


Figura 27: Correlación de los valores reales y pronósticos con el modelo de red neuronal del nivel del reservorio

La serie pronosticada mediante el aprendizaje de la red neuronal que se ha activado con la función tangente hiperbólica, se muestra en la figura 28 que los niveles presentados con la red se comportan de manera similar a los datos originales, presentando oscilaciones altas y bajas en cada uno de los años que se presentan a lo largo del periodo comprendido del año 2000 al 2018.

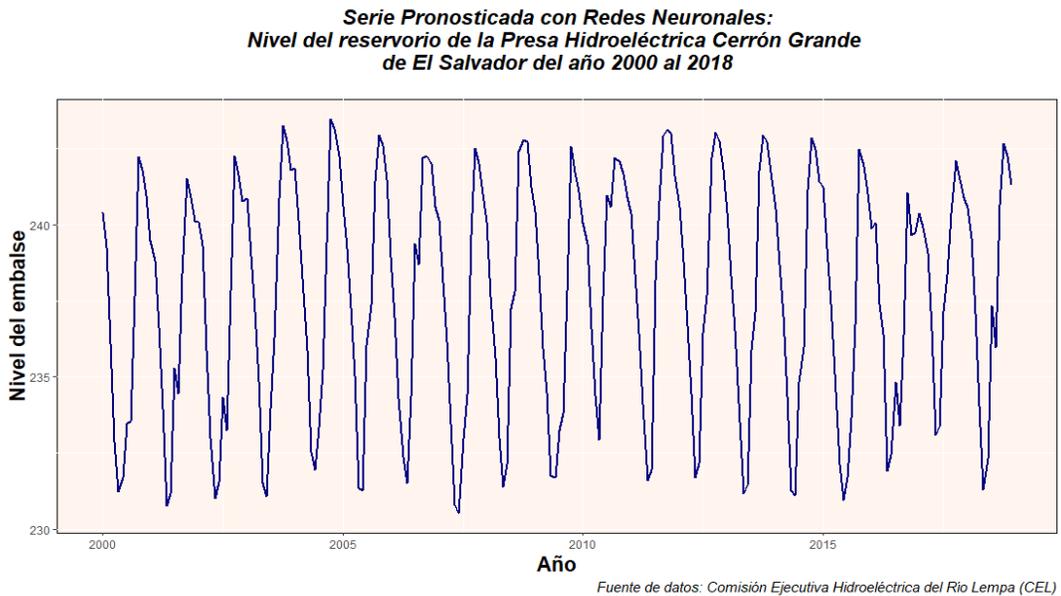


Figura 28: Serie Pronosticada utilizando el modelo de Redes Neuronales

Comparación entre los modelos de Redes Neuronales identificados

El comportamiento de los datos del nivel generados con los dos modelos seleccionados, poseen una tendencia parecida a los datos reales, observando que los modelos explican de manera efectiva a los datos. Es por ello que se realizó una comparación entre el modelo generado con la función tangente hiperbólica y sigmoide, observando cual modelo tiene un porcentaje menor de error y cual se ajusta para realizar los pronósticos.

Tabla 10: Valores reales del nivel del reservorio y valores pronosticados con los modelos de redes neuronales identificados

Comparación de los valores reales del nivel del reservorio Y los valores pronosticados con los modelos de redes neuronales						
Año	Mes	Nivel	RN Sigmoide		RN Tang Hiperbólica	
			Predicción	Error	Predicción	Error
2017	Enero	239.563	240.443	0.775203	240.356	0.6297
	Febrero	238.935	239.500	0.319094	239.898	0.9288
	Marzo	237.650	238.069	0.175684	239.063	1.9974
	Abril	235.650	235.572	0.006024	236.533	0.7807
	Mayo	232.961	232.114	0.718553	233.082	0.0145
	Junio	233.933	233.476	0.208562	233.368	0.3188
	Julio	237.302	236.109	1.424964	237.141	0.0258
	Agosto	239.169	238.454	0.511274	238.344	0.6796
	Septiembre	241.660	241.138	0.272279	240.359	1.6923
	Octubre	242.749	242.724	0.000634	242.110	0.4074
	Noviembre	242.310	242.756	0.199284	241.472	0.7024
	Diciembre	241.631	241.601	0.000943	240.886	0.5554
2018	Enero	240.778	240.391	0.150074	240.539	0.0572
	Febrero	239.386	239.439	0.002822	239.551	0.0272
	Marzo	237.390	237.576	0.034648	237.090	0.0904
	Abril	235.157	234.808	0.121825	234.039	1.2487
	Mayo	232.957	231.918	1.081064	231.298	2.7532
	Junio	233.828	232.400	2.037165	232.327	2.2506
	Julio	234.188	234.732	0.295532	237.318	9.7968
	Agosto	236.031	237.467	2.063719	235.974	0.0031
	Septiembre	240.865	240.719	0.021203	240.634	0.0533
	Octubre	242.507	242.914	0.165527	242.653	0.0214
	Noviembre	242.440	242.577	0.018585	242.240	0.0401
	Diciembre	241.440	241.693	0.063975	241.298	0.0202

Fuente: Elaboración Propia

Para ello se ha tomado los datos de los últimos dos años, el cual ayuda a realizar el contraste del comportamiento de los datos tratados. En la tabla 10 presenta en primer plano los valores reales, seguido por los valores del pronóstico de cada red neuronal, junto a sus respectivos errores.

Se aprecia que la distancia de los pronósticos de cada modelo es poca, sin embargo, existe diferencia en los errores que se presentan en el periodo del 2017 y 2018, esto debido a que en dicho intervalo el modelo con el método de la función tangente hiperbólica presenta valores cercanos a ser ceros, es decir que se mantienen congruentes a los datos que intenta asemejar, pero existe la presencia valores que presentan errores altos, en los cuales se aprecia que en el 2017 ambos modelos tenían semejanza, siendo hasta el final del periodo el modelo de red que se mantiene más cercano a los niveles reales es el que proporciona la función sigmoide (esto se observa en las gráficas 29 y 30).

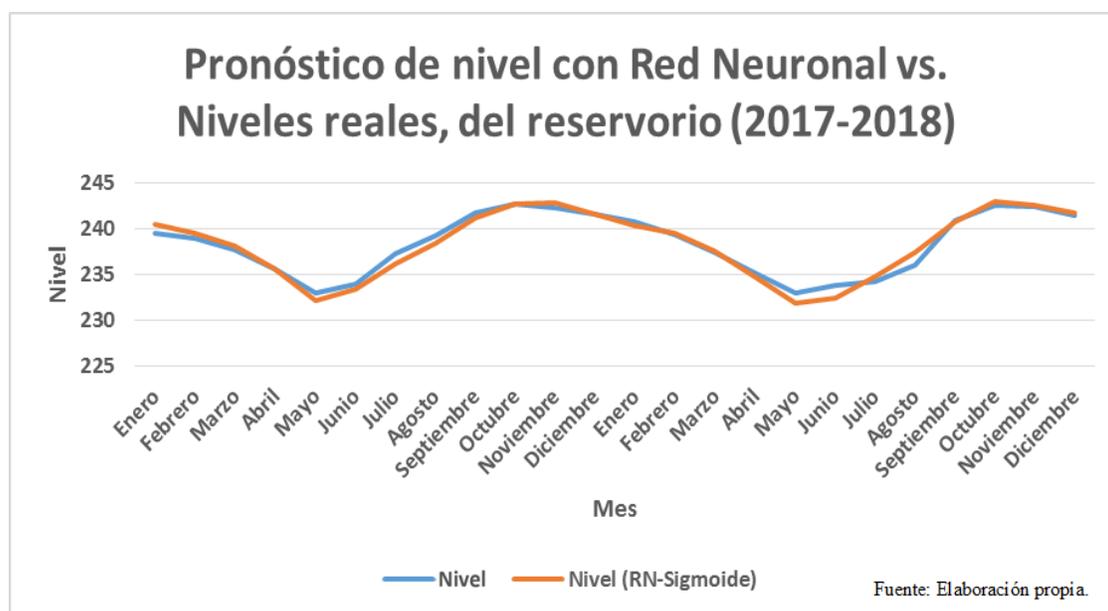


Figura 29: Datos reales vs. pronósticos con red neuronal sigmoide de los niveles del reservorio de la presa Cerrón Grande

Para la figura 30 se evidencia como a partir del segundo trimestre del 2018, la serie tuvo un comportamiento alejado de los datos originales, además se destaca que en el mes de agosto el valor fue superior a lo que venía mostrando la red neuronal.

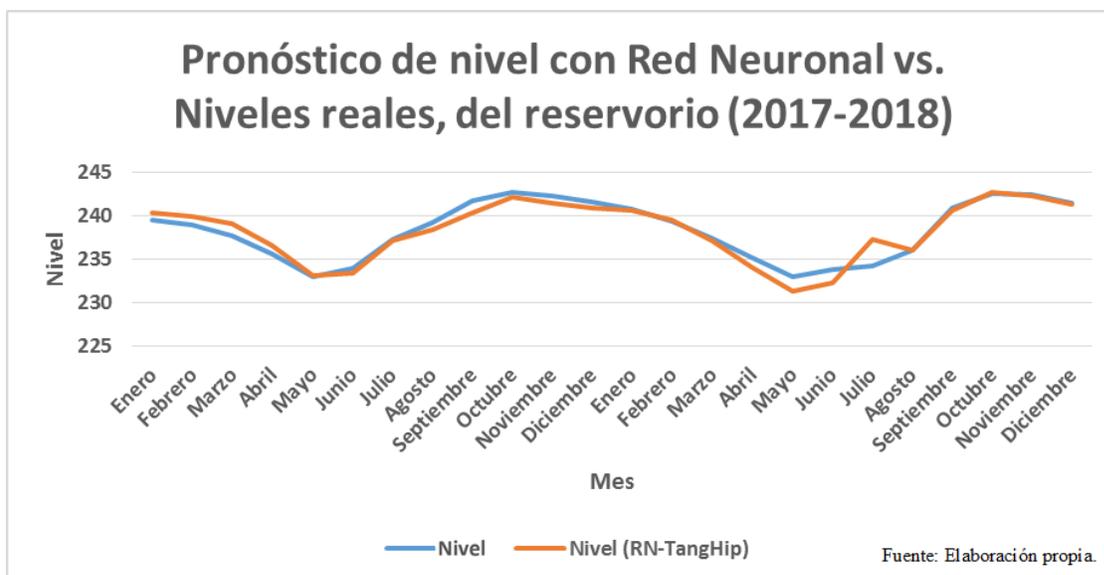


Figura 30: Datos reales vs. pronósticos con red neuronal hiperbólica de los niveles del reservorio de la presa Cerrón Grande

A partir de esta información se puede considerar la decisión de tomar como mejor modelo (mejor ajuste de los datos) el de la red neuronal sigmoide, sin embargo, antes es necesario revisar los errores estadísticos que proporciona con las predicciones.

Tabla 11: Precisión de pronóstico de los modelos de redes neuronales identificados

Errores estadísticos de predicción de los modelos de redes neuronales del nivel del reservorio				
	MSE	MAD	RMSE	MAPE
RN Sigmoide	1.325238	0.8415405	1.1511898	0.0035544
RN Tang Hiperbólica	0.7133065	0.5938535	0.8445747	0.0025074

Fuente: Elaboración Propia

De los errores del periodo comprendido del 2000 al 2018 observados en la tabla 11 se escoge el modelo que se ajusta a los datos es la red neuronal con la función tangente hiperbólica, pudiendo observar cómo es el comportamiento de la serie de este modelo con la serie original.

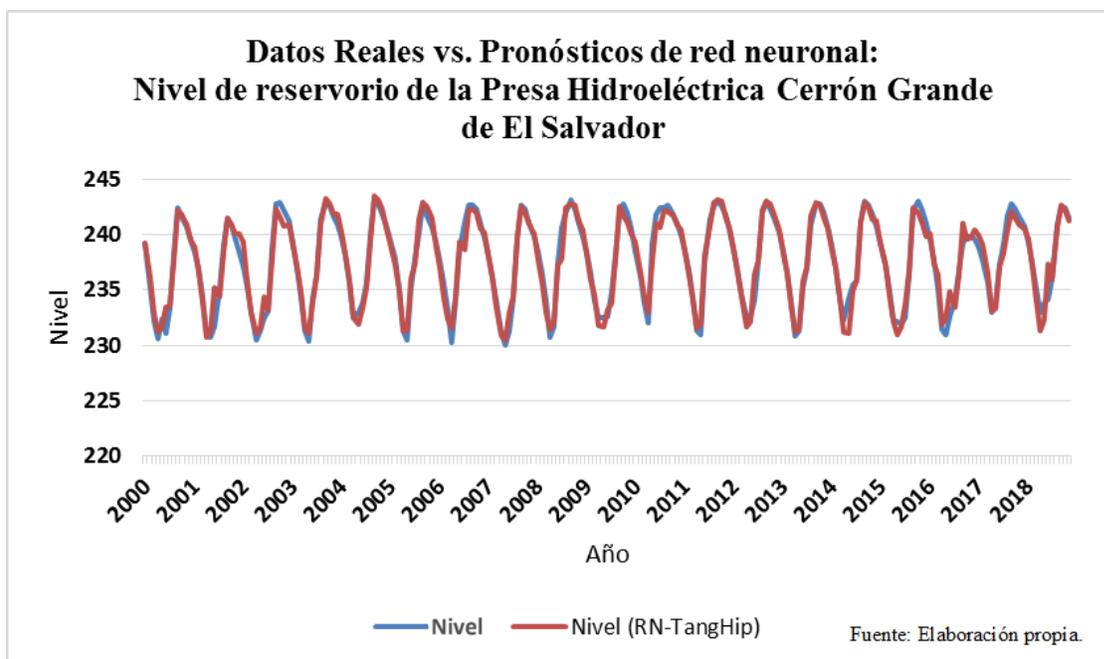


Figura 31: Datos reales vs. pronósticos de la red neuronal del nivel del reservorio de la Presa Cerrón Grande

En la figura 31 se aprecia como los datos del nivel del reservorio generados con modelo de la red neuronal se ajusta de una manera cercana a los datos originales, incluso posee similitudes en las partes altas y bajas de cada uno de los años, asemejando el comportamiento en momentos donde existen pequeños cambios en el nivel, tal y como se observa en los años 2016 y 2017. Tomando dicho modelo como el óptimo para realizar el proceso de pronóstico.

5.2.4. Pronóstico con Redes Neuronales

Con el modelo de red neuronal que se activa con la función tangente hiperbólica para realizar el pronóstico. Se realiza una comparación del último año tratado del nivel de reservorio, entre los resultados que arrojó el modelo con los valores reales en la tabla 12, donde muestra que los valores arrojados por la red neuronal se asemejan en mucho a los datos reales, únicamente el mes de julio presentó un valor alto por parte de la red neuronal, pero no afectado al comportamiento cercano de los demás meses en el año 2018.

Tabla 12: Valores reales vs. valores pronosticados con Redes neuronales del nivel del reservorio de la presa Cerrón Grande

Comparación de los valores reales con los valores generados por la red neuronal		
Mes	Valores reales 2018	Valores Pronosticados 2018
Enero	240,78	240.54
Febrero	239,39	239.55
Marzo	237,39	237.09
Abril	235,16	234.04
Mayo	232,96	231.30
Junio	233,83	232.33
Julio	234,19	237.32
Agosto	236,03	235.97
Septiembre	240,86	240.63
Octubre	242,51	242.65
Noviembre	242,44	242.24
Diciembre	241,44	241.29

Fuente: Elaboración Propia

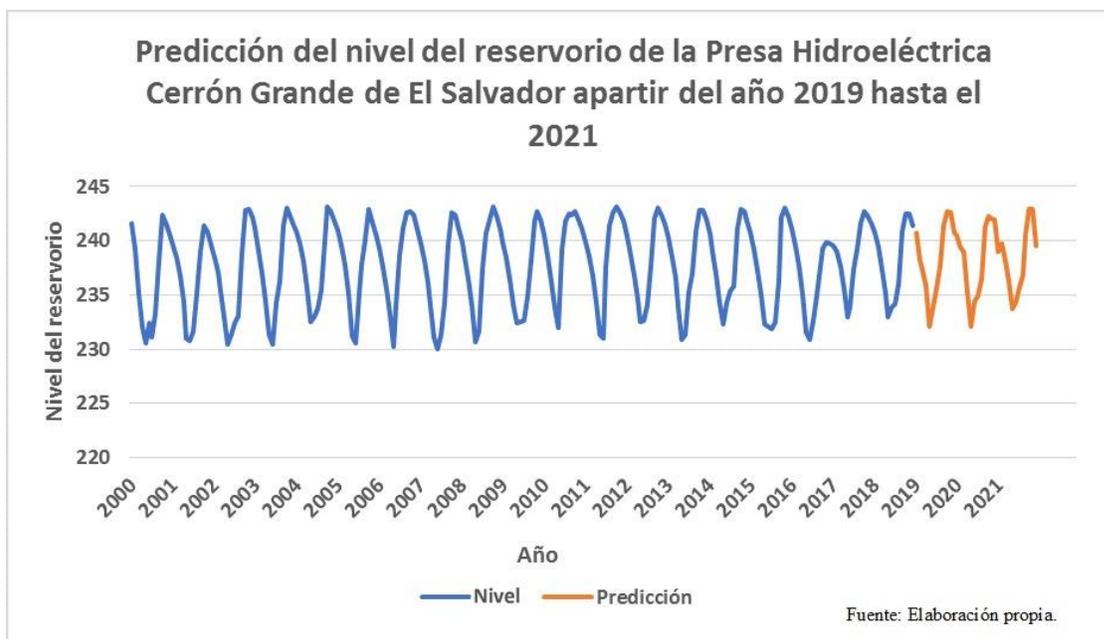


Figura 32: Valores reales y valores pronósticos del nivel del reservorio generados con Red neuronal desde el año 2000 hasta el 2021

En la figura 32, se presentan los datos originales del nivel del reservorio con los datos de pronóstico generados por la red neuronal de los tres años siguientes (2019-2021), observando que el comportamiento es similar presentando un comportamiento horizontal, sin que los valores altos de los pronósticos tengan picos más altos o más bajos que los de la serie original, además dichos datos de predicción se presentan en la tabla 13.

Tabla 13: *Pronóstico con el modelo de Red Neuronal del nivel del reservorio de la presa Cerrón Grande desde el año 2019 hasta el 2021*

Mes	2018	2019	2020	2021
Enero	240.778	240.7283	240.5584	239.0320
Febrero	239.386	238.2128	239.4290	239.7888
Marzo	237.390	237.3823	238.9797	238.2128
Abril	235.157	235.9643	235.2750	236.2868
Mayo	232.957	232.0993	233.4593	233.7777
Junio	233.828	233.8316	234.4327	234.2885
Julio	234.188	235.6066	234.9613	235.6973
Agosto	236.031	237.6527	236.4325	236.8716
Septiembre	240.865	241.4171	241.3120	240.5052
Octubre	242.507	242.6812	242.2536	242.9461
Noviembre	242.440	242.6237	242.40715	242.9326
Diciembre	241.440	240.7283	241.9533	241.4953

Fuente: Elaboración propia.

El comportamiento de los datos generados por el modelo de redes neuronales en cada uno de los tres años que se toman en cuenta para realizar el pronóstico tiene valores parecidos con el año 2018, además poseen el comportamiento de crecer en los meses de estación lluviosa y decrecer en la época seca.

5.3. Análisis de la Serie de Tiempo

La técnica de series temporales permite generar modelos para realizar pronósticos para el nivel del reservorio de la Presa Hidroeléctrica del Cerrón Grande de El Salvador, y para la construcción del modelo es necesario realizar, análisis previo de la serie, la identificación, estimación y hacer una diagnosis del modelo.

5.3.1. Análisis de la serie original

En los datos originales del nivel del reservorio presentados en la figura 33, se observa que durante el transcurso desde el año 2000 hasta 2018 pertenecen a una serie estacionaria, es decir, que los valores de los niveles se encuentran siempre alrededor del promedio de nivel con el que cuenta el embalse o reservorio, y se mantiene una variación constante a lo largo de los años presentando similitudes en periodos parecidos y no sufre cambios drásticos en el comportamiento, lo cual también evidencia que no existe la tendencia a decrecer o crecer en un determinado momento en el tiempo.

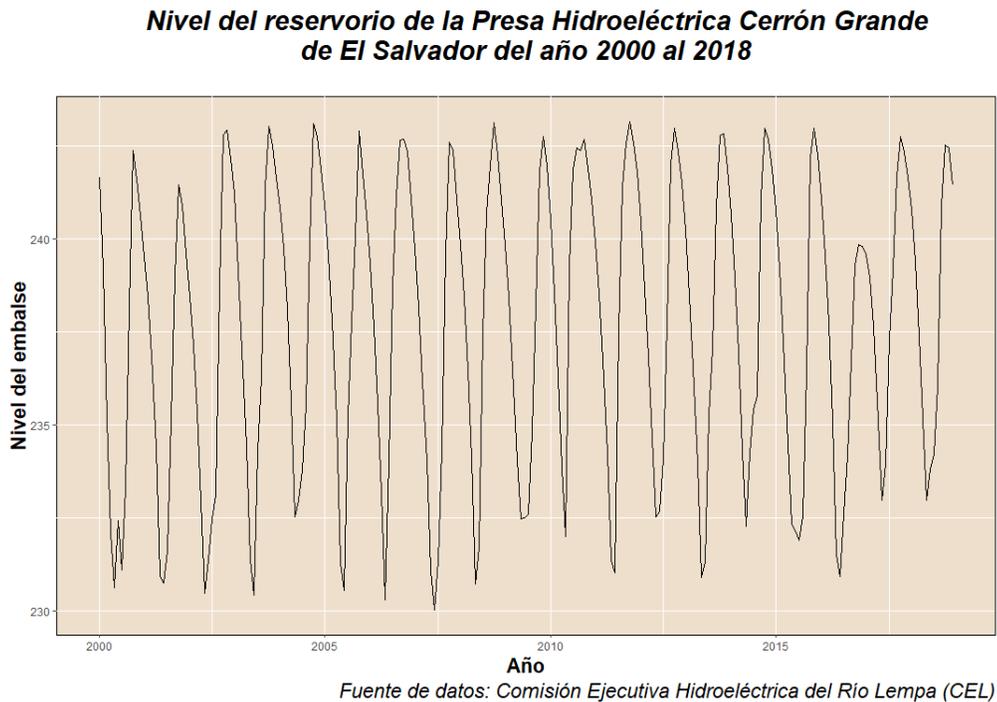


Figura 33: Gráfico Temporal de los niveles del reservorio de la presa Cerrón Grande de El Salvador

En el gráfico estacional se visualizan los mismos datos del nivel del reservorio representado en el gráfico de la serie temporal en la figura 34, muestra el comportamiento en el periodo individual por cada año en donde la información se superpone por cada temporada. Este tipo de gráfico es útil para la identificación los años que experimentaron un cambio en su patrón a lo largo del tiempo.

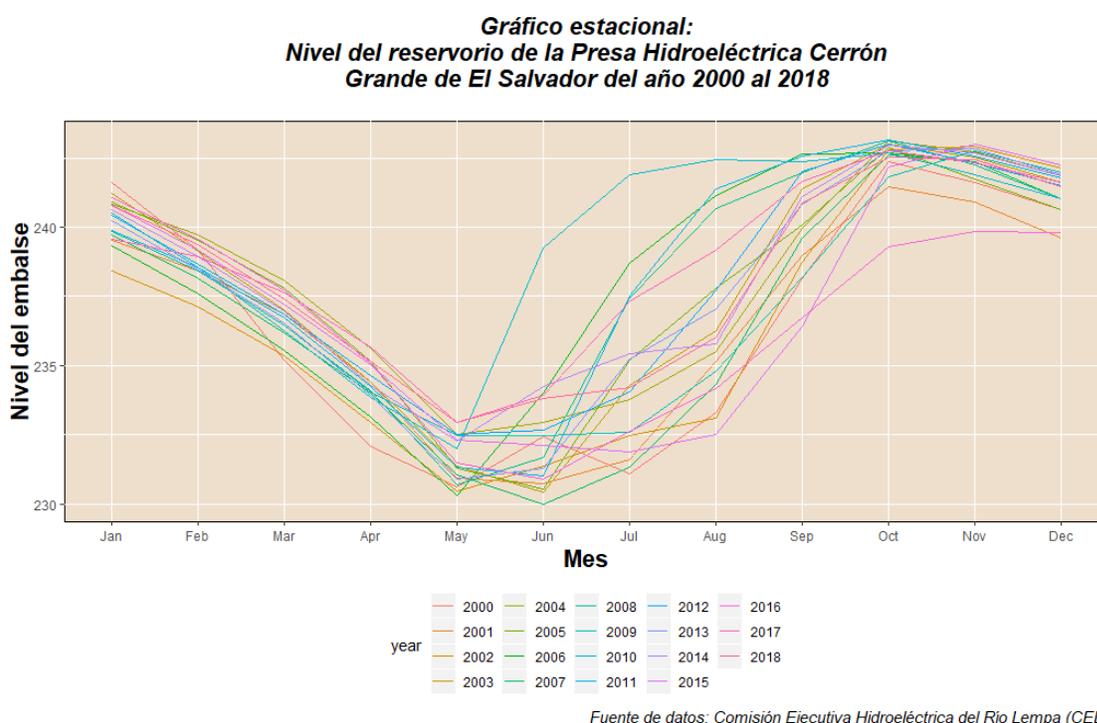


Figura 34: Gráfico estacional de los niveles del reservorio de la presa Cerrón Grande de El Salvador

Se evidencia el aumento del nivel del reservorio en los meses desde mayo hasta octubre, esto se debe a que coincide con el paso de la estación lluviosa en el territorio salvadoreño, en donde aumenta la presencia de las precipitaciones alrededor de la Presa Hidroeléctrica. En especial se muestra como en el año 2010 el nivel se elevó de manera drástica de mayo a junio, manteniéndose en un crecimiento hasta agosto, para mantenerse hasta el final de la temporada lluviosa, tal y como los demás años en esos meses. Caso contrario en la información recolectada en el año 2005, se observa que el mes de mayo decreció en comparación al mismo periodo en el resto de años, manteniendo a partir de ese momento un crecimiento bajo en los siguientes meses hasta incorporarse a cerrar de manera similar la época de lluvia.

Para el año 2016, se observa que los niveles del embalse fueron los más bajos registrados en los años presentados en la gráfica, esto se deba a que el clima en el país, pues en la época lluviosa se comportó de manera irregular, generando pocas lluvias a diferencia de otras estaciones lluviosas en los mismos periodos de meses, observando el decrecimiento de los niveles a partir de noviembre hasta los últimos días de abril. Una variación del gráfico estacional es utilizando coordenadas polares, en donde la serie de tiempo de los datos se presentan de forma circular, como se muestra en la figura 35. En la cual los datos mínimos son los que se acercan al centro del plano y caso contrario, entre mayor sea el valor del nivel, se mantendrá distante del eje polar.

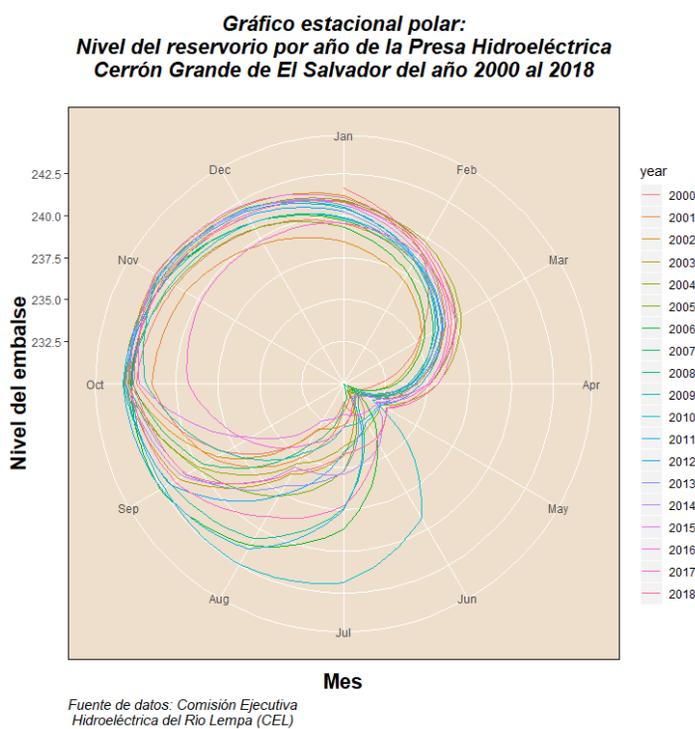


Figura 35: Gráfico estacional Polar de los niveles del reservorio de la presa Cerrón Grande de El Salvador

Además, se observa el comportamiento a través de los meses por año en el gráfico de la figura 36, que durante el primer trimestre donde la época seca tiene más presencia los niveles no presentan cambios considerables, manteniendo el mismo comportamiento los últimos tres meses siendo el punto más bajo en los últimos días de cada mes, teniendo los valores altos al finalizar la temporada seca.

**Gráfico estacional por mes:
Nivel del reservorio de la Presa Hidroeléctrica Cerrón Grande de El Salvador del año 2000 al 2018**

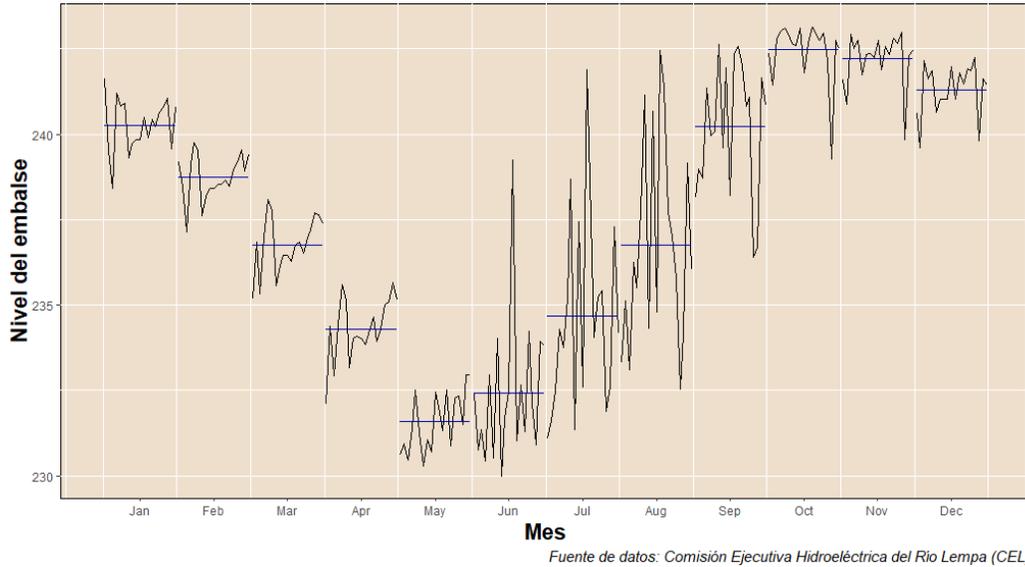


Figura 36: Gráfico estacional mensual de los niveles del reservorio de la presa Cerrón Grande de El Salvador

Gráfico Mensual Estacional: Nivel del reservorio de la Presa Hidroeléctrica Cerrón Grande de El Salvador del año 2000 al 2018

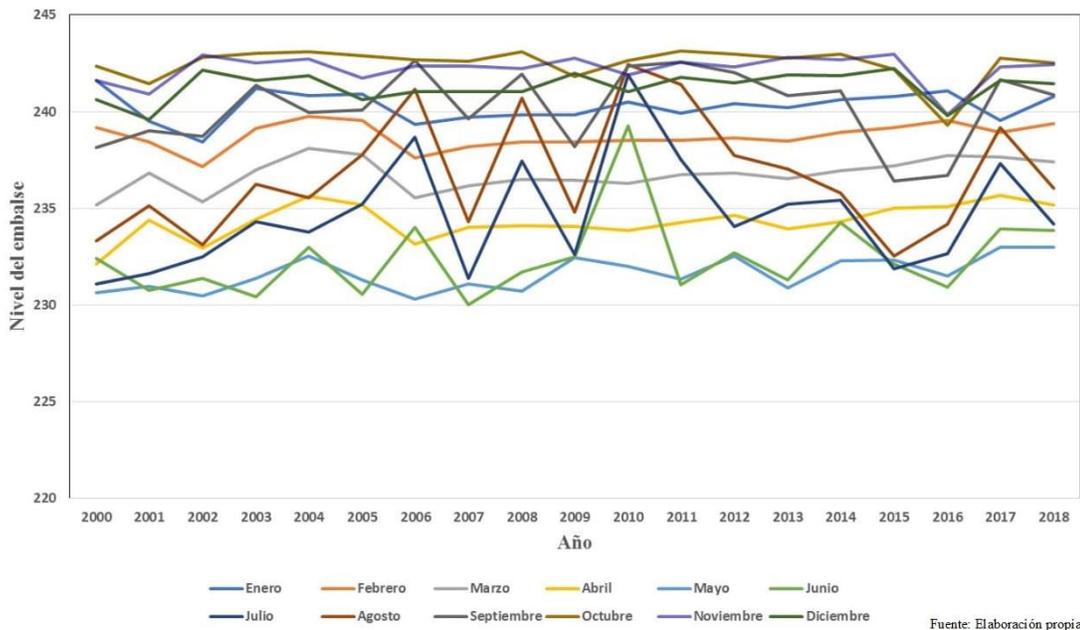


Figura 37: Gráfico estacional mensual por año de los niveles del reservorio de la presa Cerrón Grande de El Salvador

Se observa que el nivel del agua por mes posee un comportamiento similar a lo largo de los 19 años, a su vez se verifica que en los primeros y últimos meses se mantienen los niveles altos en comparación con los meses centrales del periodo anual, para el periodo de entre junio a agosto los niveles varían, teniendo como puntos máximos los años 2006, 2008, 2010 y el 2019 donde posiblemente las precipitaciones alrededor de la presa hidroeléctrica fueron mayores en comparación con los años del 2000 al 2004 junto al 2007 y 2009, donde la poca precipitación reflejó valores bajos en cuanto a la estación lluviosa dentro del territorio salvadoreño.

5.3.2. Identificación del modelo

Antes de identificar el modelo de la serie de tiempo que mejor se ajusta a los datos se procede a verificar la estacionariedad y la estacionalidad de la serie, además del análisis de los correlogramas.

5.3.2.1. Análisis de la estacionariedad de la serie

A partir de lo observado en la figura 33, en la cual se han descartado la existencia de datos atípicos mediante la descomposición de la serie temporal en todos sus componentes se puede realizar un resumen del comportamiento de la serie, previa a una búsqueda de los parámetros de los modelos, es por ello que la serie no posee una tendencia creciente o decreciente, sino que posee una tendencia horizontal a lo largo de los 18 años.

En la figura 38 se determina que la series es estacionaria en media, debido que los datos del nivel del reservorio se mantienen constante desde el 2000 al 2018, pues la serie gira alrededor de la media. Además, con dicho se verifica la estacionariedad de la varianza de la serie, porque los datos presentan un proceso con valores con alzas y bajas de la serie nivel, y se mantienen constante dentro del rango a través del tiempo. También, la aleatoriedad refleja la aparición de comportamiento irregular en los niveles de agua, es decir alzas y bajas en determinados momentos, esto posiblemente causado por los cambios de estaciones en un año, la cual presenta un patrón repetitivo en la continuidad de la serie.

Componentes de la Serie de Tiempo

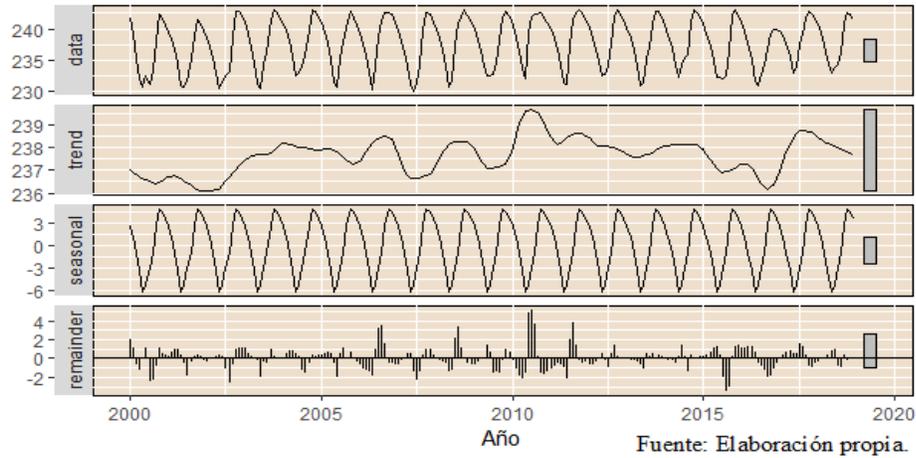
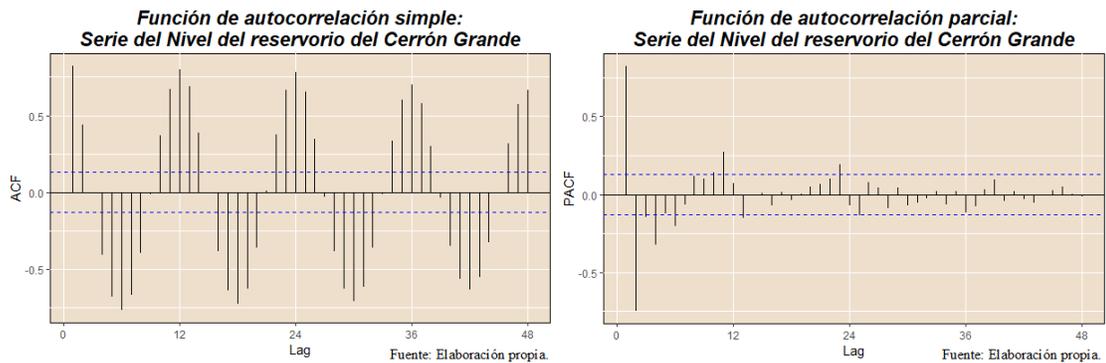


Figura 38: Componentes de la serie de los niveles del reservorio de la presa Cerrón Grande de El Salvador

Luego de la descomposición de la serie de tiempo se identifica en la función de autocorrelación simple asociada a la serie, en la figura 39a se detecta que el primer desfase está fuera de los límites de significancia de 5%, seguido por una oscilación decreciente entre correlaciones positivas y negativas. Puede presentar un término autorregresivo de orden superior y por tanto muestra una autocorrección significativa entre los desfases. Además, a partir de los desfases que se observan cada 12 observaciones, se aprecia la presencia de la estacionalidad en la serie nivel del reservorio.



(a) FAS de los datos originales

(b) FAP de los datos originales

Figura 39: Funciones de autocorrelaciones de los daos del nivel del reservorio del Cerrón Grande de El Salvador

Por otro lado, la función de autocorrelación parcial señala la presencia de dos correlaciones significativas en el primer o segundo desfases, seguidas de correlaciones alternantes poco significativas con una disminución convergente a cero. Por tanto, el término autorregresivo asociado a la serie sería de orden 1 o 2.

En la gráfica de la serie original del nivel del reservorio es estacionaria en media. Para confirmar la suposición que la serie presenta estacionariedad en media, se puede proceder a realizar el test de raíces unitarias de Dickey-Fuller (DF). Este test plantea las siguientes hipótesis:

H₀: Existencia de raíz unitaria (la serie no es estacionaria)

H_a: Ausencia de raíz unitaria (La serie es estacionaria)

Tabla 14: *Test de raíces unitarias de Dickey-Fuller*

Test		Estadístico t	Rezagos	Valor-p
Dickey-Fuller Aumentado		-10.253	6	0.01
Test critical values	1% level	-2.578476		
	5% level	-1.942688		
	10% level	-1.615.474		

Fuente: Elaboración propia

Dado que en la tabla 14 se observa un p-valor en la prueba de Dickey-Fuller Aumentado menor que 0.05, por lo tanto, existe evidencia suficiente para no aceptar la hipótesis nula, es decir hay ausencia de una raíz unitaria, por tanto, la serie es estacionaria en media.

Además, mediante la figura 33 se observa que la serie nivel del reservorio es estacionaria en varianza debido a que su comportamiento a lo largo del tiempo es compacto, manteniendo en un rango de aceptación que presentan los datos y en cada uno de los periodos anuales la variabilidad que presentan no resulta ser significativa. Por tanto, la serie es estacionaria también en varianza.

5.3.2.2. Análisis de la estacionalidad de la serie

Debido a la naturaleza de la serie se relaciona directamente a fenómenos meteorológicos, se hace énfasis en la presencia de la tendencia estacional, Y esto se puede observar que los niveles del reservorio poseen un patrón estacional, esto se puede observado en la figura 34, donde se muestra que se encuentra bien determinado el comportamiento de decrecimiento del nivel en el periodo de la estación seca (de noviembre a mayo) y presenta siempre un crecimiento en los meses que conforman la estación lluviosa.

Para poder eliminar este comportamiento estacional de la serie del nivel se puede tomar diferencias estacionales de orden 12. Entonces si se tiene que x_t es la serie que se quiere desestacionalizar, se trata de calcular $\nabla x_t = x_t - x_{t-12}$. Por ello se aplica la diferenciación estacional a la serie.

Tabla 15: Estadísticos descriptivos de la serie diferenciada estacional de los niveles del reservorio Cerrón Grande de El Salvador

Estadísticos	Valor
Media	0.08668
Valor Mínimo	-8.22267
Cuartil 1	-0.73995
Mediana	0.09097
Cuartil 3	0.78609
Valor Máximo	9.28581

Fuente: Elaboración Propia

Los estadísticos descriptivos en la tabla 15 en donde la media tiene una aproximación a cero, junto a una mediana cercana también a cero, pero sus límites presentan una distancia considerable entre sí. Y La figura 40 muestra la serie del nivel del reservorio desestacionalizada, donde se aprecia que se ha eliminado el componente estacional, junto a una reducción en los patrones de la serie, observando que la serie no cuenta con una tendencia, lo que indica que no se necesita realizar una desestacionalización regular, además de solo realizar una diferenciación debido a que la serie tiene un comportamiento línea, y por lo tanto la serie no necesita más diferenciaciones ni regulares ni estacionarias.

**Serie diferenciada (D=1):
Nivel del reservorio de la Presa Hidroeléctrica Cerrón Grande de El Salvador
del año 2000 al 2018**

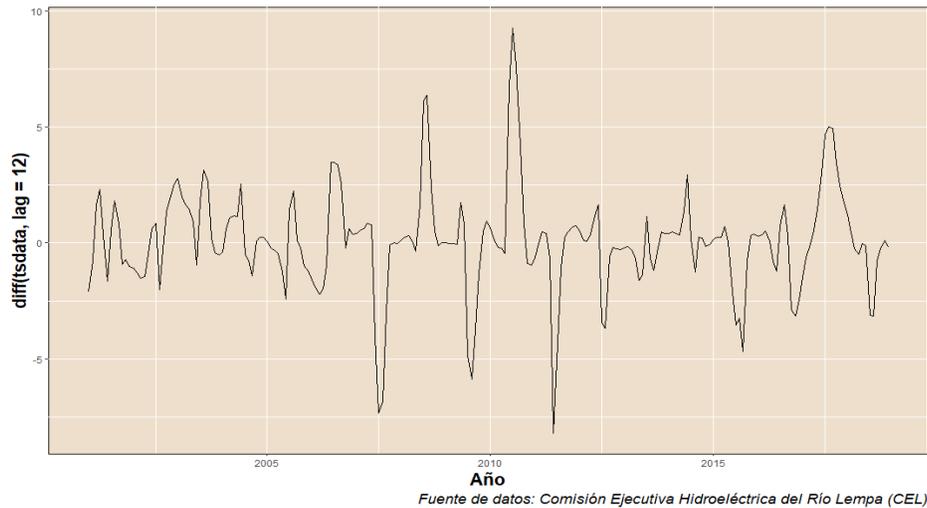
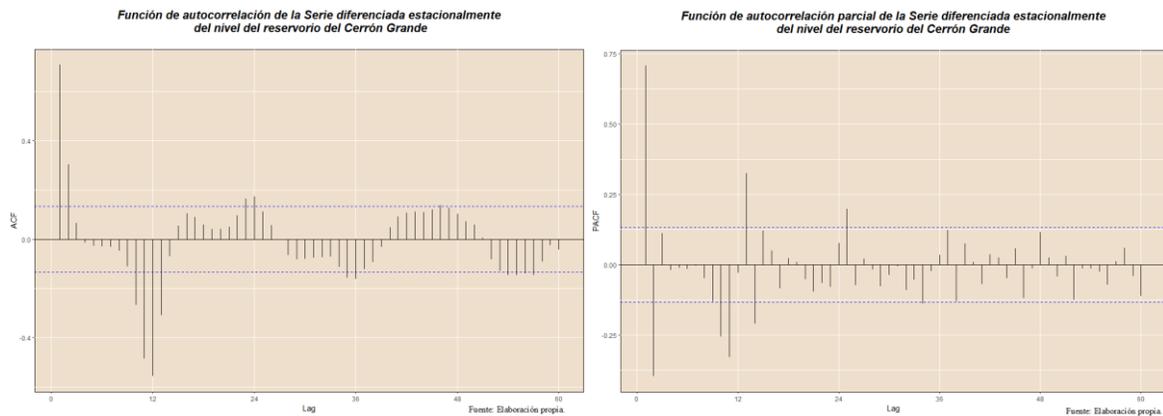


Figura 40: Serie diferenciada estacional (D=1) de los niveles del reservorio de la presa Cerrón Grande

5.3.2.3. Identificación de los posibles modelos

Para realizar la identificación de los posibles modelos, es necesario construir las gráficas de la función de autocorrelación simple y parcial de la serie diferenciada estacionalmente, proveniente de los datos originales de la serie, tal como se muestra en la figura 41.



(a) FAS de la serie nivel diferenciada estacionalmente

(b) FAP de la serie nivel diferenciada estacionalmente

Figura 41: Funciones de autocorrelación de la serie diferenciada estacionalmente del nivel del reservorio del Cerrón Grande

Mediante el comportamiento de la Función de Autocorrelación Simple se puede observar que presenta una oscilación decreciente entre correlaciones positivas y negativas, junto a la existencia de dos retardos significativos con una tendencia lenta a la anulación al aumentar el número de retardos. Por otra parte, en la Función de Autocorrelación Parcial posee una tendencia lenta a anularse, teniendo 1 retardo significativo.

A partir del análisis de las autocorrelaciones, se puede establecer que el modelo que se busca es un ARIMA estacional (o SARIMA) que posee la forma $ARIMA(p, d, q)(P, D, Q)_s$, donde

$$ARIMA \quad \underbrace{(p, d, q)}_{\text{Parte regular}} \quad \underbrace{(P, D, Q)}_{\text{Parte estacional}}$$

p, P número de parámetros **AR**

d, D número de diferenciaciones **I**

q, Q número de parámetros **MA**

Expresada por medio de la ecuación:

$$\Phi_p(L)\Phi_p(L)(1 - L^s)^D(1 - L)^d(Y_t - \mu) = \theta_q(L)\theta_q(L^s)\varepsilon_t$$

Para la identificación de los modelos ARIMA se toman en cuenta las funciones de autocorrelación, para la parte de media móvil (MA) se usa la función de autocorrelación simple, mientras que la función de autocorrelación parcial se usa para identificar los valores de la parte del proceso autoregresivo (AR).

Para considerar la parte no estacional (la parte regular) se examinan las primeras correlaciones de éstos correlogramas ($k=1$ a 11) y así determinar el modelo adecuado para $AR(p)$ y $MA(q)$. Además, cuando se considera la parte estacional, se verifican en los rezagos múltiplos de s (*para* $s=12$, $k=24$, $K=36$) y se determina el modelo adecuado para $AR(P)_s$ y $MA(Q)_s$.

Verificando las correlogramas presentados en la figura 41, iniciando con la parte regular se puede tomar en cuenta por medio de la FAS un $MA(1)$ o un $MA(2)$ debido a que existen dos rezagos que salen del límite de confianza, además por medio de la FAP se puede tomar en cuenta un $AR(1)$ o $AR(2)$ debido a la existencia de dos retardos que sobresalen. Para la parte estacional también se puede tomar un $MA(1)_{12}$, $MA(2)_{12}$ o un $MA(3)_{12}$, debido a que en los rezagos $k = 12, k = 24, K = 36$.

A partir de eso se identificaron algunos modelos ARIMA, que pueden ser propuestos para realizar predicciones; sin embargo, se tomaron en cuenta los siguientes modelos:

- $ARIMA(2,0,0)(0,1,1)_{12}$
- $ARIMA(2,0,0)(1,1,1)_{12}$
- $ARIMA(2,0,1)(0,1,1)_{12}$
- $ARIMA(1,0,2)(1,1,1)_{12}$
- $ARIMA(2,0,2)(0,1,1)_{12}$
- $ARIMA(2,0,2)(1,1,0)_{12}$

5.3.3. Estimación del modelo

Para realizar la estimación de los modelos identificados, se puede verificar la significancia estadística de los parámetros estimados, así como examinar la estacionariedad e invertibilidad de dichos parámetros.

5.3.3.1. Estimación de los parámetros

En este apartado se estiman los parámetros de los modelos identificados y propuestos anteriormente:

Modelo Integrado de Medias Móviles estacional, ARIMA (2,0,0)(0,1,1)₁₂

La ecuación generada de este modelo viene dada por:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})w_t = (\theta_1 B^{12} + 1)v_t$$

La salida que se obtiene al correr el modelo en RStudio se muestra a continuación:

Tabla 16: Parámetros del modelo ARIMA (2,0,0)(0,1,1)₁₂

Parámetro		Estimación	DE	Prueba Z	p-valor
AR	Lag 1	1.032592	0.062256	16.5863	$2.2e^{-16}$
	Lag 2	-0.399486	0.061951	-6.4484	$1.13e^{-10}$
Diferencia Estacional		1			
MA Estacional	Lag1	-0.999968	0.092963	-10.7567	$2.2e^{-16}$

Fuente: Elaboración propia

Los componentes AR(1) y AR(2), con sus p-valor $2.2e^{-16}$ y $1.13e^{-10}$, respectivamente, junto con el componente estacional de medias móviles (SMA(1)), poseen un menor valor menor a 0.05. Estos parámetros se ajustan mejor al modelo propuesto al cumplir con los requerimientos, el cual quedaría de la siguiente forma:

$$(1 - (1.032592)B + (0.399486)B^2)(1 - B^{12})w_t = (1 - 0.999968B^{12})v_t$$

Modelo Integrado de Medias Móviles estacional, ARIMA (2,0,0)(1,1,1)₁₂

La ecuación generada de este modelo viene dada por:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^{12})(1 - B^{12})w_t = (\theta_1 B^{12} + 1)v_t$$

Al evaluar los parámetros que se obtiene en la tabla 17, se puede reducir el modelo a la siguiente forma:

$$(1 - (1.037283)B + (0.40065)B^2)(1 - B^{12})w_t = v_t$$

La salida que se obtiene al correr el modelo en RStudio se muestra a continuación:

Tabla 17: *Parámetros del modelo ARIMA (2,0,0)(1,1,1)₁₂*

Parámetro		Estimación	DE	Prueba Z	p-valor
AR	Lag 1	1.037283	0.06269	16.5462	$2.2e^{-16}$
	Lag 2	-0.40065	0.061942	-6.4682	$1.13e^{-10}$
Diferencia Estacional		1			
MA Estacional	Lag 1	-0.045707	0.073754	-0.6197	0.5354
	Lag 2	-0.999951	0.112369	-8.8988	$2.2e^{-16}$

Fuente: Elaboración propia

Modelo Integrado de Medias Móviles estacional, ARIMA (2,0,1)(0,1,1)₁₂

La ecuación generada de este modelo viene dada por:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})w_t = (1 + \theta_1 B + \theta_2 B^{12})v_t$$

La salida que se obtiene al correr el modelo en RStudio se muestra a continuación:

Tabla 18: *Parámetros del modelo ARIMA (2,0,1)(0,1,1)₁₂*

Parámetro		Estimación	DE	Prueba Z	p-valor
AR	Lag 1	0.84052	0.13642	6.1612	$7.22e^{-10}$
	Lag 2	-0.25971	0.11366	-2.285	0.02231
Ma	Lag 1	0.23378	0.13442	1.7392	0.08199
Diferencia Estacional		1			
MA Estacional	Lag 1	-0.99997	0.08777	-11.3931	$2.2e^{-16}$

Fuente: Elaboración propia

Quedando el modelo de la siguiente forma:

$$(1 - (0.84052)B + (0.25971)B^2)(1 - B^{12})w_t = (1 + (0.23378)B - (0.99997)B^{12})v_t$$

Modelo Integrado de Medias Móviles estacional, ARIMA (1,0,2)(1,1,1)₁₂

La salida que se obtiene al correr el modelo en RStudio se muestra en la tabla 19, de donde se establece la forma del modelo siguiente:

$$(1 - (0.35651)B)(1 - B^{12})w_t = (1 + (0.729857)B + (0.279617)B^2 - (0.99996)B^{12})v_t$$

Tabla 19: *Parámetros del modelo ARIMA (1,0,2)(1,1,1)₁₂*

Parámetro		Estimación	DE	Prueba Z	p-valor
AR	Lag 1	0.356515	0.121741	2.9285	0.003406
Ma	Lag 1	0.729857	0.122103	5.9774	2.267e ⁻⁰⁶
	Lag 2	0.279617	0.104359	2.6794	0.007376
Ar Estacional	Lag1	-0.053528	0.073797	-0.7253	0.468247
Diferencia Estacional		1			
MA Estacional	Lag 1	-0.99996	0.101	-9.9005	2.2e ⁻¹⁶

Fuente: Elaboración propia

La ecuación generada de este modelo viene dada por:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})w_t = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^{12})v_t$$

Modelo Integrado de Medias Móviles estacional, ARIMA(2,0,2)(0,1,1)₁₂

La ecuación generada de este modelo viene dada por:

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B^{12})w_t = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^{12})v_t$$

De donde se establece el modelo de la forma:

$$(1 - B^{12})w_t = (1 - 0.999973B^{12})v_t$$

Y la salida que se obtiene al correr el modelo en RStudio se muestra en la tabla 20.

Tabla 20: *Parámetros del modelo ARIMA(2,0,2)(0,1,1)₁₂*

Parámetro		Estimación	DE	Prueba Z	p-valor
AR	Lag 1	0,381	0,36513	1,0435	0,29673
	Lag 2	-0,018583	0,232517	-0,0799	0,9363
Ma	Lag 1	0,69985	0,35944	1,9471	0,05153
	Lag 2	0,261702	0,174537	1,4994	0,13377
Diferencia Estacional		1			
MA Estacional	Lag 1	-0,999973	0,08484	-11,7866	$2.2e^{-16}$

Fuente: Elaboración propia

Modelo Integrado de Medias Móviles estacional, ARIMA (2,0,2)(1,1,0)₁₂

La ecuación generada de este modelo viene dada por:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^{12})(1 - B^{12})w_t = (1 + \theta_1 B + \theta_2 B^2 - \theta_3 B^{12})v_t$$

La salida se muestra en la tabla 21, generando la forma del modelo:

$$(1 - B^{12})w_t = (1 - (0.541775)B^{12})v_t$$

Tabla 21: *Parámetros del modelo ARIMA (2,0,2)(1,1,0)₁₂*

Parámetro		Estimación	DE	Prueba Z	p-valor
AR	Lag 1	0,436654	0,393827	1,1087	0,29675
	Lag 2	-0,082847	0,23166	-0,3576	0,7206
Ma	Lag 1	0,578814	0,390026	1,484	0,1378
	Lag 2	0,21296	0,187414	1,1363	0,2558
Diferencia Estacional		1			
MA Estacional	Lag 1	-0,541775	0,055815	-9,7066	$2.2e^{-16}$

Fuente: Elaboración propia

Después de revisar la significancia de los parámetros de los modelos identificados y propuestos, se puede expresar los modelos con sus respectivas expresiones donde se han excluido los parámetros que no resultaron ser significativos.

Tabla 22: *Expresiones de los modelos ARIMA identificados*

N.º	Modelo identificado	Expresión
1	ARIMA (2,0,0)(0,1,1) ₁₂	$w_t = \frac{(1 - 0.999B^{12})v_t}{(1 - 1.0326B + 0.399B^2)(1 - B^{12})}$
2	ARIMA (2,0,0)(1,1,1) ₁₂	$w_t = \frac{v_t}{(1 - 1.037B + 0.401B^2)(1 - B^{12})}$
3	ARIMA (2,0,1)(0,1,1) ₁₂	$w_t = \frac{(1 + 0.234B - 0.999B^{12})v_t}{(1 - 0.841B + 0.259B^2)(1 - B^{12})}$
4	ARIMA (1,0,2)(1,1,1) ₁₂	$w_t = \frac{(1 + 0.730B + 0.279B^2 - 0.99996B^{12})v_t}{(1 - 0.357B)(1 - B^{12})}$
5	ARIMA (2,0,2)(0,1,1) ₁₂	$w_t = \frac{(1 - 0.999973B^{12})v_t}{(1 - B^{12})}$
6	ARIMA (2,0,2)(1,1,0) ₁₂	$w_t = \frac{(1 - 0.542B^{12})v_t}{(1 - B^{12})}$

Fuente: Elaboración propia

5.3.3.2. Estacionariedad e invertibilidad modelos ARIMA

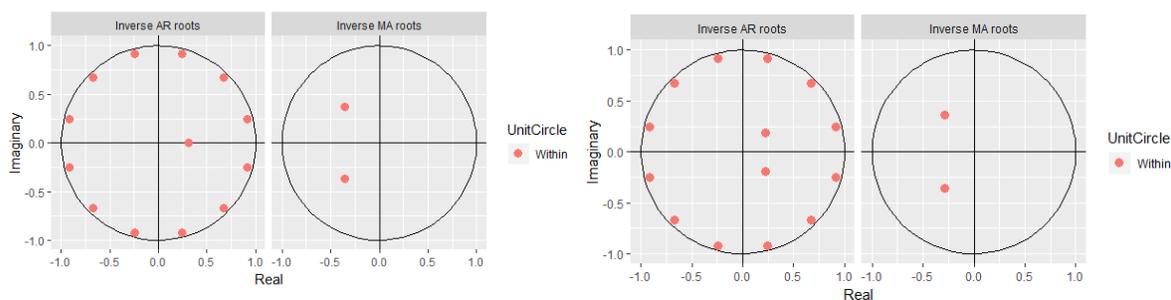
Para realizar el proceso de la revisión de los modelos identificados y seleccionar el modelo adecuado capaz de realizar predicciones con alto grado de precisión. Dado que previamente se ha revisado cuales son los estimadores de los parámetros adecuados en cada modelo propuesto, es necesario verificar si dichos modelos cumplen el proceso de estacionariedad e invertibilidad, presentados en la tabla 23.

Tabla 23: Criterios de Estacionariedad e invertibilidad de los parámetros de los modelos ARIMA identificados

N.º	Modelo	Estacionariedad	Invertibilidad
1	ARIMA (2,0,0)(0,1,1) ₁₂	Debe de cumplirse que $ \phi_1 + \phi_2 < 1$. Para este caso $ 1.0325 - 0.3994 < 1$, teniendo $ 0.6331 < 1$, cumpliendo que la serie es estacionaria	En este caso debe de cumplirse que $ \theta_1 < 1$, teniendo que $ 0.9999 < 1$, donde el modelo se puede decir que es invertible.
2	ARIMA (2,0,0)(1,1,1) ₁₂	Se cumple que $\phi_1 + \phi_2 < 1$, ya que $ 1.0372 - 0.4 < 1$, donde $ 0.6372 < 1$, por tanto, la serie es estacionaria.	Debe de cumplirse que $ \theta_1 + \theta_2 < 1$, es decir: $ -0.0457 - 0.9999 < 1$, donde $ -1.0456 > 1$ concluyendo que el proceso no es invertible.
3	ARIMA (2,0,1)(0,1,1) ₁₂	& La series estacionaria debido a que cumple que $ \phi_1 + \phi_2 + \phi_3 < 1$, ya que $ 0.84 - 0.2597 + 0.2337 < 1$, donde $ 0.814 < 1$.	Aunque el proceso se puede decir que es invertible, aunque el valor de $ \theta_1 $ es de $ -0.9999 $
4	ARIMA (1,0,2)(1,1,1) ₁₂	& Se debe de cumplir que $ \phi_1 < 1$ y $ \phi_2 < 1$, teniendo que $ 0.356 < 1$ y $ -0.0535 < 1$, cumpliendo que la serie si es estacionaria.	El proceso se puede decir que se cumple la invertibilidad
5	ARIMA (2,0,2)(0,1,1) ₁₂	Dado que se cumple $ \phi_1 + \phi_2 < 1$, debido a que $ 0.381 - 0.018 < 1$ resulta $ 0.363 < 1$, el proceso es estacionario.	El proceso se puede decir que es invertible, aunque el valor de $ \theta_1 $ es de $ -0.9999 $
6	ARIMA (2,0,2)(1,1,0) ₁₂	Para que el proceso sea estacionario debe de cumplirse que $ \phi_1 + \phi_2 < 1$, en este caso $ 0.3538 < 1$ con lo que se puede concluir que el modelo es estacionario	Para que sea invertible se debe de cumplir que $ \theta_1 + \theta_2 < 1$ y que $ \theta_3 < 1$, teniendo que $ 0.7917 < 1$ y $ -0.5417 < 1$, donde el modelo es invertible

Fuente: Elaboración propia

A partir de los resultados obtenidos, los modelos 4 y 6 son los que cumplen las condiciones necesarias de este proceso. Para los modelos 1, 3 y 5, el criterio de la invertibilidad es cercano a 1, por tanto, no es recomendable trabajar dichos modelos. También se puede observar cómo los modelos seleccionados cumplen estos criterios mediante el círculo unitario, en la figura 42 reflejando que las raíces se encuentran dentro del círculo de la parte regresiva cumpliendo la estacionariedad. Y en la parte media móvil existen raíces fuera del círculo, cumpliéndose la parte invertible.



(a) Modelo: ARIMA(1,0,2)(1,1,1)₁₂

(b) Modelo: ARIMA(2,0,2)(1,1,0)₁₂

Figura 42: Invertibilidad y estacionariedad de los modelos seleccionados

Se puede verificar con los valores de los métodos AIC y BIC cual de los dos modelos seleccionados se pueden elegir para continuar el proceso de la verificación. A partir de la tabla 24 se observa que el modelo 4, posee el valor del AIC.

Tabla 24: Análisis autocorrelativos de los modelos propuestos

N.º	Modelo identificado	AIC	BIC
1	ARIMA (2,0,0)(0,1,1) ₁₂	695.9544	709.4555
2	ARIMA (2,0,0)(1,1,1) ₁₂	695.9544	709.4555
3	ARIMA (2,0,1)(0,1,1) ₁₂	768.0005	781.5017
4	ARIMA (1,0,2)(1,1,1) ₁₂	695.4914	712.3678
5	ARIMA (2,0,2)(0,1,1) ₁₂	697.3665	717.6181
6	ARIMA (2,0,2)(1,1,0) ₁₂	697.3665	717.6181

Fuente: Elaboración propia

5.3.4. Diagnósis del modelo

5.3.4.1. Análisis de los residuos

En esta etapa se procede a realizar el análisis de los residuos que se obtiene en el modelo $ARIMA(1,0,2)(1,1,1)_{12}$, el cual debe de poseer una aproximación al comportamiento de un ruido blanco, en donde se busca la presencia de una media con valor cero, con varianza que sea constante y que posea una incorrelación dentro de los residuos. Por tanto, mediante la gráfica de los residuos de la serie se verifica si los datos se encuentran alrededor de una media cero.

Residuos del modelo $ARIMA(1,0,2)(1,1,1)$ asociados a la serie Nivel del reservorio del Cerrón Grande

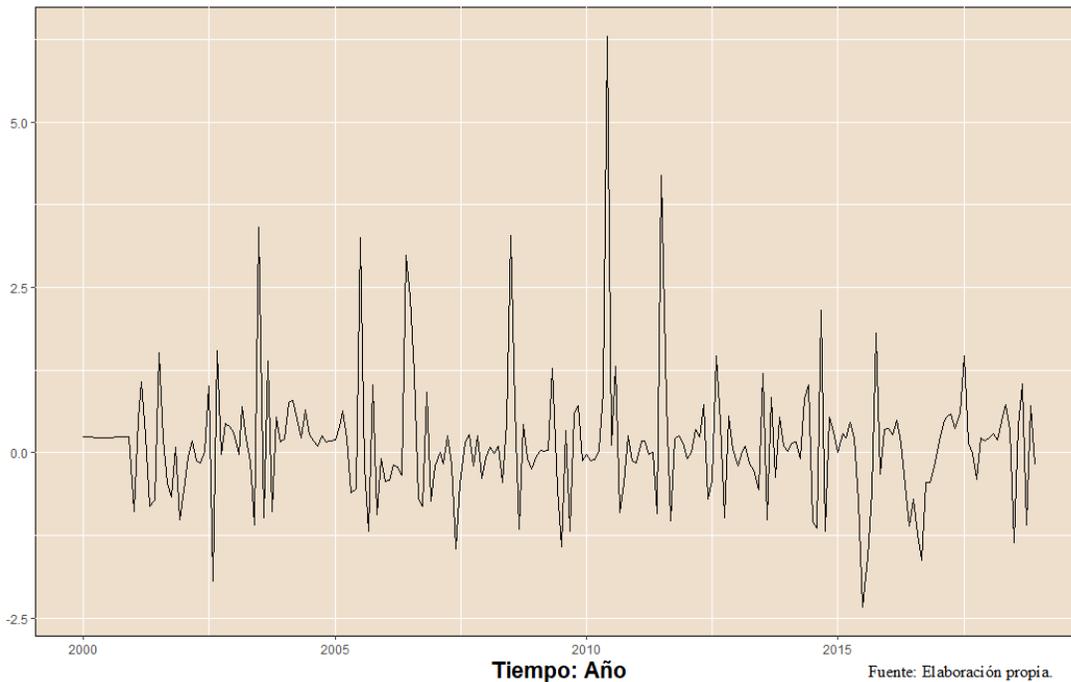
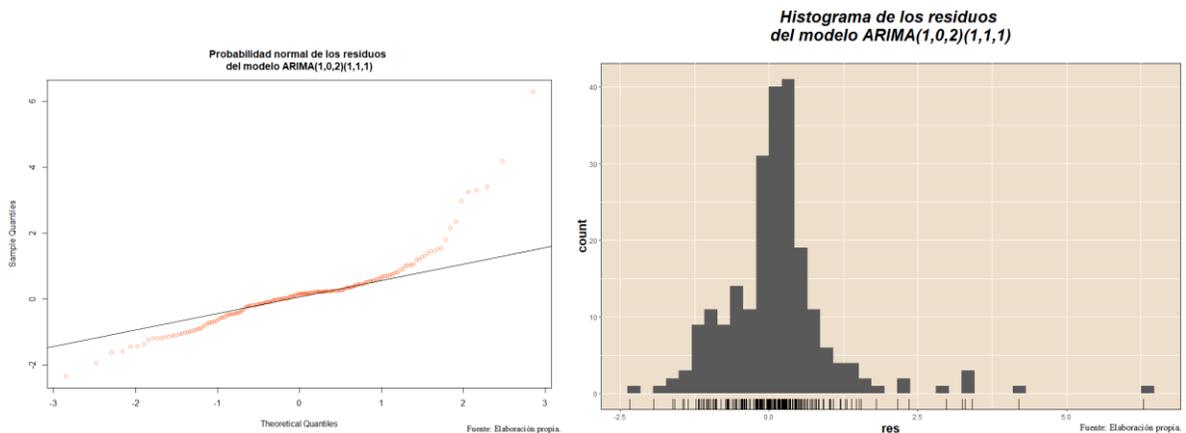


Figura 43: Residuos del modelo $ARIMA(1,0,2)(1,1,1)$, de la serie nivel del reservorio

Mediante la gráfica de los residuos puede concluirse que la media de los datos es cero ($\mu = 0$), esto gracias a que se observa el comportamiento alrededor de dicho valor. Sin embargo, existen datos que poseen valores altos en comparación con los demás residuos que se presentan en dicha figura.

Además, es de resaltar que con las gráficas del histograma de los residuos y la de probabilidad normal de los mismos se logra observar que los datos poseen comportamiento aproximadamente normal.

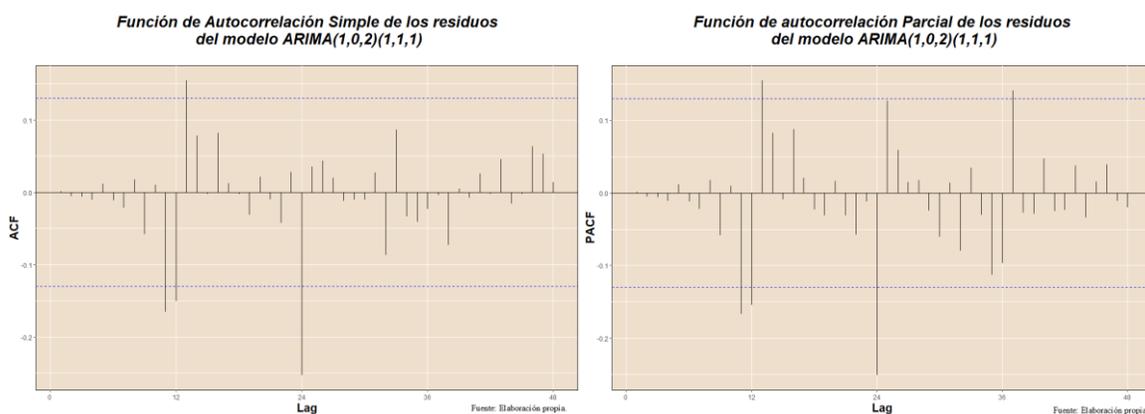


(a) Probabilidad normal de los residuos del modelo $ARIMA(1,0,2)(1,1,1)_{12}$

(b) Histograma de los residuos del modelo $ARIMA(1,0,2)(1,1,1)_{12}$

Figura 44: Normalidad de los residuos del modelo $ARIMA(1,0,2)(1,1,1)$

Para establecer si no existe correlación entre los residuos obtenidos del modelo se debe de revisar cual es el comportamiento que poseen las autocorrelaciones simples y de las autocorrelaciones parciales, es por ello que se muestran a continuación dichas gráficas.



(a) Autocorrelación simple de los residuos del modelo $ARIMA(1,0,2)(1,1,1)_{12}$

(b) Autocorrelación parcial de los residuos del modelo $ARIMA(1,0,2)(1,1,1)_{12}$

Figura 45: Funciones de autocorrelación del modelo $ARIMA(1,0,2)(1,1,1)$

Se verifica que los rezagos de las autocorrelaciones simples y parciales se encuentran en el intervalo de aceptación propuesto, es decir, aproximadamente las autocorrelaciones de los residuos se encuentran dentro de dicho intervalo. Por lo tanto, se puede establecer que las autocorrelaciones de los residuos son no significativas y se acepta que no existe correlación entre los residuos del modelo propuesto.

Antes de llevar a cabo la predicción por medio del modelo ARIMA, es necesario considerar el comportamiento de los residuos para conocer si éstos se comportan como ruido blanco. Para analizar esto se realiza la prueba de autocorrelación de forma conjunta de Ljung-Box. El cual propone las siguientes hipótesis:

H_0 : los residuos son un ruido blanco ($\epsilon_t \sim RB(0, \sigma^2)$).

H_a : los residuos no son ruido blanco.

Se puede observar en la siguiente tabla el resumen del test de Ljung-Box.

Tabla 25: *Ljung-Box Test*

Ljung-Box Test			
Test	X-squared	df	Valor-p
Box-Pierce	1.0845	10	0.9998

Fuente: Elaboración propia

Dado que la prueba del p-valor es mayor que 0.05 que es el nivel de significancia de la prueba, se puede asumir que los valores del residuo son ruido blanco, y por tanto exista la incorrelación en los residuos indicando que no son dependientes.

5.3.4.2. Análisis de la Serie Temporal con Outliers

Residuos del modelo ARIMA(1,0,2)(1,1,1) asociados a la serie Nivel del reservorio del Cerrón Grande

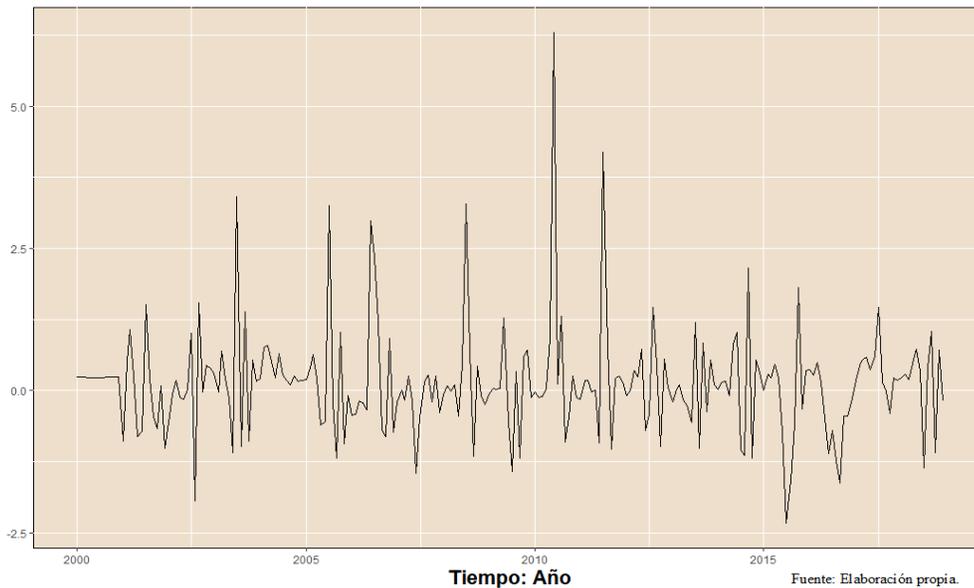


Figura 46: Residuos del modelo ARIMA(1,0,2)(1,1,1) de la serie nivel del reservorio

De la figura 46, donde se presentan los residuos asociados al modelo de la serie temporal que posee un proceso de diferenciación estacional, en la que se detecta posibles outliers (datos atípicos) que pueden representar un cambio significativo en los parámetros del modelo ARIMA y de esta manera generar un modelo de predicción que no sea capaz de ajustarse al comportamiento de los datos originales

De lo anterior, cuando se detectan outliers en el proceso de construcción de una serie temporal es necesario realizar un análisis adicional, a fin de determinar si dichos parámetros resultan ser significativos para éste. Por tanto, es preciso la aplicación de un análisis de intervención para ajustar el comportamiento de los residuos consintiendo en:

1. Identificación del modelo ARIMA,
2. Detección de posibles outliers y sus respectivos tipos (Innovador, transitorio, Aditivo),
3. Análisis de los residuos con intervención,

4. Comparación del modelo con intervención y sin intervención.

Análisis de intervención del modelo ARIMA

Tabla 26: *Modelo de la serie temporal*

Descripción del modelo	
Tipo de modelo	
Modelo_1	ARIMA(1,0,2)(1,1,1)

Fuente: Elaboración propia

En el cuadro 26, se observa el modelo ARIMA(1,0,2)(1,1,1) que cumple con los criterios de selección del modelo mejor ajustado, con componentes autorregresivas de orden uno y una componente estacional de orden uno, con una diferenciación en la componente estacional de orden uno, además una componente de media móvil en la parte autorregresiva de orden dos y una componente estacional de orden uno.

Tabla 27: *Valores atípicos del modelo ARIMA(1,0,2)(1,1,1) de la serie temporal*

Valores atípicos				
	Tipo	t	Sig.	
nivel-Modelo_1	jul-03	Transitorio	4.109	0.63
			0.861	0.055
	jul-05	Innovador	4.29	0.718
	jun-06	Innovador	3.252	0.688
	jul-06	Innovador	3.474	0.758
	jul-08	Innovador	4.384	0.693
	jun-10	Aditivo	6.136	0.65
	jul-10	Aditivo	6.727	0.915
	ago-10	Aditivo	3.951	0.652
	jul-11	Innovador	5.225	0.693
	jul-13	Aditivo	1.446	0.402
	ago-14	Aditivo	-1.616	0.4

Fuente: Elaboración propia

Mediante la tabla 27, se verifica la presencia de 11 outliers principalmente en los meses de junio, julio y agosto, entre los años 2003, 2005, 2006, 2008, 2011, 2013, 2014, de tipo; transitorio, Innovador, Aditivo. Por tanto, se debe aplicar una intervención para establecer si estos valores atípicos interfieren en el modelo de predicción.

Al momento de presentar las funciones de autocorrelación simple y parcial del modelo propuesto, en la figura 47 se observa que los residuos se mantienen dentro del intervalo de confianza que presenta las dos funciones, identificando que los residuos un ruido blanco.

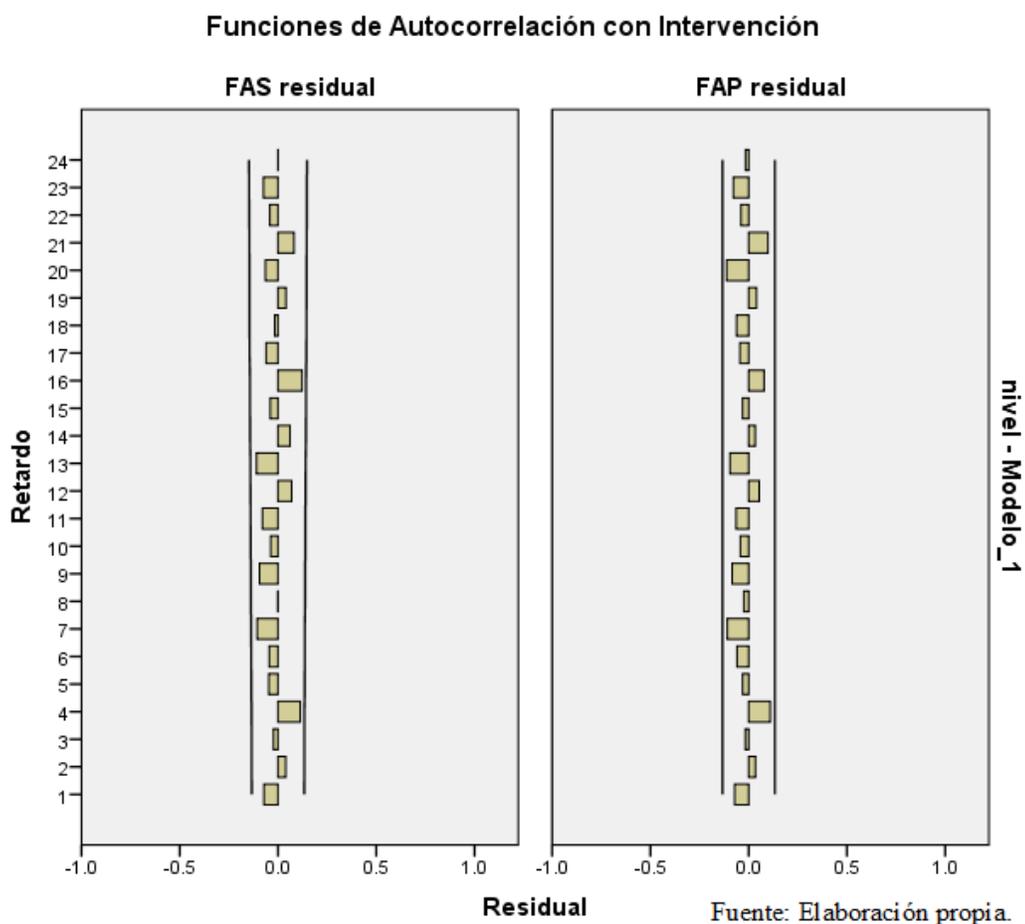


Figura 47: Funciones de Autocorrelación del modelo ARIMA(1,0,2)(1,1,1) con análisis de intervención

Además, se puede hacer una comparación entre los estadísticos del modelo ARIMA(1,0,2)(1,1,1) con el análisis de intervención y con intervención.

Tabla 28: Resumen del modelo ARIMA(1,0,2)(1,1,1) sin intervención

Ajuste del modelo ARIMA(1,0,2)(1,1,1) sin intervención			
Estadísticos de ajuste	Media	Mínimo	Máximo
R cuadrado estacionaria	0.898	0.898	0.898
R cuadrado	0.968	0.968	0.968
RMSE	0.73	0.73	0.73
MAPE	0.205	0.205	0.205
MaxAPE	1.061	1.061	1.061
MAE	0.486	0.48	0.48
MaxAE	2.473	2.473	2.473
BIC normalizado	-0.156	-0.156	-0.156

Fuente: Elaboración propia

Los datos contenidos en la tabla 28, presenta la información del modelo sin realizar intervención, que muestra un coeficiente de correlación alto, (96.8%) lo cual expresa una explicación excelente por parte de dicho modelo. Además, es de mencionar que se observa en la tabla 29, se ha colocado la información del modelo ajustado con la intervención.

Tabla 29: Resumen del modelo ARIMA(1,0,2)(1,1,1) con intervención

Ajuste del modelo ARIMA(1,0,2)(1,1,1) sin intervención			
Estadísticos de ajuste	Media	Mínimo	Máximo
R cuadrado estacionaria	0.869935049	0.869935049	0.869935
R cuadrado	0.959350851	0.959350851	0.9593509
RMSE	0.823709877	0.823709877	0.8237099
MAPE	0.227449823	0.227449823	0.2274498
MaxAPE	1.527485915	1.527485915	1.5274859
MAE	0.539944208	0.539944208	0.5399442
MaxAE	3.628109182	3.628109182	3.6281092
BIC normalizado	0.060066063	0.060066063	0.0600661

Fuente: Elaboración propia

Sin embargo, los parámetros estadísticos de ajuste que presenta el modelo que se encuentra intervenido, tienen una cercanía con los valores de los mismos parámetros, pero que no se ha realizado la intervención, por ende, es posible que el pronóstico del nivel del reservorio con el modelo ARIMA(1,0,2)(1,1,1) seleccionado no cambia, por tanto para la etapa de predicción se toma el modelo sin intervención.

2.3.5. Pronóstico con el Modelo de Series de Tiempo

Anteriormente se ha obtenido el modelo ARIMA(1,0,2)(1,1,1) adecuado para la predicción y además se comprobó su idoneidad para poder predecir, tomando dicho modelo sin intervención, debido a que se observó que no existe diferencia significativa con el modelo con intervención. Y se puede observar una comparación entre los resultados que arrojó el modelo con los valores reales en la tabla 30.

Tabla 30: *Valores reales vs. valores pronosticados de los niveles del reservorio con series temporales*

Comparación de los valores reales con los Valores pronósticos del modelo de serie de tiempo		
Mes	Valores reales 2018	Valores Pronosticados 2018
Enero	240,78	240.67
Febrero	239,39	239.27
Marzo	237,39	237.40
Abril	235,16	234.85
Mayo	232,96	231.42
Junio	233,83	233.21
Julio	234,19	235.14
Agosto	236,03	235.94
Septiembre	240,86	239.89
Octubre	242,51	243.71
Noviembre	242,44	241.81
Diciembre	241,44	241.97

Fuente: Elaboración Propia

Dado los resultados de la tabla 30 se puede observar que los valores arrojados en el último año por el modelo de serie de tiempo tienen una semejanza con los datos reales, determinado que los valores del modelo son menores con respecto a los valores reales presentando una disminución leve entre ambos, en consideración a esto el modelo planteado se ajusta a la serie original del nivel del reservorio.

Posteriormente se pueden presentar los valores pronosticados mediante el modelo que se ha elegido, observando en la figura 48, la predicción del nivel del reservorio desde el año 2019 hasta el 2021, donde el comportamiento que tiene la variable al respecto de los años que se han pronosticados posee oscilaciones periódicas parecidas a los años anteriores, teniendo un comportamiento lineal con una media constante con una varianza constante manteniéndose dentro del rango de niveles máximos y mínimos que presenta la serie original.

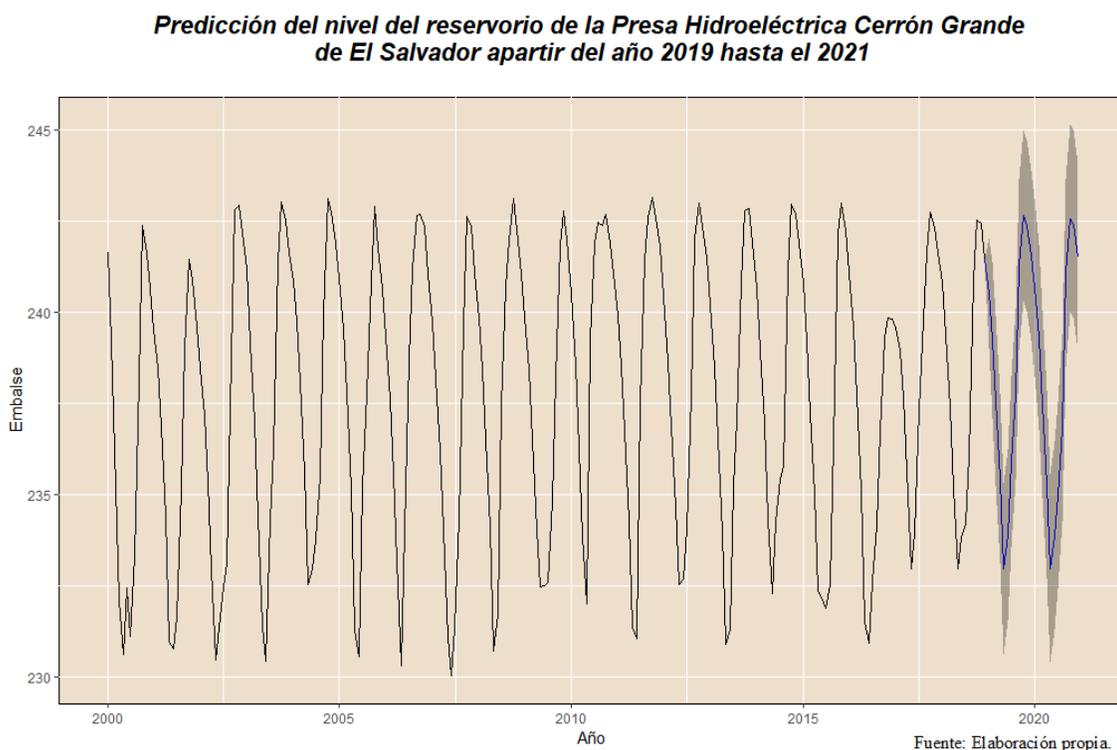


Figura 48: Pronósticos del nivel del reservorio del Cerrón Grande generados con el modelo de series temporales

Además, los valores de los pronósticos del nivel del reservorio para el periodo de los años 2019, 2020 y 2021, se observan en la tabla 31, donde se establece fluctuaciones alternantes entre niveles altos y bajos de agua en la presa, sin embargo, dichos datos pueden estar influenciados por factores externos que de alguna manera pueden generar variaciones a corto mediano o largo plazo.

Tabla 31: *Valores pronosticados con el modelo de series de tiempo del nivel del reservorio del 2019 al 2021*

Valores pronosticados para el nivel del reservorio para los años del 2019 al 2021				
Mes	2018	2019	2020	2021
Enero	240,78	240,5112	240,6560	240,5774
Febrero	239,39	239,2371	239,3180	239,2741
Marzo	237,39	237,5607	237,4683	237,5184
Abril	235,16	235,4333	235,2832	235,3646
Mayo	232,96	232,9624	232,9597	232,9611
Junio	233,83	233,8855	233,8541	233,8711
Julio	234,19	235,8776	234,9614	235,4583
Agosto	236,03	237,7328	236,8096	237,3103
Septiembre	240,86	241,2959	241,0620	241,1889
Octubre	242,51	242,6380	242,5668	242,6054
Noviembre	242,44	242,3696	242,4080	242,3872
Diciembre	241,44	241,5438	241,4875	241,5180

Fuente: Elaboración Propia

5.4. Comparación de modelos

Con el objetivo de proporcionar una forma más factible de comparar el rendimiento predictivo de las dos aproximaciones, se muestra en la tabla el MAPE (Error medio del porcentaje absoluto) asociado a cada estimación, de esta manera se tiene un indicador fiable para escoger el mejor

Tabla 32: Comparación de los valores pronosticados de los modelos de pronóstico con los valores reales

Valores pronosticados con el modelo de red neuronal y de la serie temporal en los años 2017 y 2018						
Año	Mes	Nivel	Nivel (REDES)	MAPE (REDES)	Nivel (ARIMA)	MAPE (ARIMA)
2017	Ene	239,563	240,356	0,0033	239,556	0,0000
	Feb	238,935	239,898	0,0040	238,584	0,0015
	Mar	237,650	239,063	0,0059	237,279	0,0016
	Abr	235,650	236,533	0,0037	235,207	0,0019
	May	232,961	233,082	0,0005	232,798	0,0007
	Jun	233,933	233,368	0,0024	233,009	0,0040
	Jul	237,302	237,141	0,0007	235,322	0,0083
	Ago	239,169	238,344	0,0034	239,498	0,0014
	Sep	241,660	240,359	0,0054	241,566	0,0004
	Oct	242,749	242,110	0,0026	243,437	0,0028
	Nov	242,310	241,472	0,0035	242,189	0,0005
	Dic	241,631	240,886	0,0031	241,659	0,0001
2018	Ene	240,778	240,539	0,0010	240,674	0,0004
	Feb	239,386	239,551	0,0007	239,273	0,0005
	Mar	237,390	237,090	0,0013	237,398	0,0000
	Abr	235,157	234,039	0,0048	234,851	0,0013
	May	232,957	231,298	0,0071	232,425	0,0023
	Jun	233,828	232,327	0,0064	233,208	0,0027
	Jul	234,188	237,318	0,0134	235,137	0,0041
	Ago	236,031	235,974	0,0002	235,936	0,0004
	Sep	240,865	240,634	0,0010	239,885	0,0041
	Oct	242,507	242,653	0,0006	243,709	0,0050
	Nov	242,440	242,240	0,0008	241,809	0,0026
	Dic	241,440	241,298	0,0006	241,970	0,0022

Fuente: Elaboración propia

Presentando los dos últimos años de cada uno de los modelos de pronóstico se observa cómo los errores que posee el modelo ARIMA son menores con respecto a la red neuronal, en muchos casos se acerca incluso a los datos originales, sin embargo, hacer la comparación se toma en cuenta la totalidad de los datos, de los cuales se toman el error MAPE y el RMSE (La raíz del error cuadrático medio) por cada modelo. Del cual se presenta el resumen en la tabla 33.

Tabla 33: *Resumen de los errores de los modelos de redes neuronales y series temporales*

Errores de predicción de los modelos		
Errores		
Modelo	RMSE	MAPE
Red Neuronal	0.8445747	0.0025074
Serie de Tiempo	0.823709877	0.3067957

Fuente: Elaboración propia

Mediante los análisis de los métodos de redes neuronales y de la serie de tiempo que se realizan de forma individual, se generaron modelos que permiten realizar pronósticos para los niveles del reservorio de la presa Hidroeléctrica Cerrón Grande de El Salvador, para realizar la comparación y contrastar dichos modelos y determinar cuál modelo es el mejor para dicha tarea. Es por ello que a partir de observar los errores en la tabla 33 de lo generado por las dos técnicas, y se puede elegir como el mejor modelo para realizar los pronósticos de los niveles el generado por la red neuronal, debido a que se posee errores menores que el modelo de la serie temporal.

Además, se observa de modo gráfico en la figura 49 como el modelo que presenta la Red neuronal se ajustan al comportamiento de los niveles originales de los niveles del reservorio, esto, debido a que se han colocado tanto la serie original de los niveles del reservorio, con los datos generado por cada uno de modelos de pronósticos, y en donde se verifica que la serie producida por el modelo de redes neuronales se acopla de mejor manera que el arroja por la serie temporal ARIMA(1,0,2)(1,1,1).

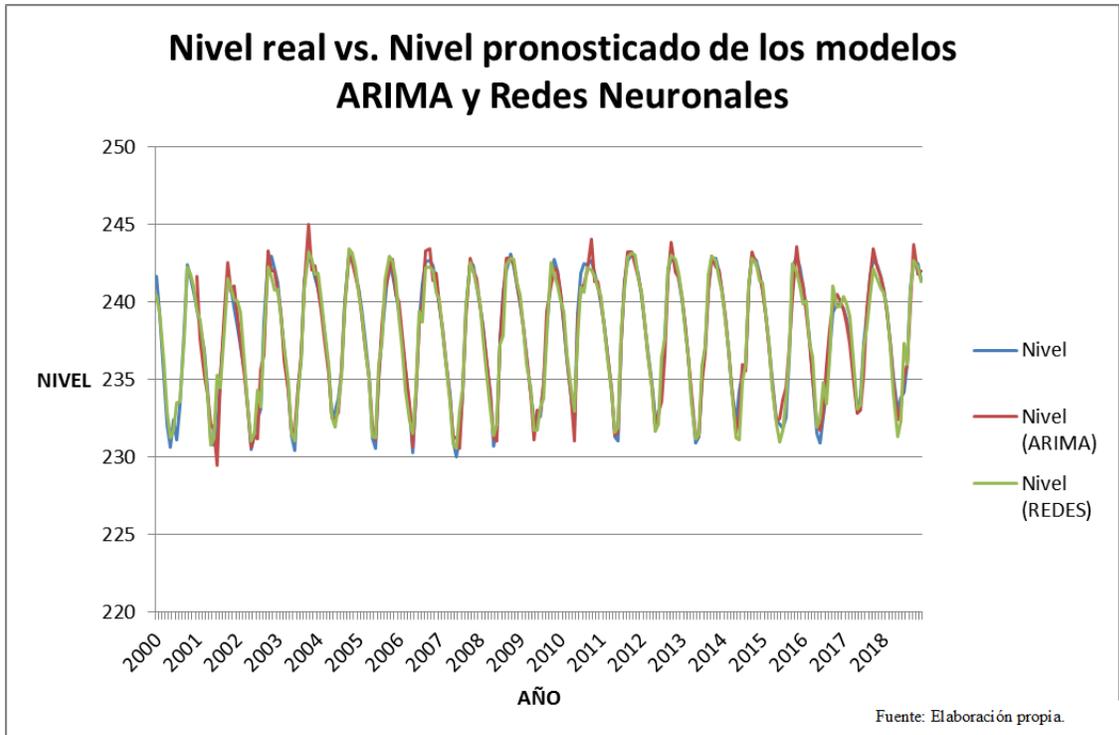


Figura 49: Comparación de los niveles del reservorio con los niveles creados con los modelos de pronóstico

Conclusiones y Recomendaciones

Conclusiones

- Por medio de los métodos redes neuronales y series de tiempo contemplados en el desarrollo de la investigación se determinan y generan modelos estadísticos capaces de ajustarse a los datos y realizar un pronóstico acertada con respecto a los niveles del reservorio, que puede permitir el monitorear los procesos operativos o de descarga dentro de la Presa Hidroeléctrica Cerrón Grande de El Salvador.
- La construcción de los modelos de redes neuronales artificiales para generar pronósticos en el nivel del reservorio de la Presa Hidroeléctrica Cerrón Grande de El Salvador, se determina que los modelos con esta técnica presentan diferencias significativas durante el proceso de evaluación mediante el contraste de los errores de cada uno de los modelos generados, al contrastar las redes neuronales que se activan mediante dos funciones de activación (Sigmoide y tangente hiperbólica).

Además, se comprueba que a pesar del fundamento teórico que sugiere la utilización con más frecuencia las redes neuronales artificiales con funciones de transferencia sigmoide, en la arquitectura perceptrón multicapa, porque tiene mayor versatilidad para explicar los datos, éste tipo de red no resultó ser el adecuado, sin embargo, aunque realiza procesos de aprendizaje con respecto al comportamiento de los niveles del reservorio, al momento de evaluar cada uno de los modelos, se establece como el modelo adecuado para obtener pronósticos se utiliza un modelo de red neuronal artificial que es activado con la función de tangente hiperbólica, permitiendo observar el comportamiento que tiene en comparación con los datos reales proporcionados por la CEL.

- En el análisis de los modelos de pronósticos generados por las series temporales se identificó que la serie presenta anomalías, las cuales fueron tratadas mediante la implementación de la metodología propia de la técnica estadística para determinar el

modelo adecuado. Es por ello que durante el análisis de la serie no se identifican modelos de autoregresión (AR), de media móvil autoregresiva (ARMA), sin embargo, se identifica que el Modelo promedio móvil progresivo autorregresivo estacional (SARIMA) o ARIMA estacional, se establece que es el modelo adecuado para observar y pronosticar el nivel del agua del reservorio de la Presa Hidroeléctrica Cerrón Grande de El Salvador. Estableciendo el modelo estacional ARIMA(1,0,2)(1,1,1), el cual cumplió con los criterios teóricamente establecidos.

- Al contrastar los resultados procedentes de los análisis de redes neuronales y de series de tiempo, se verifica que los procedimientos y los recursos necesario para la implementación de éstos varían significativamente, presentando una diferencia considerable en el ajuste de los pronósticos proporcionado por cada uno de los modelos. Por consiguiente, la selección del mejor para realizar el pronóstico del nivel del reservorio se realiza por medio de la comparación de los errores MAPE y el RMSE, para buscar el menor error, que corresponde al proceso de redes neuronales artificiales.

Recomendaciones

A la Presa Hidroeléctrica Cerrón Grande de El Salvador:

- Implementar el uso de las redes neuronales para conocer los niveles del reservorio y así controlar las descargas que realiza en sus procesos.
- Realizar un análisis independiente tomando valores diarios del nivel del reservorio, para los meses comprendidos entre junio y agosto, correspondientes a los años 2003, 2005, 2006, 2008, 2011, 2013 y 2014, debido a que éstos presentan fluctuaciones diferentes en comparación con respecto a los datos del nivel del reservorio.

A la Universidad de El Salvador:

- Se pueden realizar estudios relacionados con los factores que influyen en el comportamiento del nivel del agua, estudiando las descargas que realiza la Presa Hidroeléctrica, y analizando el flujo de los caudales que alimentan los niveles del reservorio, el cual se encuentra influenciado por las precipitaciones que ocurren en la zona.
- Proponer el uso de la red neuronal para el análisis de los datos mediante el uso de software libres como RStudio, Weka o Phyton que ayuden a implementar técnicas de minería de datos.

Bibliografía

- Arce, J. (2001). *Modelos ARIMA*. España: Depto. econometria aplicada.
- Benítez, R., Escudero, G., & Kanaan, S. (2009). *Inteligencia Artificial Avanzada*. Universidad de Catalunya.
- Berzal, F. (2018). *Redes Neuronales & Deep Learning*. Edición Independiente.
- Bishop, C., & Jordan, M. (1996). *Neural Networks*. CRC Handbook of Computer Science, CRC Press. Tucker (ed).
- Cano, E. A. (2012). *Desarrollo de redes neuronales para el cálculo del transporte neoclásico en reactores de fusión*. Madrid: Escuela Politécnica Superior de la Universidad Carlos III de Madrid.
- Chinrungrueng, C. (1993). *Evaluation of Heterogeneous Architectures of Artificial Neural Networks*. Computer Science Department, University of California at Berkeley.
- Cruz, P. P. (2010). *Inteligencia artificial con aplicaciones a la ingeniería*. México: Alfaomega Grupo Editor.
- Díaz, F. F. (2011). *Selección de modelos mediante criterios de información de análisis factorial: Aspectos teóricos y computacionales*. Universidad de Granada.
- Dorffner, G. (2008). *Neural Network for Times Series Processing*. Department of Medical Cybernetics and Artificial Intelligence, University of Vienna and Austria Research Institute for Artificial Intelligence.
- Falco, I. D., Cioppa, A. D., Natale, P., & Tarantino, E. (1997). *Artificial Neural Networks Optimization by means of Evolutionary Algorithms*. En *Soft Computing in Engineering Design and Manufacturing*. P. Chawdhry, R. Pant, R. Roy (eds). Springer Verla.
- Fayyad, U. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press.

- Fernández, S. P. (2000). Técnicas de regresión: Regresión Lineal Simple. *Cuadernos de atención primaria*, 7, 91-94.
- Gutiérrez, J. M., Cofiño, R. C., & Sordo, C. M. (2004). *Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas*. Madrid: Ministerio de Medio Ambiente.
- Hernández, J. (2007). *Análisis de series temporales económicas II*. España: Esic Editorial.
- Hurtado, C. (2008). *Tópicos en Minería de Datos: Series de Tiempo*. Universidad de Chile.
- Hurtado, J. (2000). *Metodología de la Investigación Holística*. Caracas: Instituto Universitario de Tecnología Caripito.
- Kaastra, L., & Boyd, M. (1996). *Designing a Neural Network for forecasting Financial and Economic Times Series*. Neurocomputing.
- Llamas, R. V., & Ponce, R. A. (2011). *ECONOMETRÍA PRÁCTICA: Fundamentos de Series de Tiempo*.
- Matich, D. J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. Universidad Tecnológica Nacional, Facultad Regional Rosario.
- Ministerio de Agricultura y Riego, M. (2017). *Norma y Reglamento de seguridad de presas en el Perú*. Lima, Perú.
- Mitra, T. A. (2003). *Data mining: multimedia, soft computing and bioinformatics*. John Wiley and Son.
- Mogni, A. P. (2013). *Modelos de Series de Tiempo con aplicaciones en la industria aerocomercial*. Editorial de Buenos aires.
- Moreno, F. L. (2018). *Desarrollo de diferentes modelos predictivos basados en Inteligencia Artificial para su uso en escenarios de inteligencia visual y series temporales financieras*. Escuela Técnica Superior de Ingeniería de Telecomunicación (Universitat Politècnica de València).

- Munt, A. M. (2018). *Introducción a los modelos de redes neuronales artificiales: El perceptrón Simple y Multicapa*. Universidad de Zaragoza.
- Olabe, X. B. (2012). *Redes neuronales artificiales y sus aplicaciones*. Bilbao: Escuela Superior de Ingeniería de Bilbao.
- Ramón, S., & Agüeras, C. (1992). *Textura del sistema nervioso del hombre y de los vertebrados* (Vols. v. 2, parte 2). Verlag nicht ermittelbar. Obtenido de <https://books.google.com.sv/books?id=tb6wmgEACAAJ>
- Roig, J. G., Alfonso, J. M., Quiles, R. C., & Roma, J. C. (2017). *Minería de datos: Modelos y algoritmos*. Editorial UOC.
- Roig, J. G., Alfonso, J. M., Quiles, R. C., & Roma, J. C. (2017). *Minería de datos: Modelos y algoritmos*. Editorial UOC.
- Rumelhart, D., & McClelland, J. L. (1992). *Introducción al Procesamiento Distribuido en Paralelo*. Madrid: Alianza Psicología.
- Serrano, A. J., Soria, E., & Martín, J. D. (2010). *Redes Neuronales Artificiales*. Valencia: Universidad de Valencia.
- Torres, I. (1997). *Aplicaciones de redes neuronales en Economía*. Actas V Jornadas ESEPUMA.
- Üneş, Demirci, F. a., Kişi, M. a., & Özgür. (2015). Prediction of Millers Ferry Dam Reservoir Level in USA Using Artificial Neural Network. *Periodica Polytechnica Civil Engineering*, 59(3), 309-318. doi:<https://doi.org/10.3311/PPci.7379>
- Villavicencio, J. (2010). *Introducción a Series de Tiempo*. Gobierno de Puerto Rico. Obtenido de https://www.estadisticas.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D
- Yao, X. (1993). *Evolutionary Artificial Neural Networks*. En *Encyclopedia of Computer Science and Technology*. Marcel Dekker Inc.
- Yao, X. (1995). *Evolutionary Artificial Neural Networks*. En *International Journal of Neural Systems*. Marcel Dekker Inc.

Anexos

Anexo 1. Datos originales de los niveles del reservorio

Tabla 34: Datos originales de los niveles del reservorio de la presa Hidroeléctrica Cerrón Grande

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Agos	Sep	Oct	Nov	Dic
2000	241.6	239.2	235.2	232.1	230.6	232.4	231.1	233.3	238.1	242.4	241.6	240.6
2001	239.5	238.4	236.8	234.4	231.0	230.8	231.6	235.1	239.0	241.4	240.9	239.6
2002	238.4	237.1	235.3	232.9	230.5	231.4	232.5	233.1	238.7	242.8	242.9	242.1
2003	241.2	239.1	237.0	234.4	231.4	230.4	234.3	236.3	241.4	243.0	242.5	241.6
2004	240.8	239.7	238.1	235.6	232.5	233.0	233.8	235.5	240.0	243.1	242.7	241.9
2005	240.9	239.5	237.8	235.2	231.3	230.5	235.2	237.8	240.1	242.9	241.7	240.6
2006	239.3	237.6	235.6	233.2	230.3	234.0	238.7	241.2	242.6	242.7	242.3	241.0
2007	239.7	238.2	236.2	234.0	231.1	230.0	231.4	234.3	239.6	242.6	242.4	241.0
2008	239.9	238.4	236.5	234.1	230.7	231.7	237.4	240.7	242.0	243.1	242.3	241.0
2009	239.9	238.4	236.5	234.0	232.5	232.5	232.6	234.8	238.2	241.8	242.8	242.0
2010	240.5	238.5	236.3	233.8	232.0	239.3	241.9	242.5	242.4	242.7	241.9	241.0
2011	239.9	238.5	236.8	234.3	231.3	231.0	237.5	241.4	242.6	243.2	242.6	241.8
2012	240.4	238.6	236.8	234.6	232.5	232.7	234.1	237.7	242.0	243.0	242.3	241.5
2013	240.2	238.5	236.5	233.9	230.9	231.3	235.2	237.0	240.8	242.8	242.8	241.9
2014	240.6	238.9	237.0	234.3	232.3	234.3	235.4	235.8	241.1	243.0	242.7	241.9
2015	240.8	239.2	237.2	235.0	232.3	232.1	231.9	232.5	236.4	242.2	243.0	242.2
2016	241.1	239.6	237.7	235.1	231.5	230.9	232.6	234.2	236.7	239.3	239.9	239.8
2017	239.6	238.9	237.7	235.7	233.0	233.9	237.3	239.2	241.7	242.8	242.3	241.6
2018	240.8	239.4	237.4	235.2	233.0	233.8	234.2	236.0	240.9	242.5	242.4	241.4

Fuente: Elaboración propia

Anexo 2. Nivel del reservorio generados por la red neuronal con la función Tangente Hiperbólica

Tabla 35: Valores del nivel del reservorio generados por la red neuronal con la función de activación Tangente Hiperbólica

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Agos	Sep	Oct	Nov	Dic
2000	240.4	239.3	236.1	232.9	231.2	231.7	233.5	233.6	237.6	242.2	241.8	240.9
2001	239.5	238.8	236.6	234.0	230.8	231.2	235.3	234.4	238.4	241.5	240.9	240.1
2002	240.1	239.3	235.9	233.0	231.0	231.5	234.3	233.3	237.8	242.3	241.6	240.8
2003	240.9	239.0	237.2	234.8	231.5	231.1	234.0	236.3	240.8	243.3	242.8	241.8
2004	241.9	240.2	238.2	235.9	232.5	231.9	233.3	235.2	239.4	243.5	243.2	242.3
2005	240.6	239.2	237.3	235.0	231.3	231.3	236.0	237.4	241.4	242.9	242.6	241.5
2006	239.0	237.0	234.3	232.4	231.5	234.4	239.4	238.7	242.2	242.3	242.0	240.6
2007	240.2	238.1	236.2	233.5	230.8	230.5	233.0	234.4	239.7	242.5	242.0	241.0
2008	240.1	237.6	235.6	233.0	231.4	232.1	237.2	237.8	242.4	242.8	242.7	241.2
2009	240.4	238.3	235.9	234.5	231.8	231.7	233.3	233.8	237.8	242.5	241.7	241.1
2010	240.1	239.4	236.6	234.4	232.9	237.6	241.0	240.6	242.2	242.1	241.7	240.9
2011	240.4	238.4	236.4	234.1	231.6	231.9	238.2	240.8	242.9	243.1	243.0	241.6
2012	240.6	238.8	236.7	234.5	231.7	232.2	236.4	237.7	242.2	243.0	242.8	241.8
2013	240.4	238.3	236.5	233.5	231.2	231.5	235.9	237.1	241.8	242.9	242.7	241.5
2014	240.5	238.7	237.0	234.4	231.3	231.1	234.8	236.0	241.1	242.8	242.5	241.4
2015	241.2	239.2	237.4	234.5	232.1	231.0	231.7	233.6	236.7	242.5	241.9	241.0
2016	239.9	240.0	237.4	236.4	231.9	232.5	234.8	233.4	236.4	241.1	239.6	239.7
2017	240.4	239.9	239.1	236.5	233.1	233.4	237.1	238.3	240.4	242.1	241.5	240.9
2018	240.5	239.6	237.1	234.0	231.3	232.3	237.3	236.0	240.6	242.7	242.2	241.3

Fuente: Elaboración propia

Anexo 3. Nivel del reservorio generados por la red neuronal con la función Sigmoide

Tabla 36: Valores del nivel del reservorio generados por la red neuronal con la función de activación Sigmoide

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Agos	Sep	Oct	Nov	Dic
2000	239.8	237.8	236.9	234.7	230.8	232.7	235.1	237.2	239.6	241.2	241.3	240.3
2001	240.0	237.2	235.7	232.7	230.4	231.3	232.0	235.4	238.6	241.4	241.1	240.5
2002	238.5	236.8	234.4	232.0	230.6	230.1	232.7	234.0	238.2	242.4	242.4	241.1
2003	239.8	239.0	236.8	233.6	231.8	231.6	234.1	236.4	239.7	243.1	242.9	241.8
2004	240.3	239.2	237.6	235.1	231.7	232.7	235.1	237.8	240.6	242.6	242.4	241.3

2005	240.9	239.2	237.7	234.9	231.7	232.8	234.7	237.9	240.7	242.8	242.4	241.6
2006	240.2	239.2	237.2	235.1	231.3	233.1	237.3	238.2	241.0	242.3	242.4	240.9
2007	239.3	237.7	236.2	233.2	231.0	231.2	233.1	235.5	238.8	242.2	241.9	240.9
2008	240.2	239.2	237.7	235.8	231.9	233.6	237.0	238.3	240.9	242.4	242.6	241.4
2009	240.0	237.9	235.9	233.4	231.4	231.6	233.9	235.6	238.6	243.0	242.6	241.5
2010	240.8	239.7	237.8	236.7	233,0	235.0	239.0	239.1	241.5	242.7	242.7	241.6
2011	240.9	239.9	237.7	235.2	231.9	232.7	236.4	238.4	241.6	243.3	243.2	242.0
2012	240.5	238.7	237.0	233.9	231.1	231.8	234.5	237.1	240.2	242.6	242.4	241.3
2013	240.4	239.0	236.3	233.6	231.3	231.4	234.3	236.5	240.5	243.2	243.0	242.0
2014	240.5	239.3	237.6	234.6	232.2	233.4	235.4	237.8	240.4	242.8	242.7	241.7
2015	240.4	238.3	236.3	233.4	231.1	231.7	233.5	236.3	239.6	242.6	242.3	241.3
2016	239.9	237.2	236.5	233.4	230.6	232.0	232.8	236.2	238.5	240.9	240.9	240.3
2017	240.4	239.5	238.1	235.6	232.1	233.5	236.1	238.5	241.1	242.7	242.8	241.6
2018	240.4	239.4	237.6	234.8	231.9	232.4	234.7	237.5	240.7	242.9	242.6	241.7

Fuente: Elaboración propia

Anexo 4. Arquitectura de la Red Neuronal

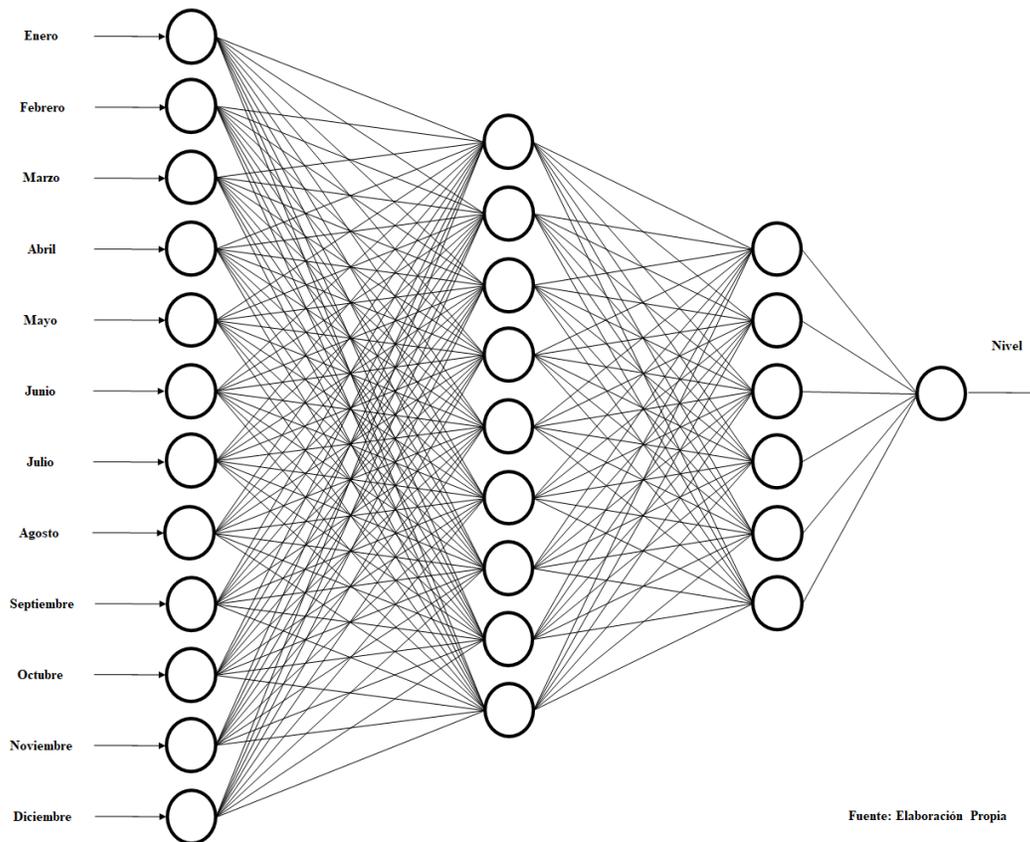


Figura 50: Arquitectura de la Red Neuronal

Anexo 5. Estimaciones de los parámetros generados por el modelo de la red neuronal

Tabla 37: Estimaciones de los parámetros generados por el modelo de la red neuronal artificial

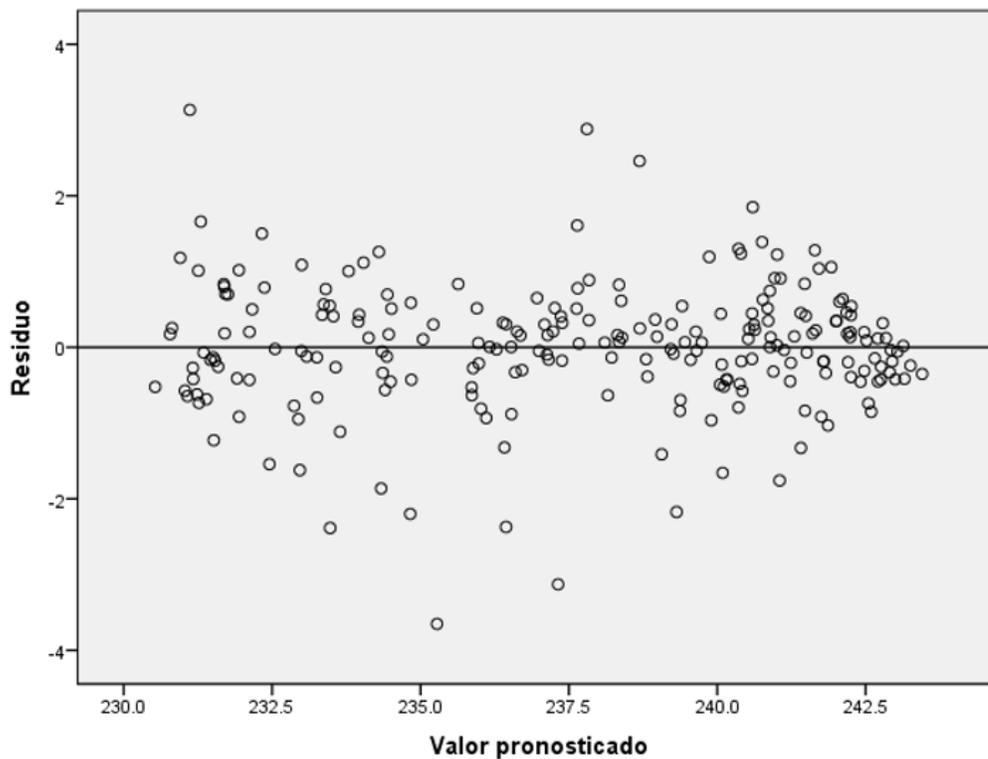
Predictor		Estimaciones de parámetro								
		Pronosticado								
		Capa oculta 1								
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)
Capa de entrada	(Sesgo)	-.512	-.010	.254	.050	-.471	-.273	-.764	.179	-.141
	[año=2000]	.124	.142	.631	-.243	-.241	-.040	-.289	-.168	.137
	[año=2001]	-.064	-.539	-.409	.322	-.439	.575	.277	-.143	.263
	[año=2002]	.418	-.247	.064	.003	-.573	-.135	-.349	-.012	.203
	[año=2003]	.169	.229	.222	.009	.524	.041	.288	-.199	.010
	[año=2004]	.405	.051	.351	-.117	.474	-.362	-.654	-.252	.453
	[año=2005]	.506	-.286	-.199	.156	.187	-.018	.180	.078	-.116
	[año=2006]	-.472	.216	.130	.274	-.626	-.123	.414	1168	.368
	[año=2007]	.125	.520	.147	.416	.173	.322	.188	-.296	-.110
	[año=2008]	.417	.095	-.035	.518	-.022	-.102	.133	.669	.282
	[año=2009]	.122	.171	-.286	-.239	-.172	-.349	-.276	-.684	.353
	[año=2010]	-.126	-.045	.420	-.492	-.766	.161	.296	1210	-.315
	[año=2011]	.272	-.267	.197	.385	.548	-.311	.663	.733	-.160
	[año=2012]	.386	.311	-.066	-.247	.259	.343	.413	.252	.553
	[año=2013]	-.275	.534	.102	.343	.171	.082	.275	.243	-.028
	[año=2014]	.177	.216	-.196	.412	.155	.241	.141	.043	.209
	[año=2015]	-.112	.248	.737	.601	.474	-.282	-.805	-.596	.177
	[año=2016]	.374	-.324	-.586	-.138	-.885	.767	-.719	-.856	.162
	[año=2017]	-.590	-.684	-.303	-.302	-.054	.368	.085	-.419	-.783
	[año=2018]	.176	-.062	-.201	.086	-.439	-.248	-.204	.243	-.096
	[mes=1]	-.017	-.028	.455	.266	-.014	.460	.272	1264	-.655
	[mes=2]	-.435	.181	.797	-.354	-.520	.068	.789	1273	-.399
	[mes=3]	-.054	-.277	.400	.834	-.201	-.367	.427	1332	.735
	[mes=4]	-.624	-.206	-.259	-.710	.254	.089	.686	-.1948	-.612
[mes=5]	-.120	-.158	.036	.286	.187	-.272	.476	-.1436	.500	
[mes=6]	.124	.105	.368	-.366	-.1059	.321	.433	-.869	.232	
[mes=7]	-.065	.055	-.409	-.249	-.1054	-.326	-.476	-.479	.045	
[mes=8]	.473	-.363	-.143	-.172	.413	-.605	-.759	-.695	.345	
[mes=9]	-.256	.374	-.731	.419	.737	-.256	-.974	-.560	-.425	
[mes=10]	-.127	.156	-.762	-.338	-.045	-.222	-.699	1037	-.799	
[mes=11]	-.245	-.294	-.832	-.436	.441	.426	-.567	.360	-.688	
[mes=12]	-.130	-.248	-.176	.093	.004	.512	-.272	.932	-.272	
Capa oculta 1	(Sesgo)									
	H(1:1)									
	H(1:2)									
	H(1:3)									
	H(1:4)									

	H(1:5) H(1:6) H(1:7) H(1:8) H(1:9)						
Capa oculta 2	(Sesgo) H(2:1) H(2:2) H(2:3) H(2:4) H(2:5) H(2:6)						
Estimaciones de parámetro							
Predictor		Pronosticado					Capa de salida nivel
		Capa oculta 2					
		H(2:1)	H(2:2)	H(2:3)	H(2:4)	H(2:5)	H(2:6)
Capa de entrada	(Sesgo)						
	[año=2000]						
	[año=2001]						
	[año=2002]						
	[año=2003]						
	[año=2004]						
	[año=2005]						
	[año=2006]						
	[año=2007]						
	[año=2008]						
	[año=2009]						
	[año=2010]						
	[año=2011]						
	[año=2012]						
	[año=2013]						
	[año=2014]						
	[año=2015]						
	[año=2016]						
	[año=2017]						
	[año=2018]						
	[mes=1]						
	[mes=2]						
	[mes=3]						
	[mes=4]						
	[mes=5]						
	[mes=6]						
	[mes=7]						
	[mes=8]						
	[mes=9]						
	[mes=10]						
	[mes=11]						
	[mes=12]						
Capa oculta I	(Sesgo)	.087	.296	.044	-.833	-.146	-.376
	H(1:1)	.278	-.365	-.134	.122	.105	.071

	H(1:2)	-.159	-.577	.230	.199	.145	-.510
	H(1:3)	-.257	-.906	-.152	-.200	-.088	.156
	H(1:4)	.174	-.293	-.512	.357	.156	.824
	H(1:5)	-.573	.721	.366	.817	-.553	-.763
	H(1:6)	.109	.699	-.329	-.548	-.243	-.204
	H(1:7)	-.358	-.590	-.254	-.656	.220	.433
	H(1:8)	.170	1260	.189	.243	-.889	-1231
	H(1:9)	.198	-.733	.600	-.515	-.063	.497
Capa oculta 2	(Sesgo)						-.186
	H(2:1)						.134
	H(2:2)						.774
	H(2:3)						.366
	H(2:4)						.305
	H(2:5)						-.184
	H(2:6)						-.476

Anexo 6. Comportamiento de los residuos del modelo de red neuronal

Residuos del modelo de la Red Neuronal



Variable dependiente: nivel

Fuente: Elaboración propia.

Figura 51: Comportamiento de los residuos del modelo de red neuronal