

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Facultad de Ciencias e Ingeniería



Análisis predictivo de series temporales de temperatura corporal

Tesis para optar por el Título de Ingeniera Industrial, que presenta la bachillera:

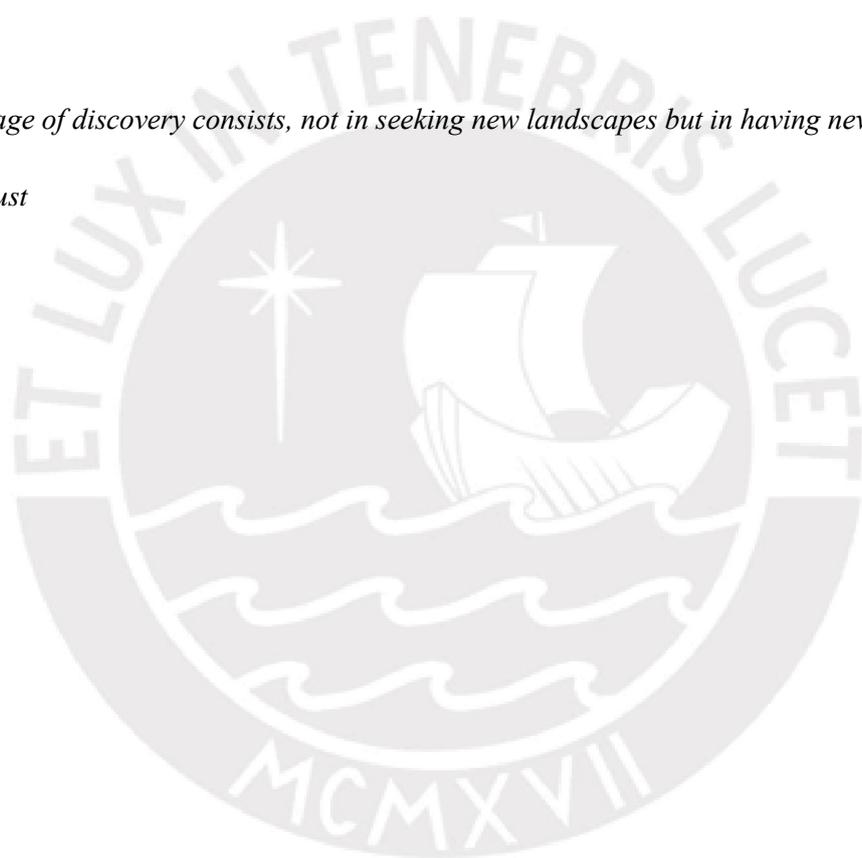
Nikol Alexandra Cotrina Araujo

Asesor: José Rau Alan Álvarez

Lima, octubre de 2020

“The real voyage of discovery consists, not in seeking new landscapes but in having new eyes”

— *Marcel Proust*



Agradecimientos

A mis padres y a mi hermana por impulsarme siempre a lograr mis objetivos y motivarme en cada momento a dar lo mejor de mí.

A mis asesores en la Universidad Politécnica de Madrid José Mira McWilliams y María Jesús Sánchez Naranjo por su paciencia y buena disposición al resolver mis dudas.

Al equipo de investigación de la clínica del Hospital Universitario de Móstoles en Madrid y en especial a Borja por darse el tiempo para atender mis preguntas

A Luis Alberto Gutiérrez por ayudarme en los trabajos que hemos realizado juntos en esta carrera universitaria, pero sobre todo por motivarme a ser mejor. A Fiorella Pacheco por acompañarme en la aventura del doble diploma y en esta nueva vida en Madrid.

A todos los amigos que hice en mi carrera universitaria entre la PUCP y la UPM, en especial a Valery Salas, Erick Prado, Manuel Chauca, Yamile Aylas y Diego Cisneros H. por sus buenos deseos y constantes éxitos ya que todos ellos me enseñan y me motivan cada día a mejorar de manera personal y profesional.

También a mis profesores por su dedicación al enseñar y por demostrarme que es posible hacer de tu pasión tu trabajo

¡Gracias a todos!

Resumen

Como un ser homeopático el ser humano presenta mecanismos de regulación de la temperatura que aseguran un óptimo funcionamiento del sistema fisiológico. Estos mecanismos de regulación tienen su centro de control e integración en el hipotálamo, que se encarga de mantener la temperatura corporal funcionando a $\pm 1^\circ\text{C}$ de la temperatura en reposo. Cuando la temperatura se encuentra fuera de este rango los sistemas biológicos comienzan a verse afectados.

A efectos prácticos, el organismo puede dividirse en una parte central o nuclear, el cual estaría constituido por la cabeza y las cavidades torácica y abdominal; y una parte superficial, que está constituido por la piel, el tejido celular subcutáneo y el grueso de la masa muscular. De manera que, las temperaturas del núcleo y la piel serían las temperaturas central y periférica, respectivamente.

La temperatura central (representada por las temperaturas oral, rectal, esofágica, membrana del tímpano, hipotalámica o de la sangre al pasar por cualquiera de los órganos de la parte central o nuclear) permanece relativamente constante, gracias a que es regulada y se mantiene dentro de límites bastante estrechos.

Además, dentro de la práctica clínica, la temperatura es una de las variables que se encuentran en constante monitoreo pues resulta fundamental en la atención de pacientes con patologías agudas, ya que permite determinar si el paciente presenta un aumento en su temperatura central o fiebre que usualmente es asociado con el desarrollo de una infección.

La fiebre es una respuesta fisiológica adaptativa frente a un agente patógeno, que mejora la respuesta del sistema inmune y evita la propagación de los agentes infecciosos. De hecho, “la elevación de la temperatura corporal en algunos grados puede aumentar la eficiencia de los macrófagos para destruir los microorganismos invasores, dificultando la replicación de diferentes microorganismos y otorgando una ventaja adaptativa al sistema inmune” (Gómez, 2008).

Debido a lo anterior, la predicción del desarrollo de fiebre es importante pues, según el horizonte de predicción de este estado en el paciente, se puede permitir la obtención de un cultivo de sangre cuando el recuento de bacterias está en su punto máximo y, de esa manera, precisar el diagnóstico del paciente.

Durante la práctica clínica, se registra la temperatura de los pacientes cada 8-12 h, sin embargo, tales mediciones no proporcionan información significativa en el diagnóstico de enfermedades. Por ello, el monitoreo continuo de la temperatura y el análisis de los registros obtenidos utilizando métodos analíticos podrían ayudar a revelar respuestas únicas de fiebre de los pacientes y en diferentes condiciones clínicas.

Esta tesis analiza series de temperatura central de pacientes aplicando técnicas estadísticas de series temporales como modelos lineales ARIMA y modelos no lineales de redes neuronales recurrentes para predecir futuros incrementos de la temperatura central que permitiría anticipar el diagnóstico y tratamiento que podría recibir un paciente.

Los modelos ARIMA y de Redes neuronales recurrentes fueron caracterizados con un análisis univariante, donde la variable estudiada es la Temperatura central. Por otro lado, los datos utilizados fueron recogidos con un equipo de Monitorización Continua de Temperatura, con un termómetro timpánico de infrarrojos denominado Thercom.

Estos dispositivos se configuraron para realizar una determinación de temperatura central por minuto durante aproximadamente 24 horas. No obstante, debido a que la toma de datos de temperatura podría ser incómoda para los pacientes, hay momentos en las series en que los datos obtenidos contienen mediciones que podrían no ser fiables.

La primera técnica empleada para el análisis de series temporales fue el modelo de Box-Jenkins o también llamados procesos autorregresivos integrados con media móvil (modelos ARIMA, por sus siglas en inglés) estos modelos pueden capturar las tendencias a corto plazo y las variaciones periódicas en las series temporales, sin embargo, estos no se ajustan a tendencias no lineales que pudiera tener la variable evaluada.

Posteriormente, se probaron los modelos no lineales de redes neuronales recurrentes de Elman y se comparó la precisión de las predicciones de los modelos ARIMA, frente a los modelos de Elman usando como estadístico el error absoluto porcentual medio o MAPE para horizontes de 15, 30 y 60 minutos. Ambos modelos fueron desarrollados usando el software estadístico Rstudio con las librerías: “tseries”, para modelos ARIMA; “forecast”, para determinar las predicciones en modelos ARIMA y “RSNNS”, para modelar redes neuronales.

Debido a la naturaleza de la variable temperatura central la precisión de los modelos debe ser tal que el error de predicción del modelo no debe ser mayor a 1°C, esto en promedio, implica que las predicciones del modelo deben tener un MAPE máximo admisible de 2.7%.

Las predicciones de los modelos ARIMA para los 3 pacientes produjeron MAPEs menores a 3% en todos los horizontes de predicción, no obstante, los residuos obtenidos no siguen una distribución normal, aunque en todos los casos cumplían la hipótesis fundamental de que eran independientes. Además, se produjeron predicciones muy satisfactorias para los 3 pacientes en el horizonte de 15 minutos. Para los otros horizontes de tiempo, los modelos presentaban un intervalo de confianza con amplitud mayor a 2°C, aunque el MAPE producido seguía siendo aceptable.

Para los modelos de redes neuronales se utilizaron como variables de entrada, el primer y el segundo retardo de la temperatura central. Como resultado, las predicciones de los modelos de redes de Elman se ajustaron a los datos de tal manera que se obtuvieron MAPEs inferiores a 0.5% para los 3 pacientes en los 3 horizontes de tiempo estudiados (15, 30 y 60 minutos).

A pesar de la complejidad de la regulación fisiológica de la temperatura central, el usar retardos de la temperatura central como valores de entrada para el modelo de redes neuronales hace que la estructura del modelo se simplifique y que la predicción de los futuros valores pueda estimarse con modelos de una sola capa de manera satisfactoria, ajustándose a los datos en su escala real y produciendo errores mínimos.

Contenido

1.	Introducción	1
1.1	Regulación de la temperatura corporal	2
1.2	La fiebre	5
1.3	Por qué es importante predecir la fiebre	6
2.	Estado del arte.....	7
3.	Objetivos.....	8
4.	Técnicas estadísticas y algorítmicas	9
4.1	Series temporales	9
4.1.1	Clasificación de una serie temporal	9
4.1.2	Proceso de ruido blanco	11
4.1.3	Funciones de autocorrelación simple y parcial	12
4.1.4	Modelos univariantes de series temporales.....	13
4.2	Redes Neuronales Artificiales.....	14
4.2.1	Neuronas o unidades de procesamiento	15
4.2.2	Función de activación	16
4.2.3	Topologías de redes neuronales	16
4.2.4	Entrenamiento de redes neuronales.....	18
4.2.5	Redes neuronales recurrentes (RNN).....	22
5.	Metodología	26
5.1	Recogida de datos	26
5.2	Selección de datos.....	27
5.3	Variable utilizada	28
5.4	Análisis de Series Temporales	29
5.4.1	Identificación del modelo.....	30
5.4.2	Estimación de parámetros y contrastes paramétricos.....	34
5.4.3	Diagnóstico del modelo.....	36
5.4.4	Predicción	37
5.5	Redes Neuronales Recurrentes.....	38
5.5.1	Transformación de los datos	38
5.5.2	Selección del número de capas	38
5.5.3	Número de Neuronas por capa.....	39
5.5.4	Número de epochs.....	39

5.6	Medidas de Error.....	39
5.6.1	Error Absoluto Medio o MAE	39
5.6.2	Raíz del error cuadrático medio o RMSE	40
5.6.3	Promedio del error absoluto en porcentaje o MAPE.....	40
6.	Resultados y discusión.....	41
6.1	Modelos ARIMA	41
6.1.1	Paciente 1	41
6.1.2	Paciente 2.....	48
6.1.3	Paciente 3.....	55
6.1.4	Discusión sobre los resultados de los 3 pacientes	61
6.1.5	Conclusiones	61
6.2	Redes Neuronales Recurrentes.....	62
6.2.1	Paciente 1	62
6.2.2	Paciente 2.....	68
6.2.3	Paciente 3.....	72
6.2.4	Discusión sobre los resultados de los 3 pacientes	76
6.2.5	Conclusiones sobre los modelos de redes de Elman	76
7.	Conclusiones Generales	77
8.	Líneas futuras.....	79
9.	Planificación Temporal.....	80
9.1	Estructura de Descomposición del Trabajo.....	80
9.2	Planificación	82
10.	Bibliografía	84
11.	Anexos	87

Índice de tablas

Tabla 1	Mecanismos para equilibrar la ganancia y pérdida de calor. Fuente: Trastornos de la temperatura corporal (Gómez 2007).....	3
Tabla 2	Valores de entrada y salida de la red de ejemplo.....	20
Tabla 3	Cálculo de los valores de entrada para cada nodo de la red de ejemplo.....	25
Tabla 4	Resumen sobre el comportamiento de las ACF y PACF según el modelo.....	34
Tabla 5	Comparación de modelos para el paciente 1.....	46
Tabla 6	Comparativa entre las predicciones a 15, 30 y 60 minutos $ARI(1,1)$ - paciente 1.....	47
Tabla 7	Comparación de modelos para el paciente 2.....	52
Tabla 8	Comparativa entre las predicciones a 15, 30 y 60 minutos $IMA(1,1)$ - paciente 2.....	54
Tabla 9	Comparación de modelos para el paciente 3.....	59
Tabla 10	Comparativa entre las predicciones a 15, 30 y 60 minutos $ARI(1,1)$ - paciente 3.....	60
Tabla 11	Comparativa de modelos – paciente 1.....	63
Tabla 12	Nodos de la red de Elman generada – paciente 1.....	64
Tabla 13	Pesos de las conexiones para la red generada.....	65
Tabla 14	Cálculo de los valores de entrada o de salida para cada nodo de la red.....	66
Tabla 15	Comparativa de modelos – paciente 2.....	69
Tabla 16	Nodos de la red de Elman generada – paciente 2.....	69
Tabla 17	Pesos de las conexiones para la red generada – paciente 2.....	70
Tabla 18	Comparativa de modelos – paciente 3.....	73
Tabla 19	Nodos de la red de Elman generada – paciente 3.....	73
Tabla 20	Pesos de las conexiones para la red generada – paciente 3.....	74
Tabla 21	Comparación de MAPE para las predicciones de los modelos ARIMA y RNN.....	77

Índice de figuras

Figura 1 Evolución de la temperatura corporal, central y periférica en función de la temperatura ambiental. Fuente: Fernández-Tresguerres 2010.....	3
Figura 2 Número de pasajeros de avión (mensual) USA, enero 1949 a diciembre 1960. Fuente: Apuntes M.J.Sánchez.....	10
Figura 3 Proceso Autorregresivo AR (1).....	11
Figura 4 Proceso de Ruido Blanco.....	11
Figura 5 Componentes básicos de una red neuronal artificial. Fuente: Kröse 1996.....	15
Figura 6 Capas de una red Neuronal. Fuente: Researchgate.net.....	15
Figura 7 Funciones de activación para una unidad de procesamiento.....	16
Figura 8 Red neuronal de una sola capa con dos entradas y una salida.....	17
Figura 9 Esquema general de una red neuronal.....	17
Figura 10 Red de ejemplo para la estimación de parámetros usando Backpropagation.....	20
Figura 11 Representación de una red neuronal recurrente.....	22
Figura 12 Estructura general de las redes de Jordan. Fuente: Kröse 1996.....	23
Figura 13 Estructura general de las redes de Elman. Fuente: Kröse 1996.....	24
Figura 14 Red de ejemplo para explicar las fórmulas de Elman.....	25
Figura 15 Flujoograma para modelar el análisis de una serie temporal.....	29
Figura 16 Serie X, modelo Autorregresivo $p=1$	31
Figura 17 Autocorrelación simple y parcial de un proceso AR 1.....	31
Figura 18 Serie z modelo de Media Móvil $q=1$	32
Figura 19 Autocorrelación simple y parcial de un modelo de MA 1.....	32
Figura 20 Prueba Ljung-Box para los residuos.....	36
Figura 21 Termometría del paciente 1: T. Central.....	41
Figura 22 Serie de temperatura transformada del paciente 1.....	42
Figura 23 Autocorrelación simple (izquierda) y parcial (derecha) de la T. central paciente 1.....	42
Figura 24 Residuos obtenidos para el modelo ARI(1,1) – paciente 1.....	44
Figura 25 Normalidad de los residuos ARI(1,1) – paciente 1.....	44
Figura 26 Residuos del modelo ARI (2,1) – paciente 1.....	45
Figura 27 Residuos del modelo ARIMA (1,1,1) – paciente 1.....	45
Figura 28 Predicciones a un plazo de 15 minutos.....	46
Figura 29 Predicciones a un plazo de 30 minutos.....	47
Figura 30 Predicciones a un plazo de 60 minutos.....	47
Figura 31 Termometría del paciente 2: T. Central.....	48
Figura 32 Serie de temperatura transformada.....	49
Figura 33 Autocorrelación simple (izquierda) y parcial (derecha) de la T. central paciente 2.....	49
Figura 34 Prueba Ljung-Box para los residuos del modelo IMA(1,1) – paciente 2.....	50
Figura 35 Normalidad de los residuos para el modelo IMA(1,1) – paciente 2.....	50
Figura 36 Residuos del modelo ARIMA(1,1,1) - paciente 2.....	51
Figura 37 Residuos del modelo ARI(1,1) - paciente 2.....	52
Figura 38 Predicciones a un plazo de 15 minutos.....	53
Figura 39 Predicciones a un plazo de 30 minutos.....	53

Figura 40	Predicciones a un plazo de 60 minutos.....	54
Figura 41	Termometría del paciente 3: T. Central.....	55
Figura 42	Serie de temperatura transformada.....	55
Figura 43	Autocorrelación simple (izquierda) y parcial (derecha) de la T.Central paciente 3.....	56
Figura 44	Prueba Ljung-Box para los residuos del modelo ARI (1,1) – Paciente 3.....	57
Figura 45	Normalidad de los Residuos obtenidos para el modelo ARI(1,1).....	57
Figura 46	Residuos del modelo ARI(2,1) - paciente 3.....	58
Figura 47	Residuos del modelo ARIMA(1,1,1) - paciente 3.....	58
Figura 48	Predicciones a un plazo de 15 minutos.....	59
Figura 49	Predicciones a un plazo de 30 minutos.....	60
Figura 50	Predicciones a un plazo de 60 minutos.....	60
Figura 51	Análisis descriptivo de la termometría del paciente 1.....	63
Figura 52	Representación de la red de Elman – paciente 1.....	65
Figura 53	Gráfica del ajuste del modelo vs los datos reales.....	67
Figura 54	Predicciones a 60 minutos – paciente 1.....	67
Figura 55	Análisis descriptivo de la termometría del paciente 2.....	68
Figura 56	Representación de la red de Elman – paciente 2.....	70
Figura 57	Gráfica del ajuste del modelo vs los datos reales.....	71
Figura 58	Predicciones a 60 minutos – paciente 2.....	71
Figura 59	Análisis descriptivo de la termometría del paciente 3.....	72
Figura 60	Representación de la red de Elman – paciente 3.....	74
Figura 61	Gráfica del ajuste del modelo vs los datos reales.....	75
Figura 62	Predicciones a 60 minutos – paciente 3.....	75
Figura 63	Estructura de Descomposición del Trabajo para el desarrollo de la tesis.....	81
Figura 64	Planificación de tareas.....	82
Figura 65	Diagrama de Gantt de las actividades indicadas previamente.....	83

Índice de ecuaciones

Ecuación 1 Características de un proceso de ruido blanco (Peña, 2010).....	11
Ecuación 2 Función de autocorrelación simple	12
Ecuación 3 Función de autorrelación parcial.....	12
Ecuación 4 Representación de un modelo univariante de una serie temporal.	13
Ecuación 5 Forma simplificada para representar un modelo univariante	13
Ecuación 6 Función sigmoide o en forma de S.....	16
Ecuación 7 Representación de una Red feedforward. Fuente Connor 1994.....	16
Ecuación 8 Resultado de la red neuronal planteada en la figura 8.....	17
Ecuación 9 Cálculo del error en la capa de salida.....	20
Ecuación 10 Variación del error de la capa de salida en función del valor de entrada.....	21
Ecuación 11 Variación del error en función de los valores de entrada en las capas ocultas.....	21
Ecuación 12 Ajuste de pesos.....	21
Ecuación 13 Representación de una red neuronal recurrente	22
Ecuación 14 Formulación de la red de Elman	24
Ecuación 15 Diferenciación de una serie	30
Ecuación 16 Representación de un Proceso autorregresivo AR 1	30
Ecuación 17 Correlación simple para el retardo k	31
Ecuación 18 Representación matemática de un modelo Autorregresivo AR	31
Ecuación 19 Representación de un modelo de Media móvil MA 1.....	32
Ecuación 20 Autocorrelación parcial para el retardo k.....	32
Ecuación 21 Representación matemática de un modelo con media móvil MA.....	33
Ecuación 22 Transformación de un proceso AR en un proceso MA.....	33
Ecuación 23 Representación matemática del modelo Box-Jenkins.	34
Ecuación 24 Expresión del criterio de Akaike	35
Ecuación 25 Representación del criterio de información de Bayes.....	35
Ecuación 26 Predicción del valor t+2 para un modelo AR (1)	37
Ecuación 27 Predicción del valor t+k para un modelo AR (1)	37
Ecuación 28 Método min-max para escalar los datos.....	38
Ecuación 29 Escalado estándar	38
Ecuación 30 Error absoluto medio.....	39
Ecuación 31 Error cuadrático medio.....	40
Ecuación 32 Media del error absoluto en porcentaje.....	40
Ecuación 33 Estimación de los datos del paciente 1 con el modelo ARI(1,1).....	43
Ecuación 34 Estimación de los datos del paciente 2 con el modelo IMA(1,1).....	49
Ecuación 35 Estimación de los datos del paciente 3 con el modelo ARI(1,1).....	56

Capítulo 1

1. Introducción

Esta tesis fue desarrollada en colaboración con el grupo de investigación de la clínica del Hospital Universitario de Móstoles en Madrid. Dicho grupo está al servicio de Medicina Interna del hospital y trabaja desde hace más de 15 años en el análisis de complejidad de señales biológicas y su aplicación en el ámbito clínico.

En la actualidad, el grupo es financiado a través de la convocatoria del Fondo de Investigaciones Sanitarias del Instituto de Salud Carlos III del año 2017, para el desarrollo de un proyecto que pretende utilizar el análisis de series temporales mediante la aplicación de modelos predictivos para la detección precoz de la fiebre y la optimización del rendimiento diagnóstico de los hemocultivos, una estrategia fundamental en la atención a los pacientes con infecciones graves.

Con respecto a los análisis de la temperatura corporal mediante dispositivos de monitorización continua, el grupo ha publicado trabajos que ofrecen una información relevante sobre los mecanismos de termorregulación que sugieren posibles aplicaciones de la termometría en la práctica clínica, entre los que destacan “La pérdida de complejidad de la regulación de la temperatura corporal como marcador de mal pronóstico en pacientes críticos”, y “Predicción de picos febriles mediante modelos de regresión logística con un horizonte de unos minutos de antelación”.

En esta investigación se analizarán series de temperatura de diferentes pacientes que se modelarán utilizando métodos lineales como los modelos ARIMA y algoritmos no lineales como redes neuronales recurrentes (RNN) con el objetivo de predecir futuros valores de temperatura en los pacientes, en diferentes horizontes de tiempo y de acuerdo con ello brindarles un tratamiento adecuado. En los siguientes apartados, se dará una introducción sobre la regulación de la temperatura corporal, la fiebre y la importancia de detectarla oportunamente.

En la práctica clínica es rutinario monitorear ciertas variables como la presión arterial, frecuencia cardiaca, temperatura, entre otras; debido a que el empleo de estas variables resulta fundamental en la atención del paciente con patologías agudas. No obstante, estas variables continuas, suelen determinarse solo de manera puntual o periódica. Y aunque este es un enfoque muy práctico, no tiene en consideración la evolución de estas variables y, por tanto, permite que ciertos cambios relevantes puedan pasar desapercibidos.

El análisis y monitoreo de la temperatura corporal busca determinar si el paciente presenta una elevada temperatura corporal o fiebre, ya que este fenómeno suele implicar el desarrollo de un proceso infeccioso. No obstante, a pesar de que la fiebre es un concepto utilizado ampliamente, no existe un consenso en la comunidad médica sobre un límite universal para determinar que un paciente tiene fiebre.

1.1 Regulación de la temperatura corporal

Los humanos son seres homeotermos endotérmicos, de ahí que su principal fuente de calor sea su propio organismo, es decir, producen calor y son capaces de regular su temperatura corporal. En ese sentido, la temperatura corporal se logra a través del metabolismo, mientras que la disipación de calor, mediante procesos de conducción y convección a través de superficies como la piel y los pulmones. Por otro lado, a medida que la temperatura central se incrementa, el mecanismo más efectivo para la disipación de calor es la evaporación (Tansey & Johnson, 2014).

Además, la sensibilidad a la temperatura y la termorregulación forman una parte esencial de la homeostasis fisiológica y la supervivencia del ser humano y resultan clave para operar satisfactoriamente en sus entornos (Gisolfi y Mora, 2000). Adicionalmente, se sabe que la temperatura corporal suele estar alrededor de los 37°C, aunque puede variar en el rango de 33.2 a 38.2 °C según el lugar de las mediciones (central o periférica). Cabe resaltar que, también existen fluctuaciones normales de la temperatura a lo largo del día, del mes o por el envejecimiento.

Las desviaciones fuera de estos rangos pueden ser fatídicas; por ejemplo, si la temperatura central se eleva por encima de 42°C, empieza la citotoxicidad con la desnaturalización de las proteínas y la síntesis de ADN deteriorado, lo cual finalmente conlleva al fallo generalizado de órganos y deterioro neuronal (Mallet, 2002). Asimismo, si la temperatura cae por debajo de 30° los cambios neuromusculares, cardiovasculares y respiratorios pueden resultar igual de fatales (Lepock, 2003).

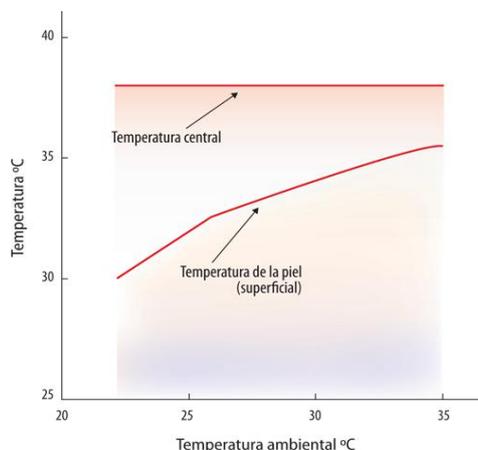
A efectos de entender estos conceptos el organismo puede dividirse en una parte central o nuclear y una parte superficial o más en contacto con el medio externo ambiental. El centro o núcleo estaría constituido por la cabeza y las cavidades torácica y abdominal. En contraste, la piel, el tejido celular subcutáneo y el grueso de la masa muscular constituyen la superficie. De manera que, las temperaturas de una y otra parte del organismo serían las temperaturas central y superficial. La Figura 1 muestra el efecto de la exposición del organismo humano en reposo a diferentes temperaturas ambientales.

Nótese que el promedio de la temperatura de la piel (representativa de la temperatura superficial) aumenta con el incremento de la temperatura ambiental. En cambio, la temperatura central (representada por las temperaturas oral, rectal, esofágica, membrana del tímpano, hipotalámica o de la sangre al pasar por cualquiera de los órganos de la parte central o nuclear) permanece relativamente constante, gracias a que es regulada y se mantiene constante dentro de límites bastante estrechos

No obstante, no todos los seres humanos tienen la misma temperatura en estado de reposo, debido a que los mecanismos de autorregulación son influenciados por variables antropomórficas como edad, género, peso, estatura y niveles hormonales u otras variables externas como la medicación del individuo, el ambiente y el lugar del cuerpo en donde se mida la temperatura pueden afectar la temperatura del paciente.

Figura 1 Evolución de la temperatura corporal, central y periférica en función de la temperatura ambiental.

Fuente: Fernández-Tresguerres 2010



- **Variaciones intraindividuo:**

Las variaciones de la temperatura dentro de un individuo pueden producirse por varias razones: ejercicio físico, el estrés, el calor externo; y se puede reducir en la inactividad, el sueño y el frío externo. En condiciones normales, el valor mínimo de temperatura se alcanza de madrugada y el valor máximo, a mitad de tarde.

- **Variaciones interindividuos:**

Las variaciones de temperatura entre distintos individuos pueden producirse también por varias razones: la edad (T^a media menor en ancianos), por el género (T^a media ligeramente más elevada en mujeres que en hombres), el desarrollo de actividad física o la ingesta de fármacos antipiréticos o inmunosupresores.

- **Métodos y lugares de medida:**

La temperatura corporal suele medirse en estas zonas: axila, cavidad oral, recto y membrana timpánica. De todos los lugares de medición, aquel que muestra una temperatura más próxima a la temperatura central es la medición rectal, con un valor medio normal de 36.8 °C.

Entonces, para asegurar un funcionamiento fisiológico óptimo los humanos deben ser capaces de preservar la temperatura central (cabeza, tórax y abdomen) de cara a los cambios que puede haber en su entorno. De manera que la ganancia de temperatura sea igual que su pérdida. El cuerpo humano logra esto gracias a unos mecanismos de regulación de la temperatura que afectan tanto a la producción como a la disipación del calor, como se puede ver en la Tabla 1.

Tabla 1 Mecanismos para equilibrar la ganancia y pérdida de calor. Fuente: Trastornos de la temperatura corporal (Gómez 2007)

Producción de calor	Disipación de calor
<ul style="list-style-type: none"> • Contracción muscular • Reacciones metabólicas • Ingesta de alimentos • Acción de hormonas • Lipólisis de grasa 	<ul style="list-style-type: none"> • Radiación • Convección • Conducción • Evaporación del sudor

Los mecanismos mencionados previamente, tienen su centro de control e integración en el hipotálamo. Esta región del cerebro se encarga de mantener la temperatura central funcionando a $\pm 1^{\circ}\text{C}$ de la temperatura en reposo, fuera de estos límites el funcionamiento de los sistemas biológicos podría verse comprometido (Gómez, 2007). Por otro lado, según Romanovsky (2007), el control de temperatura se realiza principalmente desde el núcleo preóptico anterior del hipotálamo (POA), el cual, recibe información de termorreceptores periféricos y centrales. Por otro lado, la regulación de la temperatura implica los siguientes agentes:

- Vías aferentes o Termorreceptores están localizados en la piel y en el núcleo preóptico del hipotálamo, permiten la recogida de información térmica que la dirige a los centros de integración.
- Vías eferentes o Efectos termorreguladores son los distintos tipos de sistemas que permiten elevar o reducir la temperatura corporal y están basados en la sudación y la vasodilatación periférica.
- Centro de integración o Área de control se localiza en el cerebro y es el centro encargado de integrar la información de las vías aferentes y desencadenar una respuesta a través de las vías eferentes (Guyton, 2004)

La desviación de la temperatura corporal en reposo afecta varios sistemas fisiológicos en el cuerpo, lo cual es indicativo del alcance de las funciones biológicas y disfunciones que interactúan con los mecanismos termorreguladores (El-Radhi, 2006). Por ejemplo, un incremento de la temperatura corporal puede desencadenar sudoración, dilatación de los vasos sanguíneos de la piel o sudoración, o, por otro lado, si la temperatura baja genera mecanismos para conservar el calor: escalofríos, taquicardia o vasoconstricción. (Gómez, 2007).

Existe evidencia de que las temperaturas centrales y periféricas influyen en los circuitos efectores individuales de forma independiente (Morrison SF, 2011). Las neuronas termorreceptivas se activan cuando se alcanza el umbral de temperatura apropiado para esa neurona, y los potenciales de acción ascienden al POA. Estas señales, junto con las señales termorreceptivas que surgen dentro del núcleo preóptico anterior del hipotálamo, actúan sobre varias salidas efectoras (Nakamura, 2011).

1.2 La fiebre

La fiebre es un estado experimentado con frecuencia por profesionales de la salud y pacientes por igual. Se supone que su significado es claro y universalmente entendido, cuando en realidad la interpretación es a menudo exclusivamente personal. De acuerdo con Mackowiak et al. (1997), "después de más de un milenio de investigación clínica, ni siquiera existe una definición [clínica] generalmente aceptada de fiebre"

Del mismo modo, la definición de fiebre dentro de la literatura médica y de enfermería varía ampliamente. En numerosos artículos, se define u operacionaliza simplemente como una elevación de la temperatura corporal (O'Grady et al.1998, Volker 1998, Creechan et al.2001, Dinarello y Gelfand 2004). El nivel de elevación de la temperatura que garantiza el diagnóstico de "fiebre" también varía ampliamente dentro de esa literatura.

En el 2008, un creciente grupo de investigación en inmunología y neurofisiología llegó a la conclusión de que la fiebre es generalmente una respuesta fisiológica adaptativa a alguna amenaza (Mackowiak 1997, Rowsey 1997, Kluger et al.1998). Esta respuesta sería provocada por el propio organismo con el fin de mejorar la respuesta del sistema inmune y evitar la propagación de microorganismos patógenos.

Una revisión de la ontología de la fiebre revela que se refiere a un proceso fisiológico: un aumento de la temperatura corporal por encima de lo normal que forma parte de una respuesta coordinada de fase aguda que incluye respuestas automáticas de comportamiento y neuroendocrinas. De manera que no existe una temperatura universal que defina el estado febril, sino que este estado está determinado por la temperatura corporal del individuo durante el día.

Clásicamente se ha aceptado que el límite superior de la normalidad, descrito Wunderlich, era 38°C. Sin embargo, Mackowiak planteaba que, dada la variabilidad interindividual, las diferencias entre regiones del organismo y las oscilaciones a lo largo del día de la temperatura corporal, la aplicación de medias poblacionales a cada individuo no era adecuada. Con sus resultados planteó que, en individuos jóvenes y con determinaciones de temperatura oral, se debía considerar como fiebre una temperatura mayor de 37.2°C a primera hora de la mañana o de 37.7°C por la tarde (Mackowiak, 2010).

Se admite que la aparición de fiebre es consecuencia de la estimulación de la producción de pirógenos endógenos, polipéptidos elaborados por diversas células y macrófagos titulares (parte del sistema inmunitario) frente a agentes externos. La patogénesis de la fiebre tiene el mismo mecanismo fisiopatológico para procesos de muy diversa etiología, por lo que la fiebre es un signo totalmente inespecífico que puede producirse por procesos infecciosos y no infecciosos debido a la presencia de bacterias, virus, protozoarios, entre otros (Ramón-Romero 2014).

De hecho, la elevación de la temperatura corporal en algunos grados puede aumentar la eficiencia de los macrófagos para destruir los microorganismos invasores, dificultando la replicación de diferentes microorganismos y otorgando una ventaja adaptativa al sistema inmune (Gómez, 2008). Esto se acompaña a nivel periférico de la activación de los sistemas termoelectores generando calor y disminuyendo su pérdida (Nakamura 2011). En un sentido práctico, se admite que previo al pico febril, la temperatura central suele incrementarse, mientras que la temperatura periférica disminuye.

1.3 Por qué es importante predecir la fiebre

Según Bennett (1954) y Mylotte (2000), la fiebre no es un fenómeno inmediato, sino que se trata de un proceso esmerado y metabólicamente exigente que requiere muchas operaciones de por medio: desde la producción de lipopolisacáridos y otras sustancias pirogénicas, la activación de macrófagos, liberación de interleuquinas y la activación del sistema de termorregulación aferente y eferente antes de que la temperatura comience a elevarse. Todo ello conlleva un tiempo, además, hay estudios (Bennet y Beeson, 1954) que demuestran que previo al estado febril la cantidad de bacterias en la sangre es elevada.

En línea con lo anterior, si se logra predecir el desarrollo de fiebre en un momento específico, entonces, según el lapso de predicción de este estado en el paciente se puede permitir la obtención de un cultivo de sangre cuando el recuento de bacterias está en su punto máximo, y de esa manera mejorar el diagnóstico. También puede provocar medidas preventivas o terapéuticas para prevenir o mitigar el episodio febril en pacientes en los que la fiebre puede ser especialmente indeseable (Dinarello 2012).

Adicionalmente, desde un punto de vista metabólico, la fiebre es un proceso costoso, debido a que, incrementa la demanda de oxígeno de forma significativa, según Dinarello (2012) por cada 1°C por encima de 37°C el consumo de oxígeno aumenta un 13%. Este incremento en la demanda de oxígeno podría ser fatal para pacientes con disfunción cardíaca, respiratoria o del sistema nervioso. (Vargas, 2017) De manera que, predecir picos febriles en los pacientes podría traer como consecuencia que los pacientes reciban un tratamiento más oportuno.

Finalmente, aunque se deben tomar decisiones clínicas y, por lo tanto, necesitamos algún tipo de umbral, sería ingenuo suponer que hay un límite prefijado que separa a los pacientes febriles de los afebriles. Entonces, determinar el umbral clínico varía según el paciente y el contexto y es posiblemente un ejemplo típico de juicio clínico. Obviamente, admitir los límites borrosos de la temperatura no cuestiona la existencia de fiebre. La fiebre es un fenómeno clínico indiscutible que, aunque originalmente es un mecanismo de defensa, puede ser una prueba de fragilidad o de inestabilidad en los pacientes. (Jordán et al. 2017)

2. Estado del arte

Como ya se ha explicado previamente, la fiebre es una respuesta compleja del huésped a muchos agentes externos e internos y es un contribuyente potencial a muchas condiciones clínicas. No obstante, a pesar de ser uno de los principales signos vitales, la temperatura, su análisis y sus variaciones durante muchas condiciones patológicas aún no se han examinado en detalle utilizando técnicas matemáticas.

Se han desarrollado patrones clásicos de fiebre basados en registros obtenidos cada 8-12 h, pero tales patrones no proporcionan información significativa en el diagnóstico de enfermedades. Además, debido a que la fiebre es una respuesta del paciente, es probable que pueda haber una respuesta única a etiologías específicas. Por ello, el monitoreo continuo de la temperatura a largo plazo y el análisis de patrones utilizando métodos analíticos desarrollados en ingeniería podrían ayudar a revelar respuestas únicas de fiebre de los pacientes y en diferentes condiciones clínicas.

En ese sentido, en el 2017 (Vargas) se llevó a cabo el desarrollo de un modelo estadístico basado en la monitorización continua de temperatura, que usaba dos esquemas de modelado: un análisis discriminante y una regresión logística. En dichos modelos, se realizaron análisis univariantes y bivariantes para seleccionar el modelo que más se ajustara a la serie de temperaturas. Cabe resaltar que la variable de salida en ambos modelos es una variable cualitativa que describe si el paciente tiene fiebre o no, mientras que las variables de entrada fueron cuantitativas.

En el estudio mencionado, se estableció que un pico de fiebre ocurría cuando la temperatura central era superior a 38°C con al menos una medida menor a 37.5 en los 30 minutos previos. El modelo logístico de este estudio fue capaz de predecir picos de fiebre con ~85% de precisión, asimismo, el modelo discriminante; ambos con una media de anticipación de unos 50 minutos a los picos febriles en la serie de validación. Por otro lado, este estudio sugiere que los modelos podrían aportar más información clínica si en lugar de usar umbrales adecuados se pudieran identificar alteraciones en la termorregulación de los pacientes.

Adicionalmente, dicho estudio incluye una medida de la complejidad de la regulación de temperatura en los pacientes, denominada ApEn, que relaciona la Temperatura Central y la temperatura periférica en la que se concluye que existe una correlación negativa significativa entre la complejidad y la temperatura corporal entre los pacientes enfermos, mas no entre los individuos sanos. Se añade, también, que 1.25 picos de fiebre por paciente y por día de monitorización pasaron desapercibidos con las determinaciones habituales, en comparación con el sistema de monitorización continua.

3. Objetivos

En esta tesis se pretende desarrollar un modelo que permita describir las series temporales de temperatura corporal ingresados en la unidad de cuidados intensivos y que presentaron fiebre en las 24 horas previas. Posteriormente estos modelos se utilizarán para realizar predicciones sobre valores que podrían tomar las variables de temperatura central en los próximos 15, 30 y 60 minutos.

Como ya se ha indicado previamente, la predicción de futuros picos de fiebre permite brindarle un diagnóstico oportuno al paciente y, por tanto, darle un tratamiento adecuado, evitándole futuros padecimientos. No obstante, la consideración de la fiebre en la práctica clínica habitual es poco consistente, ya que no se puede identificar un único criterio válido para cualquier paciente (B. Vargas, 2017). De ahí que sea necesario un estudio de temperaturas con modelos regresivos que permitan analizar las termometrías de cada paciente para determinar si este tiene fiebre o no.

En esta investigación se desarrollarán las siguientes actividades para lograr el objetivo definido previamente:

- Aplicar fundamentos sobre planeamiento y desarrollo de proyectos para realizar el desarrollo del estudio: Diagramas de Flujo, Estructura de Descomposición del trabajo (EDT/ WBS) y diagramas de Gantt
- Realizar un análisis sistémico de los datos recogidos para la selección de datos a emplear con los modelos estadísticos y algorítmicos
- Desarrollar modelos estadísticos sobre series de temporales a una serie de temperaturas corporales: modelos Autorregresivos Integrados de Media Móvil, Método Box-Jenkins
- Desarrollar modelos algorítmicos sobre series temporales: Redes Neuronales Recurrentes, Redes de Elman.
- Comparar el poder predictivo de las series temporales y las redes neuronales en series de temperatura corporal.

Capítulo 2

4. Técnicas estadísticas y algorítmicas

En este apartado se explicarán las técnicas estadísticas y algorítmicas empleadas para analizar las series de temperaturas recogidas. En primer lugar, se explicarán las series temporales y su clasificación. A continuación, se presentan algunos conceptos básicos y luego, modelos de series temporales univariantes.

Posteriormente, se presentan las redes neuronales y su composición, y la explicación de cada uno de estos componentes. Luego, se profundiza en las redes neuronales recurrentes y se presentan 2 topologías distintas: las Redes de Elman, las redes de Jordan.

4.1 Series temporales

Una serie temporal es la realización de un proceso estocástico y puede ser definida como la sucesión de observaciones de una característica (series univariantes) o sobre varias características (serie multivariante o vectorial) tomadas en varios instantes de tiempo equidistantes cronológicamente.

Además, el análisis de series temporales permite realizar predicciones utilizando el histórico de dicha serie, es decir, valores previos de una variable. De manera que, los valores predichos estarán condicionados por sus valores pasados

4.1.1 Clasificación de una serie temporal

Las series temporales se pueden clasificar como estacionarias o no estacionarias, para analizar si una serie sigue uno de estos comportamientos se analiza el gráfico de la serie en búsqueda de patrones o ciclos repetitivos, tendencias en el gráfico o el nivel que sigue una serie de datos.

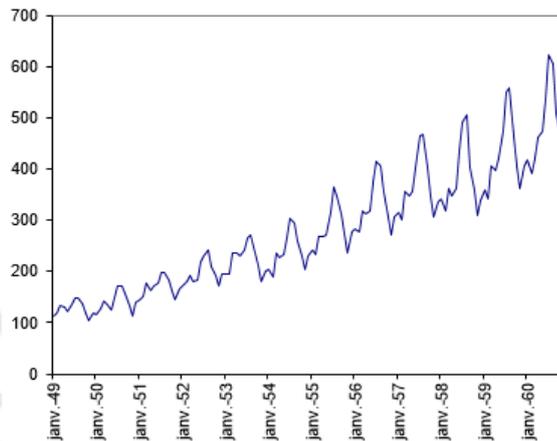
4.1.1.1 Series no estacionarias

Se define una serie como no estacionaria a aquella cuyas propiedades estadísticas se ven alteradas con el paso del tiempo. Estas propiedades incluyen, pero no se limitan a: media,

varianza, covarianza y correlación t . En términos matemáticos se puede expresar que un proceso es no estacionario cuando las propiedades estadísticas de una secuencia finita $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ ($n \geq 1$) difiere de la serie $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}$ para al menos un entero $h > 0$

En un gráfico el cambio de la varianza se observa como la dispersión de los datos. El cambio en la media, por otro lado, implica que los datos experimentan tendencias crecientes o decrecientes y que la serie no oscila alrededor de un nivel constante. Adicionalmente, las series no estacionarias pueden presentar efectos estacionales, es decir patrones que se repiten periódicamente.

Figura 2 Número de pasajeros de avión (mensual) USA, enero 1949 a diciembre 1960. Fuente: Apuntes M.J.Sánchez



La Figura 2 muestra una serie de pasajeros de avión de una aerolínea internacional recogida mensualmente desde enero de 1949 a diciembre de 1960. En esta figura, se puede observar que a medida que aumenta el número de pasajeros se incrementa la variabilidad, es decir, es una serie no estacionaria en varianza. Además, con respecto a la media de pasajeros se aprecia que esta sigue una tendencia creciente, serie no estacionaria en media y que sigue un patrón estacional.

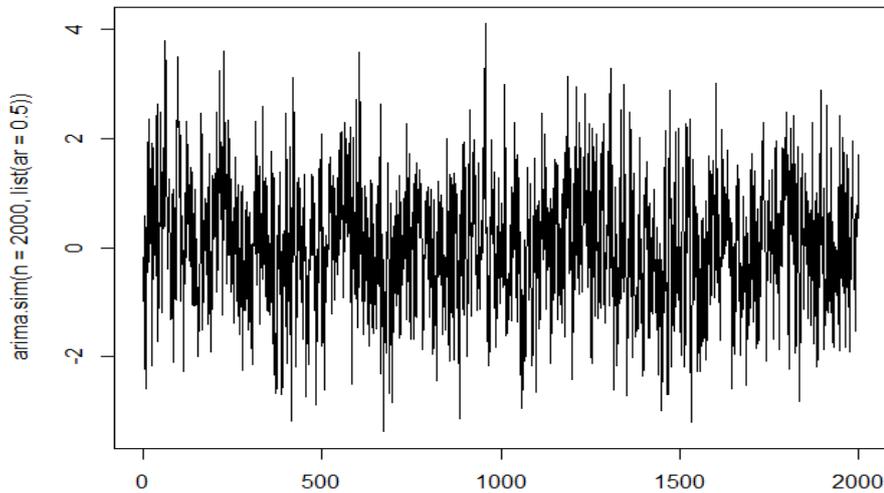
4.1.1.2 Series estacionarias

Se define una serie como estacionaria a aquella cuyas propiedades estadísticas se mantienen constantes a lo largo del tiempo. Debido a esta característica, las series estacionarias son predecibles. Matemáticamente se define un proceso estacionario cuando las propiedades estadísticas de cualquier secuencia finita $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ ($n \geq 1$) son semejantes a las de la secuencia $Y_{t_1+h}, Y_{t_2+h}, \dots, Y_{t_n+h}$ para cualquier valor entero $h = \pm 1, \pm 2, \dots$

Además, si una serie es estacionaria se pueden predecir intervalos de confianza para las predicciones. En un gráfico variable vs tiempo, se puede identificar una serie estacionaria cuando los valores oscilan alrededor de un nivel constante: media constante y la dispersión de los datos se mantiene con el tiempo: varianza constante (homocedasticidad).

En la Figura 3 se muestra una serie el proceso autorregresivo con parámetro $p = 1$, en este gráfico se puede apreciar que los valores oscilan alrededor del nivel 0 que es constante y no existen tendencias crecientes o decrecientes que se mantengan con el tiempo. Además, presenta una variabilidad constante alrededor de dicho nivel.

Figura 3 Proceso Autorregresivo AR (1)



4.1.2 Proceso de ruido blanco

Un proceso de ruido blanco es un proceso estacionario definido por las siguientes condiciones:

$$E [X_t] = 0 \quad t = 1, 2, 3 \dots$$

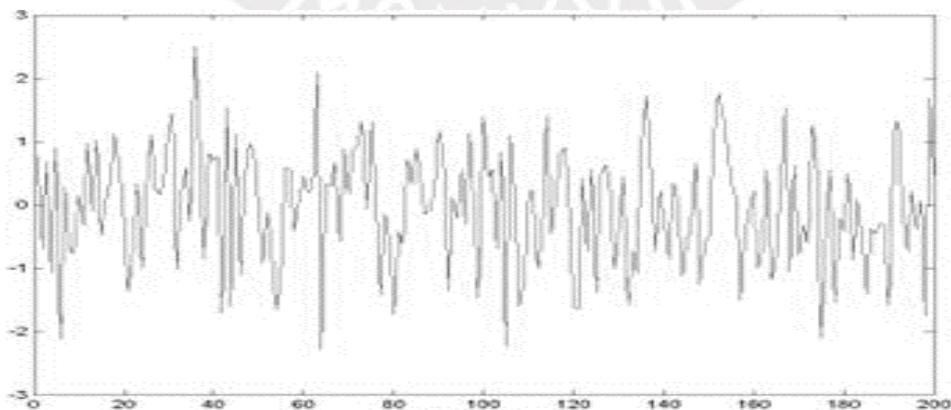
$$Var [X_t] = \sigma^2 \quad t = 1, 2, 3 \dots$$

$$Cov[X_t, X_{t-k}] = 0 \quad k = \pm 1, \pm 2$$

Ecuación 1 Características de un proceso de ruido blanco (Peña, 2010)

La primera condición establece que la esperanza es constante e igual a cero, la segunda condición, es que la variabilidad es constante y la tercera es que las variables del proceso no deben estar correlacionadas para ningún retardo, esto es que cada una de las observaciones es independiente. En un proceso de ruido blanco conocer los valores pasados no proporciona ninguna información sobre el futuro, en ese sentido se dice que el proceso “no tiene memoria” (Peña, 2010). Por ejemplo, un modelo en un ordenador que genere números aleatorios con media cero y varianza constante como se muestra en la Figura 4.

Figura 4 Proceso de Ruido Blanco



4.1.3 Funciones de autocorrelación simple y parcial

Cuando se analizan series temporales es necesario analizar si un valor determinado depende de valores previos, también llamados retardos. Para analizar si existe esta dependencia se utilizan dos herramientas: la Función de Autocorrelación Simple (FAS o ACF por sus siglas en inglés) y la Función de Autocorrelación Parcial (FAP o PACF por sus siglas en inglés).

4.1.3.1 Función de Autocorrelación simple, FAS

La función de autocorrelación simple de una serie proporciona la estructura de dependencia entre los valores X_t y $X_{t-1}, X_{t-2}, X_{t-3}, \dots$. En ese sentido, la función de autocorrelación proporciona el coeficiente de correlación entre observaciones separadas un número de k de periodos, entonces la FAS es una sucesión de números ρ_k .

Así ρ_1 indica cómo influye una observación X_t sobre la siguiente X_{t+1} , mientras que ρ_2 , la influencia de la observación X_t sobre la observación generada luego de dos periodos X_{t+2} y así sucesivamente.

$$\rho_k = \text{cor}(X_j, X_{j-k}) = \frac{\text{cov}(X_j, X_{j-k})}{\sqrt{V(X_j)} * \sqrt{V(X_{j-k})}}$$

Ecuación 2 Función de autocorrelación simple

Entonces si el valor ρ_1 es distinto de cero, significa que X_1 está relacionado con X_2 y X_2 con X_3 . De ahí, se infiere que X_1 está relacionado con X_3 , es decir la autocorrelación ρ_2 también será distinta de cero. Análogamente, con $\rho_3, \rho_4 \dots$ luego, se hace necesario identificar cuál es la influencia de X_1 sobre X_3 directamente, es decir sin tener en cuenta el valor X_2 . Para ello es necesario estudiar la Función de Autocorrelación Parcial o FAP.

4.1.3.2 Función de Autocorrelación parcial, FAP

La función de autocorrelación parcial proporciona la relación directa que existe entre dos observaciones separadas por k retardos, sin tomar en cuenta los retardos intermedios. De este modo se puede dar el caso de que π_1 sea significativo mientras que π_2 y π_3 no.

$$\pi_k = \text{cor}(X_j, X_{j-k} / X_{j-1}, X_{j-2}, \dots, X_{j-k+1})$$
$$\pi_k = \frac{\text{cov}(x_j - \hat{x}_j, x_{j-k} - \hat{x}_{j-k})}{\sqrt{V(X_j - \hat{X}_j)} \sqrt{V(X_{j-k} - \hat{X}_{j-k})}}$$

Ecuación 3 Función de autocorrelación parcial

4.1.4 Modelos univariantes de series temporales

Los modelos univariantes de series temporales consideran una sola variable endógena y tratan de construir un modelo empleando los valores previos de esta variable para predecir valores futuros de la misma.

Estos modelos suelen estar compuestos de una función de relación, que describe la relación causal entre la variable para un tiempo T con respecto a valores previos que tomó esa variable y una parte de error, que toma en cuenta consideraciones externas a la variable analizada.

$$X_t = f(X_{t-1}, X_{t-2}, \dots) + a_t$$

Ecuación 4 Representación de un modelo univariante de una serie temporal.

Fuente: Series Temporales Univariantes, Peña (2010)

La Ecuación 4 representa a la variable temporal X_t como una función de sus valores en los instantes de tiempo previos, mientras que el valor de a_t representa la componente de ruido o innovación y su valor no está condicionado por sus valores previos.

$$X_t = X_t^* + a_t$$

Ecuación 5 Forma simplificada para representar un modelo univariante

Fuente: Series Temporales Univariantes, Peña (2010)

En la Ecuación 5, X_t^* representa la componente predecible y a_t es el término de innovación y representa un proceso de ruido blanco con media 0 y varianza σ^2 . Además, la covarianza entre dos innovaciones en distinto instante de tiempo es igual a cero, lo cual implica que dichas innovaciones son independientes.

Para obtener la Ecuación 5 existen dos enfoques básicos:

- Métodos clásicos: Modelar X_t^* , donde X_t^* es la variable a predecir
- Método Box-Jenkins: Obtener a_t , donde a_t es ruido blanco o factor de innovación

4.2 Redes Neuronales Artificiales

Las redes neuronales artificiales son modelos matemáticos para el aprendizaje automático inspirados en la estructura del cerebro biológico. Incluso con muestras limitadas, las redes neuronales pueden generalizar y tratar con éxito ejemplos no vistos. Las redes neuronales comenzaron con los perceptrones simples de una sola capa, pero ahora están representados por un conjunto diverso de arquitecturas que incluyen múltiples capas e incluso conexiones recurrentes para implementar la retroalimentación.

Los modelos de redes neuronales pueden ser mejor caracterizados como modelos computacionales con la habilidad de adaptarse o ‘aprender’, esto con el objetivo de predecir o clasificar las respuestas de una variable. Además, como su nombre indica, este modelo intenta imitar las conexiones neuronales que existen en nuestro cerebro, que realiza una computación masiva de unidades concurrentes, con redundancia de tareas y conexiones que proporcionan robustez frente a numerosos estímulos complejos y, a menudo, con mucho ruido.

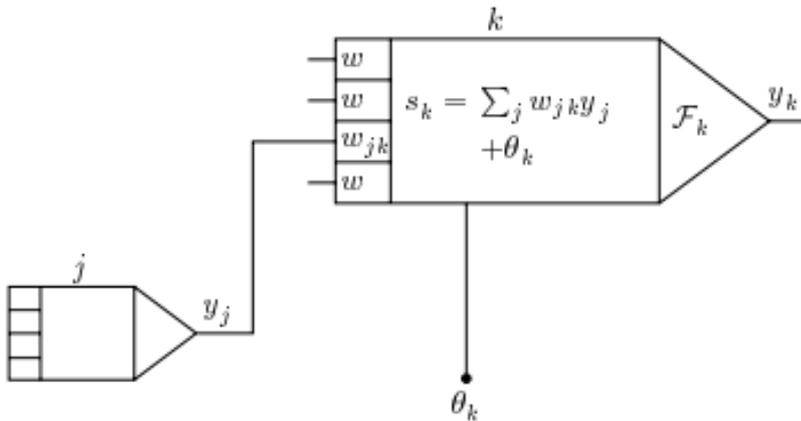
Para imitar este comportamiento las redes neuronales artificiales se estructuran como un conjunto de unidades computacionales simples, denominadas neuronas o nodos, que adquieren capacidad de procesamiento creciente a medida que aumenta su número e interconectividad. Estos nodos se comunican enviando señales sobre un gran número de conexiones con un peso atribuido. (Kröse 1996).

En breve, se puede describir a las redes neuronales artificiales como redes que comunican señales, a través de pesos y funciones de activación que activan neuronas o nodos. Estas redes ajustan estos pesos usando algoritmos de entrenamiento para resolver un problema dado. El resultado de estas redes se calcula como la suma ponderada de señales, a la que se le aplican sesgos a través de funciones de activación. (Jones, 2017)

Las redes neuronales artificiales constan de los siguientes elementos:

- Un conjunto de unidades de procesamiento o nodos
- Un estado de activación para cada unidad, y_k es decir el resultado de la unidad k
- Conexiones entre las unidades, estas conexiones se definen por un peso w_{jk} , que determina el efecto que tiene la unidad j en la unidad k
- Regla de propagación, determina el input efectivo s_k de los inputs externos sobre una unidad k
- Función de activación \mathcal{F}_k , determina el nivel de activación en función del input efectivo s_k
- Inputs externos o término de sesgo, θ_k , para cada unidad

Figura 5 Componentes básicos de una red neuronal artificial. Fuente: Kröse 1996

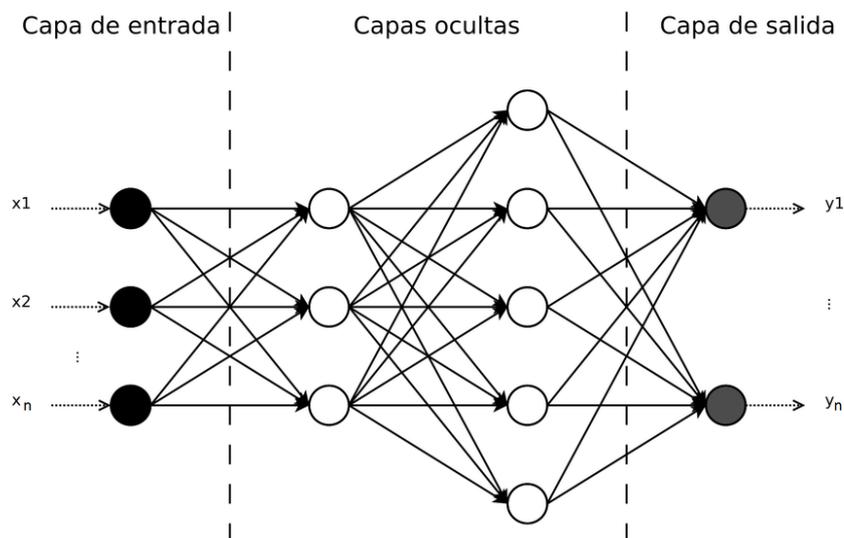


4.2.1 Neuronas o unidades de procesamiento

Cada unidad de procesamiento recibe una señal de entrada de nodos próximos o de fuentes externas y las utiliza para computar una señal de salida que luego es propagada hacia otras unidades. Según la intensidad de esta señal, el siguiente nodo puede ser activado o no. Además, de este procesamiento, cada neurona realiza la tarea de ajustar los pesos de las conexiones con otras unidades de procesamiento de manera simultánea.

En los sistemas neuronales es necesario distinguir tres tipos de unidades: unidades de entrada, que reciben señales desde fuera de la red neuronal; unidades de salida, que envían la data fuera de la red neuronal; y unidades ocultas, cuyas entradas y salidas permanecen dentro de la red neuronal.

Figura 6 Capas de una red Neuronal. Fuente: Researchgate.net



4.2.2 Función de activación

La función de activación \mathcal{F}_k se aplica sobre combinaciones lineales del resultado obtenido en los nodos previos. Usualmente la función de activación es una función no decreciente, aunque no se limita solo a ese tipo de funciones. Generalmente, se establece una función con un umbral que puede ser una función muy restrictiva como la función signo (sgn), una función lineal o semilineal, o una función suavizada como la función sigmoide que limita los resultados en un rango de $[0, 1]$ o la tangente hiperbólica que limita los resultados en un rango de $[-1, +1]$ como muestra la Figura 7. En la mayoría de los casos la misma función de activación se usa para cada neurona de la red.

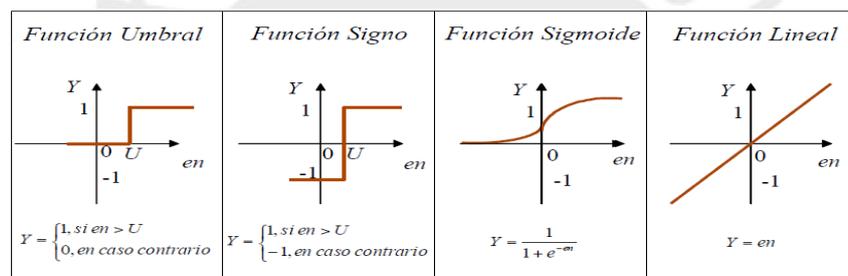
$$y_k = F(s_k) = \frac{1}{1 + e^{-sk}}$$

Donde:

$$s_k = \sum_J w_{jk} y_j + \theta_k$$

Ecuación 6 Función sigmoide o en forma de S

Figura 7 Funciones de activación para una unidad de procesamiento.



4.2.3 Topologías de redes neuronales

Según la topología de las redes neuronales estas pueden clasificarse en redes neuronales *feedforward* y redes neuronales recurrentes.

4.2.3.1 Redes neuronales *feedforward* o de avance

Las redes *de* avance inicialmente fueron propuestas como un modelo autorregresivo no lineal para la predicción de series temporales (Lapedes y Farber, 1987), de ahí que una red de avance sea una aproximación de h dada por la Ecuación 7. Entonces, se dice que la red neuronal de avance es un mapeo no lineal de observaciones previas para la predicción de observaciones futuras. En la Ecuación 7 la función $f(x)$ es una función monótona acotada, típicamente una sigmoide.

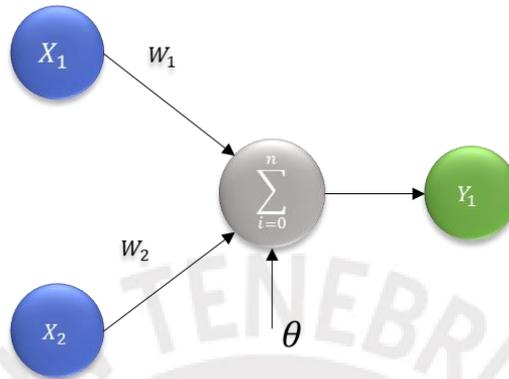
$$\hat{x}_t = h(x_{t-1}, \dots, x_{t-p}) = \sum_{i=1}^I W_i f\left(\sum_{j=1}^p w_{ij} x_{t-j} + \theta_i\right)$$

Ecuación 7 Representación de una Red de avance. Fuente Connor 1994

Los parámetros W_i , w_{ij} y θ_i son estimados de una muestra de entrenamiento, obteniendo un estimado \hat{h} de la función h . Estos estimados son obtenidos minimizando la suma de cuadrados de los residuos (SSE) obtenidos de comparar x y \hat{x} .

En las redes neuronales *feedforward* el flujo de información desde las unidades de entrada hacia las unidades de salida fluye estrictamente hacia adelante. Es decir, el procesamiento de información se extiende sobre múltiples nodos o capas de nodos, sin ninguna conexión que permita la retroalimentación entre las conexiones desde los nodos de salida hacia los nodos de entrada en la misma capa o capas previas.

Figura 8 Red neuronal de una sola capa con dos entradas y una salida.



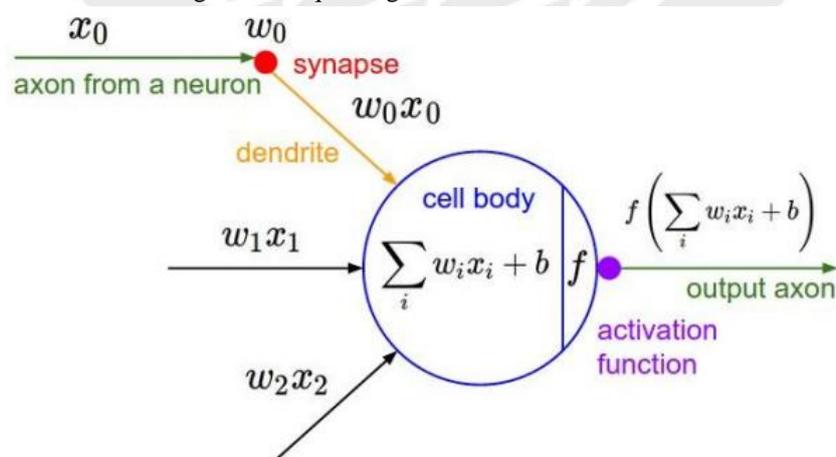
En la Figura 8 se puede apreciar el caso más simple de una red neuronal *feedforward*, esta consiste en dos neuronas de entrada, cada una con un factor de peso w_{i0} conectada a los inputs i y una sola salida, además de los valores de entrada se añade un término de sesgo θ . El resultado, entonces, queda afectado por este valor y una función de las dos entradas. Si consideramos esta función de activación como una función lineal, entonces el resultado estaría dado por la Ecuación 8.

$$y = \sum_{j=1,2} w_j x_j + \theta$$

Ecuación 8 Resultado de la red neuronal planteada en la figura 8

De manera general, se puede indicar entonces que una red neuronal con n valores de entrada que se transforman con una función de activación en un valor de salida tiene la estructura mostrada en la Figura 9.

Figura 9 Esquema general de una red neuronal



Cada uno de estos n elementos de entrada pasa a través de una transformación afín que básicamente combina los efectos de dos o más valores de entrada que luego pasan a través de una función de activación, de la cual se obtiene una probabilidad. Esta probabilidad determina si la neurona se activará o si el resultado debe conectarse con la función de error para evaluar el rendimiento del algoritmo.

En algunos modelos de redes neuronales son necesarias varias transformaciones para llegar al resultado final, por ello se insertan capas ocultas en los modelos que mejoran su capacidad de predicción; no obstante, el ajuste del peso de las capas ocultas resulta complicado.

4.2.4 Entrenamiento de redes neuronales

Para los algoritmos de *Machine Learning*, el entrenamiento de los modelos equivale a lo que en estadística aplicada se conoce como la estimación de los parámetros del modelo. En ese sentido, una red neuronal se debe configurar de manera que la aplicación de las señales de entrada produzca el conjunto de señales de salida deseadas. Existen muchos algoritmos de aprendizaje, no obstante, la idea general de estos algoritmos es modificar los pesos de las conexiones de manera iterativa hasta que el error producido disminuya por debajo de un valor predeterminado. Para lograr dicho resultado, una forma es ajustar los pesos manualmente utilizando conocimientos previos sobre las relaciones entre variables, otra forma es estimar las redes neuronales alimentándolas con patrones y dejando que el modelo ajuste sus pesos.

4.2.4.1 Entrenamiento de redes neuronales según el tipo de aprendizaje:

- Aprendizaje supervisado o Aprendizaje asociativo en el cual las redes neuronales se entrenan, atribuyendo una serie de valores de entrada con una o varias variables de respuesta asociadas.
- Aprendizaje no supervisado o autoorganizado, en este caso un resultado es entrenado para responder a uno o varios clústeres de patrones de entrada. En este paradigma el sistema encuentra características estadísticas de las señales de entrada y desarrolla las categorías en las que deben ser clasificadas según el estímulo de entrada.

4.2.4.2 Algoritmo de propagación hacia atrás o Backpropagation

El algoritmo de propagación hacia atrás (*backpropagation*, en adelante) es un algoritmo de aprendizaje muy popular, que calcula los cambios en los pesos de las conexiones con el método de gradiente descendiente. Estos cambios se evalúan para cada nodo en todas las capas de una red neuronal y a su vez para cada *epoch*.

Un *epoch* es una iteración completa, que implica el paso de los errores hacia adelante y hacia atrás para todos los valores de entrenamiento. El gradiente descendiente es esencialmente el ajuste de un cálculo diferencial para minimizar el error entre el resultado producido por la red y el resultado deseado. (Lewis, 2016)

El algoritmo de *backpropagation*, aplica el error de propagación de las señales de salida hacia las señales de entrada y gradualmente ajusta los pesos de los nodos para minimizar la SSE (suma de los errores al cuadrado) usando la técnica de la gradiente descendiente. Este algoritmo de aprendizaje funciona de la siguiente manera:

- Paso 1: Inicialización de la red

Lo primero es establecer los valores iniciales de los pesos para cada conexión. Generalmente, se establecen de manera aleatoria.

- Paso 2: Propagación hacia adelante

La información se transfiere hacia adelante, desde las señales de entrada pasando por los nodos ocultos y la capa de salida de la red neuronal, a través de las funciones de activación y los pesos en las conexiones. La función de activación es típicamente una función sigmoideal aplicada a la suma ponderada de los nodos de entrada.

- Paso 3: Análisis del error

El resultado de la red se compara con los valores de salida esperados. Si el error está por debajo de los parámetros preestablecidos la red está entrenada y el algoritmo se detiene.

- Paso 4: Propagación hacia atrás

El error en la capa de salida es usado para modificar nuevamente los pesos de las conexiones. Luego, el algoritmo propaga el error a través de la red hacia las señales de entrada, capa por capa y calcula el gradiente de cambio del error con respecto a los cambios en los pesos de las conexiones.

- Paso 5: Ajuste

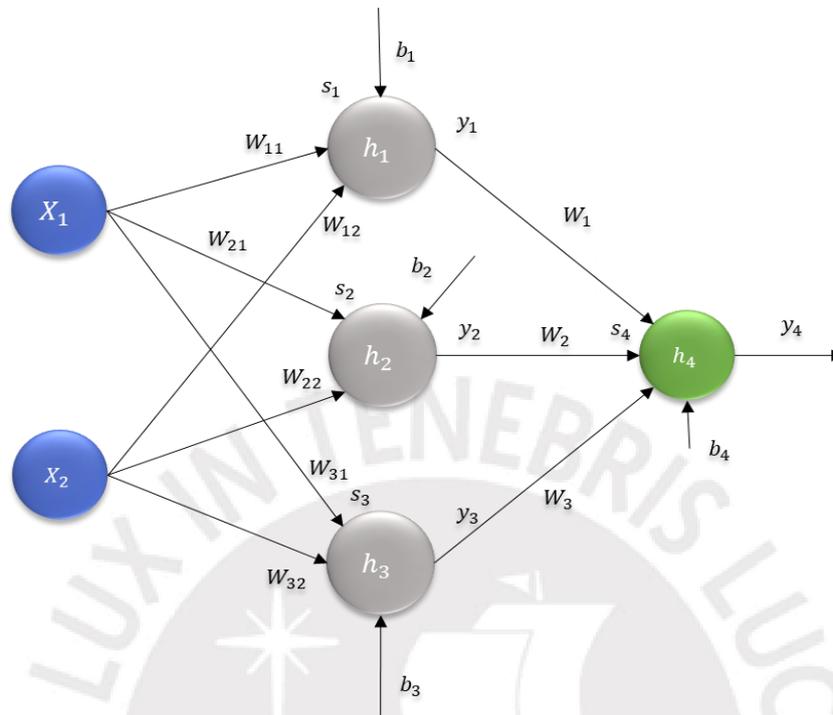
Se realizan los ajustes en los pesos de las conexiones, con el resultado de la gradiente de cambio con el objetivo de minimizar el error. Luego los pesos y los sesgos de cada nodo son ajustados por un factor basado en la derivada de la función de activación, las diferencias entre las salidas de la red neuronal y los valores objetivos. Si la derivada parcial del error con respecto al cambio en los pesos es positiva, entonces, el peso se reduce; si en cambio esta derivada es negativa, entonces el peso se incrementa. Como ya se ha indicado antes, cada ciclo a través de este proceso se denomina *epoch*.

Al pasar por este proceso de cálculo de errores y ajuste de pesos se dice que la red “aprende”. Este proceso se aplica a toda la muestra de entrenamiento hasta que los pesos se mantienen constantes o se alcanza el tiempo máximo de análisis establecido.

No obstante, encontrar una solución global que evite los mínimos locales de las funciones de error puede ser complicado debido a que una típica red neuronal suele tener cientos de pesos (Redes neuronales profundas o *Deep Neural Networks*) que al combinarse generan una solución altamente no lineal, debido al efecto acoplado de diversas funciones no lineales, haciendo el proceso de optimización complejo.

A continuación, se demostrará un ejemplo del funcionamiento de este algoritmo, para ello se debe tener en cuenta el gráfico de la Figura 10 :

Figura 10 Red de ejemplo para la estimación de parámetros usando Backpropagation



Paso 1– Inicializar los pesos y bias de la red: En un principio se establecen los pesos $[w_{11}, w_{21}, w_{12}, w_{22}, w_{31}, w_{32}, w_1, w_2, w_3]$ y los bias $[b_1, b_2, b_3, b_4]$

Paso 2 – Propagación hacia adelante:

Tabla 2 Valores de entrada y salida de la red de ejemplo

Valores de entrada	Valores de salida
$s_1 = w_{11}x_1 + w_{12}x_2 + b_1$	$y_1 = F_1(s_1) \Rightarrow 1/(1 + e^{-i_1})$
$s_2 = w_{21}x_1 + w_{22}x_2 + b_2$	$y_2 = F_2(s_2) \Rightarrow 1/(1 + e^{-i_2})$
$s_3 = w_{31}x_1 + w_{32}x_2 + b_3$	$y_3 = F_3(s_3) \Rightarrow 1/(1 + e^{-i_3})$
$s_4 = w_1y_1 + w_2y_2 + w_3y_3 + b_4$	$y_4 = F_4(s_4) = 1 / (1 + e^{-i_4})$

Paso 3 – Cálculo del error, donde o es el número de nodos de salida, en este caso sería igual a 1, d es el valor deseado (valor observado). La función de error se define como la suma de cuadrados del error producido,

$$E = \frac{1}{n} \sum_{o=1}^{o=1} (d_o - y_4)^2$$

Ecuación 9 Cálculo del error en la capa de salida

Paso 4 – Propagación hacia atrás

Se expresa el error en función de los pesos y luego se aplica la regla de la cadena para minimizar la función de error donde γ es una constante de proporcionalidad:

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial s_k} \cdot \frac{\partial s_k}{\partial w_{jk}} = \frac{\partial E}{\partial s_k} \cdot y_j$$

$$\Rightarrow \Delta w_{jk} = -\gamma \cdot \frac{\partial E}{\partial s_k} \cdot y_j$$

$$\delta = \frac{\partial E}{\partial s_k} = \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial s_k} = \frac{\partial E}{\partial y_k} \cdot F'(s_k)$$

Para el nodo de salida (o):

$$\delta_o = \frac{\partial E}{\partial y_o} \cdot F'(s_o) = (d_o - y_o) \cdot F'_o(s_o)$$

Ecuación 10 Variación del error de la capa de salida en función del valor de entrada

Para los nodos ocultos (h):

El error puede escribirse como una función de los nodos de entrada desde las capas ocultas hasta la capa de salida

$$\frac{\partial E}{\partial y_h} = \sum_{o=1}^o \frac{\partial E}{\partial s_o} \frac{\partial s_o}{\partial y_h} = \sum_{o=1}^o \frac{\partial E}{\partial s_o} \frac{\partial}{\partial y_n} \sum_{j=1}^{N_h} w_{ko} y_j = \sum_{o=1}^{N_o} \frac{\partial E}{\partial s_o} \cdot w_{ho}$$

$$\delta_h = F'(s_h) \sum_{o=1}^o \delta_o w_{ho}$$

Ecuación 11 Variación del error en función de los valores de entrada en las capas ocultas

Paso 5 – Ajuste

Los pesos se ajustarán según la siguiente fórmula

$$\Delta w_{jk} = -\gamma \cdot \frac{\partial E}{\partial w_{jk}} \cdot y_j$$

Ecuación 12 Ajuste de pesos

Donde γ es una constante de proporcionalidad

4.2.5 Redes neuronales recurrentes (RNN)

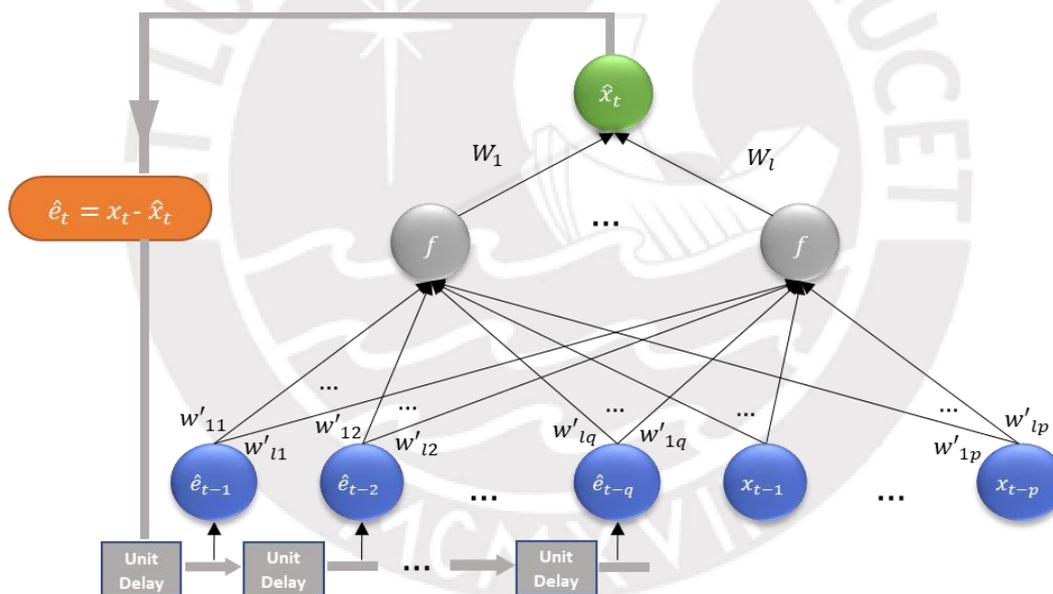
Las redes neuronales recurrentes fueron presentadas en los 80's y su capacidad de almacenar información mientras procesan nuevas señales de entrada, abrió una nueva categoría de problemas para las redes neuronales. Por ejemplo, las aplicaciones de series temporales requieren un tipo de topología que tenga en cuenta el histórico de las señales de entrada. En estos casos, la memoria de las redes neuronales recurrentes las hace ideales para procesar este tipo de datos en los cuales valores previos de los datos pueden afectar a valores futuros (Jones, 2017).

Un modelo de red neuronal recurrente puede ser aproximado por la Ecuación 13. En dicha ecuación, los parámetros W_i , w_{ij} , w'_{ij} son estimados por el método de mínimos cuadrados; la topología de esta red se muestra en la Figura 11

$$\hat{x}_t = \sum_{i=1}^I W_i f \left(\sum_{j=1}^p w_{ij} x_{t-j} + \sum_{j=1}^q w'_{ij} (x_{t-j} - \hat{x}_{t-j}) + \theta_i \right)$$

Ecuación 13 Representación de una red neuronal recurrente

Figura 11 Representación de una red neuronal recurrente



La ventaja de una red recurrente sobre una red *feedforward* recae en las dimensiones de la red, ya que para una red *feedforward* las dimensiones se multiplican con el número de señales de entrada y se obtiene una red muy grande (Kröse 1996). Por otro lado, redes neuronales recurrentes requieren menos señales de entrada ya que la red aprende de la influencia de los resultados anteriores y, por lo tanto, pueden entrenarse de manera más rápida y sencilla que una red *feedforward*.

Las RNN agrupan diferentes topologías que pueden ser aplicados a distintos problemas. Uno de los enfoques más simples es introducir una capa de estados en la red. Esta capa o capas de estado influyen a las siguientes señales que se reciben. Dos enfoques muy populares para la introducción de estados son las redes de Elman y las redes de Jordan. En estas redes los valores de activación de los nodos ocultos pueden retroalimentarse de un conjunto adicional de valores de entrada (Red de Elman) o los valores de salida retroalimentar a los nodos ocultos (Red de Jordan).

A continuación, se describirá con más detalles las redes de mencionadas previamente.

4.2.5.1 Redes de Jordan

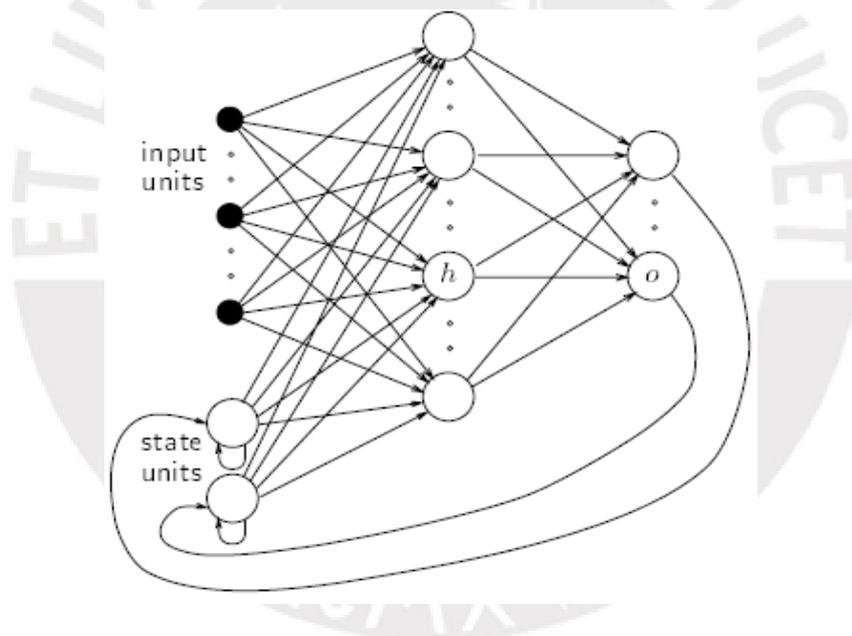
Las redes de Jordan fueron introducidas en 1986 y fueron uno de los primeros modelos de redes neuronales recurrentes. En este tipo de redes los valores de activación de las señales de salida son recurrentemente incluidas en la capa de entrada como unidades adicionales denominadas unidades de estado o de contexto, este proceso permite a la red tener una memoria de su valor previo.

En las redes de Jordan, se definen tantos nodos de estado como el número de unidades de salida en la red, además, las conexiones entre los nodos de salida y los nodos de estado tienen un peso fijo de +1. Adicionalmente, el aprendizaje toma lugar en las conexiones de los nodos de entrada y las unidades ocultas, así como desde las unidades ocultas a las unidades de salida.

En la Figura 12 se muestra un esquema generalizado de una red de Jordan, en la que se muestran los nodos de estado como parte de la capa de entrada y como tal mantienen una conexión con los nodos de las capas ocultas.

Cabe destacar que en este modelo se permiten auto conexiones para los nodos de estado

Figura 12 Estructura general de las redes de Jordan. Fuente: Kröse 1996

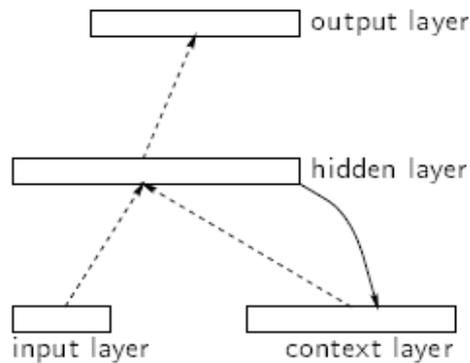


4.2.5.2 Redes de Elman

Las redes de Elman fueron introducidas en 1990, y de manera similar a las redes de Jordan, añaden una capa de estado o contexto que sirve como un dato de entrada y cuyos valores provienen de la retroalimentación de las capas ocultas. Además, las unidades ocultas conectadas con los nodos de contexto tienen un peso fijo de +1. En la

Figura 13 se aprecia la estructura general de una red de Elman, en este caso se muestran rectángulos ya que estos representan las capas, de manera que, puede haber tantos nodos en las capas de contexto como nodos existan en las capas ocultas.

Figura 13 Estructura general de las redes de Elman. Fuente: Kröse 1996



Se dice que las redes de Elman tienen memoria en el sentido de que los nodos de contexto almacenan los valores de activación de unidades ocultas de un paso previo y las devuelven a la red como nodos de entrada, como indica la Ecuación 14

$$h_t = F(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = F_y(W_y h_t + b_y)$$

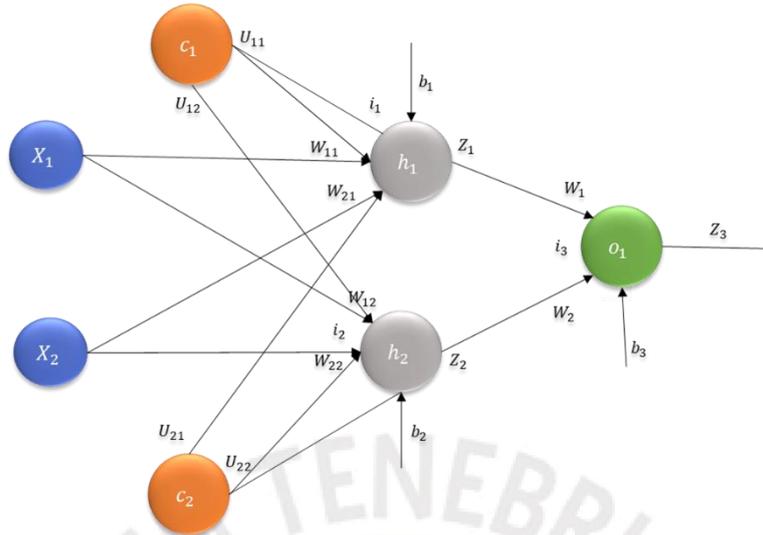
Ecuación 14 Formulación de la red de Elman

En este tipo de redes el aprendizaje se realiza como se explica a continuación:

- Paso 1: Para un $t=1$ se inicializan los valores de las unidades en la capa de contexto con el valor de 0.
- Paso 2: La capa de entrada es inicializada con el primer patrón y se realizan los cálculos de los pesos y valores de activación
- Paso 3: Se aplica el algoritmo de *backpropagation*
- Paso 4: Las unidades de contexto toman el valor de activación de las unidades ocultas
- Paso 5: Se cargan los siguientes valores de inicio.

Cabe destacar que las unidades de contexto en el tiempo t , toman el valor de activación de las unidades ocultas en el tiempo $t-1$. A continuación, se mostrará un ejemplo básico para explicar el funcionamiento de la red de Elman mostrada en la Figura 14.

Figura 14 Red de ejemplo para explicar las fórmulas de Elman



Para esta red, se calculan los valores de entrada (i) y los valores de salida (z) de cada nodo en dos epoch (p) consecutivos.

Tabla 3 Cálculo de los valores de entrada para cada nodo de la red de ejemplo

Valores de entrada y salida por nodo		
$h_{1,t}$	$i_{1,t} = w_{11}x_{1,t} + w_{12}x_{2,t} + U_{11} \cdot z_{1,t-1} + U_{21} \cdot z_{2,t-1} + b_1$	$z_{1,t} = F(i_1) = 1 / 1 + e^{-i_1}$
$h_{2,t}$	$i_{2,t} = w_{21}x_{1,t} + w_{22}x_{2,t} + U_{12} \cdot z_{1,t-1} + U_{22} \cdot z_{2,t-1} + b_2$	$z_{2,t} = F(i_2) = 1 / 1 + e^{-i_2}$
$o_{1,t}$	$i_{3,t+1} = w_1 z_1 + w_2 z_2 + b_3$	$z_{3,t} = F(i_3) = 1 / 1 + e^{-i_3}$

Valores de entrada y salida por nodo		
$h_{1,t+1}$	$i_{1,t+1} = w_{11}x_{1,t+1} + w_{12}x_{2,t+1} + U_{11} \cdot z_{1,t} + U_{21} \cdot z_{2,t} + b_1$	$z_{1,t+1} = F(i_{1,t+1}) = 1 / 1 + e^{-i_{1,t+1}}$
$h_{2,t+1}$	$i_{2,t+1} = w_{21}x_{1,t+1} + w_{22}x_{2,t+1} + U_{12} \cdot z_{1,t} + U_{22} \cdot z_{2,t} + b_2$	$z_{2,t+1} = F(i_{2,t+1}) = 1 / 1 + e^{-i_{2,t+1}}$
$o_{1,t+1}$	$i_{3,t+1} = w_1 z_{1,t+1} + w_2 z_{2,t+1} + b_3$	$z_{3,t+1} = F(i_{3,t+1}) = 1 / 1 + e^{-i_{3,t+1}}$

En las tablas mostradas previamente w_{ik} indica el peso del nodo k sobre el nodo i, U_{ij} denota el efecto del nodo de contexto j sobre el nodo i y b_k , denota el sesgo del nodo k.

Capítulo 3

5. Metodología

Se abordaron dos esquemas de modelado: análisis de series temporales y redes neuronales recurrentes que fueron caracterizadas con un análisis univariante. En ambos casos se busca predecir valores posteriores de la temperatura central.

Las variables de entrada escogidas son series de temperaturas correspondientes a pacientes ingresados en el servicio de Medicina Interna del Hospital Universitario de Móstoles en Madrid entre noviembre de 2008 y octubre de 2010. Además, los registros de temperatura corresponden a pacientes con edades entre 40 y 74 años, que presentaron una temperatura de 38°C el día previo a empezar las mediciones.

Aunque algunas variables pueden no seguir una distribución normal, se puede aplicar series temporales ARIMA con la premisa de que la serie sigue una distribución normal condicionada, es decir que el valor que toma una variable para cualquier instante está condicionado a sus k valores previos ($k > 0$).

5.1 Recogida de datos

Los datos fueron recogidos con un equipo de Monitorización Continua de Temperatura, que incluye un termómetro timpánico de infrarrojos denominado *Thercom* de la empresa *Innovatec Sensorización y Comunicación*. Dicho dispositivo, permite la medición, almacenamiento y traspaso inalámbrico de una gran cantidad de datos sin la necesidad de la supervisión de un personal sanitario. Cabe resaltar que este equipo ya ha sido utilizado en otros estudios similares sobre temperatura como Varela (2006) y Jordán (2017)

Previo a la medición de datos se realizó una otoscopia a los pacientes, para comprobar la integridad de la membrana timpánica, posteriormente, se colocó a los pacientes una sonda óptica de temperatura en el conducto auditivo externo. Estos dispositivos fueron configurados para realizar una determinación de temperatura central por minuto.

Se monitorizó la temperatura de cada uno de los pacientes por al menos 24 horas, aunque en algunos casos se tuvo que quitar el dispositivo antes de este periodo, ya que se presentaban continuas desconexiones.

Características técnicas del dispositivo de monitorización

El dispositivo de monitorización continua de temperatura Thercom empleado, presenta las siguientes características, según el manual del equipo.

- Batería de litio 3,7V con una autonomía de entre 48 y 168 horas midiendo en tiempo real, dependiendo de la frecuencia de muestreo. La batería se recarga por USB en 60 minutos.
- Precisión de la medida de temperatura $\pm 0,2^{\circ}\text{C}$, depende del sensor que se emplee.
- Resolución de la medida de temperatura $\pm 0,01^{\circ}\text{C}$.
- Frecuencia de muestreo seleccionable entre 30 segundos y 15 minutos a intervalos
- 4 canales de toma de datos. Cada uno de los canales puede tomar una medida independiente.
- Comunicación a PC mediante USB o Bluetooth
- Tamaño del dispositivo 28x66x66 mm
- Peso del dispositivo 83 g.
- Rango máximo de medida de temperatura entre $-10,0^{\circ}\text{C}$ y $50,0^{\circ}\text{C}$.
- Sensores de temperatura YSI-400 e YSI-700, ambos con conexión hembra compatibles a Thercom.
- Memoria interna de almacenamiento de datos para 6527 lecturas (acelerómetro y 4 canales). Útil si se corta la comunicación, ya que los datos se almacenan en la memoria interna y se transmiten al reanudarse la comunicación.
- Permite la configuración de alarmas visuales, así como el aviso por envío de SMS o correo electrónico. El umbral de la alarma puede ser por encima o por debajo de límites.
- El equipo dispone de auto calibración.

5.2 Selección de datos

Debido a que la toma de datos de temperatura podría ser incómoda para los pacientes, hay momentos en los datos en que se observan mediciones no fiables. Por ello, se procedió a la revisión manual de cada serie en búsqueda de desconexiones o alteraciones del registro, se usaron series que presentaban todos los datos. Para el análisis, es fundamental la continuidad de toda la serie temporal; en algunos casos los registros fueron correctos durante varias horas seguidas, por lo que algunas de las series pudieron incluirse parcialmente en el análisis.

De las series disponibles, se seleccionó una muestra representativa de los pacientes, tomando las termometrías de pacientes con edades de 48, 63 y 72 años, del género femenino y masculino.

Además, para poner a prueba los modelos, se utilizó el método de validación de “Holdout”, en el que se separan los datos en una muestra de entrenamiento, que son los datos usados para entrenar a los diferentes modelos y una muestra de prueba, que es usada para evaluar cómo se desempeña el modelo frente a datos nuevos. Generalmente, los datos se dividen en un 80% para la muestra de entrenamiento y un 20% para la muestra de prueba. No obstante, debido a limitaciones de los datos y las mejoras de los modelos al usar un set de datos más grande, los datos fueron divididos en una proporción de 85% y 15%.

5.3 Variable utilizada

- Temperatura Central

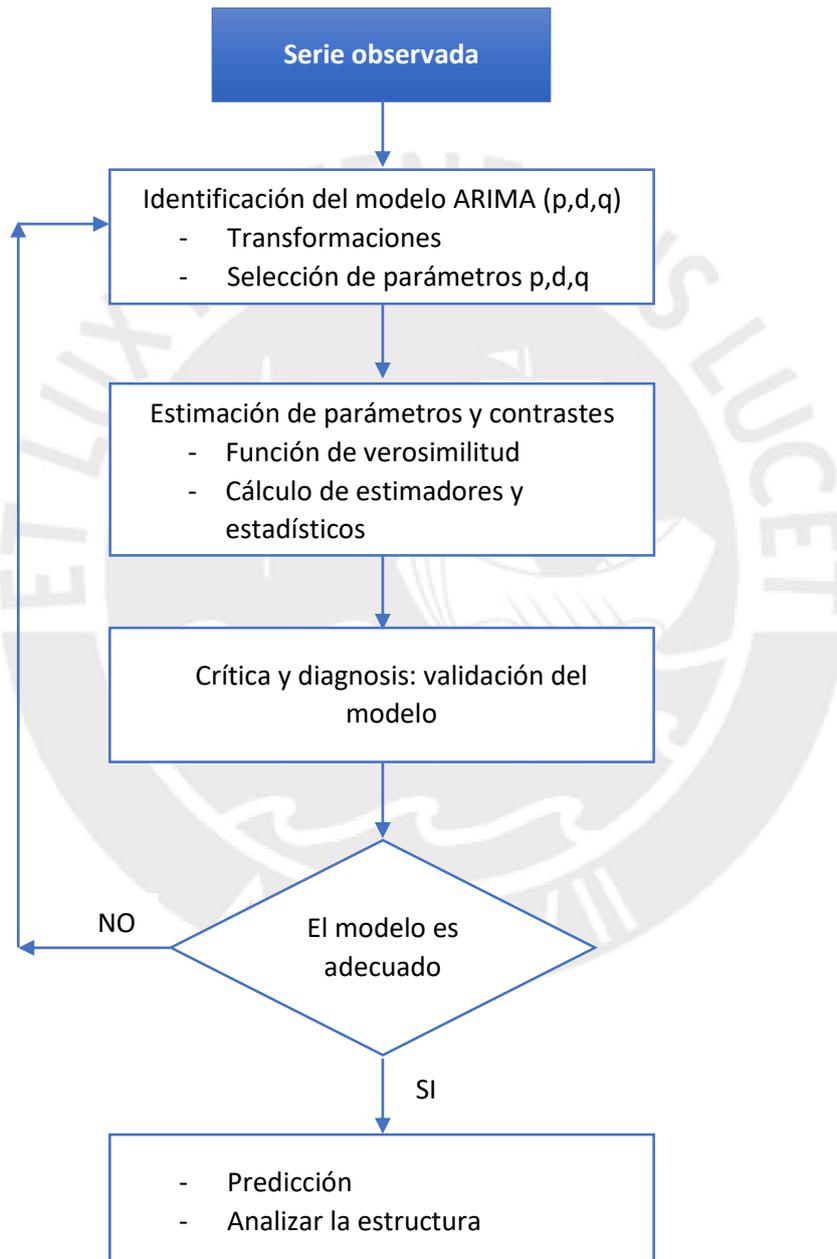
En la práctica de medicina se suele utilizar la temperatura central como un sustituto de la temperatura interna del cuerpo (núcleo), aunque pueden existir algunas diferencias según donde se mida: cavidad oral, rectal, esofágica, membrana del tímpano, hipotalámica o de la sangre al pasar por cualquiera de los órganos de la parte central o nuclear. En este caso los datos sobre temperatura central recogidos fueron realizados con una medición de temperatura del canal auditivo externo.



5.4 Análisis de Series Temporales

En primer lugar, se realizó un modelo de series temporales según el modelo Box-Jenkins para analizar los datos de temperatura y obtener predicciones sobre futuros valores de la serie. Para el análisis se siguieron los pasos indicados en el flujograma de la Figura 15. En los siguientes apartados se presentarán las diferentes etapas que llevarán al modelo correcto

Figura 15 Flujograma para modelar el análisis de una serie temporal



5.4.1 Identificación del modelo

5.4.1.1 Transformaciones

Para analizar y modelizar una variable según el modelo Box-Jenkins es necesario determinar si la serie es estacionaria y si tiene alguna estacionalidad. Si se determina que la serie no es estacionaria se transforman los valores de la serie hasta obtener una serie con varianza y media constantes.

En primer lugar, se toma el logaritmo a los valores, con el fin de estabilizar la varianza y garantizar la homocedasticidad de los datos. Luego, se eliminan las tendencias de las series tomando una o varias diferencias de la serie. Para diferenciar una serie a cada observación se le resta la observación anterior esto es:

$$W_t = X_t - X_{t-1}$$

Ecuación 15 Diferenciación de una serie

Donde la serie W_t es la serie diferenciada.

5.4.1.2 Selección de parámetros p, d y q

Determinación del orden de diferenciación regular d es el indicador de la Integración del modelo ARIMA. Este indicador se calcula como el número de diferencias que hay que realizar en un conjunto de datos hasta convertirla en estacionaria. Una vez realizado este proceso, se grafica la serie W_t y se comprueba que los valores en el diagrama de autocorrelación simple se amortigüen y que se haya perdido la tendencia en los datos, en el caso de que se sigan visualizando tendencias, será necesario diferenciar la serie una segunda vez. Por lo general solo es necesario realizar este procedimiento 1 o 2 veces.

Una vez obtenida una serie estacionaria en media y varianza y haber tratado la estacionalidad, se procederá a identificar el orden los procesos autorregresivos y de medial móvil, es decir p y q. Para ello, la principal herramienta, son los diagramas de autocorrelación simple (ACF) y de autocorrelación parcial (PCF). Una vez obtenidos, estos gráficos se comparan con el comportamiento teórico de las autocorrelaciones de series con ordenes conocidos.

5.4.1.3 Proceso Autorregresivo, AR (1)

Por ejemplo, tomemos una serie cuyo valor actual está determinado por el último valor que tomó aquella serie, es decir:

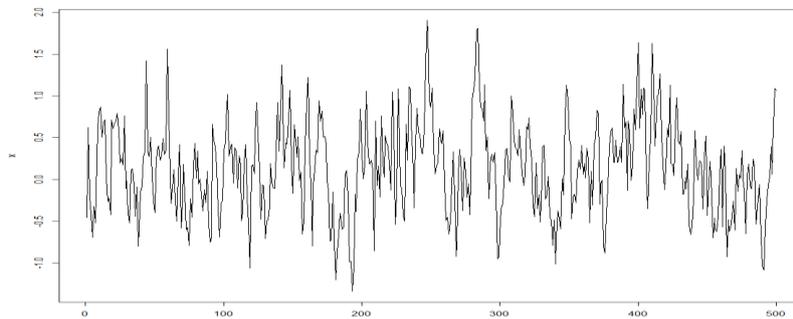
$$X_t = \phi X_{t-1} + a_t$$

Ecuación 16 Representación de un Proceso autorregresivo AR 1

Donde

- $a_t \sim N(0, \sigma_a^2)$
- $E[X_t] = 0, \forall t$
- $Var[X_t] = \frac{\sigma_a^2}{1-\phi^2}$

Figura 16 Serie X, modelo Autorregresivo p=1



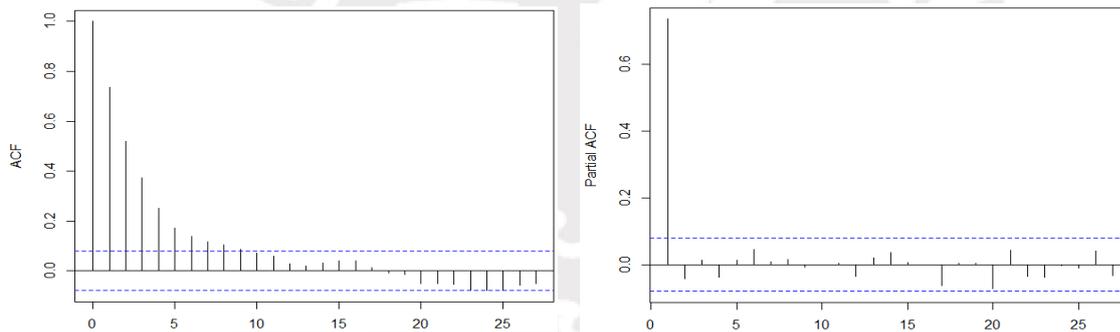
Para este proceso y reemplazando en la ecuación de autocorrelaciones simples, se obtiene lo siguiente:

$$\rho_k = \phi^k$$

Ecuación 17 Correlación simple para el retardo k

Cómo ϕ es menor que 1 entonces se espera que el gráfico de Autocorrelación simple decrezca de manera exponencial con los retardos. Por otro lado, ya que para cada valor solo influye directamente el valor previo, entonces luego del primer retardo la autocorrelación parcial no será significativa.

Figura 17 Autocorrelación simple y parcial de un proceso AR 1



5.4.1.4 Proceso Autorregresivo, AR (p)

Entonces de manera general, se puede decir que el modelo autorregresivo explica el valor actual como una regresión lineal de sus valores anteriores. De esta manera, un modelo autorregresivo de orden p, denominado AR(p), puede definirse según:

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t$$

Ecuación 18 Representación matemática de un modelo Autorregresivo AR

Donde:

- X_t es la serie temporal,
- a_t es el ruido blanco o las innovaciones y
- $\delta = (1 - \sum_{i=1}^p \phi_i) \mu$, con μ denotando la media del proceso

Asimismo, la autocorrelación simple un modelo AR (p) decrece con los retardos de forma exponencial mientras que la autocorrelación parcial no es significativa para retardos mayores que p.

En resumen, para calcular el orden de un proceso autorregresivo AR(p) el diagrama de autocorrelación simple (ACF) debe complementarse con un gráfico de autocorrelación parcial (PCF), en este se observa que el valor de la autocorrelación parcial de un proceso AR (p) se convierte en cero desde el retardo p + 1 en adelante.

5.4.1.5 Modelo de Media Móvil, MA (1)

Ahora consideremos un modelo cuyo valor actual esté determinado por los errores de sus valores previos, es decir:

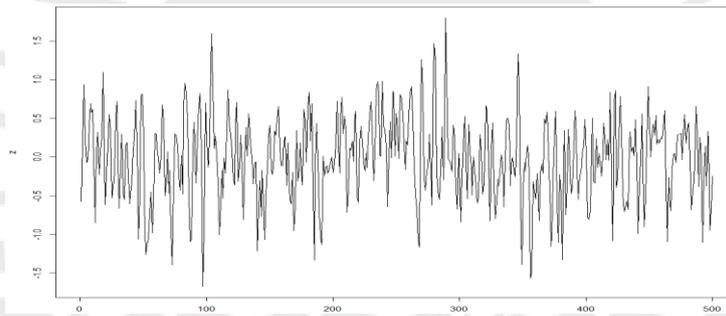
$$X_t = a_t - \theta_1 a_{t-1}$$

Ecuación 19 Representación de un modelo de Media móvil MA 1

Donde

- $a_t \sim N(0, \sigma_a^2)$
- $E[X_t] = 0, \forall t$
- $Var[X_t] = \sigma_a^2(1 + \theta_1^2)$

Figura 18 Serie z modelo de Media Móvil q=1

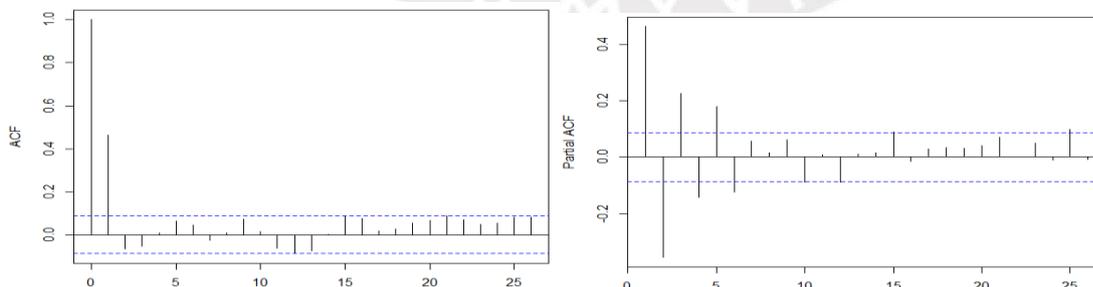


En este caso se obtiene que el valor de la autocorrelación parcial depende del retardo k como muestra la Ecuación 20, es decir que decrece exponencialmente mientras que la autocorrelación simple no es significativa para retardos mayores a 1 como se muestra en la Figura 19

$$\pi_k = \theta^k$$

Ecuación 20 Autocorrelación parcial para el retardo k

Figura 19 Autocorrelación simple y parcial de un modelo de MA 1



5.4.1.6 Modelo de Media Móvil, MA (q)

De manera general, en un modelo de media móvil el valor actual se explica conceptualmente como una regresión del ruido blanco o errores aleatorios de uno o más valores previos. En este modelo se asume que los valores aleatorios de cada registro previo siguen la misma distribución; típicamente, se trata de una distribución normal con media cero y varianza constante.

El modelo de Media móvil se distingue del proceso autorregresivo en que la componente aleatoria se propaga en valores futuros de la serie. Un modelo de media móvil de orden q, denominado MA(q), puede definirse según la Ecuación 21.

$$X_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Ecuación 21 Representación matemática de un modelo con media móvil MA.

Fuente: Time Series Handbook itl.nist.gov

Donde:

- X_t es la serie temporal,
- μ es la media de los valores de la serie
- a_{t-i} es el término de innovación o ruido blanco

Para calcular el orden del proceso de media móvil MA(q) se observa el gráfico de la función de autocorrelación simple (ACF) que el valor de autocorrelación se convierte en cero en el retraso q + 1 en adelante. La función de autocorrelación parcial muestral generalmente no es útil para identificar el orden del proceso de media móvil.

Los modelos MA tienen una interpretación menos obvia que los modelos AR, pero pueden representar un modelo AR utilizando menos parámetros. Esto es debido a que existe una dualidad entre los procesos MA y AR, que permite expresar un modelo AR (1) como un proceso MA(∞) y viceversa, es decir, se puede expresar un modelo MA(1) con un modelo AR(∞).

$$X_t = \phi_1 X_{t-1} + a_t \quad ; \quad X_{t-1} = \phi_1 X_{t-2} + a_{t-1}$$

$$X_t = \phi_1 (\phi_1 X_{t-2} + a_{t-1}) + a_t$$

$$X_t = \phi_1^2 X_{t-2} + \phi_1 X_{t-1} + a_t$$

$$X_t = \sum_{i=0}^{\infty} \phi_1^i * a_{t-i} \rightarrow MA(\infty)$$

Ecuación 22 Transformación de un proceso AR en un proceso MA

Para diferenciar un proceso autorregresivo AR de un proceso con media móvil MA, se suelen estudiar las autocorrelaciones simples (ACF) y las autocorrelaciones parciales (PACF).

5.4.1.7 Modelo Box – Jenkins o ARMA

G. E.P Box y G.M. Jenkins (1994), fueron dos estadísticos que desarrollaron este modelo en el que se combinan el proceso autorregresivo y el modelo de media móvil, lo innovador de este modelo es que desarrolla una forma compacta de mezclar los dos modelos previos para identificar y estimar modelos. La Ecuación 23 es una representación del modelo y combina los términos de un modelo autorregresivo y un modelo de media móvil.

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - \dots - \theta_q A_{t-q}$$

Ecuación 23 Representación matemática del modelo Box-Jenkins.

Fuente Box, Jenkins and Reinsel, 1994

Basado en la descomposición del teorema de Wold, cualquier proceso estacionario puede ser aproximado por un modelo ARMA. Para conseguir un modelo estacionario este modelo recomienda diferenciar aquellas series no estacionarias una o más veces hasta alcanzar la estacionariedad. Una serie se diferencia restando cada observación con la observación anterior. Al realizar este procedimiento y aproximarlos con un modelo ARMA, se produce un modelo Autorregresivo Integrado con Media Móvil (ARIMA, por sus siglas en inglés).

Los modelos ARIMA suelen ajustarse muy bien cuando se utilizan para modelizar fenómenos lineales, además al utilizar componentes autorregresivas y de media móvil, permite expresar un modelo con menos parámetros haciendo el modelo más parco. No obstante, un ajuste efectivo de Box – Jenkins requiere obtener una serie moderadamente larga. Chatfield (1996) recomienda al menos 50 observaciones, aunque otros sugieren al menos 100 observaciones.

En la Tabla 4 se muestra a modo de resumen, cómo identificar un modelo ARIMA según las gráficas de autocorrelación:

Tabla 4 Resumen sobre el comportamiento de las ACF y PACF según el modelo.

Fuente: Apuntes de Series temporales M.J. Sánchez

	ACF	PACF
AR (p)	Muchos coeficientes distintos a 0	Igual a 0 excepto para los primeros p valores
MA (q)	Igual a 0 excepto los primeros q valores	Muchos coeficientes distintos a 0
ARMA (p,q)	Muchos coeficientes distintos a 0	Muchos coeficientes distintos a 0

5.4.2 Estimación de parámetros y contrastes paramétricos

La dificultad de la selección de un modelo reside en obtener la verosimilitud del modelo a partir de la información contenida en los datos (Hirotoyu, 1974). Por ello, Luego de identificar el modelo se procede a realizar pruebas sobre la verosimilitud del modelo identificado, es decir, se genera el modelo y se obtienen los coeficientes del modelo. Luego, se comprueba que los parámetros son suficientes para representar la serie, es decir son no nulos e invertibles y, posteriormente, se evalúa el ajuste del modelo con estos nuevos parámetros.

Lo siguiente es identificar el mejor modelo en función de ciertas medidas de error. En este proyecto se evaluará el criterio de información de Akaike (AIC) y el criterio de información Bayesiano (BIC) para evaluar el ajuste de los modelos con respecto a los datos observados.

5.4.2.1 Criterio de Información de Akaike o AIC

El criterio de información de Akaike permite seleccionar modelos estimando la medida de la calidad relativa de un modelo estadístico para un conjunto de datos. Este criterio utiliza como bondad de ajuste al criterio de máxima verosimilitud y viene dado por la Ecuación 24.

$$AIC = 2k - 2 \ln(L)$$

Ecuación 24 Expresión del criterio de Akaike

Donde:

- k es el número de parámetros
- L máximo valor de la función de verosimilitud

Se entiende la verosimilitud como la probabilidad $p(y|\theta_k)$ de los datos condicionada a los k parámetros del modelo, que es una medida inversa a la suma de las distancias euclídeas de cada uno de los datos al modelo. De esta manera, mientras más pequeño sea el valor AIC, el modelo tendrá un mejor ajuste.

Por otro lado, el criterio de Akaike toma al número de parámetros k como la medida de la complejidad del modelo, siguiendo el criterio de parcidad: si el número k de parámetros de un modelo se incrementa, el modelo se hace más complejo. Esto tiene como consecuencia el incremento del término de penalización, no obstante, al mismo tiempo el desajuste disminuye. Por lo tanto, el valor final del AIC supone un equilibrio entre reducir la complejidad y mantener un valor mínimo entre el modelo teórico y el modelo estimado.

5.4.2.2 Criterio de Información de Bayes o BIC

El criterio de información Bayesiano fue desarrollado por Schwarz (1978) y es una medida de bondad de ajuste de un modelo estadístico, por tanto, es utilizado como un criterio para la selección de modelos y viene dado por la Ecuación 25.

$$BIC = k \ln(n) - 2 \ln(L)$$

Ecuación 25 Representación del criterio de información de Bayes

Donde:

- k es el número de parámetros,
- L es el valor de máxima verosimilitud y
- n es el número de datos

En este caso la complejidad del modelo se penaliza más que en el caso del AIC. Por otro lado, se mantiene que el modelo con menor valor del criterio de información será mejor aceptado ya que implica un menor número de variables explicativas o un mejor ajuste del modelo. Además, el criterio de BIC usualmente selecciona el modelo más sencillo y que hace predicciones a menor detalle

5.4.3 Diagnóstico del modelo

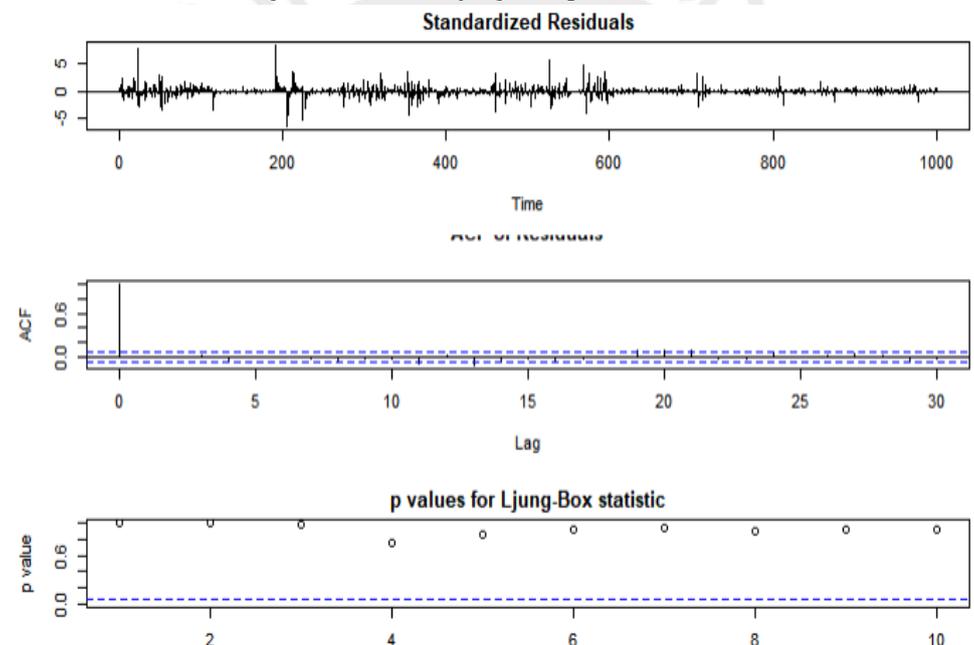
5.4.3.1 Análisis de coeficientes

Los coeficientes del modelo deben ser tales que representen un modelo estacionario y, además, sean invertibles. Por lo general esto indica que los coeficientes son menores a uno.

5.4.3.2 Análisis de residuos

Una vez que se obtienen los valores de los parámetros estimados, se obtienen los errores que genera el modelo. De acuerdo con el modelo de series univariantes estos residuos, deben ser un proceso de ruido blanco, es decir, deben seguir una distribución normal con media cero y varianza constante, además de ser independientes, esto se puede demostrar con una prueba de Ljung-Box (1978) como se observa en la Figura 20.

Figura 20 Prueba Ljung-Box para los residuos



5.4.4 Predicción

La función de predicción de un modelo ARIMA tiene una estructura simple en donde el término de integración (diferenciación), si es que existe, determina la predicción en un largo plazo, mientras que los operadores estacionarios, AR y MA, determinan la predicción a corto plazo. (Peña, 2010)

Por ejemplo, supongamos un proceso AR(1) $X_T = \mu + \phi_1 X_T + a_T$. La ecuación para representar la predicción del siguiente periodo sería: $\hat{X}_{T+1} = \mu + \phi_1 X_T$ y para dos periodos después:

$$\hat{X}_{T+2} = \mu + \phi_1 \hat{X}_T(1) = \mu(1 + \phi_1) + \phi_1^2 X_T$$

Ecuación 26 Predicción del valor t+2 para un modelo AR (1)

Para cualquier periodo $k > 0$ se tiene

$$\hat{X}_T(k) = \mu + \phi_1 \hat{X}_T(k-1) = \mu(1 + \phi_1 + \dots + \phi_1^k) + \phi_1^k \hat{x}_T$$

Ecuación 27 Predicción del valor t+k para un modelo AR (1)

Como $|\phi_1| < 1$, para un k muy grande

$$\hat{X}_T(k) = \mu / (1 - \phi_1) \approx \mu$$

De manera que para un proceso estacionario AR(p) se puede generalizar que las predicciones para un horizonte k grande tienden tomar el valor de la media del proceso.

5.5 Redes Neuronales Recurrentes

5.5.1 Transformación de los datos

El primer paso para construir una red neuronal es la normalización de los datos. De esta manera se evitan dificultades en el entrenamiento de los datos que pueden resultar en problemas de convergencia. Adicionalmente, los valores iniciales del modelo determinan la efectividad del ajuste de los pesos en la capa de entrada y podría tener un gran efecto en la calidad de las predicciones (Hastie, 2009).

$$x_{escalado} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Ecuación 28 Método min-max para escalar los datos

La opción de escalado con el método Min-Max mantiene los valores entre 0 y 1 y ajusta a todos los valores en este rango. Cabe resaltar que bajo este método el ruido podría amplificarse tras la transformación de datos.

Otra opción de escalado es el método de escalado estándar o normal, en donde se transforman los valores usando como parámetros la media y la desviación estándar, como se muestra en la Ecuación 29. Con este método la dispersión de la variable transformada se mantiene. No obstante, este tipo de transformación es sensible a valores atípicos en los datos.

$$x_{n-escalado} = \frac{x - \bar{x}}{\hat{\sigma}}$$

Ecuación 29 Escalado estándar

5.5.2 Selección del número de capas

En 1989 Hornik et al. demostraron que una sola capa oculta era suficiente para modelar cualquier función continua con un error tolerable. No obstante, en la práctica una sola capa oculta puede no ser suficiente ya que no todas las variables siguen una distribución continua. Adicionalmente, muchos modelos pueden mejorar su capacidad predictiva al añadir más capas ocultas. En ese sentido, la selección del número de capas es esencialmente un problema de selección de modelos ya que puede afectar significativamente los resultados de predicción. (Lewis, 2016)

Para la selección del número de capas, el método más usado suele ser el de prueba y error, teniendo en cuenta que mientras mayor sea el número de capas, el modelo será capaz de detectar y aprender más características de los datos. Siendo lo más común escoger un número razonable entre 5 y 100 unidades, dependiendo del tamaño de la muestra analizada.

De manera general se prefiere un modelo con muchas capas ocultas ya que un modelo con muy pocas capas no tendrá suficiente flexibilidad para capturar la no linealidad de los datos (Hastie, 2008). Las redes neuronales profundas explotan la propiedad de que muchas señales naturales que muestran características de alta complejidad son obtenidas por la composición de pequeñas características.

5.5.3 Número de Neuronas por capa

Establecer el número de neuronas por capa es un aspecto clave del modelo, para determinar una estrategia sugiere identificar a los factores que causan variabilidad en la característica (o variable) observada. Dichos factores de variabilidad pueden ser aleatorios o debidos al ruido, mientras que otros pueden ser sistemáticos o inherentes del proceso que genera la variable observada. Bajo este criterio, añadir más neuronas podría hacer que el modelo detecte patrones de la data de manera muy precisa. No obstante, mientras mayor sea el número de neuronas ocultas el riesgo de un sobreajuste es mucho mayor.

El sobreajuste (u *Overfitting*) implica que el modelo ha aprendido no solo los patrones de los datos sino también el ruido que hay en los mismos. En consecuencia, los modelos sobre ajustados serán muy precisos para expresar valores dentro de la muestra de entrenamiento, pero perderán precisión al predecir un valor fuera de esta muestra. En ese sentido, se deben generar modelos con el menor número de neuronas posible, pero con un error que sea tolerable en la muestra de entrenamiento.

Adicionalmente, es importante tener en cuenta que el poder de las redes neuronales yace en que cada neurona funciona como un detector de características independiente; no obstante, en la práctica puede ocurrir que dos o más neuronas empiecen a detectar la misma característica. Este fenómeno es denominado coadaptación e implica que se están generando gastos computacionales innecesarios al calcular funciones de activación sobre neuronas que identifican la misma característica. La coadaptación es similar a la idea de colinealidad en las regresiones lineales e implica que el resultado de un nodo puede ser linealmente predicho por otro nodo con un error muy pequeño.

5.5.4 Número de epochs

El número de epochs, es un hiperparámetro que se define antes de entrenar un modelo. Un epoch implica que el conjunto de datos ha sido analizado por la red neuronal usando algoritmos de propagación hacia adelante y hacia atrás. Este parámetro es muy importante ya que podría permitir lograr un mejor desempeño en los modelos. Para estimar este hiperparámetro se realiza una estimación del porcentaje de mejora de un modelo al incrementar el número de epochs. Cuando el costo computacional del incremento de epochs no es superado por la mejora del modelo entonces, se puede decir que se ha alcanzado el número óptimo de epochs.

5.6 Medidas de Error

En este estudio se utilizaron las siguientes medidas de error, para analizar las comparaciones entre las muestras de prueba y los valores estimados

5.6.1 Error Absoluto Medio o MAE

Este estadístico se calcula como el promedio del error obtenido (diferencia entre el valor observado y el valor estimado) en valor absoluto y se calcula según la Ecuación 30.

$$EM = \frac{1}{n} \sum_{i=1}^n |\hat{x}_i - x_i|$$

Ecuación 30 Error absoluto medio

El error absoluto medio determina qué tan grande es el error que se puede esperar en promedio.

5.6.2 Raíz del error cuadrático medio o RMSE

Este estadístico penaliza de manera más drástica los errores grandes y se calcula según la siguiente ecuación.

$$RECM = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2}$$

Ecuación 31 Error cuadrático medio

Al comparar este estadístico con el anterior se puede identificar si la muestra evaluada contiene errores grandes, mientras más grande sea la diferencia entre el MAE y el RMSE entonces el tamaño del error será mayor.

5.6.3 Promedio del error absoluto en porcentaje o MAPE

Este estadístico se calcula como el promedio del error absoluto en porcentaje y se obtiene según la siguiente ecuación.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{x}_i - x_i}{x_i} \right|$$

Ecuación 32 Media del error absoluto en porcentaje

El MAPE se emplea para poder distinguir qué puede ser considerado como un error grande o como un error pequeño. El MAPE es muy sensible a la escala de los datos y no debería emplearse si es que entre los valores de respuesta se encuentran valores iguales a cero, ya que esto generaría que un valor de MAPE infinito.

Capítulo 4

6. Resultados y discusión

6.1 Modelos ARIMA

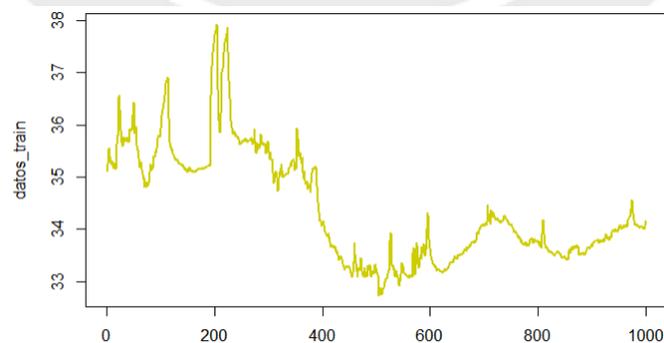
Para realizar el análisis con modelos ARIMA se usaron las librerías `tseries` y `forecast` del software libre RStudio. Para iniciar, se transformaron las series hasta obtener una serie estacionaria. Luego se obtuvieron las funciones de autocorrelación simple y autocorrelación parcial para estimar los parámetros p y q de cada serie de datos, en este caso se propone un modelo para cada serie de temperatura por paciente. Posteriormente se realizó un análisis del modelo generado y finalmente, se procedió a realizar una comparación entre las predicciones realizadas por el modelo para horizontes de 15, 30 y 60 minutos. Para realizar esto último se utilizaron los siguientes 15, 30 y 60 valores como muestra de prueba y se compararon con los valores que el modelo predecía, obteniendo el MAPE.

6.1.1 Paciente 1

6.1.1.1 Características generales:

Se trata de una paciente femenina de 63 años de edad, que ingresó a UCI con una infección viral. Para el análisis de sus termometrías se consideró una muestra de entrenamiento con los 1000 primeros valores de la serie que se presenta en la Figura 21 y el tramo restante como muestra de prueba. La temperatura mínima observada es de 32.7°C y la máxima de 37.93°C .

Figura 21 Termometría del paciente 1: T. Central



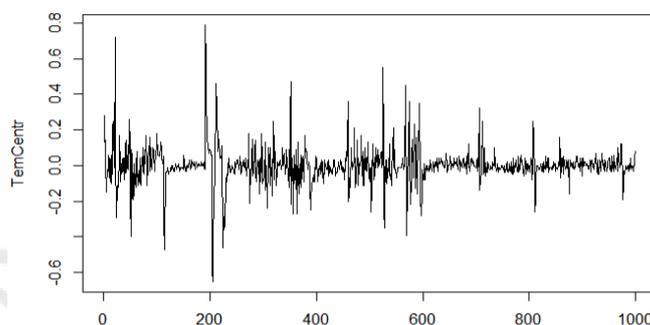
Del gráfico de la Figura 21 se aprecia que luego de los 400 valores la temperatura comienza a descender por debajo de 35°C . Además, los datos son estacionarios en varianza, no obstante, oscilan entre distintos niveles, es decir, no son estacionarios en media; para conseguir que la serie sea estacionaria se realizarán transformaciones de los datos

6.1.1.2 Identificación del modelo

6.1.1.2.1 Transformaciones

Luego de diferenciar los datos una vez ($d=1$), se han eliminado las tendencias crecientes y decrecientes de las series, además, se comprobó en la FAS que los valores se amortiguaban con los retardos. De manera que se ha obtenido una serie estacionaria para los valores de temperatura central como se muestra en la Figura 22.

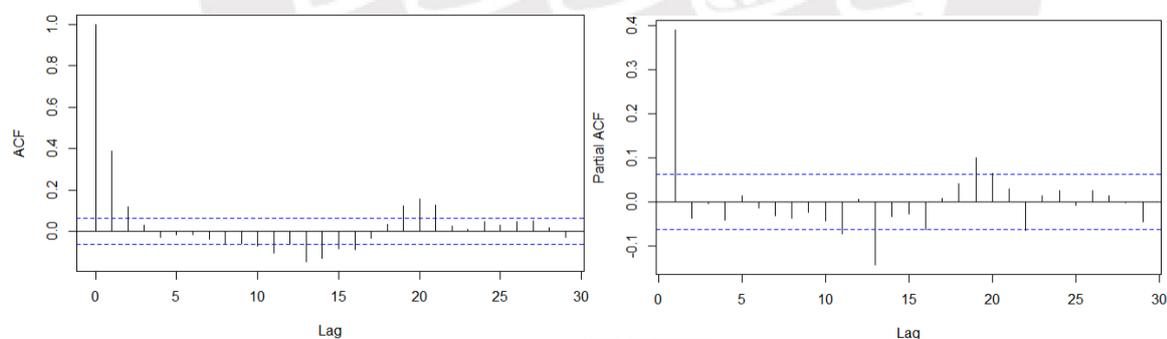
Figura 22 Serie de temperatura transformada del paciente 1



6.1.1.2.2 Selección de p , d y q

Del apartado anterior se obtuvo que d era igual a 1, luego, para seleccionar los parámetros p y q se grafican los diagramas de autocorrelación simple y autocorrelación parcial.

Figura 23 Autocorrelación simple (izquierda) y parcial (derecha) de la T. central paciente 1



De la Figura 23, en el diagrama de autocorrelación simple se observa que los valores se amortiguan rápidamente con los retardos. Además, se observa que la autocorrelación parcial solo toma un valor significativo para los 10 primeros retardos, con lo que se identifica un modelo autorregresivo con un parámetro p igual a 1. Además, previamente se diferenció la serie una vez, por ello se identifica el modelo como un $ARI(1,1)$.

6.1.1.3 Estimación de parámetros y contrastes:

6.1.1.3.1 Temperatura central

Una vez identificado el modelo se procede a estimar los coeficientes y residuos que genera el modelo ARI(1,1) al ajustarse a la termometría del paciente 1:

$$Z_t = a_t + 1.3908 Z_{t-1} - 0.3908 Z_{t-2}$$

Ecuación 33 Estimación de los datos del paciente 1 con el modelo ARI(1,1)

Con:

- $a_t \sim N(\hat{\mu}_a = 0, \hat{\sigma}_a^2 = 0,00901)$
- $\text{Log} - \text{verosimilitud} = 932.17$
- $\text{AIC} = -1860.34$

Se realizó un contraste z sobre el coeficiente obtenido para analizar la nulidad de este y se obtuvo un valor de ese contraste $z = 13.419$, entonces, a un nivel de confianza del 95% se admite que el coeficiente es no nulo.

Además, se analizaron los residuos obtenidos con la prueba de Ljung-Box. Cabe destacar, que la prueba de Ljung-Box es una prueba de contraste que en este caso se usa para probar la independencia de los residuos y, por lo tanto, probar que estos se comportan como ruido blanco. Para dicha prueba se establece en la hipótesis nula que los residuos del modelo deben tener autocorrelaciones simples iguales a cero para cualquier retardo ($t \geq 1$).

Como se muestra en la Figura 24 la correlación simple de los residuos es poco significativa para los retardos mayores a cero, además, el promedio de valor p es igual a 0.7947. Por lo anterior, se admite que los residuos obtenidos son independientes y, por lo tanto, que el modelo ARI(1,1) se ajusta adecuadamente a los datos. A continuación, se realizará un análisis de la normalidad de los residuos, para evaluar la normalidad de los datos.

Figura 24 Residuos obtenidos para el modelo ARI(1,1) – paciente 1

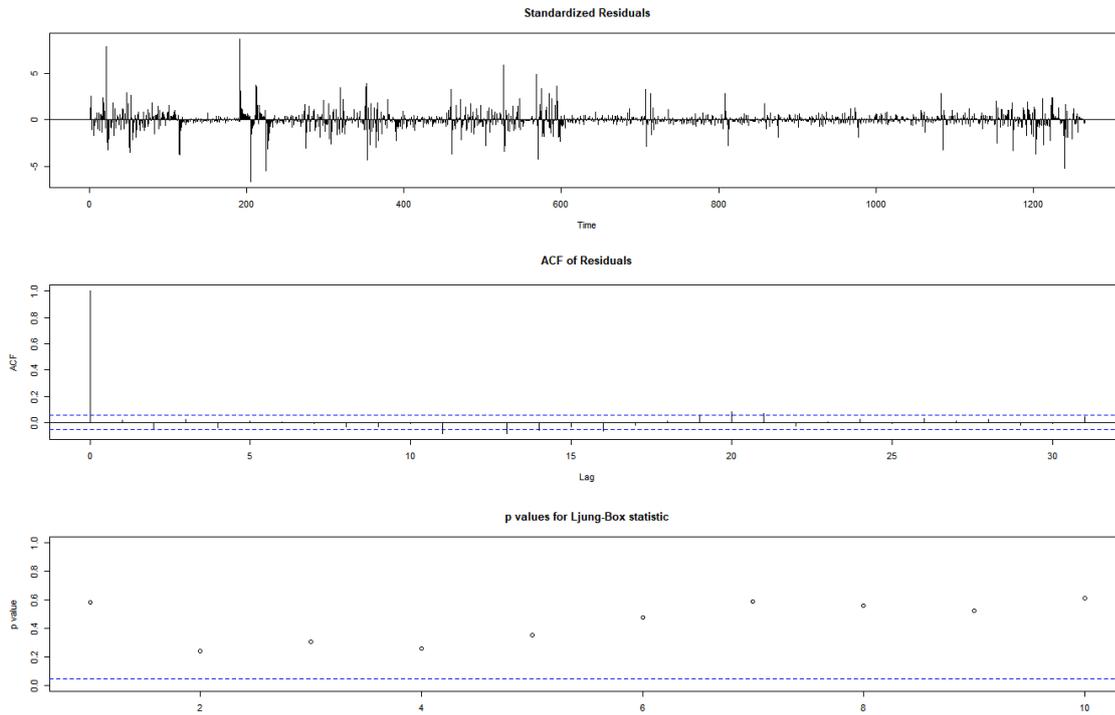
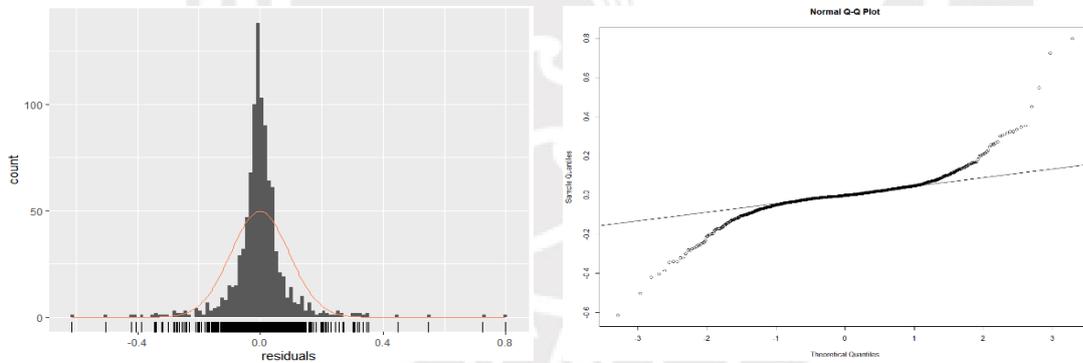


Figura 25 Normalidad de los residuos ARI(1,1) – paciente 1

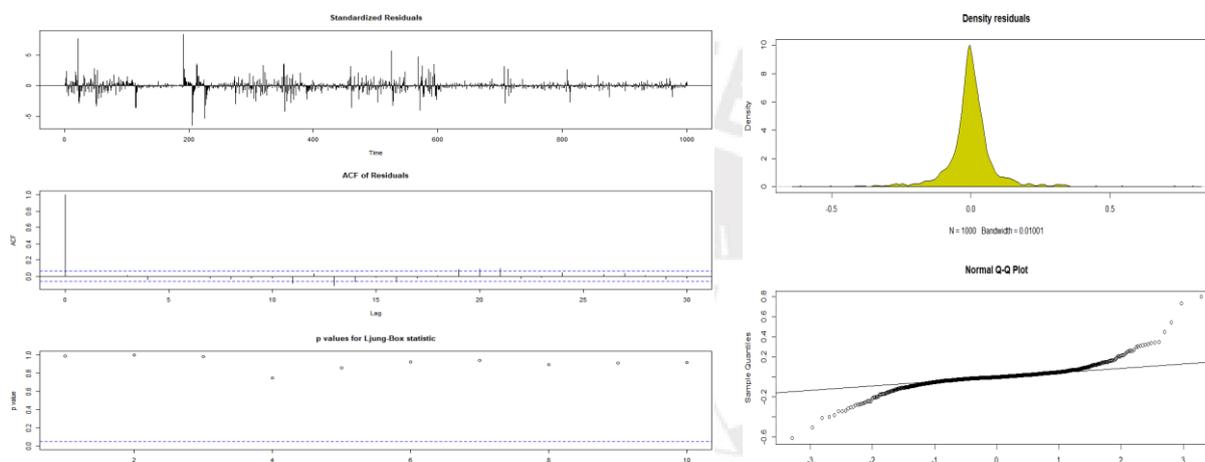


En el histograma de la Figura 25 se observa que los residuos tienen varios valores atípicos. Además, los residuos forman una campana de Gauss centrada ligeramente a la izquierda de cero, aunque con muchos valores alrededor de cero; para comprobar la normalidad de los residuos vemos el gráfico Q-Q. En este gráfico se observa que los residuos no se comportan siguiendo una distribución normal, hay muchas desviaciones con respecto a la línea marcada.

6.1.1.4 Diagnósis del modelo

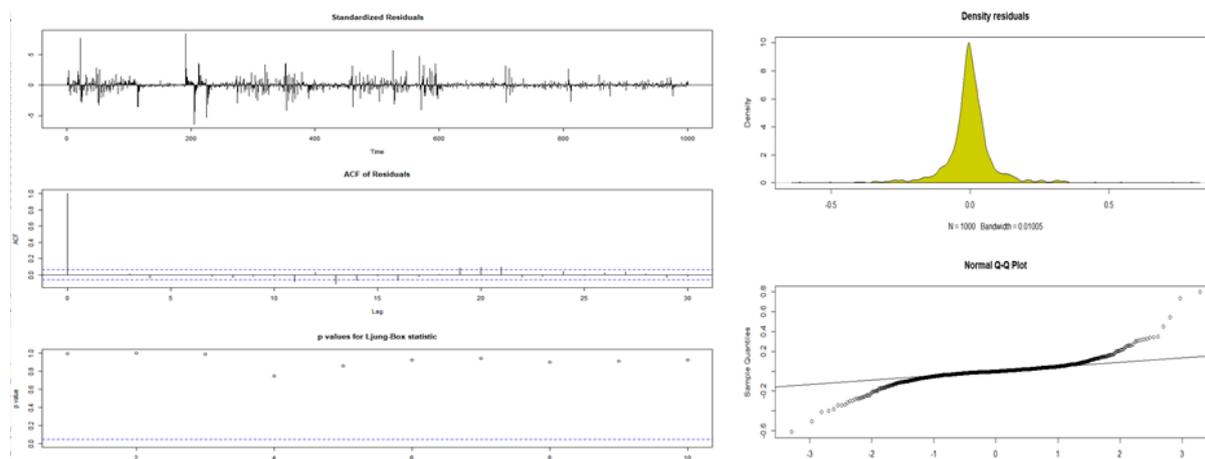
El análisis de los coeficientes indica que estos son no nulos e invertibles. Además, de la Figura 25 se concluye que los residuos obtenidos del modelo no se ajustan a una distribución normal. Sin embargo, los valores p del modelo eran superiores a 0.05, entonces, se puede afirmar con un 95% del nivel de confianza que los residuos son independientes y que el modelo se ajusta a los datos. De modo, que se selecciona el modelo ARI (1,1). Como parte del diagnóstico y para descartar que el modelo se encuentre infra dimensionado se evaluó también el modelo ARI (2,1) y el modelo ARIMA(1,1,1).

Figura 26 Residuos del modelo ARI (2,1) – paciente 1



Los coeficientes del modelo ARI(2,1) cumplen la condición de que son invertibles [coeficientes = $ar(0.4049; 0.0365)$] no obstante, luego del contraste de nulidad se determinó que el coeficiente ϕ_2 tenía un $z = -1.1517$, entonces, no se puede demostrar que ϕ_2 sea no nulo. Por otro lado, de la Figura 26 se observa que la prueba de Ljung-Box tiene un comportamiento similar al modelo ARI(1,1) con un ρ significativo en $t=0$ para la FAS, y valores p significativos para todos los retardos, por tanto, se concluye que el modelo se ajusta a los datos. No se observan mejoras con respecto a la normalidad de los residuos, sigue habiendo muchos valores atípicos.

Figura 27 Residuos del modelo ARIMA (1,1,1) – paciente 1



Para el modelo ARIMA(1,1,1) se obtuvieron coeficientes invertibles [coeficientes= ar(0.3120, ma(0.0932)], pero al aplicar el contraste z para identificar algún coeficiente nulo se obtuvo que el valor de θ_1 no era significativo ($z= 1.1468$). Además, en la Figura 27 no se observa ningún valor significativo en la FAS excepto para $t=0$, además para la prueba de Ljung-Box se obtuvieron valores p significativos para todos los retardos, con ello, se comprueba la independencia de los residuos y el ajuste del modelo a los datos. Por otro lado, en la parte derecha de la figura se observa que los residuos no siguen una distribución normal.

Tabla 5 Comparación de modelos para el paciente 1

	Valor p – Ljung-Box	AIC	BIC
ARI(1,1)	0.446	-1860.342	-1850.529
ARI(2,1)	0.809	-1859.668	-1844.948
ARIMA(1,1,1)	0.813	-1859.657	-1844.937

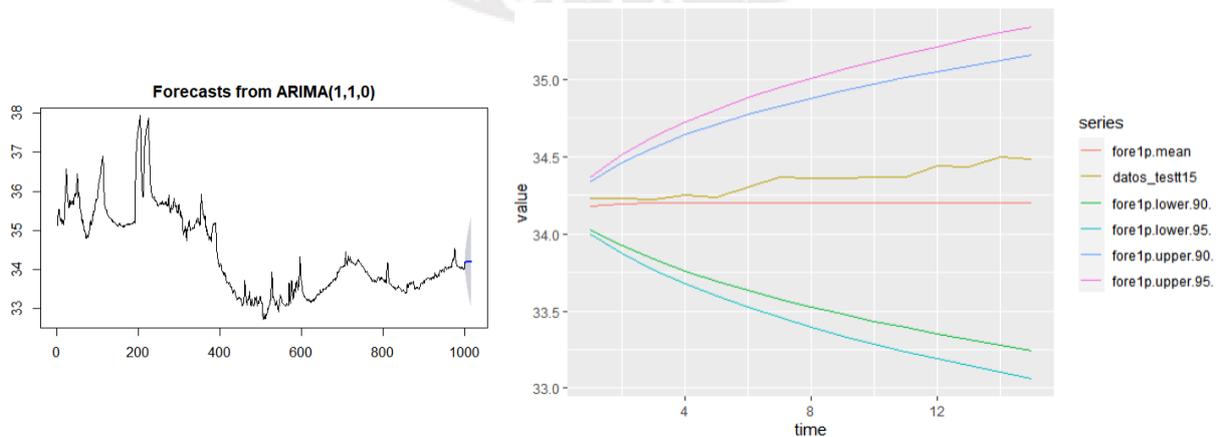
Al comparar los modelos se puede apreciar que los resultados del Criterio de información de Akaike no indican diferencias significativas entre los 3 modelos, aunque el modelo ARI (1,1) tiene un mejor valor. Esta observación se ve confirmada por el Criterio de Información Bayesiano que indica que el modelo con mejor ajuste a los datos es el modelo ARI(1,1), adicionalmente, se observa que las diferencias entre el modelo ARI(2,1) y el modelo ARIMA(1,1,1) no son significativas.

6.1.1.5 Predicciones a 15, 30 y 60 minutos

En este modelo los valores posteriores de la temperatura solo dependen de la última observación realizada, por lo tanto, se espera que los valores predichos converjan con este valor. Se probó el modelo para predecir los siguientes 15 valores de la serie de termometría, se observa que el modelo predice valores (color rojo) muy cercanos con diferencias

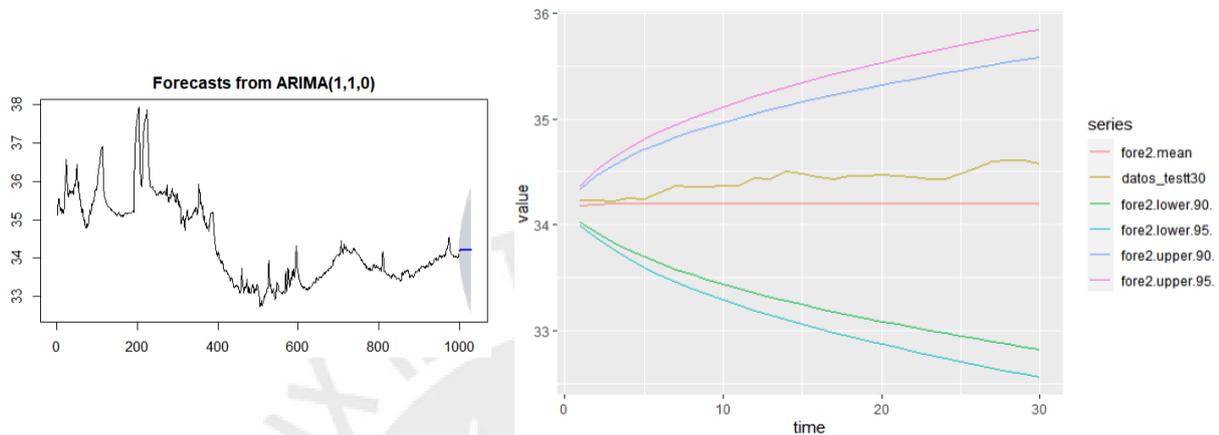
imperceptibles que convergen a 34.20°C . Aunque al inicio hay similitud entre las predicciones y los datos reales (color mostaza) luego las diferencias se hacen más notorias, aun así, los valores reales se encuentran dentro de los intervalos de confianza para niveles de confianza de 90 y 95. Se observan amplitudes del nivel de confianza de 1.5°C y 1.9°C para los niveles de confianza de 90 y 95 respectivamente.

Figura 28 Predicciones a un plazo de 15 minutos



Para el horizonte de predicción de 30 minutos, los valores predichos convergen a 34.20°C. Los datos reales (color mostaza) tienen mucha más variabilidad, no obstante, en este caso los primeros 5 valores predichos coinciden con los datos observados, luego de estos la temperatura real se eleva, pero se mantiene dentro de los intervalos de confianza. Finalmente, se puede observar una amplitud de 3.5°C y 3.74°C para los intervalos de confianza de 90 y 95 respectivamente.

Figura 29 Predicciones a un plazo de 30 minutos



Adicionalmente se utilizó el modelo para predecir los siguientes 60 valores, de igual manera que en los horizontes anteriores, el modelo converge a 34.2°C. Se observa una amplitud de 4°C y 4.7°C para los niveles de confianza de 90 y 95 respectivamente.

Figura 30 Predicciones a un plazo de 60 minutos

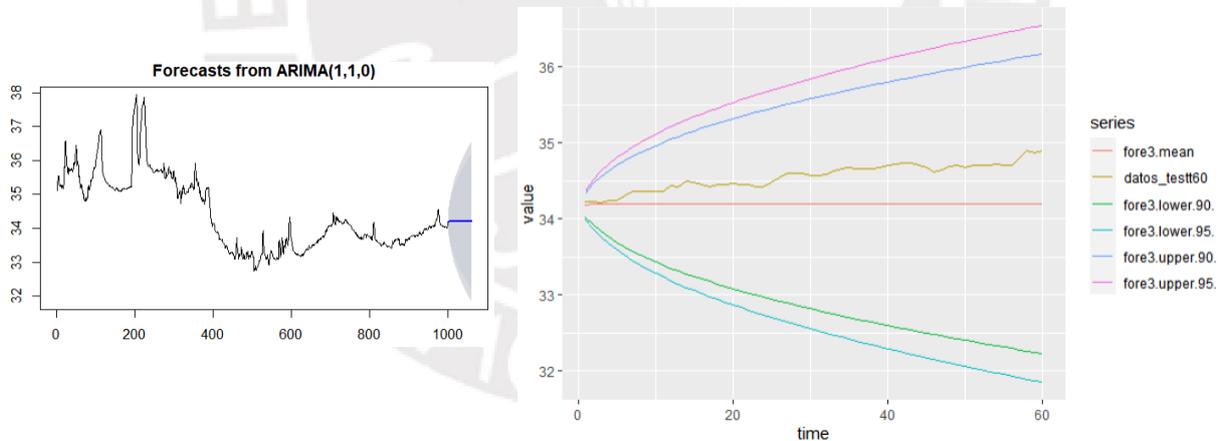


Tabla 6 Comparativa entre las predicciones a 15, 30 y 60 minutos ARI(1,1) - paciente 1

	MAE	RMSE	MAPE	Amplitud NC 90	Amplitud NC 95
Predicciones 15 min	0.144%	0.169%	0.419%	1.5°C	1.9°C
Predicciones 30 min	0.220%	0.246%	0.638%	3.5°C	3.8°C
Predicciones 60 min	0.362%	0.400%	1.045%	4.0°C	4.7°C

Es importante resaltar que como se están prediciendo valores de temperatura corporal central el error del modelo no puede ser mayor a 1° , además, de la termometría de este paciente se obtuvo un valor medio de 34.58°C , entonces, para estos datos el máximo de error permitido en las predicciones es de 2.85%.

Asimismo, se ha definido una amplitud del intervalo de confianza máxima de las predicciones de 2°C debido a que amplitudes mayores no cumplirían con la variación máxima aceptable de $\pm 1^\circ\text{C}$. El modelo debe detectar variaciones mayores de temperatura central ya que de superar $\pm 1^\circ\text{C}$ las condiciones del paciente podrían ser críticas.

De la Tabla 6 se observa que el modelo ARI(1,1) muestra resultados bastante acertados para los 3 horizontes de tiempo, en todos los casos se obtiene un MAPE menor al 2.85%, el error porcentual máximo aceptado. En el caso de las predicciones a 15 minutos se observa que MAPE es de 0.419%, que es un valor muy satisfactorio, además, el modelo plantea intervalos de confianza bastante acotados con una amplitud menor a 2°C , entonces se concluye que el modelo es preciso para este horizonte de tiempo.

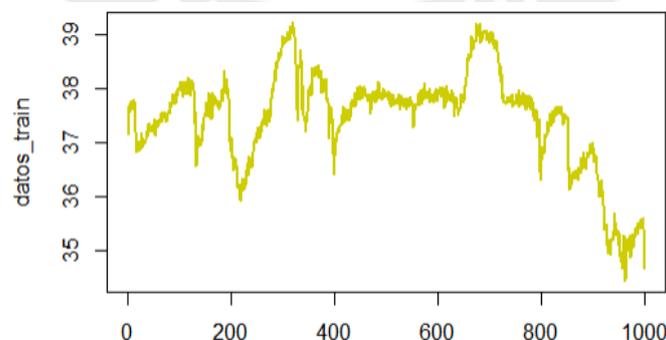
Para las predicciones a 30 minutos, la amplitud del intervalo de confianza se incrementa a 3.5°C y 3.8°C , mientras que para las predicciones a un horizonte de 60 minutos tienen una amplitud de 4°C y 4.7°C , y a pesar de tener buenas predicciones, la precisión del intervalo de confianza escapa de lo aceptable.

6.1.2 Paciente 2

6.1.2.1 Características generales

Paciente masculino de 48 años de edad, ingresado a la UCI con endocarditis, se tomó una porción de toda la serie de temperaturas como muestra de entrenamiento ($n=1000$) y el tramo restante para el contraste de predicciones. Como se muestra en la Figura 31, la temperatura del paciente 2 oscila entre valores de 34.8°C y 39.2°C de temperatura máxima. En este caso se observa que el paciente inicia con temperaturas altas y tiene hasta 2 picos de temperatura de 39.2°C , aunque finalmente, la temperatura comienza a descender, esto puede indicar que el paciente habría bajado la fiebre, o que el dispositivo de medición se hubiera movido y no estuviera enviando señales correctas.

Figura 31 Termometría del paciente 2: T. Central



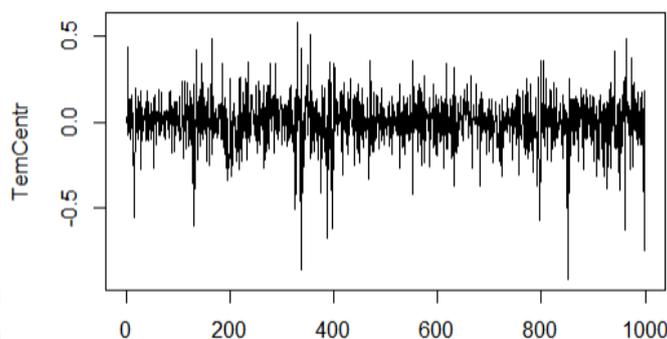
De la Figura 31 se puede observar que la dispersión de los datos en el tiempo se mantiene constante. Sin embargo, se observan muchas tendencias por lo que en este caso se realizarán diferenciaciones de la serie hasta obtener una serie estacionaria.

6.1.2.2 Identificación del modelo

6.1.2.2.1 Transformaciones

Luego de diferenciar los datos una vez ($d=1$), se han eliminado las tendencias crecientes y decrecientes de la serie como se observa en la Figura 32 De manera que se ha obtenido una serie estacional para los valores de temperatura central.

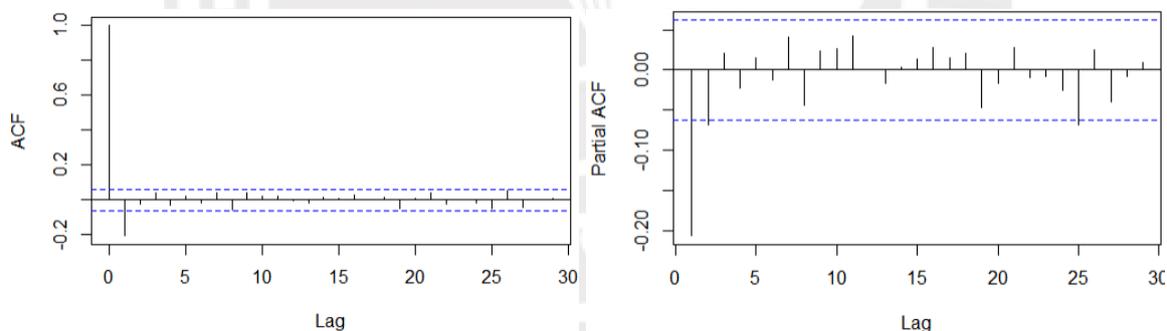
Figura 32 Serie de temperatura transformada



6.1.2.2.2 Selección de p, d y q

Del apartado anterior se obtuvo que d era igual a 1. Luego es necesario graficar los diagramas de autocorrelación simple y de autocorrelación parcial para seleccionar los parámetros p y q .

Figura 33 Autocorrelación simple (izquierda) y parcial (derecha) de la T. central paciente 2



De la Figura 33 en el diagrama de autocorrelación simple se observa que luego del primer valor las autocorrelaciones son no significativas, además, se observa que el diagrama de autocorrelaciones parciales se amortiguan con los retardos y solo muestra un valor significativo para el primer retardo, entonces se puede inferir que se trata de un proceso de media móvil con $q = 1$. Luego, se concluye que se trata de un proceso de media móvil y diferenciado una vez: IMA (1,1)

6.1.2.3 Estimación de parámetros y contrastes

Luego de identificar el modelo se procede a estimar los coeficientes y residuos que genera el modelo IMA(1,1) al ajustarse a la termometría del paciente 2.

$$Z_t = Z_{t-1} - 0.2213a_{t-1} + a_t$$

Ecuación 34 Estimación de los datos del paciente 2 con el modelo IMA(1,1)

Con:

- $a_t \sim N(\hat{\mu}_a = 0; \hat{\sigma}_a^2 = 0,0244)$
- $\text{Log - verosimilitud} = 436.9$
- $\text{AIC} = -869.79$

Se realizó un contraste z sobre el coeficiente obtenido para analizar la nulidad de este y se obtuvo un valor de contraste $z = -7.1464$, con lo que a un nivel de confianza del 95% se admite que el coeficiente es no nulo. Además, se analizaron los residuos obtenidos con la prueba Ljung-Box como se observa en la Figura 34, donde se aprecia que la correlación simple de los residuos es poco significativa para los retardos mayores a cero, además, el promedio de valor p es igual a 0.5967, que indica que los residuos son independientes

Por lo anterior, se admite que los residuos obtenidos se comportan como ruido blanco y, por lo tanto, que el modelo IMA(1,1) se ajusta adecuadamente a los datos. A continuación, se realizará un análisis de la normalidad de los residuos, para evaluar la normalidad de los datos.

Figura 34 Prueba Ljung-Box para los residuos del modelo IMA(1,1) – paciente 2

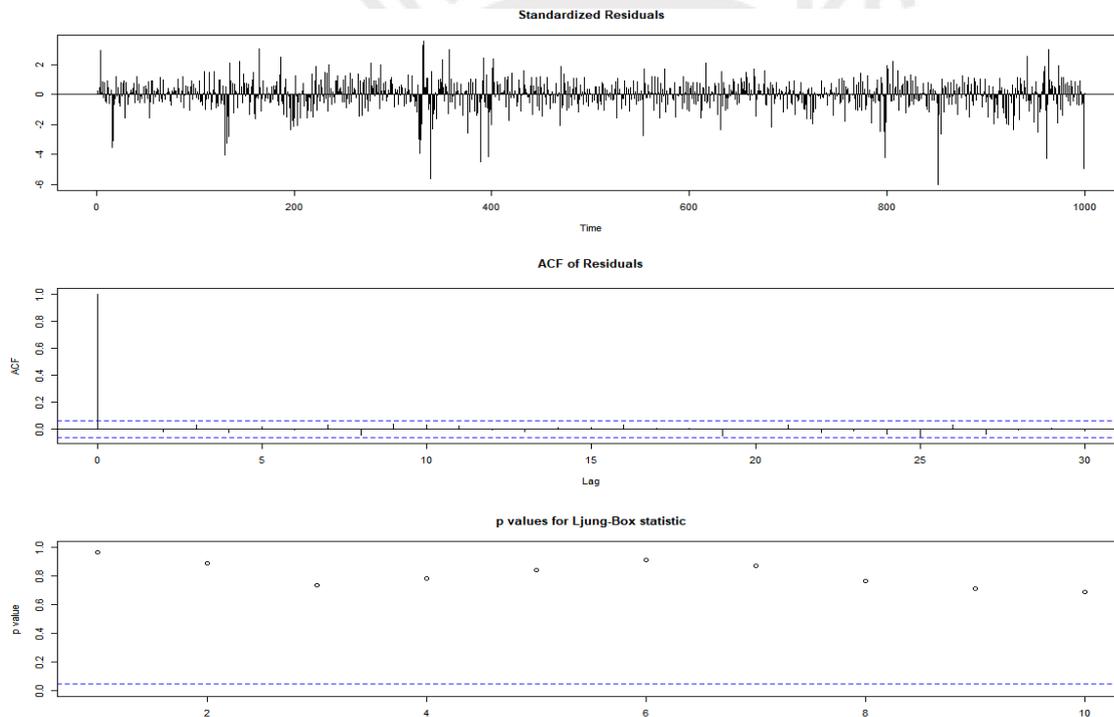
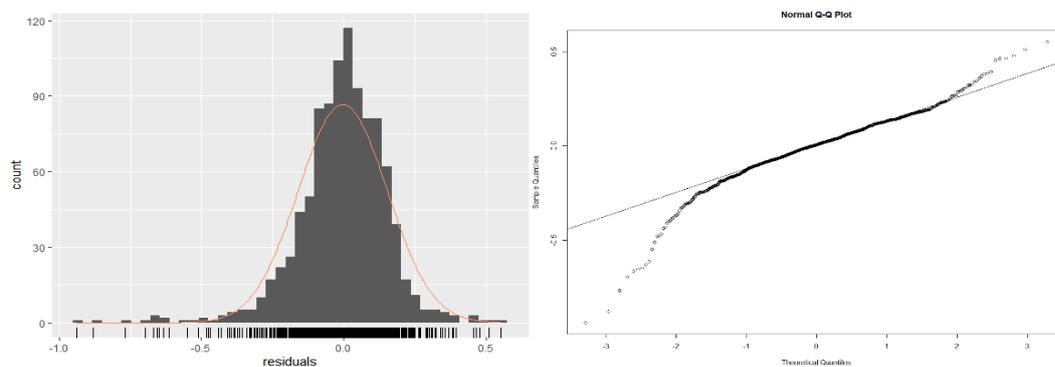


Figura 35 Normalidad de los residuos para el modelo IMA(1,1) – paciente 2

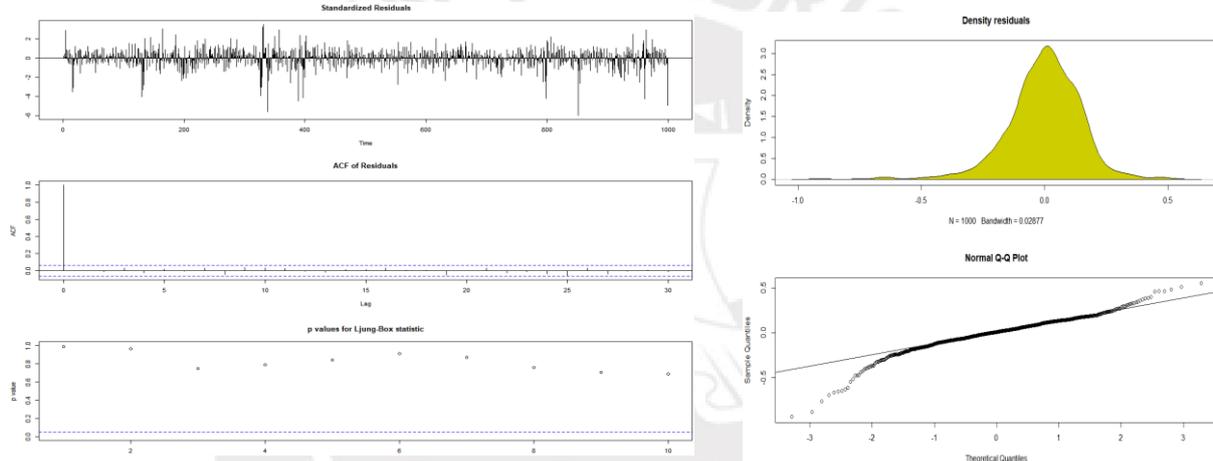


En el histograma de la Figura 35 se observa que los residuos forman una campana de Gauss centrada ligeramente a la derecha de cero, aunque con muchos valores acumulados alrededor de cero y una cola izquierda más larga; para comprobar la normalidad de los residuos vemos el gráfico Q-Q (derecha). En este, se observa que los residuos no se comportan siguiendo una distribución normal, hay muchas desviaciones con respecto a la línea marcada.

6.1.2.4 Diagnósis del modelo

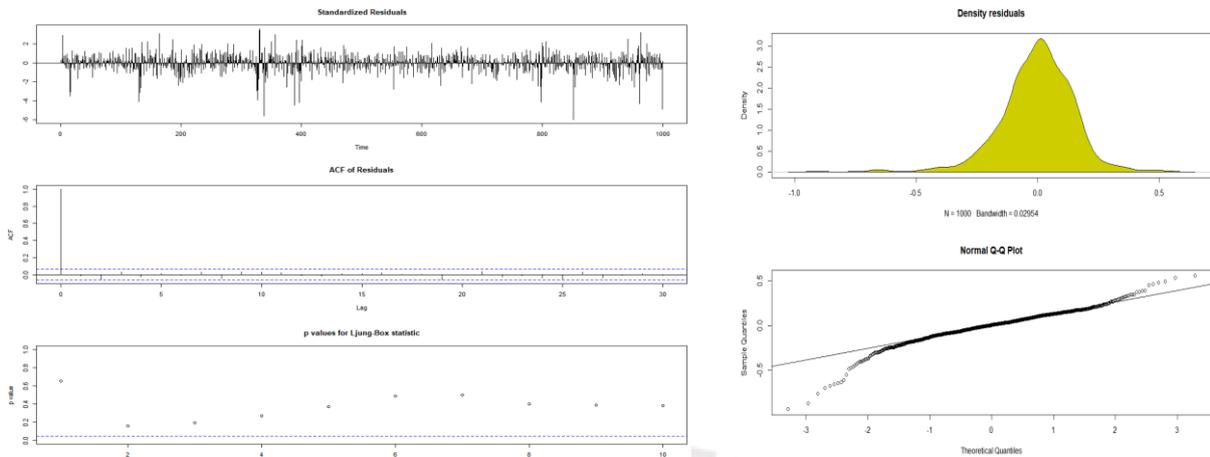
El análisis de coeficientes indica que estos son no nulos e invertibles, además, de la Figura 35 se puede concluir que los residuos no se comportan de manera normal. No obstante, de la prueba de Ljung-Box se obtuvieron valores p superiores a 0.05, entonces se comprueba que los residuos son independientes y el modelo se ajusta a los datos. Para descartar que el modelo se encuentre infra dimensionado se evaluaron los modelos ARIMA (1,1,1) y el modelo ARI(1,1).

Figura 36 Residuos del modelo ARIMA(1,1,1) - paciente 2



Para el modelo ARIMA (1,1,1) se obtuvieron coeficientes invertibles [coeficientes= ar(-0.0343) ma(-0.2536)], aunque luego la prueba de nulidad [$Z = (0.2605, -1.996)$] se obtuvo que solo el coeficiente de la MA (θ) era no nulo con un 95% del nivel de confianza. Adicionalmente, se analizaron los residuos, Figura 36, con la prueba Ljung-Box, se observa que la FAS solo toma un p significativo en $t=0$, además se obtuvieron valores p significativos para todos los retardos. Por tanto, se considera que los residuos son independientes y que el modelo se ajusta a los datos.

Figura 37 Residuos del modelo ARI(1,1) - paciente 2



Para el modelo ARI(1,1) se obtuvo un coeficiente invertible [coeficientes= ar(-0.2062)] y significativo ($z = -6.6594$). Además, en la

Figura 37 se observa que los residuos cuentan con varios valores atípicos a lo largo de la serie que afectan la variabilidad. Luego no se observa ningún valor significativo en la FAS, además para la prueba de Ljung-Box se obtuvieron valores p significativos para todos los retardos. Con ello, se comprueba la independencia de los residuos y el ajuste del modelo a los datos. Por otro lado, en la parte derecha de la figura se observa que los residuos no siguen una distribución normal.

Tabla 7 Comparación de modelos para el paciente 2

	Promedio Valor p – Ljung-Box	AIC	BIC
ARI(1,1)	0.297	-866.226	-856.413
ARIMA(1,1)	0.494	-867.963	-853.142
IMA(1,1)	0.597	-869.795	-859.981

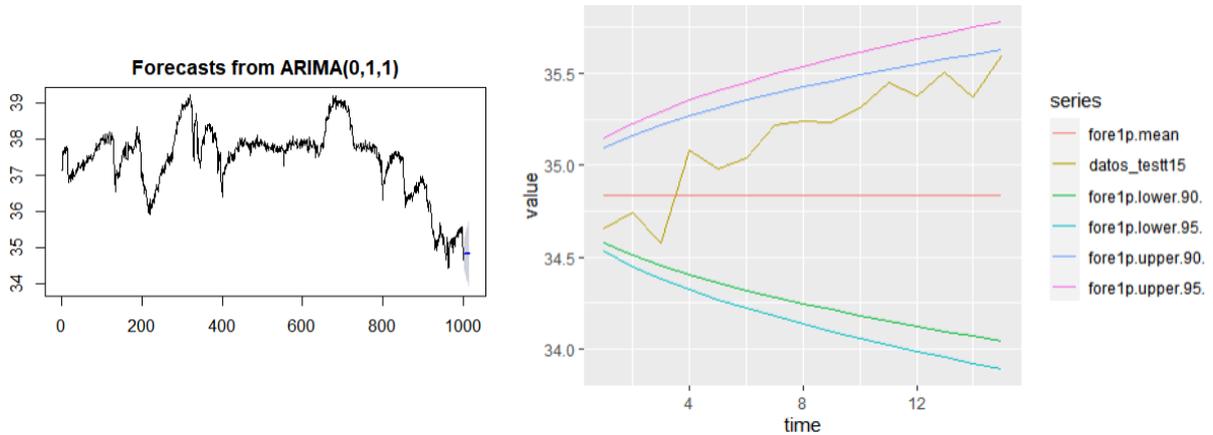
Al comparar los modelos se puede apreciar que los resultados del Criterio de información de Akaike y el Criterio de Información Bayesiano indican que el modelo que mejor se ajusta a los datos es el modelo IMA(1,1).

6.1.2.5 Predicciones a 15, 30 y 60 minutos

Se probó el modelo escogido para predecir los siguientes 15 valores de la serie de termometría y se observa que el modelo predice valores (color rojo) que convergen a 34.83°C. Aunque en los 3 primeros valores hay una similitud entre las predicciones y los datos reales (color mostaza) luego la temperatura real del paciente se va elevando hasta llegar a tomar el valor de 35.59°C en $t=15$, cabe destacar que los valores reales se encuentran dentro de los intervalos de confianza. La variación entre el valor máximo y mínimo de los intervalos de confianza es de 1.6°C y 1.9°C para los niveles de 90 y 95 respectivamente.

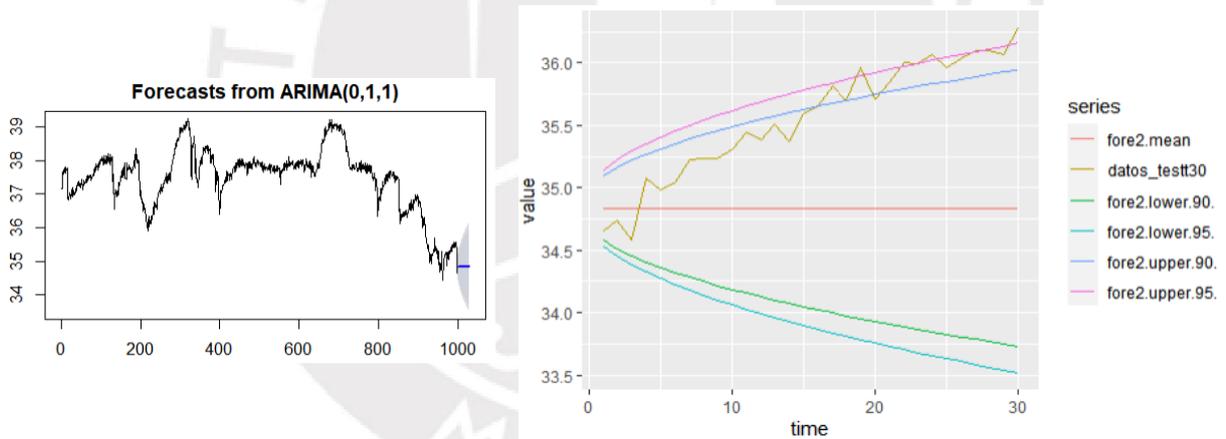
También se resalta que este incremento sostenido de temperatura puede ser debido a una posible reconexión de la sonda óptica.

Figura 38 Predicciones a un plazo de 15 minutos



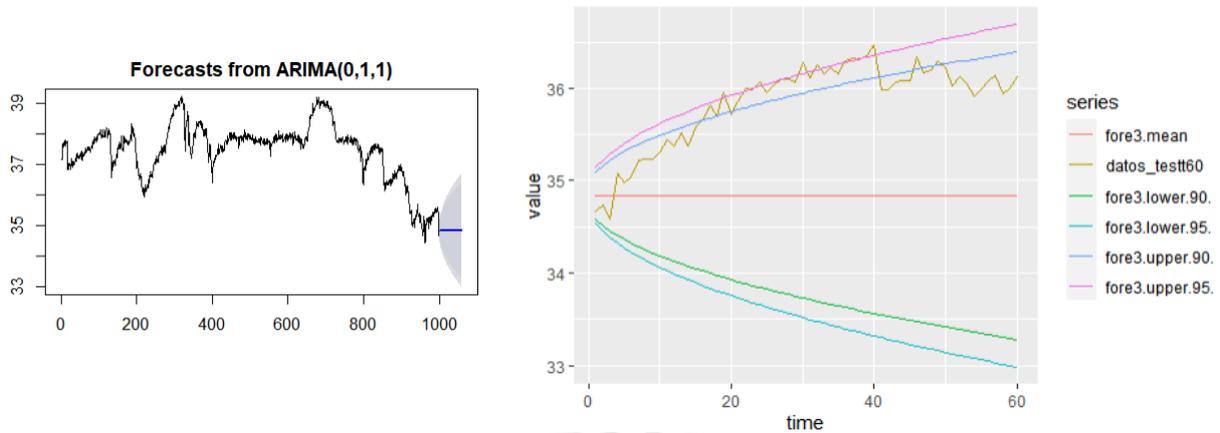
Para el horizonte de 30 minutos, los valores predichos convergen a 34.83°C (color rojo). Además, se observa que los datos reales (color mostaza) muestran una tendencia creciente hasta alcanzar los 36.13°C. En este plazo de tiempo se observa que los valores de temperatura continúan elevándose en el tramo de 15 a 30 minutos llegando a salir de los intervalos de confianza. Para este plazo de predicción se tiene una amplitud de los intervalos de confianza de 2.2°C y 2.6°C para los niveles de confianza de 90 y 95, respectivamente.

Figura 39 Predicciones a un plazo de 30 minutos



Adicionalmente se utilizó el modelo para predecir los siguientes 60 valores, como en los casos anteriores, el modelo converge a 34.83°C y los valores reales se salen de los intervalos de confianza hasta $t = 40$, momento en el cual la temperatura desciende ligeramente. Sin embargo, se observa que la diferencia entre el valor mínimo y máximo de los valores posibles dentro de los intervalos de confianza es de 3.2°C y 3.7°C para un nivel de confianza de 90 y 95 respectivamente.

Figura 40 Predicciones a un plazo de 60 minutos



El caso de este paciente es particular, ya que es probable que la temperatura del paciente haya descendido debido a algún tratamiento para la fiebre que luego se suspendiera o que el dispositivo de medición se haya movido por varios minutos para luego volver a colocarse adecuadamente. Esto explicaría que la temperatura observada se eleve en los plazos analizados, para la muestra de prueba.

Tabla 8 Comparativa entre las predicciones a 15, 30 y 60 minutos IMA(1,1) - paciente 2

	MAE	RMSE	MAPE	Amplitud NC 90	Amplitud NC 95
Predicciones 15 min	0.321%	0.438%	1.111%	1.6°C	1.9°C
Predicciones 30 min	0.718%	0.856%	2.106%	2.2°C	2.6°C
Predicciones 60 min	1.015 %	1.112%	2.867%	3.2°C	3.7°C

El paciente 2 tiene una temperatura media de 37.46°C, y se ha establecido una variación máxima de $\pm 1^\circ\text{C}$, entonces para este modelo se admite un error máximo de 2.6%. De la Tabla 8 se tiene que las predicciones del modelo son muy buenas, por debajo del error máximo, para los horizontes de 15 y 30 minutos.

Para las predicciones a 15 minutos se observa que el MAPE es 1.11%, además, el modelo plantea intervalos de confianza bastante acotados con una amplitud menor a 2°C por lo que se concluye que el modelo es aceptable para este horizonte.

Para el periodo de 30 minutos las predicciones del modelo salían del intervalo de confianza (se explica esto por una posible reconexión del equipo), no obstante, se observa un MAPE= 2.1% que se considera aceptable. A pesar de que en los horizontes de 30 minutos los MAPE son satisfactorios la precisión de las predicciones es muy grande (amplitud>2°C), en consecuencia, se descarta el modelo para estos dos horizontes.

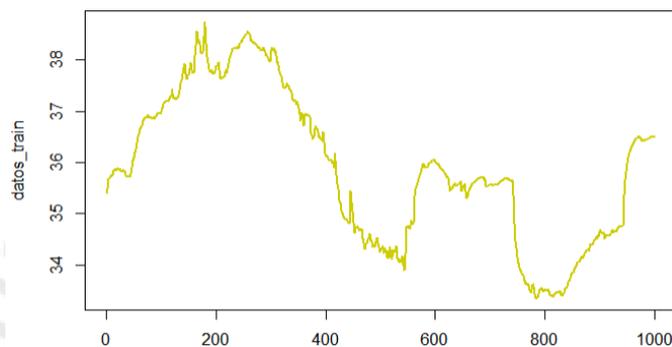
Por otro lado, para el periodo de 60 minutos se tiene un MAPE (2.87%) superior al aceptable, y una amplitud mayor a 3°C, entonces se admite que las predicciones del modelo son acertadas solo para el horizonte de 15 minutos.

6.1.3 Paciente 3

6.1.3.1 Características generales

Paciente masculino de 72 años de edad ingresado por una infección en el tracto urinario. La serie de temperaturas recogida incluía una desconexión al inicio de la serie, de manera que se eliminaron las primeras 30 observaciones. Posteriormente se observa que los valores de temperatura disminuyen incluso por debajo de 34°C para luego volver a incrementarse.

Figura 41 Termometría del paciente 3: T. Central



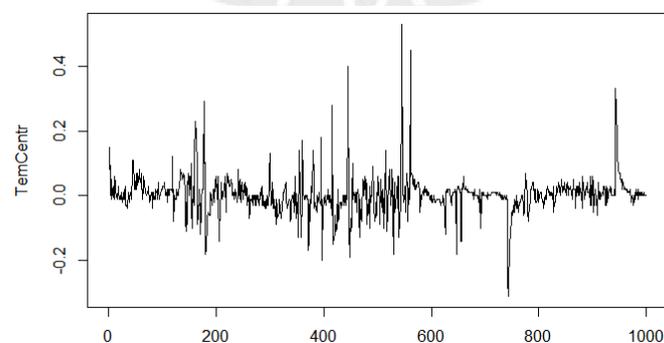
De la Figura 41 en la se puede observar que la dispersión de los datos en el tiempo se mantiene constante. Sin embargo, se observan muchas tendencias por lo que se realizarán diferenciaciones de la serie hasta obtener una serie estacionaria.

6.1.3.2 Identificación del modelo

6.1.3.2.1 Transformaciones

Se diferenció la serie una vez hasta eliminar las tendencias crecientes y decrecientes ($d=1$), para evaluar si hiciera falta realizar una diferenciación adicional se observó en la FAS del modelo que los valores se amortiguaban hasta quedar delimitados en la banda de no significación. De esta manera se ha obtenido una serie estacional para los valores de temperatura central, como muestra la Figura 42

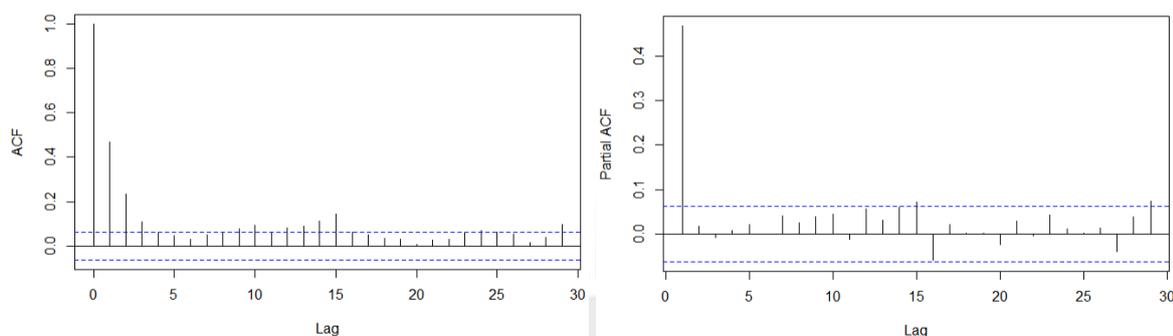
Figura 42 Serie de temperatura transformada



6.1.3.2.2 Selección de p, d y q

Del apartado anterior se obtuvo que d era igual a 1. Luego es necesario graficar los diagramas de autocorrelación simple y de autocorrelación parcial para seleccionar los parámetros p y q se grafican los diagramas de autocorrelación simple y autocorrelación parcial

Figura 43 Autocorrelación simple (izquierda) y parcial (derecha) de la T.Central paciente 3



En la Figura 43 se observa que el diagrama de autocorrelación simple se amortigua con el tiempo y en el diagrama de autocorrelación parcial se observa que el único valor significativo lo toma el primer retardo y luego la autocorrelación parcial toma valores no significativos, esto indica que se trata de un proceso autorregresivo con valor p=1. En consecuencia, se infiere que los datos siguen un modelo ARI (1,1)

6.1.3.3 Estimación de parámetros y contrastes

Luego de identificar el modelo se procede a estimar los coeficientes y residuos que genera el modelo ARI(1,1) al ajustarse a la termometría del paciente 3.

$$Z_t = a_t + 1.4687 Z_{t-1} - 0.4687 + Z_{t-2}$$

Ecuación 35 Estimación de los datos del paciente 3 con el modelo ARI(1,1)

Con:

- $a_t \sim N(\hat{\mu}_a = 0; \hat{\sigma}_a^2 = 0,0274)$
- $\text{Log} - \text{verosimilitud} = 1529.16$
- $\text{AIC} = -3054.32$

Se realizó un contraste z sobre el coeficiente obtenido para analizar la nulidad de este y se obtuvo un valor de contraste $z = 16.776$, entonces, con un nivel de confianza del 95% se admite que el coeficiente es no nulo. Además, se analizaron los residuos obtenidos con la prueba Ljung-Box como se observa en la

Figura 44, en donde se aprecia que la correlación simple de los residuos es poco significativa para los retardos mayores a cero, además, el valor p promedio es igual a 0.6828, por ello, se admite que los residuos obtenidos son independientes y, por lo tanto, que el modelo se ajusta adecuadamente a los datos. A continuación, se realizará un análisis de la normalidad de los residuos, para evaluar la normalidad de los datos.

Figura 44 Prueba Ljung-Box para los residuos del modelo ARI (1,1) – Paciente 3

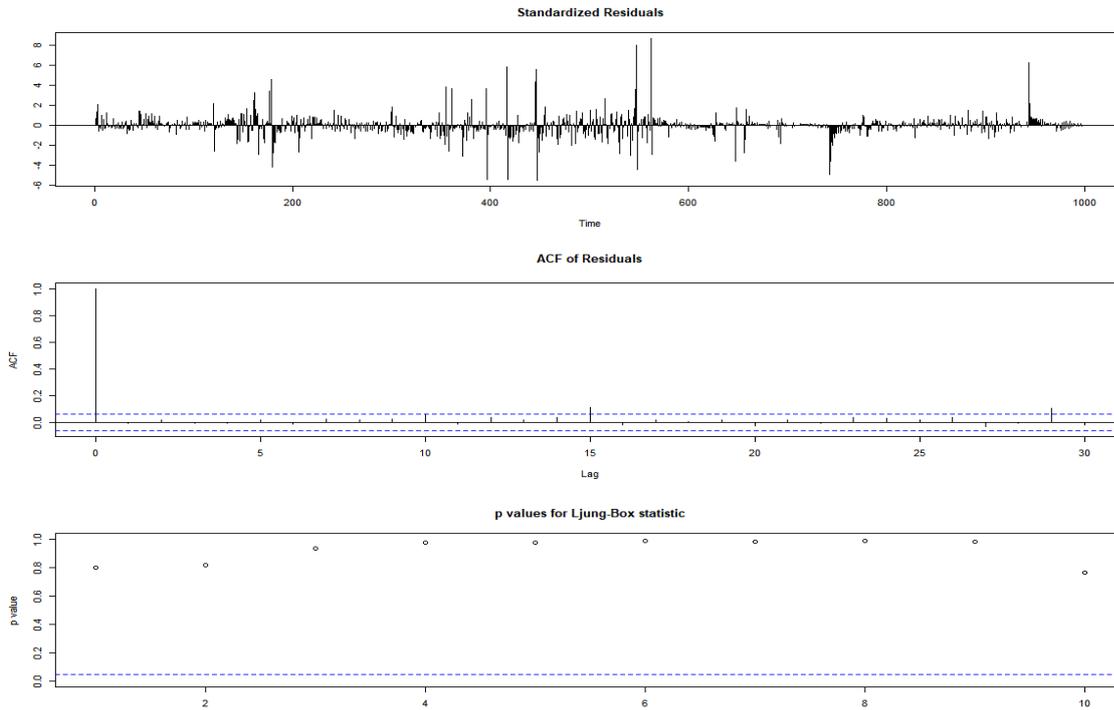
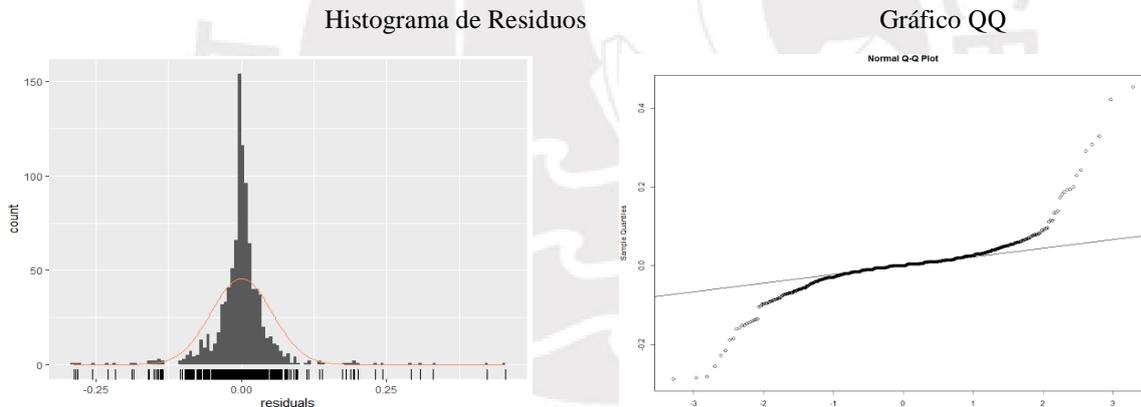


Figura 45 Normalidad de los Residuos obtenidos para el modelo ARI(1,1)



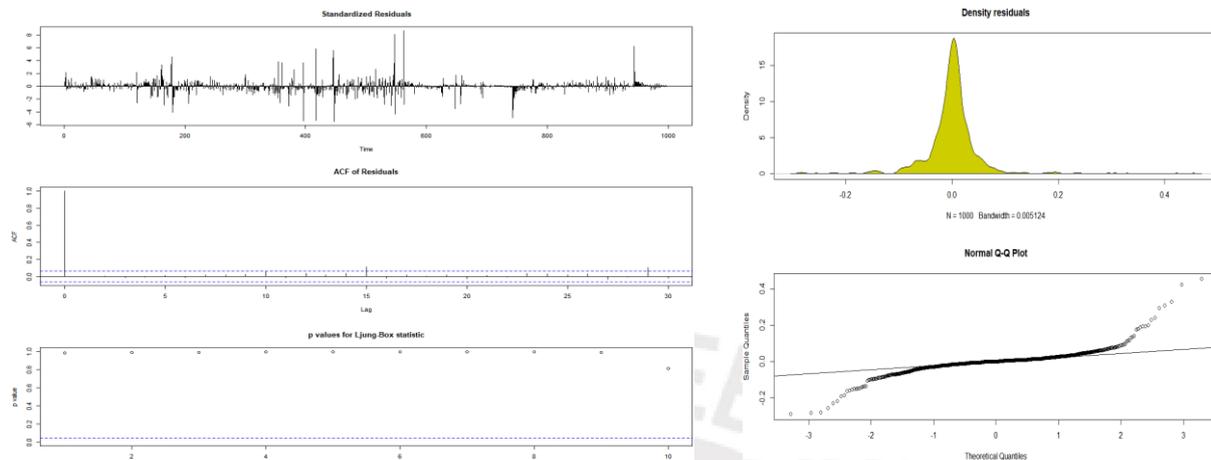
En el histograma de la Figura 45 se observa que los residuos forman una campana de Gauss centrada en cero, aunque una cola derecha más larga; para comprobar la normalidad de los residuos vemos el gráfico Q-Q (derecha). En este, se observa que los residuos no se comportan siguiendo una distribución normal, ya que hay muchas desviaciones con respecto a la línea marcada.

6.1.3.4 Diagnósis del modelo

El análisis de coeficientes indica que este es no nulo e invertible, además, de la Figura 45 se puede concluir que los residuos no se comportan de manera normal. No obstante, de la prueba de Ljung-Box se obtuvo valores p cercanos a 1, entonces se comprueba que los residuos son independientes y el modelo se ajusta a los datos. Para descartar que el modelo se encuentre infra dimensionado se evaluaron los modelos ARI (2,1) y el modelo ARIMA (1,1,1).

Para el modelo ARI(2,1) se realizó un contraste para analizar la nulidad de los coeficientes y se obtuvo un valor z para el coeficiente ϕ_2 de $0.589 \in (-2,2)$, entonces no se rechaza que sea no nulo.

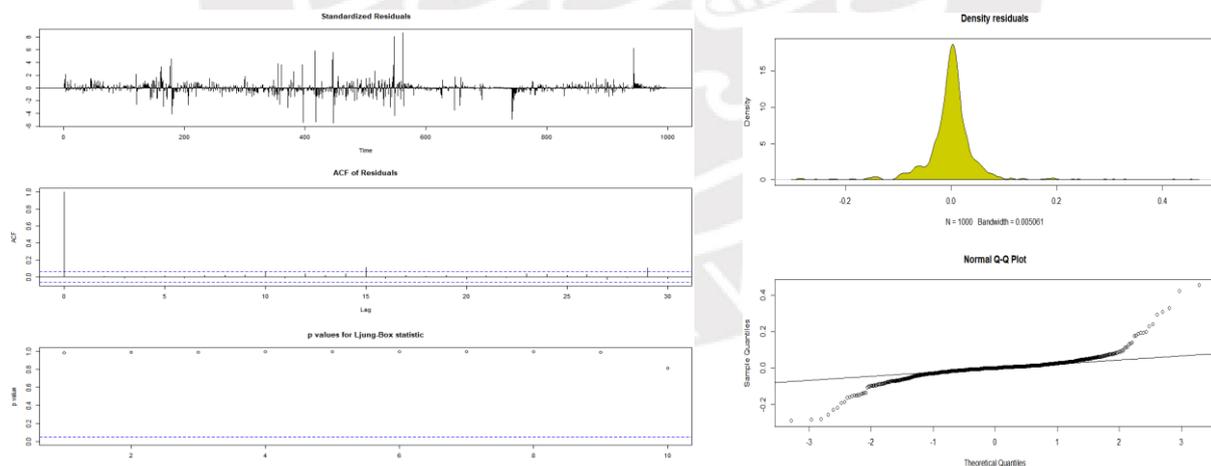
Figura 46 Residuos del modelo ARI(2,1) - paciente 3



En la Figura 46 se aprecia que los residuos aún cuentan con muchos valores atípicos, además para la prueba Ljung-Box se tiene una FAS con un valor significativo solo para $t = 0$, adicionalmente, se obtienen los valores p muy cercanos a 1, se concluye que los residuos son independientes y el modelo se ajusta a los datos. Finalmente, en el diagrama Normal QQ se muestra que los residuos no se comportan de manera normal, aunque su distribución es semejante a una campana de Gauss.

Para el modelo ARIMA(1,1,1) se observa que el coeficiente para de la parte de media móvil es no significativo, valor $z = -0.5766 \in (-2,2)$, entonces no se rechaza que sea no nulo.

Figura 47 Residuos del modelo ARIMA(1,1,1) - paciente 3



En la Figura 47 se aprecia que los residuos siguen teniendo valores atípicos en toda la serie, con la prueba de Ljung-Box se comprueba que los residuos son independientes, ya que todos los valores p son mayores a 0.05. Además, en el histograma y el diagrama QQ se puede ver que los residuos no se ajustan a una distribución normal.

Tabla 9 Comparación de modelos para el paciente 3

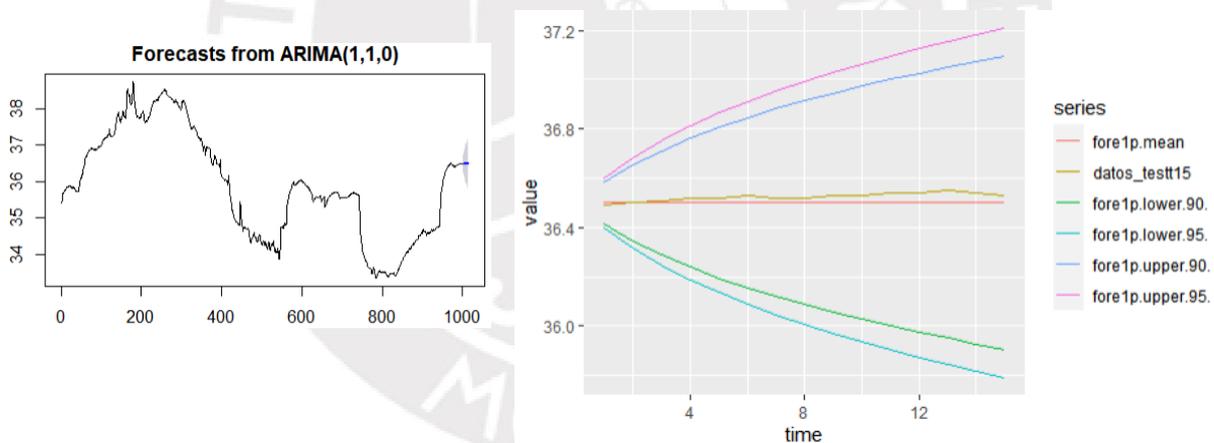
	Promedio de Valor p – Ljung-Box	AIC	BIC
ARI(1,1)	0.6828	--3054.323	-3044.510
ARI(2,1)	0.6472	-3052.671	-3037.951
ARIMA(1,1,1)	0.6455	-3052.660	-3037.94

Al comparar los modelos se puede apreciar que los resultados del Criterio de información de Akaike y el Criterio Bayesiano indican que el mejor modelo es el ARI(1,1), además, no se indican resultados significativamente distintos para los modelos ARI (2,1) y ARIMA(1,1,1). Finalmente, para los modelos ARI(2,1) y ARIMA (1,1,1) se obtuvieron coeficientes significativos solo para ϕ_1 , con lo que se concluye que el modelo con mejor ajuste es el ARI(1,1).

6.1.3.5 Predicciones a 15, 30 minutos y 60 minutos

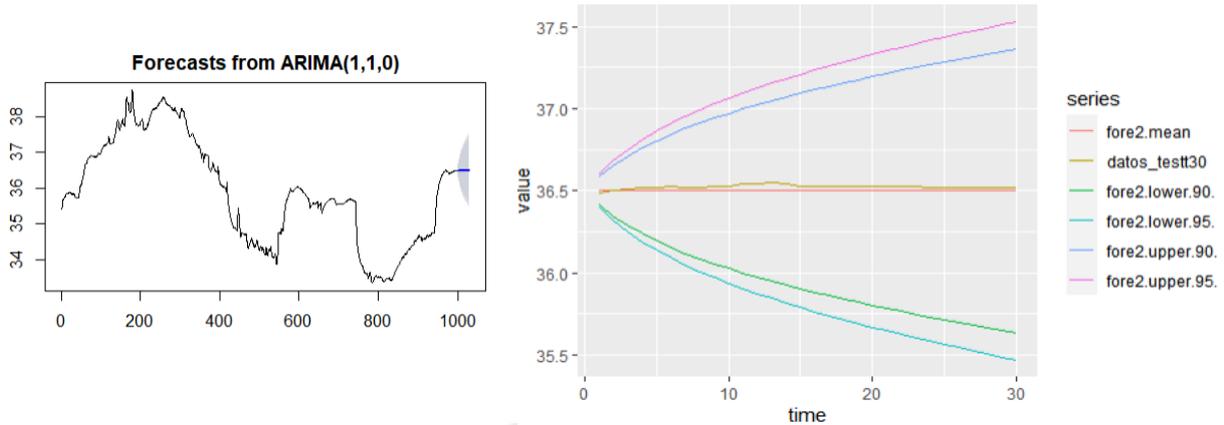
Se probó el modelo escogido para predecir los siguientes 15 valores de la serie de termometría y se observa que el modelo predice valores (color rojo) que convergen a 36.5°C, siendo una buena estimación de los datos reales (color mostaza) La variación entre el valor máximo y mínimo de los intervalos de confianza es de 1.2°C y 1.5°C para los niveles de 90 y 95 respectivamente

Figura 48 Predicciones a un plazo de 15 minutos



Asimismo, se probó el modelo para predecir los siguientes 30 valores de la serie de temperatura, se observa que, de manera similar al caso anterior, los valores predichos convergen a 36.5°C (color rojo). Como en el caso anterior se observa que los datos predichos se aproximan bastante bien a los datos reales (color mostaza). Para este periodo se tiene un rango de 1.7°C y 2.1°C para los intervalos de confianza de 90 y 95, respectivamente.

Figura 49 Predicciones a un plazo de 30 minutos



Además, se utilizó el modelo para predecir los siguientes 60 valores; como en los casos anteriores, el modelo converge a 36.5°C siendo una aproximación muy buena de los datos reales. También, se observa que la diferencia entre el valor mínimo y máximo de los valores posibles dentro de los intervalos de confianza es de 2.5°C y 2.9°C grados para un nivel de confianza de 90 y 95 respectivamente.

Figura 50 Predicciones a un plazo de 60 minutos

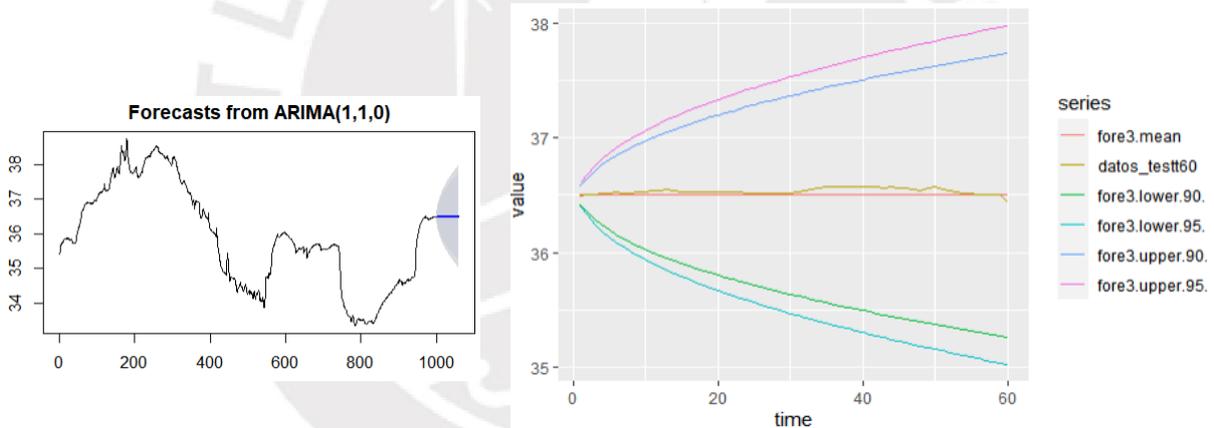


Tabla 10 Comparativa entre las predicciones a 15, 30 y 60 minutos ARI(1,1) - paciente 3

	EAM	RMSE	MAPE	Rango NC 90	Rango NC 95
Predicciones 15 min	0.0247%	0.0291%	0.0712%	1.2°C	1.5°C
Predicciones 30 min	0.0250%	0.0275%	0.0257%	1.7°C	2.1°C
Predicciones 60 min	0.0347 %	0.0425%	0.101%	2.5°C	2.9°C

El paciente 3 tiene una temperatura media de 35.89°C, se acepta una variación máxima de 1°C, en consecuencia, para este modelo se admite un error máximo de 2.78%

De la Tabla 10 se tiene que las predicciones del modelo son muy buenas, por debajo incluso de 0.5% para todos los periodos de predicción. Además, en el periodo de predicción de 15 minutos las amplitudes de los niveles de confianza son inferiores a 2°C para los niveles de confianza de 90 y 95%. En el periodo de 30 minutos la amplitud del intervalo de confianza de 90 es menor a 2°C mientras que para un nivel de confianza de 95 es de 2.1°C. Por otro lado, las predicciones para un intervalo de 60 minutos tienen un rango mayor a 2°. En conclusión, para este paciente las predicciones son aceptadas hasta un plazo de 30 minutos.

6.1.4 Discusión sobre los resultados de los 3 pacientes

- En base a lo presentado en la Introducción se ha establecido que el valor máximo de la variación de temperatura central para que un paciente se mantenga estable no debe superar $\pm 1^\circ\text{C}$. Esto implica que la precisión del modelo elegido debe ser tal que el error de predicción del modelo no debe ser mayor que 1°C .
- Para los pacientes 1 y 3 se ha observado que los datos se ajustan muy bien a un proceso ARI (1,1), mientras que para el paciente 2 el modelo con mejor ajuste era IMA(1,1).
- A pesar de que los residuos no fueron normales para ningún modelo estudiado; en los 3 modelos se cumple la hipótesis fundamental de la independencia de los residuos; además, los modelos para cada paciente tuvieron buenos ajustes ya que todos los modelos tienen un MAPE inferior a 3% en los 3 plazos de tiempo estudiados (15, 30 y 60 minutos).
- Debido a las restricciones de la variable temperatura central, se ha descartado la utilidad de modelos rango del intervalo de confianza para las predicciones mayor a 2°C.
- Por lo anterior, los modelos identificados para el paciente 1 y para el paciente 2 son adecuados para predicciones hasta un plazo de 15 minutos. Mientras que el modelo del paciente 3 produce predicciones aceptables para un horizonte de hasta 30 minutos.

6.1.5 Conclusiones

Es importante tener en cuenta la naturaleza de la variable observada, si bien, en las muestras de prueba de las series utilizadas había rangos de temperatura de 5.2°C, 3.4°C y 4.5°C para los pacientes 1, 2 y 3, respectivamente. Para que el modelo sea útil en un ámbito médico, es necesario acotar el intervalo de confianza de la predicción al máximo, ya que no detectar variaciones de $\pm 1^\circ\text{C}$ en la temperatura central pueden resultar fatales para el paciente.

Desde la perspectiva de modelos ARIMA se cumple la hipótesis fundamental de independencia de los residuos para los 3 modelos identificados mas no la hipótesis de la distribución normal de los residuos. Por ello se analizarán los datos con algoritmos no lineales, redes neuronales, para modelar los datos.

6.2 Redes Neuronales Recurrentes

Para el análisis de los datos se realizaron redes de Elman utilizando la librería de RSNNS en RStudio. Además, para poder obtener resultados comparables de redes neuronales recurrentes con los resultados de los modelos ARIMA se utilizará el mismo conjunto de datos para cada paciente tanto para las muestras de entrenamiento como las muestras de prueba.

En todos los casos se transformaron los datos utilizando el método estándar explicado en el capítulo de Metodología, ya que se quería mantener la dispersión de los datos. Para los distintos modelos se usó como variables de entrada hasta 2 retardos de la temperatura central. Además, por limitaciones computacionales se plantearon modelos con 1 y 2 capas ocultas entre estos, se evaluaron hasta 8 modelos por cada paciente, con diferentes configuraciones de nodos en las capas ocultas.

Para predecir la variable de temperatura central de los pacientes se decidió usar dos retardos la variable de temperatura central, esto debido a que usar 2 retardos en lugar de solo uno mejoraba los resultados obtenidos. Sin embargo, cuando se seleccionan 2 retardos se debe tener en cuenta que no se dispone de datos para el instante de tiempo -2, entonces, se tienen que asignar los 2 primeros valores de la serie a los respectivos retardos.

Para evaluar el desempeño de los modelos generados se usó la técnica de holdout, y se obtuvo el MAPE para la muestra de entrenamiento y para la muestra de prueba en cada uno de los modelos, esto con el fin de evitar seleccionar modelos que tengan un buen ajuste para los datos en la muestra de entrenamiento, pero no en las muestras de prueba, de este modo, se evita el sobreajuste del modelo.

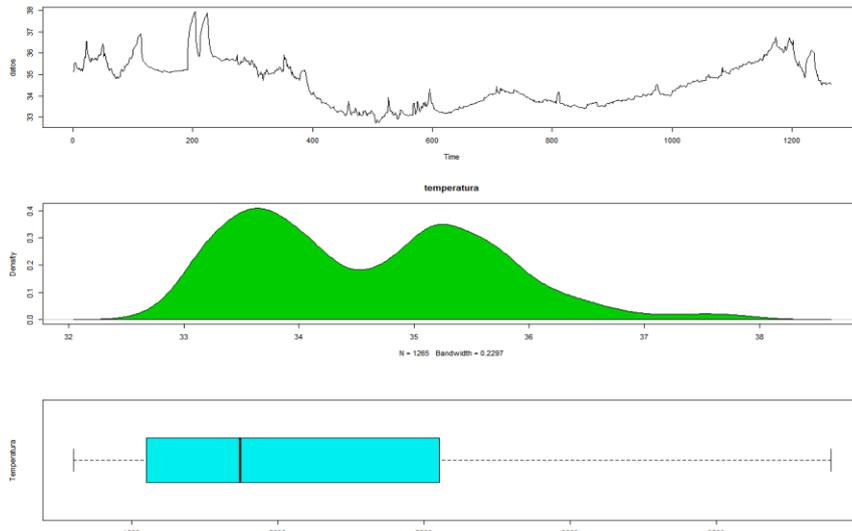
Se estableció la muestra de prueba de acuerdo con los horizontes de predicción, al igual que en modelos ARIMA, es decir, para el plazo de 15 minutos se utilizaron los siguientes 15 datos observados como muestra de prueba, para un horizonte de 30 minutos, los siguientes 30 datos y para un horizonte de 60 minutos, los siguientes 60 datos.

6.2.1 Paciente 1

6.2.1.1 Características generales

A modo de recordatorio y para facilitar la lectura se recuerda que la paciente uno es una mujer de 63 años de edad que ingresó a UCI con una infección viral. Para el análisis de sus termometrías se consideró una muestra de entrenamiento con los 1000 primeros valores de la serie y el tramo restante como muestra de prueba. La temperatura mínima observada es de 32.7°C y la máxima de 37.93°C.

Figura 51 Análisis descriptivo de la termometría del paciente 1



6.2.1.2 Selección del modelo

Al modelar redes de Elman usando el paquete RNSSN de RStudio, se pueden configurar los siguientes parámetros: número de nodos en las capas ocultas, número máximo de iteraciones y la tasa de aprendizaje. Para este estudio se configuró el número máximo de iteraciones en 500. Para establecer la tasa de aprendizaje se tuvo cuidado de no establecer este valor muy alto ya que podría causar que el modelo converja a un mínimo local. Luego de un proceso de prueba y error con los datos se estableció la tasa de aprendizaje en 0.1

A continuación, se establecieron diferentes configuraciones del número de unidades por capa y para cada uno de estos se comparó el ajuste del modelo a través del MAPE, se presenta un resumen de los modelos y los MAPE obtenidos en la Tabla 11.

Tabla 11 Comparativa de modelos – paciente 1

#Modelo	Capas Ocultas	Número de nodos en cada capa oculta	Datos de entrenamiento	Datos de prueba		
			MAPE datos reales	MAPE 15 minutos	MAPE 30 minutos	MAPE 60 minutos
1	1	2	0.783%	0.202%	0.277%	0.388%
2	1	5	0.497%	0.117%	0.136%	0.152%
3	1	4	0.341%	0.088%	0.092%	0.108%
4	1	3	0.362%	0.091%	0.097%	0.116%
5	2	(1,3)	0.773%	0.200%	0.270%	0.339%
6	2	(2,1)	1.314%	0.112%	0.147%	0.306%
7	2	(2,3)	0.612%	0.151%	0.134%	0.172%
8	2	(3,4)	0.551%	0.097%	0.122%	0.150%

En este estudio se considerará solo el MAPE a escala real (MAPE datos reales), si bien las transformaciones de los datos ayudan a mejorar el ajuste del modelo y evitar la convergencia, la finalidad del modelo es explicar los valores de entrenamiento (o, en su defecto, valores de prueba) en su escala real. Cabe destacar que el MAE y el RMSE son menos sensibles a la estandarización de los datos, sin embargo, se prefiere usar el MAPE como medida de comparación ya que ofrece una comparación relativa de los errores.

De todos los modelos generados, se observa que el modelo 3 tiene un mejor ajuste (señalado en amarillo) ya que en general muestra un menor MAPE en las muestras de entrenamiento y en las muestras de prueba.

6.2.1.3 Descripción del mejor modelo

- Número de nodos: 11
- Número de conexiones: 36
- Definición de cada nodo:
 - Inp: denota nodo de entrada, hid, nodo de capa oculta, out, nodo de salida y con, nodo de contexto

Tabla 12 Nodos de la red de Elman generada – paciente 1

	Tipo	Nombre	Sesgo	Función de activación
1	Entrada	inp1	0.21255	F. Identidad
2	Entrada	inp2	-0.86841	F. Identidad
3	Oculto	hid1	-0.25492	F. Sigmoide
4	Oculto	hid2	-0.50836	F. Sigmoide
5	Oculto	hid3	-1.02672	F. Sigmoide
6	Oculto	hid4	-2.00727	F. Sigmoide
7	Salida	out1	1.77923	F. Identidad
8	Contexto	con1	0.5	F. Identidad
9	Contexto	con2	0.5	F. Identidad
10	Contexto	con3	0.5	F. Identidad
11	Contexto	con4	0.5	F. Identidad

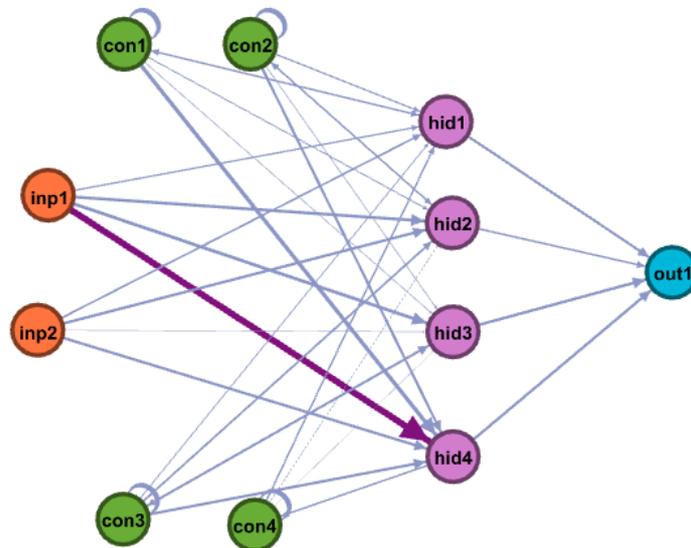
- Conexiones entre nodos

Tabla 13 Pesos de las conexiones para la red generada

Objetivo	Origen: peso					
3 – hid1	11: 0.94960	10:-0.56097	9: 0.87379	8:-1.00811	2: 1.09059	1: 0.87755
4 – hid2	11: 0.24965	10:-1.16216	9:-0.91714	8:-0.60641	2: 1.58395	1:-1.74635
5 – hid3	11:-0.15606	10:-1.41096	9:-0.43466	8: 0.40110	2:-0.28597	1:-1.87159
6 – hid4	11: 0.29654	10:-1.55141	9:-1.46662	8:-2.23728	2:-1.57704	1: 3.75557
7 – out1	6: 1.54639	5:-1.62955	4:-1.05043	3: 1.42573		
8 – con1	8: 0.30000	3: 1.00000				
9 – con2	9: 0.30000	4: 1.00000				
10 – con3	10: 0.30000	5: 1.00000				
11 – con4	11: 0.30000	6: 1.00000				

En la Figura 52 se ilustra la red generada utilizando el software libre Gephi v0.9.2. En este gráfico se han definido diversos colores para los distintos tipos de nodos: nodos de entrada: naranja, nodos ocultos: morado, nodos de contexto: verdes y nodo de salida: celeste, además, las diferentes tonalidades en los enlaces se asignaron en función de los pesos en valor absoluto de cada enlace, en ese sentido, los colores más claros indican pesos más bajos mientras que colores más oscuros indican pesos más altos. Los pesos se muestran de manera resumida en la Tabla 13.

Figura 52 Representación de la red de Elman – paciente 1



De la Figura 52 se observa que el nodo de entrada “inp1”, que hace referencia a un retardo de la variable temperatura central, aporta más información que el nodo de entrada “inp2”, que hace referencia a dos retardos de la variable.

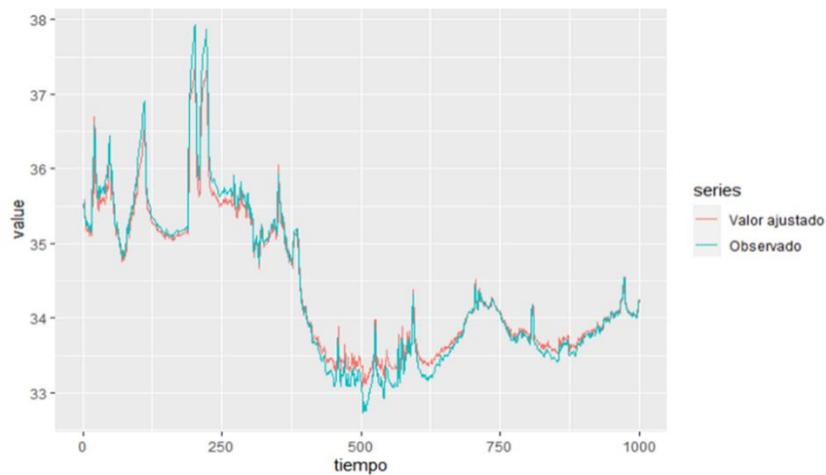
A continuación, se calculan los valores de entrada y salida de cada nodo, en esta ecuación los valores de entrada para cada nodo se representan con una i , mientras que los valores de salida con z .

Tabla 14 Cálculo de los valores de entrada o de salida para cada nodo de la red

Nodo		Valores de entrada y salida de cada nodo	
3	hid_1	$i_{3,t} = 0.878x_{1,t} + 1.091x_{2,t} - 1.008.u_{1,t} \\ + 0.874u_{2,t} - 0.561u_{3,t} \\ + 0.949.u_{4,t} - 0.255$	$z_{3,t} = F(i_{3,t}) = 1 / 1 + e^{-i_{3,t}}$
4	hid_2	$i_{4,t} = -1.746x_{1,t} + 1.584x_{2,t} - 0.606u_{1,t} \\ - 0.917u_{2,t} - 1.162u_{3,t} \\ + 0.249u_{4,t} - 0.508$	$z_{4,t} = F(i_{4,t}) = 1 / 1 + e^{-i_{4,t}}$
5	hid_3	$i_{5,t} = -1.872x_{1,t} - 0.286x_{2,t} + 0.401u_{1,t} \\ - 0.435u_{2,t} - 1.411u_{3,t} \\ - 0.156u_{4,t} - 1.027$	$z_{5,t} = F(i_{5,t}) = 1 / 1 + e^{-i_{5,t}}$
6	hid_4	$i_{6,t} = 3.756x_{1,t} - 1.578x_{2,t} - 2.237u_{1,t} \\ - 1.467u_{2,t} - 1.551u_{3,t} \\ + 0.297u_{4,t} - 2.01$	$z_{6,t} = F(i_{6,t}) = 1 / 1 + e^{-i_{6,t}}$
7	out_1	$i_{7,t} = 1.426z_{3,t} - 1.050z_{4,t} - 1.629z_{5,t} \\ + 1.546z_{6,t} + 1.779$	$z_{6,t} = i_{7,t}$
8	con_1	$c_{1,t} = z_{3,t-1} + 0.5$	$u_{1,t} = 0.3(c_{1,t})$
9	con_2	$c_{2,t} = z_{4,t-1} + 0.5$	$u_{2,t} = 0.3(c_{2,t})$
10	con_3	$c_{3,t} = z_{5,t-1} + 0.5$	$u_{3,t} = 0.3(c_{3,t})$
11	con_4	$c_{4,t} = z_{6,t-1} + 0.5$	$u_{4,t} = 0.3(c_{4,t})$

Previamente, ya se había calculado que el MAPE del modelo seleccionado para la muestra de entrenamiento era de 0.34%, esto implica que el modelo representa a los datos observados con mucha certeza. En la Figura 53 se aprecia cómo el modelo se adapta a las tendencias y picos de los datos reales, aunque subestima los cambios, por lo que se puede decir que se trata de un modelo conservador para predecir los cambios de temperatura central. Se espera un comportamiento similar para las predicciones.

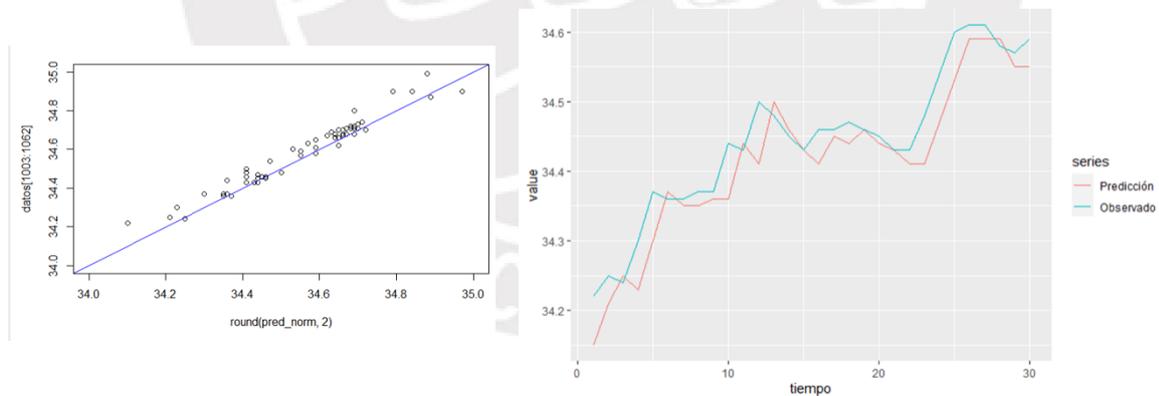
Figura 53 Gráfica del ajuste del modelo vs los datos reales



6.2.1.4 Predicciones a 15, 30 y 60 minutos

A continuación, se muestra la estimación del modelo a 60 minutos, este horizonte ya incluye los horizontes de 15 y 30 minutos.

Figura 54 Predicciones a 60 minutos – paciente 1



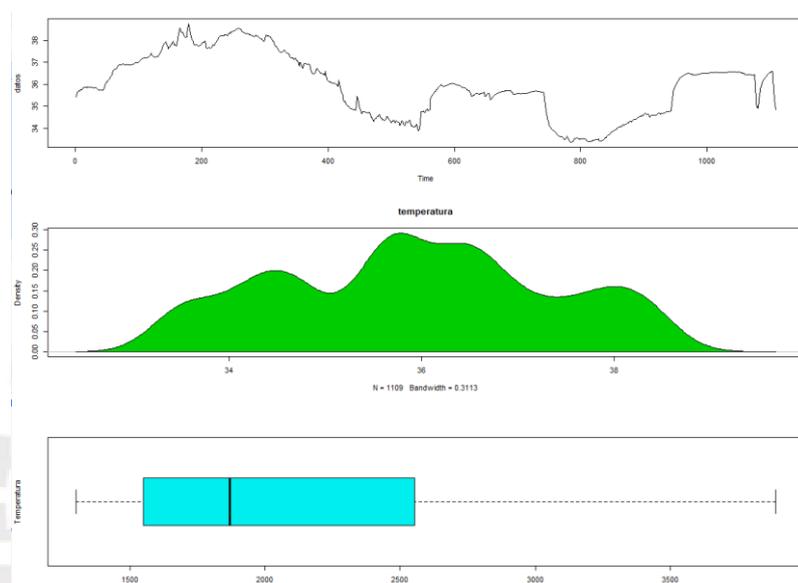
En la Figura 54 se puede observar que los valores predichos (en rojo) han logrado predecir las tendencias de los valores observados de temperatura (azul), sin embargo, para todos los horizontes se ha subestimado la temperatura real. Con respecto al estadístico del MAPE cabe resaltar, que como el límite de variación es de 1° el MAPE máximo permisible sería de 2.85%. y en este caso los errores generados son 0.088%, 0.092% y 0.108% para los horizontes de 15, 30 y 60 minutos, respectivamente.

6.2.2 Paciente 2

6.2.2.1 Características generales

El paciente dos es un hombre de 48 años de edad, que ingresó a UCI con una infección viral. Para el análisis de sus termometrías se consideró una muestra de entrenamiento con los 1000 primeros valores de la serie y el tramo restante como muestra de prueba. La temperatura mínima observada es de 32.7°C y la máxima de 37.93°C.

Figura 55 Análisis descriptivo de la termometría del paciente 2



6.2.2.2 Selección del modelo

De manera análoga al paciente 1, se analizaron diversos modelos de redes de Elman, variando el número de capas por modelo y en cada caso se evaluó el MAPE obtenido por cada modelo en la muestra de entrenamiento y en las muestras de prueba. Para este modelo, se usó una tasa de aprendizaje de 0.1 en los modelos de 1 capa y de 0.05 para los modelos de 2 capas, esto debido a que con la tasa de aprendizaje en 0.1 los modelos mostraban MAPE más grandes, tanto en la muestra de entrenamiento como en la muestra de prueba. Adicionalmente, se fijó el número máximo de iteraciones en 500 debido a que números mayores no mostraban mejoras significativas.

A continuación, se muestran los MAPE obtenidos para las predicciones de los modelos evaluados, el mejor modelo se ha resaltado en color amarillo.

Tabla 15 Comparativa de modelos – paciente 2

#Modelo	Capas Ocultas	Número de nodos en cada capa oculta	Datos de entrenamiento	Datos de prueba		
			MAPE datos reales	MAPE 15 minutos	MAPE 30 minutos	MAPE 60 minutos
1	1	2	0.362%	0.501%	0.499%	0.454%
2	1	3	0.443%	0.715%	0.764%	0.605%
3	1	4	0.449%	0.694%	0.726%	0.632%
4	1	5	0.452%	0.679%	0.746%	0.674%
5	2	(2,1)	0.745%	0.991%	1.229%	1.231%
6	2	(2,3)	0.506%	0.708%	0.774%	0.750%
7	2	(3,2)	0.477%	0.68.%	0.763%	0.735%
8	2	(3,4)	0.466%	0.673%	0.661%	0.574%

De la Tabla 15 se observa que el mejor modelo es el modelo 1, ya que, en comparación con los otros modelos produce medidas de MAPE inferiores tanto en los datos de entrenamiento como en los datos de prueba. Destaca además que para todos los horizontes de predicción se produjeron MAPE inferiores a 0.5% que indican que el modelo predice con mucha certeza los valores futuros.

Cabe resaltar que el modelo 1 cuenta con solo 1 capa oculta y 2 nodos en dicha capa, es un modelo relativamente sencillo, ya que en total solo tiene 7 nodos, en el siguiente apartado se explicará en detalle el modelo.

6.2.2.3 Descripción del mejor modelo

- Número de nodos: 7
- Número de conexiones: 14

Tabla 16 Nodos de la red de Elman generada – paciente 2

Nodo	Tipo	Nombre	Sesgo	Función de activación
1	Entrada	inp1	-0.15575	F.Identidad
2	Entrada	inp2	-0.55014	F.Identidad
3	Oculto	hid1	-1.7303	F. Sigmoide
4	Oculto	hid2	-1.5444	F. Sigmoide
5	Salida	out1	3.38092	F.Identidad
6	Contexto	con1	0.5	F.Identidad
7	Contexto	con2	0.5	F.Identidad

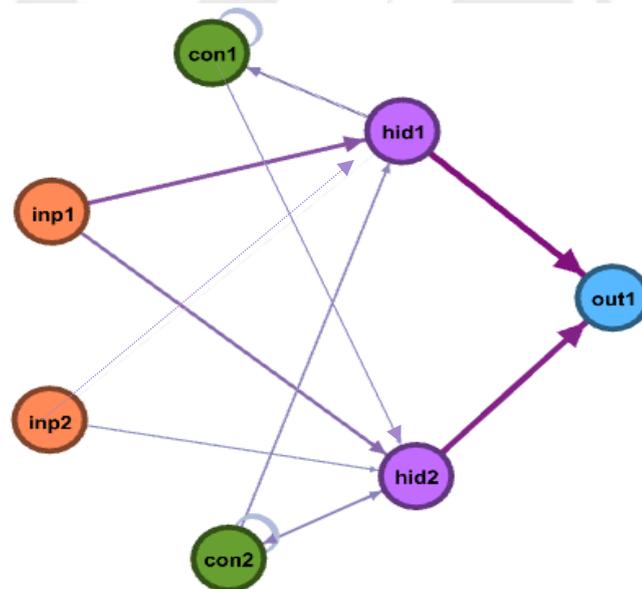
- Conexiones entre nodos

Tabla 17 Pesos de las conexiones para la red generada – paciente 2

Objetivo	Origen: peso			
3 – hid1	7:-0.80103	6: 0.05625	2:-0.11981	1:-1.58490
4 – hid2	7:-0.88828	6:-0.47063	2: 0.60633	1: 1.24662
5 – out1	4: 2.15955	3:-2.32726		
6 – con1	6: 0.30000	3: 1.00000		
7 – con2	7: 0.30000	4: 1.00000		

En la Figura 56 se ilustra la red generada, usando Gephi. En este gráfico se han definido diversos colores para los distintos tipos de nodos: nodos de entrada, color naranja, nodos ocultos(morado), nodos de contexto(verdes) y nodo de salida(azules), además, las diferentes tonalidades en los enlaces se asignaron en función de los pesos en valor absoluto de cada enlace, en ese sentido, los colores más claros indican pesos más bajos mientras que colores más oscuros indican pesos más altos. Los pesos se muestran de manera resumida en la Tabla 17.

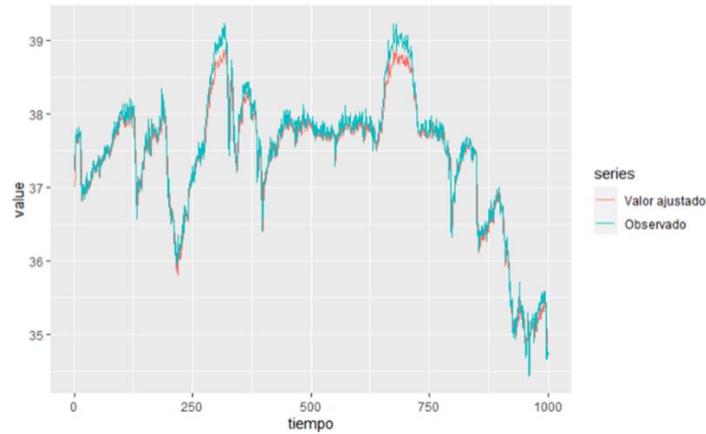
Figura 56 Representación de la red de Elman – paciente 2



De la figura anterior, se observa que el nodo de entrada “inp1” (primer retardo, t-1) contribuye en mayor medida a la obtención del valor final que la variable “inp 2” (segundo retardo, t-2)

Se eligió el modelo con un MAPE de 0.362% para la muestra de entrenamiento, la magnitud de esta medida implica que el modelo representa los datos observados acertadamente, ver Figura 57. En esta figura se observa que el modelo se adapta a las variaciones de los datos observados y estima adecuadamente los picos de temperatura en los datos de entrenamiento. Para este paciente se debe tener en cuenta que hay una posible desconexión del dispositivo de medición para los últimos 100 momentos de la serie de temperatura, y que el dispositivo se podría haber vuelto a conectar en durante los últimos minutos.

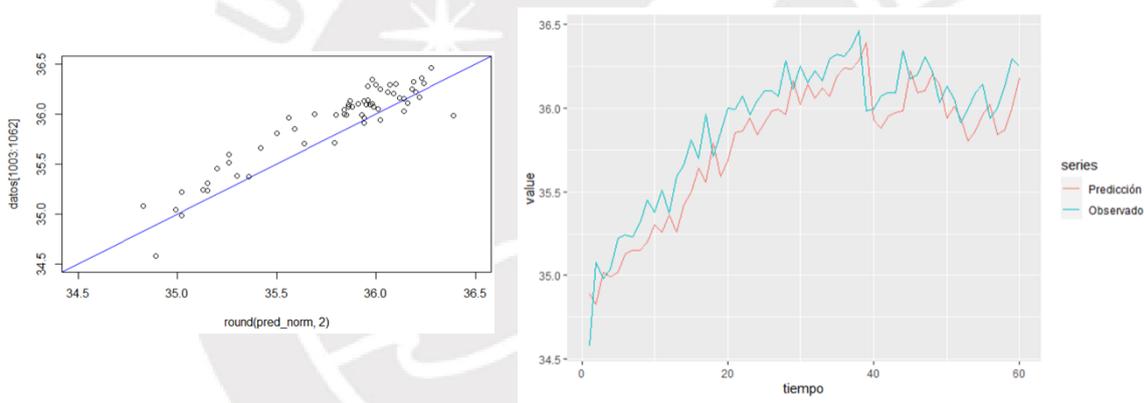
Figura 57 Gráfica del ajuste del modelo vs los datos reales



6.2.2.4 Predicciones a 15, 30 y 60 minutos

A continuación, se muestra la estimación del modelo a 60 minutos, este horizonte incluye las predicciones para los plazos de 15 y 30 minutos.

Figura 58 Predicciones a 60 minutos – paciente 2



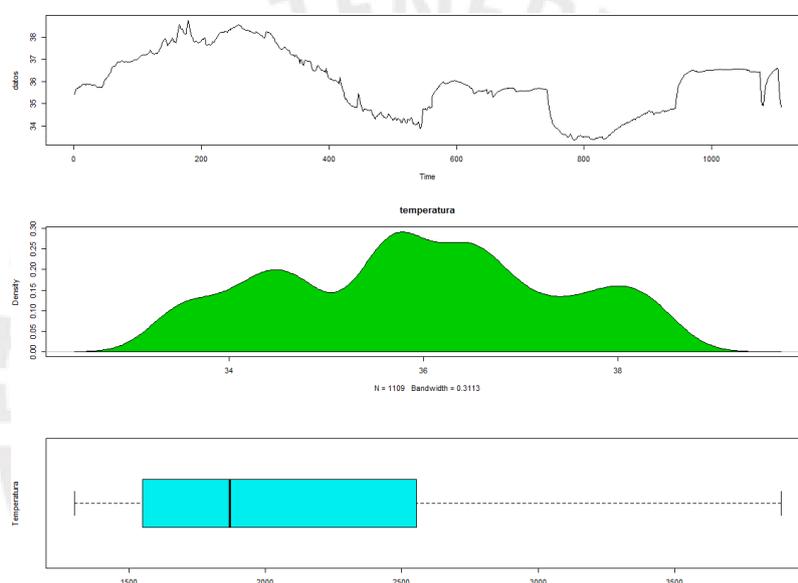
En la Figura 58 se observa como el modelo elegido predice (en rojo) el incremento de temperatura producido por la reconexión del equipo de la sonda ótica, y aunque los valores observados de temperatura (azul) son superiores, el ajuste del modelo es muy satisfactorio. Con respecto al estadístico MAPE cabe resaltar, que como el límite de variación es de 1° el MAPE máximo permisible sería de 2.6%. y en este caso los errores generados son 0.501%, 0.499% y 0.454% para los horizontes de 15, 30 y 60 minutos, respectivamente.

6.2.3 Paciente 3

6.2.3.1 Características generales

Paciente masculino de 72 años de edad ingresado en la UCI por una infección en el tracto urinario. Para el análisis de sus termometrías se consideró una muestra de entrenamiento con los 1000 primeros valores de la serie y el tramo restante como muestra de prueba. En la serie, se registra una temperatura mínima igual a 33.35°C y la máxima 38.74°C. Esta termometría tiene la particularidad de que aproximadamente las últimas 40 observaciones de la muestra de entrenamiento toman valores con diferencias mínimas, estos valores oscilan entre 36.36°C y 36.5 °C, posteriormente se observa un incremento ínfimo de la temperatura y los datos toman como máximo 36.57°C.

Figura 59 Análisis descriptivo de la termometría del paciente 3



6.2.3.2 Selección del modelo

Como ya se ha mencionado anteriormente, se seleccionará como mejor modelo al que presente un menor MAPE en las muestras de prueba y de entrenamiento. Para este modelo, se usó una tasa de aprendizaje de 0.1 en los modelos de 1 y 2 capas. Adicionalmente, se fijó el número máximo de iteraciones en 500 debido a que un número de iteraciones mayor en los modelos no mostraban mejoras significativas.

A continuación, se muestran los MAPE obtenidos para las predicciones de los modelos evaluados, el mejor modelo se ha resaltado en color amarillo.

Tabla 18 Comparativa de modelos – paciente 3

#Modelo	Capas Ocultas	Número de nodos en cada capa oculta	Datos de entrenamiento	Datos de prueba		
			MAPE datos reales	MAPE 15 minutos	MAPE 30 minutos	MAPE 60 minutos
1	1	2	0.268%	0.021%	0.016%	0.021%
2	1	3	0.289%	0.020%	0.014%	0.020%
3	1	4	0.270%	0.021%	0.017%	0.022%
4	1	5	0.308%	0.019%	0.017%	0.021%
5	2	(1,3)	0.729%	0.020%	0.017%	0.023%
6	2	(2,3)	0.784%	0.020%	0.018%	0.023%
7	2	(2,4)	0.737%	0.021%	0.021%	0.027%
8	2	(3,4)	0.715%	0.021%	0.019%	0.024%

De la Tabla 18 se observa que el mejor modelo es el modelo 2, produce medidas de MAPE inferiores a los demás modelos, tanto en los datos de entrenamiento como en los datos de prueba. Destaca además que en este modelo en todos los horizontes de predicción se produjeron MAPE inferiores a 0.02%, se atribuye esta precisión a que en el último tramo la serie se aproxima a una línea recta.

Cabe resaltar que el modelo 2 cuenta con 1 capa oculta y 3 nodos en dicha capa, en el siguiente apartado se explicará en detalle el modelo.

6.2.3.3 Descripción del mejor modelo

- Número de nodos: 9
- Número de conexiones: 24

Tabla 19 Nodos de la red de Elman generada – paciente 3

Nodo	Tipo	Nombre	Sesgo	Función de activación
1	Entrada	inp1	0.26732	F. Identidad
2	Entrada	inp2	0.5682	F. Identidad
3	Oculto	hid1	-0.19589	F. Sigmoide
4	Oculto	hid2	-0.40375	F. Sigmoide
5	Oculto	hid3	-0.9754	F. Sigmoide
6	Salida	out1	0.51062	F. Identidad
7	Contexto	con1	0.5	F. Identidad
8	Contexto	con2	0.5	F. Identidad
9	Contexto	con3	0.5	F. Identidad

- Conexiones entre nodos

Tabla 20 Pesos de las conexiones para la red generada – paciente 3

Objetivo	Origen: peso				
3 – hid1	9:-0.61197	8:-0.33094	7: 0.54993	2: 1.02452	1: 0.47117
4 – hid2	9:-0.21377	8:-1.03187	7: 0.05081	2:-0.46984	1: 1.29024
5 – out1	9:-0.16393	8:-0.00236	7: 0.64355	2: 0.46560	1:-1.41429
6 – con1	5:-2.76879	4: 1.90796	3: 0.91484		
7 – con2	7: 0.30000	3: 1.00000			
8 – con1	8: 0.30000	4: 1.00000			
9 – con2	9: 0.30000	5: 1.00000			

En la Figura 60 se ilustra la red generada en este gráfico se distingue los tipos de nodos por colores: nodos de entrada: naranja, nodos de contexto : verde, nodos ocultos: morado y nodo de salida: celeste. Además, las diferentes tonalidades en los enlaces se asignaron en función de los pesos en valor absoluto de cada enlace. Los pesos se muestran de manera resumida en la tabla 20.

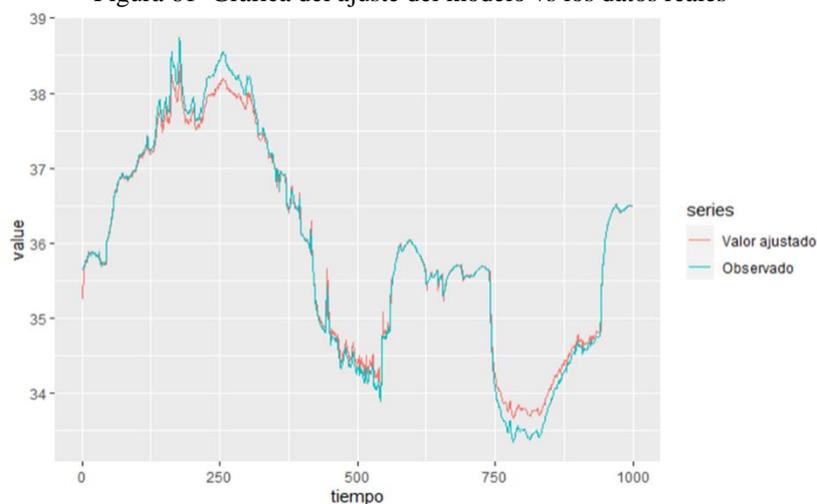
Figura 60 Representación de la red de Elman – paciente 3



Para el paciente 3, se observa que el primer retardo (“inp1”) contribuye en mayor medida al resultado final, no obstante, la contribución del nodo “inp2” segundo retardo también es significativa.

Se eligió el modelo con un MAPE de 0.289% para la muestra de entrenamiento, lo que indica gran precisión del modelo, ver Figura 61. En esta figura se observa cómo el modelo se adapta a las variaciones de los datos observados y estima adecuadamente los picos y valles de temperatura en los datos de entrenamiento.

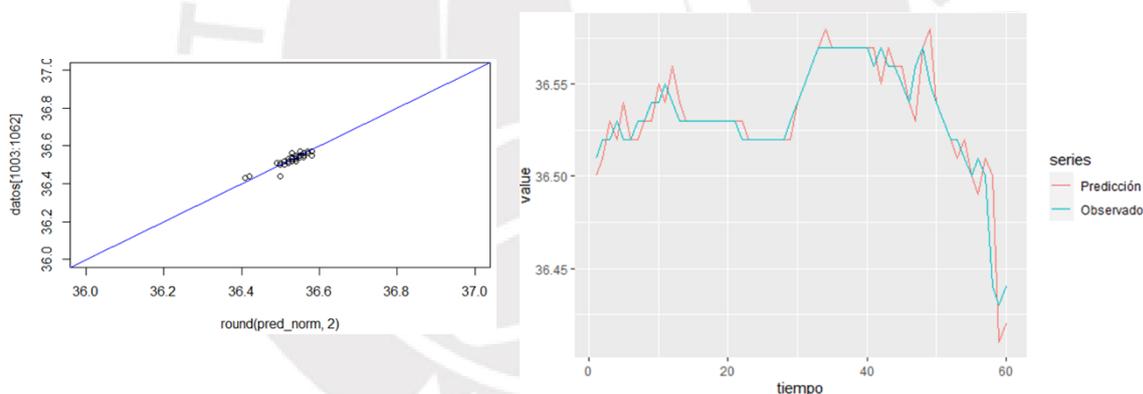
Figura 61 Gráfica del ajuste del modelo vs los datos reales



6.2.3.4 Predicciones a 15, 30 y 60 minutos

A continuación, se muestra la estimación del modelo a 60 minutos, este horizonte incluye las predicciones para los horizontes de 15 y 30 minutos.

Figura 62 Predicciones a 60 minutos – paciente 3



En la Figura 62 se observan las predicciones del modelo en rojo y los valores observados de temperatura en azul. Este modelo predice el perfil de cambios de la temperatura central para la muestra de entrenamiento. No obstante, se observa que cuando la temperatura se eleva el modelo predice incrementos ligeramente más altos que los observados. Con respecto al estadístico MAPE cabe resaltar, que el límite de variación máximo permitido es de 1°C , entonces, el MAPE máximo permisible sería de 2.78% y en este caso los errores generados son 0.02%, 0.014% y 0.02% para los horizontes de 15, 30 y 60 minutos, respectivamente.

6.2.4 Discusión sobre los resultados de los 3 pacientes

- En todos los modelos el primer retardo tenía una contribución mayor en el resultado final que el segundo retardo. Sin embargo, la introducción del segundo retardo como nodo de entrada, mejoraba la precisión del modelo.
- Para las 3 series de datos, se obtuvo mejores resultados en los modelos con solo una capa oculta.
- A pesar de que cada una de las series presenta particularidades, cambios de niveles y posibles desconexiones de los equipos, los modelos de redes neuronales de Elman representan con precisión los datos con MAPEs menores a 1% para los 3 horizontes de tiempo evaluados y para las series de temperatura de los 3 pacientes.

6.2.5 Conclusiones sobre los modelos de redes de Elman

Previamente se estableció que un modelo de predicción de temperaturas centrales de pacientes debía tener como máximo un error de predicción igual a 1°C, lo cual implica en promedio un error aceptable de 2.7%. Teniendo esto en cuenta, los modelos de redes neuronales de Elman predicen adecuadamente las temperaturas de los pacientes para los 3 horizontes de tiempo estudiados, obteniendo MAPE menores a 0.5% para las termometrías de los 3 pacientes estudiados.

A pesar de la complejidad de la regulación fisiológica de la temperatura central, el usar retardos de la temperatura central como valores de entrada para el modelo de redes neuronales hace que la estructura del modelo se simplifique y que la predicción de los futuros valores pueda estimarse con modelos de una sola capa de manera satisfactoria.

Para evaluar con mayor detalle la precisión de las redes de Elman es necesario calcular los intervalos de confianza de las predicciones de los modelos. Actualmente, aún se están estudiando métodos para el cálculo de los intervalos de confianza de las predicciones en algoritmos no lineales como las redes neuronales.

Capítulo 5

7. Conclusiones Generales

- En la introducción se estableció que la temperatura central de un paciente sano se establece aproximadamente constante y se mantiene dentro de límites estrechos ($\pm 1^\circ\text{C}$). Sin embargo, las series de temperatura central de los pacientes enfermos tienen una gran dispersión ($s_1=1.065$, $s_2=1.105$ y $s_3=1.405$, correspondientes al paciente 1, 2 y 3, respectivamente) esto puede ser debido a limitaciones del dispositivo para la medición de datos, que en ocasiones podría ser incómodo para el paciente ocasionando que la sonda de medición de temperatura central pudiera moverse y distorsionar los datos. Esta limitación se debe tener en cuenta, pues ha afectado también a los modelos generados.
- Se planteó que las redes neuronales recurrentes podrían tener una precisión ligeramente mayor que los modelos ARIMA. Como se puede observar en la Tabla 21, las predicciones de los modelos de Redes neuronales recurrentes de Elman (RNN), obtienen MAPEs hasta un 75% mejores. Cabe destacar, que ambos modelos son muy buenos para realizar las predicciones de los datos, sin embargo, se debe tener en cuenta la precisión que tiene cada modelo con respecto al horizonte de tiempo evaluado. Entonces, cabe esperar que la incertidumbre de las predicciones sea mayor cuanto mayor sea el horizonte de predicción.

Tabla 21 Comparación de MAPE para las predicciones de los modelos ARIMA y RNN

	MAPE 15 minutos		MAPE 30 minutos		MAPE 60 minutos	
	ARIMA	RNN	ARIMA	RNN	ARIMA	RNN
Paciente 1	0.419%	0.088%	0.638%	0.092%	1.045%	0.108%
Paciente 2	1.111%	0.501%	2.106%	0.499%	2.867%	0.454%
Paciente 3	0.071%	0.020%	0.026%	0.014%	0.101%	0.020%

-
- En la Tabla 21 se observa que los modelos ARIMA y las redes Elman planteados se ajustaron mejor a las termometrías del paciente 3. Esto resulta curioso ya que se trata del paciente con mayor edad (72). Según D'Hyver et-al (2019) la capacidad de regulación de la temperatura corporal se deteriora con la edad. Entonces, por la dispersión de los datos podría ser una serie un poco más difícil de estimar, no obstante, las predicciones de los modelos presentan MAPEs en todos los horizontes inferiores o iguales a 0.101%.
 - También, se observa que la termometría del paciente 2 (48 años de edad) ha tenido un MAPE más alto que los otros dos pacientes tanto para modelos ARIMA como RNN. Esto podría explicarse debido a que el dispositivo de medición parece haberse desacomodado en los últimos minutos de la muestra de entrenamiento en la termometría de este paciente, distorsionando los datos de temperatura central.
 - Para el paciente 1 se muestra que los modelos RNN tienen un error de predicción MAPE= 0.108% para un horizonte de 60 minutos. Mientras que para el modelo ARIMA, el MAPE es de 1.0405%. casi 10 veces mayor, aunque los dos son muy satisfactorios.
 - Tal como demuestran los modelos ARIMA, se observa que el primer retardo de la temperatura central es un buen predictor para futuros valores de esta variable. Entonces, a pesar de la complejidad de la regulación fisiológica de la temperatura, el usar retardos de la temperatura central como valores de entrada en lugar de otras variables relativas a las condiciones del paciente para el modelo RNN hace que la estructura del modelo se simplifique y que la predicción de los futuros valores pueda estimarse con modelos de una sola capa de manera satisfactoria, ajustándose a los datos en su escala real y produciendo MAPEs mínimos.

8. Líneas futuras

La identificación anticipada de fiebre a través de la predicción de futuras temperaturas corporales con un alto grado de precisión podría traer consigo mejoras en el tratamiento y el diagnóstico de dichos pacientes, no obstante, actualmente esta técnica no es habitual en la práctica clínica.

En ese sentido, sería interesante utilizar el método de monitorización continua de temperatura y la predicción de temperaturas corporales a pacientes con un sistema inmunológico deficiente, ya que este estudio y su aplicación podría permitir brindarles un tratamiento adecuado de manera oportuna e incrementar sus probabilidades de supervivencia.

Con el avance en los últimos años de los dispositivos de medición de temperatura y otras tecnologías se podría replicar este estudio para la monitorización continua de pacientes ingresados en UCI con un sensor o sonda que no incomode al paciente durante la toma de datos. Esto permitiría obtener termometrías de pacientes con valores más fiables y poder predecir variaciones de temperatura en horizontes de tiempo que permitan la identificación de pacientes con riesgo de desarrollar fiebre y en función a ello aplicarles el tratamiento oportuno.

Por otro lado, en un estudio de monitorización continua de pacientes puede suceder que se pierda la conexión del dispositivo de medición o el sensor de temperatura se mueva, en esos casos se podría utilizar los modelos desarrollados en esta investigación para cubrir aquellos intervalos de tiempo en donde se estima que la medición de los datos podría no ser correcta.

Finalmente, sería interesante obtener termometrías de pacientes con diversas características, edad, género, patología e identificar si existen patrones comunes de alteraciones de su temperatura corporal. Además, el estudio podría enriquecerse con un análisis multivariante incluyendo otras variables de entrada como presión arterial o frecuencia cardíaca.

9. Planificación Temporal

9.1 Estructura de Descomposición del Trabajo

Para explicar cómo se desarrolló la planificación de este trabajo es importante empezar describiendo la estructura de descomposición del trabajo. El desarrollo de este se compone de 3 Fases, según:

1. Elección del Tema de Investigación
 - Recolección de información
 - Análisis de los datos y posibles líneas de investigación
 - Elección del tema de tesis
 - Investigación del tema escogido

2. Desarrollo de la investigación
 - Investigación clínica
 - Investigación sobre mecanismos de termorregulación
 - Investigación sobre el desarrollo del proceso de fiebre
 - Investigación de estudios similares
 - Investigación técnica
 - Investigación sobre redes neuronales
 - Investigación sobre series temporales
 - Programación en R
 - Instalación e investigación de librerías
 - Simulación de modelos
 - Programación y tratamiento de datos
 - Informes sobre el desarrollo de la investigación
 - Recolección y análisis de datos
 - Análisis e interpretación de resultados
 - Recolección y redacción de la memoria
 - Corrección de informes

3. Presentación y defensa del proyecto
 - Elaboración de las diapositivas
 - Preparación de la defensa oral ante el tribunal
 - Acto de defensa

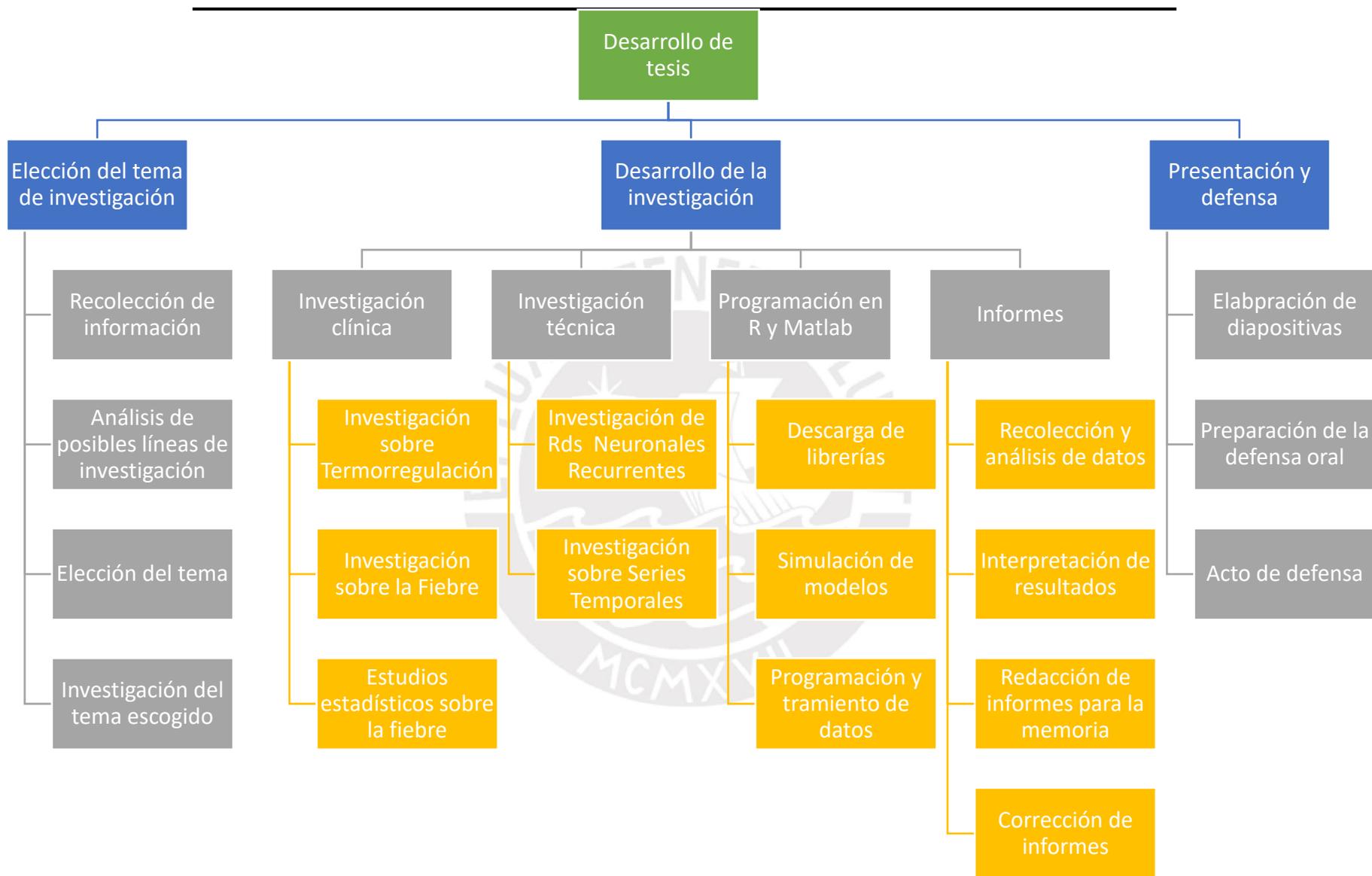


Figura 63 Estructura de Descomposición del Trabajo para el desarrollo de la tesis

9.2 Planificación

En base a la descomposición de tareas realizadas previamente, se realizó la siguiente planificación de tareas. Cabe resaltar, que los horarios destinados al avance de la tesis fueron cambiando conforme pasaba el tiempo debido a actividades laborales y horarios de clase.

En total, el desarrollo de esta tesis tomó poco más de 210 días, que se contabilizan desde el inicio del proyecto en setiembre, hasta la presentación del acto de defensa que se estima sea en la quincena de Julio, según el POD.

Para la obtención de la planificación y el diagrama de Gantt resultante, se utilizó el programa Microsoft Project 2016, ajustando los calendarios de trabajo al tiempo disponible.

Figura 64 Planificación de tareas

Nombre de tarea	Duración	Comienzo	Fin	Predecesoras
Desarrollo de Trabajo de Fin de Máster	210.92 días	vie 20/09/19	mar 14/07/20	
Elección del Tema de Investigación	30.5 días	vie 20/09/19	mar 19/11/19	
Recolección de información	20 días	vie 20/09/19	mar 29/10/19	
Análisis de datos	15 días	lun 30/09/19	mar 29/10/19	3CC+5 días
Posibles líneas de investigación	5 días	lun 21/10/19	mié 30/10/19	4FF
Investigación del tema escogido	10 días	jue 31/10/19	mar 19/11/19	5
Desarrollo de la investigación	184.92 días	lun 4/11/19	vie 10/07/20	
Investigación clínica	81.75 días	jue 5/12/19	lun 6/04/20	
Investigación sobre mecanismos de termorregulación	15 días	jue 5/12/19	mié 22/01/20	6FC+1 día
Investigación sobre la fiebre	20 días	jue 5/12/19	mar 28/01/20	6
Investigación de estudios similares	30 días	mar 3/03/20	lun 6/04/20	10
Investigación técnica	127.92 días	lun 4/11/19	vie 8/05/20	
Investigación sobre redes neuronales	20 días	jue 16/04/20	vie 8/05/20	6
Investigación sobre series temporales	15 días	lun 4/11/19	mar 3/12/19	6FC+3 días
Programación en R y Matlab	133.58 días	dom 10/11/19	mar 19/05/20	
Instalación de librerías	3 días	dom 10/11/19	vie 15/11/19	14CC+3 días
Investigación de librerías ST	10 días	sáb 16/11/19	jue 5/12/19	16
Investigación de librerías RNN	20 días	lun 23/03/20	mar 14/04/20	6
Simulación de modelos de ST	5 días	vie 6/12/19	dom 15/12/19	17
Simulación de modelos de RNN	20 días	mar 14/04/20	mié 6/05/20	18
Programación y tratamiento de datos ST	45 días	vie 6/12/19	mar 25/02/20	17
Programación y tratamiento de datos RNN	20 días	lun 27/04/20	mar 19/05/20	18
Informes sobre el desarrollo de la investigación	159.92 días	jue 16/01/20	vie 10/07/20	
Recolección y consolidación de información	80 días	jue 26/03/20	mar 23/06/20	6CC+3 días;3;9CC-10
Análisis de datos	95 días	vie 27/03/20	vie 10/07/20	19;24CC+2 días;18CC
Análisis de resultados	95 días	vie 27/03/20	vie 10/07/20	21CC+10 días;25CC
Redacción de la memoria	130 días	jue 16/01/20	mar 9/06/20	19FC+4 días
Corrección de los informes	45 días	lun 4/05/20	mar 23/06/20	27CC+20 días
Presentación y defensa del proyecto	22.5 días	vie 19/06/20	mar 14/07/20	
Elaboración de diapositivas	15 días	vie 19/06/20	mar 7/07/20	28FC-2 días
Preparación de la defensa oral	12 días	mar 30/06/20	lun 13/07/20	30
Acto de defensa	1 día	lun 13/07/20	mar 14/07/20	31

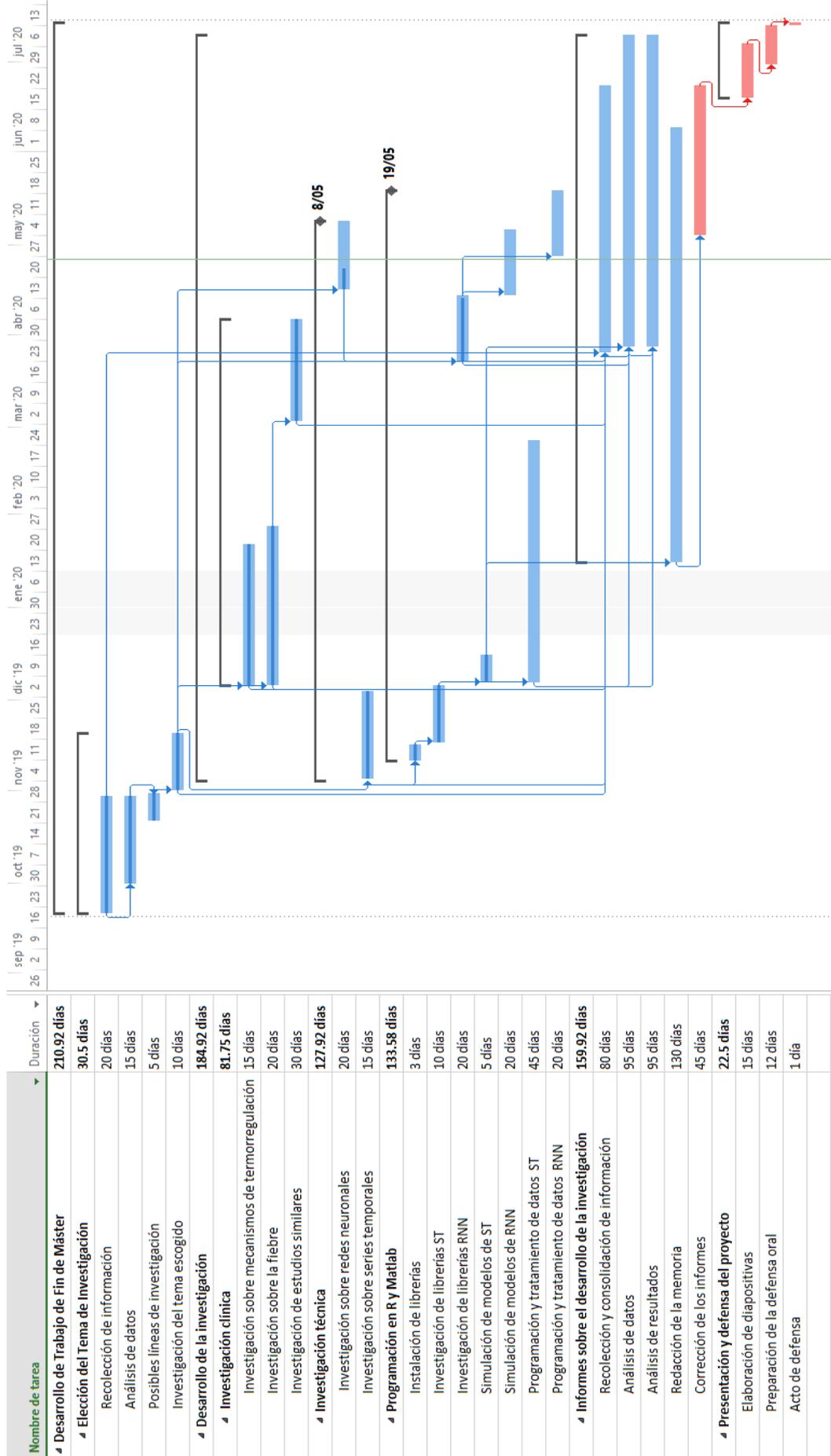


Figura 65 Diagrama de Gantt de las actividades indicadas previamente.

10. Bibliografía

Bennett IL, Beeson RB. (1954). Bacteremia: a consideration of some experimental and Clinical aspects. Yale J Biol Med pp 241–62

Box, G. E. P., Jenkins, GM., and Reinsel, G. C. (1994). Time Series Analysis, Forecasting and Control, 3 e. Prentice Hall, Englewood Cliffs, NJ.

Chatfield C.(1996). The Analysis of time series: An Introduction. 5e.pp 508 [\[DOI\]](#)

Creechan T, Vollman K, Kravutske ME. (2001). Cooling by convection vs. cooling by conduction for treatment of fever in critically ill adults. American Journal of Critical Care. pp: 52–59 [\[Google Scholar\]](#)

D'Hyver C, Gutiérrez L. M. y Zúñiga C. (2019). Geriatria 4ta Edición Manual Moderno México. [\[Google Scholar\]](#)

Dinarello CA, Gelfand JA. (2004). Fever and hyperthermia. In: Kasper DL, Braunwald E, Fauci AS, Hauser SL, Longo DL, Jameson JL, editors. Harrison's Principles of Internal Medicine 16e. McGraw-Hill; New York. pp: 104–108 [\[Google Scholar\]](#)

Dinarello CA, Porat R. (2012). Fever and hyperthermia. In: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson JL, Loscalzo J, editors. Harrison's principles of internal medicine, 18e. New York, NY: McGraw-Hill. pp: 86-92

Gisolfi CV, Mora F. (2000). The Hot Brain: Survival, Temperature and the Human Body. Massachusetts: MIT Press, pp: 1-13, 94-119, 157-63, 171-4, 191-215

Gómez A. (2007). Trastornos de la temperatura corporal. Educación sanitaria Vol. 26 Núm 7 Julio-Agosto [\[Elsevier\]](#)

Gómez A.(2008). La fiebre en la edad pediátrica. Pautas de actuación Vol. 27. Núm. 1. pp: 53-57 [\[Elsevier\]](#)

Guyton, A. C. (1971)., Basic Human Physiology: Normal Function and Mechanisms of Disease, W.B. Saunders Co.

Hastie, T, Tibshirani R, Friedman J. (2008) The Elements of Statistical Learning Data Mining, Inference and Prediction, 2e, Springer, pp: 389-403

Hirotoogu A. (1974). A new look at the statistical model identification», IEEE Transactions on Automatic Control 19 pp: 716-723

Hornik, K., Stichcombe M., y White H. (1989). Multilayer feedforward networks are universal approximators. Neural Networks, pp 359–366

Jones, T. (2017). "Recurrent neural networks deep dive" Build your own RNN, IBM Developer, IBM [\[IBM Developer\]](#)

Jordán J., Miro-Martínez P., Vargas B., Varela-Entrecanales M., Cuesta-Frau D. (2017). Statistical models for fever forecasting based on advanced body temperature monitoring, *Journal of Critical Care*, Volumen 37, pp.136-140 [\[ScienceDirect\]](#)

Kluger MJ., Kozak W., Conn CA., Leon LR, Soszynski D.(1998). Role of fever in disease. *Annals of the New York Academy of Sciences*. pp: 224–233. [\[Amazon AWS\]](#)

Kröse B. & Van der Smagt P. (1996). *An Introduction to Neural Networks*, 8e, University of Amsterdam. pp: 11-50

Lapedes A. y Farber R. (1987). "Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling", Technical Report, LA-UR87-2662, Los Alamos National Laboratory, Los Alamos, New Mexico.

Lepock JR. (2003). Cellular effects of hyperthermia: relevance to the minimum dose for thermal damage. *Int J Hyperthermia* pp: 252–266

Lewis N. D (2016). *Deep Learning made easy with R. A Gentle Introduction for Data Science*, pp: 6-120

Ljung, G. and Box, G. (1978). "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, pp: 297-303

Mackowiak PA, Durack DT (2010). Fever of unknown origin. In: Mandell GL, Bennett JE, Dolin R, editors. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 7e. New York: Churchill Livingstone Elsevier pp: 779–789

Mackowiak PA, Bartlett JG, Borden EC, Goldblum SE, Hasday JD, Munford RS, Nasraway SA, Stolley PD, Woodward TE. (1997). Concepts of fever: recent advances and lingering dogma. *Clinical Infectious Diseases* pp:119–138 [\[Google Scholar\]](#)

Mallet ML. (2002). Pathophysiology of accidental hypothermia. *Q J Med* 95 pp: 775–785 [\[Google Scholar\]](#)

Mylotte JM, Tayara A. (2000). Blood cultures: clinical aspects and controversies. *Eur J Clin. Microbiol Infect Dis* pp 157–63 [\[NCBI\]](#)

Morrison SF, Nakamura K. (2011). Central neural pathways for thermoregulation. *Front Biosci* pp: 74–104

Nakamura K.(2011). Central circuitries for body temperature regulation and fever. *Am J Physiol Regul Integr Comp Physiol*, pp: 1207–1228

O'Grady NP, Barie PS, Bartlett J, Bleck T, Garvey G, Jacobi J, Linden P, Maki DG, Nam M, Pasculle W, Pasquale MD, Tribett DL, Masur H. (1998). Practice parameters for evaluating new fever in critically ill adult patients. *Critical Care Medicine* pp: 392–408 [[Google Scholar](#)]

Peña, D. (2010). *Análisis de series temporales*, Alianza Editorial. Madrid

Ramón-Romero F., & Farías J. M. (2014). La fiebre. *Revista de la Facultad de Medicina, México*, pp: 20-33 [[Scielo](#)]

Romanovsky AA. (2007). Thermoregulation: some concepts have changed. Functional architecture of the thermoregulatory system. *Am J Physiol Regul Integr Comp Physiol*. An. pp: 37–46

Rowsey PJ.(1997). Pathophysiology of fever part 1: the role of cytokines. *Dimensions of Critical Care Nursing* pp: 202–207 [[PubMed](#)]

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics* [[Projecteuclid](#)]

Tansey E. and Johnson Christopher D. (2014). Centre for Biomedical Sciences Education, Queen's University, Belfast, Northern Ireland Recent advances in thermoregulation *Adv Physiol Educ* pp: 139–148 [[Physiology journal](#)]

Varela M., Churruca J., González A., Martín A., Ode J., Galdos P. (2006). Temperature Curve Complexity Predicts Survival in Critically Ill Patients *Am J Respir Crit Care Med* 2006. Aug; pp: 290–298

Vargas, B. (2017). *Análisis de complejidad y modelos estadísticos de predicción de la fiebre y consideración sobre la validez del concepto de fiebre en el ámbito clínico*. Madrid

Volker D. (1998). Fever of unknown origin. *Nurse Practitioner Forum*. pp: 70–176 [[Google Scholar](#)]

11.Anexos

- Modelos ARIMA

A modo de ejemplo se incluye el código utilizado con el software RStudio para el análisis de modelos ARIMA.

#Lectura de datos

```
datos = read.table("caso_1.csv", header=T, dec='.', sep=',';')
var(datos$tc)

## [1] 1.134283

sd(datos$tc)

## [1] 1.065027

#head(datos)
library(forecast)

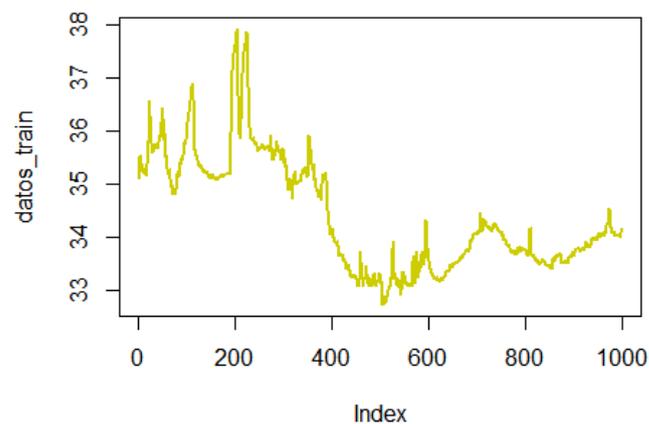
## Registered S3 method overwritten by 'quantmod':
## method      from
## as.zoo.data.frame zoo

library(tseries)
library(reshape2)
library(ggplot2)
```

Graficas de temperatura

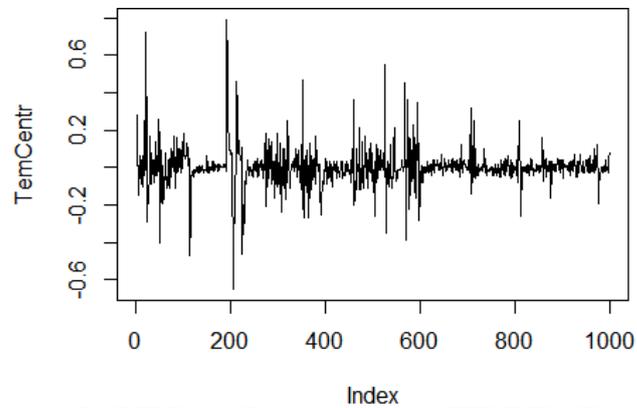
```
datos_train= window(datos$tc,start = 1,end=c(1000,1))

plot(datos_train, type="l", lty= 1, lwd=2, col='yellow3')
```



#Tratamiento de datos

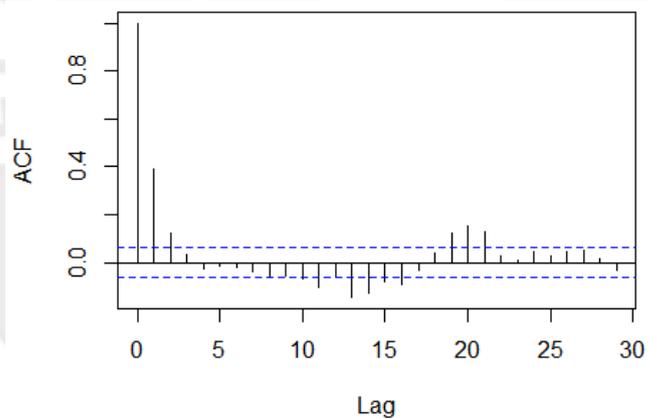
```
TemCentr=diff(datos_train)  
plot(TemCentr, type="l")
```



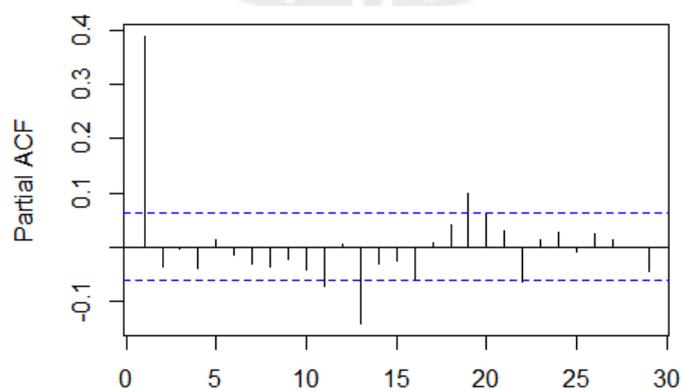
Autocorrelaciones Temperatura central

En este caso se trata de un proceso autorregresivo y diferenciado una vez:

```
acf(TemCentr)
```

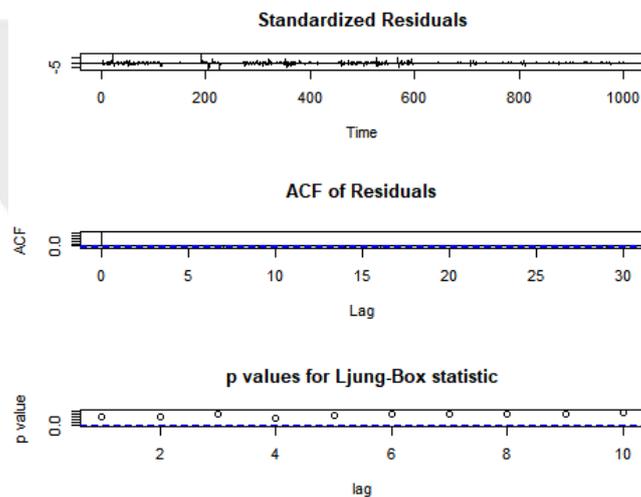


```
pacf(TemCentr)
```

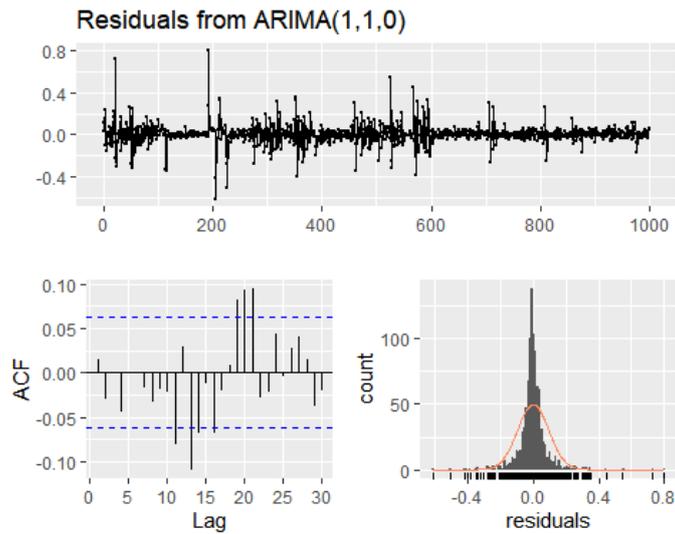


Diagnóstico de los modelos

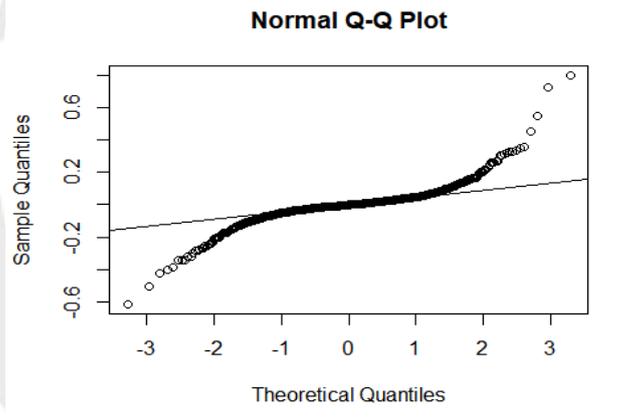
```
fit1 <- arima(datos_train, order=c(1,1,0),method="ML")
#)
summary(fit1)
## Coefficients:
##      ar1
##      0.3908
## s.e. 0.0291
## sigma^2 estimated as 0.009057: log likelihood = 932.17, aic = -1860.3
## Training set error measures:
##           ME   RMSE   MAE   MPE   MAPE
## Training set -0.0005348847 0.09512554 0.05457891 -0.001678485 0.1572499
##           MASE   ACF1
## Training set 0.9382951 0.01463273
library(lmtest)
coeftest(fit1)
##
## z test of coefficients:
##
## Estimate Std. Error z value Pr(>|z|)
## ar1 0.390808  0.029123  13.419 < 2.2e-16 ***
##
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
tsdiag(fit1)
```



```
checkresiduals(fit1)
```



```
## Ljung-Box test
## data: Residuals from ARIMA(1,1,0)
## Q* = 5.4369, df = 9, p-value = 0.7947
## Model df: 1. Total lags used: 10
qqnorm(residuals(fit1));qqline(residuals(fit1))
```



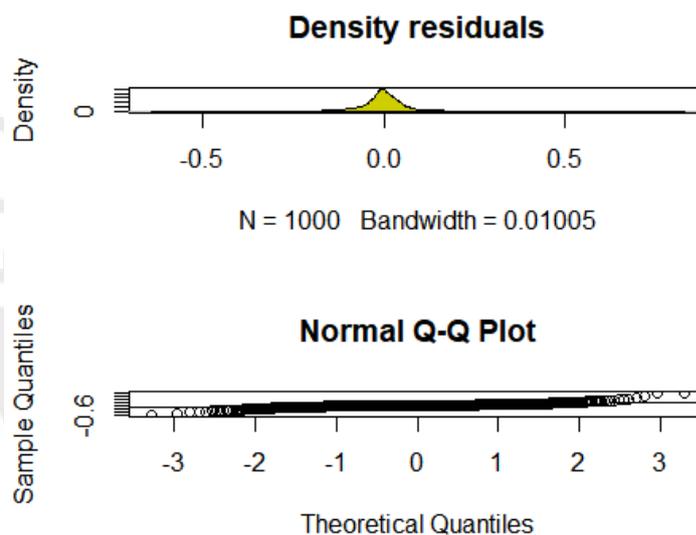
```
AIC(fit1)
## [1] -1860.342
BIC(fit1)
## [1] -1850.529
fit2 <- arima(datos_train, order=c(1,1,1),method="ML")
summary(fit2)
## Call:
## arima(x = datos_train, order = c(1, 1, 1), method = "ML")
##
## Coefficients:
##      ar1      ma1
##  0.3120  0.0932
## s.e. 0.0775  0.0813
##
## sigma^2 estimated as 0.009045: log likelihood = 932.83, aic = -1859.66
```

```

## ## Training set error measures:
##           ME   RMSE   MAE   MPE   MAPE
## Training set -0.0005561943 0.09506283 0.05475172 -0.001758582 0.1577113
##           MASE   ACF1
## Training set 0.9412659 0.0002262094
## Ljung-Box test
##
## data: Residuals from ARIMA(1,1,1)
## Q* = 4.4598, df = 8, p-value = 0.8134
##
## Model df: 2. Total lags used: 10

par(mfrow=c(2,1))
x <- density(fit2$residuals)
plot(x, main = "Density residuals")
polygon(x, col = "yellow3", border = 'black')
qqnorm(residuals(fit2));qqline(residuals(fit2))

```



```

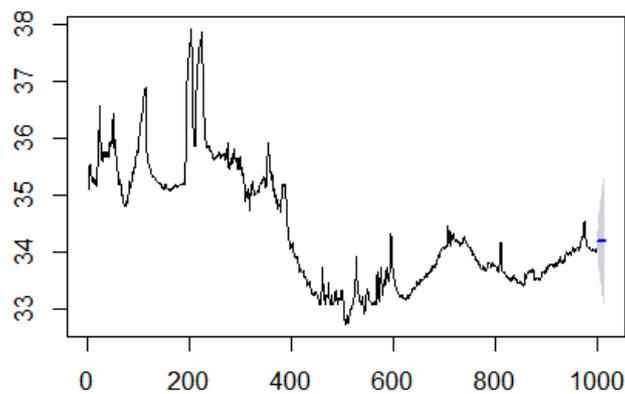
AIC(fit2)
## [1] -1859.657
BIC(fit2)
## [1] -1844.937

#Predicciones 15 minutos
datos_testt15=window(datos$tc,start = 1001,end = c(1015,1))
fit_pred <- arima(datos_train, order = c(1,1,0))
fore1p=forecast(fit_pred, h=15, level=c(90,95),col='green')
summary(fore1p)
## Forecast method: ARIMA(1,1,0)
## Model Information:
## Call:
## arima(x = datos_train, order = c(1, 1, 0))

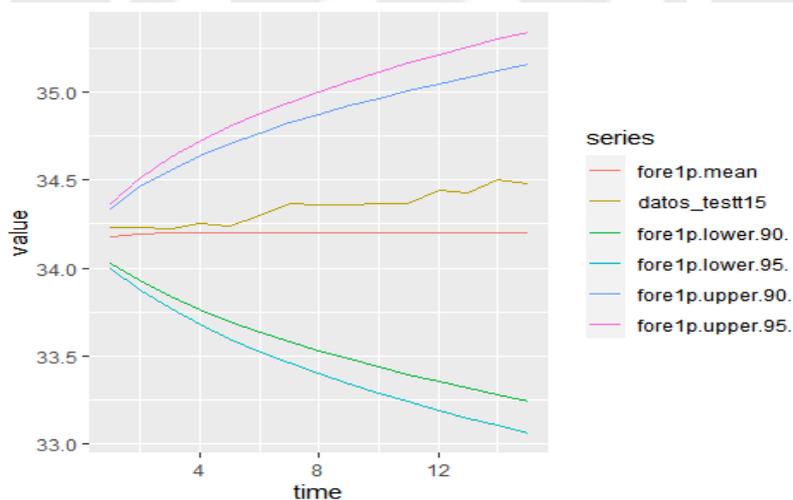
```

```
## Coefficients:
##   ar1
##   0.3908
## s.e. 0.0291
## sigma^2 estimated as 0.009057: log likelihood = 932.17, aic = -1860.34
## Error measures:
##           ME   RMSE   MAE   MPE   MAPE
## Training set -0.0005349132 0.09512554 0.05457887 -0.001678589 0.1572498
##           MASE   ACF1
## Training set 0.9382945 0.0146604
plot(fore1p)
```

Forecasts from ARIMA(1,1,0)



```
mm <- data.frame(cbind(time=1:15,fore1p$mean, datos_testt15, fore1p$lower, fore1p$upper))
mm<- melt(mm,id.vars= 'time', variable.name = 'series')
ggplot(mm, aes(time,value))+geom_line(aes(colour=series))
```



```
comparacion<- cbind(time=1:15,fore1p$mean, datos_testt15)
comparacion
```

```

## Time Series:
## Start = 1001
## End = 1015
## Frequency = 1
##   time fore1p$mean datos_testt15
## 1001  1  34.18126    34.23
## 1002  2  34.19348    34.23
## 1003  3  34.19825    34.22
## 1004  4  34.20012    34.25
## 1005  5  34.20085    34.24
## 1006  6  34.20113    34.30
## 1007  7  34.20124    34.37
## 1008  8  34.20129    34.36
## 1009  9  34.20130    34.36
## 1010 10  34.20131    34.37
## 1011 11  34.20131    34.37
## 1012 12  34.20131    34.44
## 1013 13  34.20132    34.43
## 1014 14  34.20132    34.50
## 1015 15  34.20132    34.48
## 1090 90  34.20132    35.08

```

```
accuracy(fore1p$mean,datos_testt15)
```

```

##           ME  RMSE  MAE  MPE  MAPE  ACF1 Theil's U
## Test set 0.1442123 0.16929 0.1442123 0.4192316 0.4192316 0.798932 4.603166

```

```
accuracy(fore2$mean,datos_testt30)
```

```

##           ME  RMSE  MAE  MPE  MAPE  ACF1 Theil's U
## Test set 0.2201151 0.2457763 0.2201151 0.6384722 0.6384722 0.8768473 7.308986

```

```
accuracy(fore3$mean,datos_testt60)
```

```

##           ME  RMSE  MAE  MPE  MAPE  ACF1 Theil's U
## Test set 0.3620665 0.4000947 0.3620665 1.045142 1.045142 0.9192901 11.75185

```

```

error <- mean(abs(fore1p$mean-datos_testt15)/fore1p$mean)
error

```

```
## [1] 0.004216655
```

- Modelos de redes neuronales de Elman

A modo de ejemplo se incluye el código generado en el software RStudio para el análisis de redes de Elman, cabe resaltar, que este no es necesariamente el modelo obtenido y explicado en apartados previos

```
##Lectura de datos
```

```
datos = read.table("caso_1.csv", header=T, dec='.', sep=',';)
```

```
#head(datos)
```

```
require(RSNNS)
```

```
require(quantmod)
```

```
library(ggplot2)
```

```
require(forecast)
```

```
require(deepnet)
```

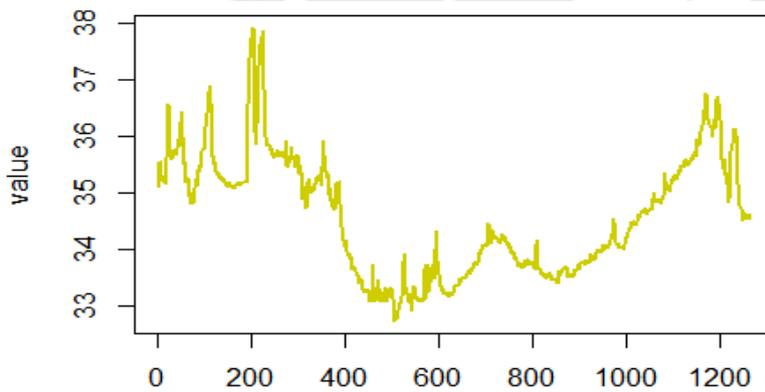
```
library(reshape2)
```

```
library(ggplot2)
```

```
Gráficas de temperatura
```

```
datos <- as.ts(datos$tc)
```

```
plot(datos, xlab="time", ylab="value", main="Total", lty= 1, lwd=2, col='yellow3')
```



```
par(mfrow=c(3, 1))
```

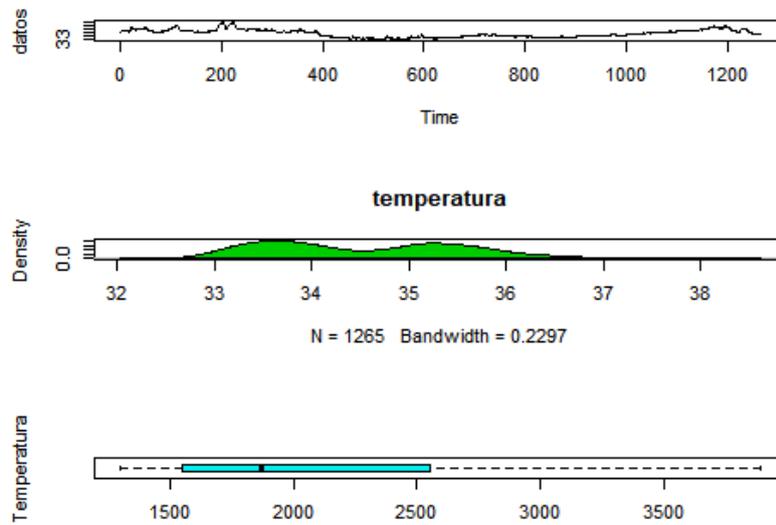
```
plot(datos)
```

```
x <- density(datos)
```

```
plot(x, main = "temperatura")
```

```
polygon(x, col = "green3", border = 'black')
```

```
boxplot(ldeaths, col= "cyan2", ylab="Temperatura", horizontal = TRUE)
```



```

mean(datos)
## [1] 34.58846
sd(datos)
## [1] 1.065027
y <- as.ts(datos)

#y <- log(y)
y <- as.ts(normalizeData(y, type="norm"))
y <- as.zoo(y)
x1 <- Lag(y, k = 1)
x2 <- Lag(y, k = 2)
data <- cbind(x1,x2,y)
colnames(data) <- c("x1", "x2", "y")

data <- data[-(1:2),]
head(data)
##      x1      x2      y
## 3 0.6211499 0.4990873 0.8840541
## 4 0.8840541 0.6211499 0.8934435
## 5 0.8934435 0.8840541 0.9028329
## 6 0.9028329 0.8934435 0.7619914
## 7 0.7619914 0.9028329 0.6587077
## 8 0.6587077 0.7619914 0.6399288

inputs <- data[,1:2]
outputs <- data[,3]
head(inputs)
##      x1      x2
## 3 0.6211499 0.4990873
## 4 0.8840541 0.6211499

```

```

## 5 0.8934435 0.8840541
## 6 0.9028329 0.8934435
## 7 0.7619914 0.9028329
## 8 0.6587077 0.7619914

head(outputs)

##      3      4      5      6      7      8
## 0.8840541 0.8934435 0.9028329 0.7619914 0.6587077 0.6399288

fit <- elman(inputs[1:1000,], outputs[1:1000,], size = c(2), learnFuncParams= c(0.1), maxit=500, shufflePatterns = FALSE, initFunc="JE_Weights", learnFunc = "JE_BP", )

#( 1.0 -1.0 0.3 1.0 0.5)
print(fit$initFuncParams)

## [1] 1.0 -1.0 0.3 1.0 0.5

summary(fit)

## network name : RSNNS_untitled
## source files :
## no. of units : 7
## no. of connections : 14
## no. of unit types : 0
## no. of site types : 0

## learning function : JE_BP
## update function : JE_Order
## unit definition section :
##
## no. | typeName | unitName | bias | st | position | act func | out func | sites
## 1 | | inp1 | 0.37987 | i | 1, 1, 0 | Act_Identity | |
## 2 | | inp2 | -0.06785 | i | 1, 2, 0 | Act_Identity | |
## 3 | | hid1 | -1.56180 | h | 7, 1, 0 |||
## 4 | | hid2 | -0.61517 | h | 7, 2, 0 |||
## 5 | | out1 | 23.11168 | o | 13, 1, 0 | Act_Identity | |
## 6 | | con1 | 0.50000 | sh | 4, 4, 0 | Act_Identity | |
## 7 | | con2 | 0.50000 | sh | 4, 5, 0 | Act_Identity |

##
## connection definition section :
##
## target | site | source:weight
## 3 | | 7: 1.74246, 6:-1.66811, 2:-0.58402, 1: 1.93551
## 4 | | 7: 1.50031, 6:-0.02226, 2:-2.59269, 1: 0.89746
## 5 | | 4:-1.62502, 3: 3.06378
## 6 | | 6: 0.30000, 3: 1.00000
## 7 | | 7: 0.30000, 4: 1.00000

```

```
an<- accuracy(as.ts(fit$fitted.values), outputs[1:1000])
```

```
print(weightMatrix(fit))
```

```
## inp1 inp2 hid1 hid2 out1 con1 con2
## inp1 0 0 1.9355056 0.89745975 0.000000 0.0 0.0
## inp2 0 0 -0.5840157 -2.59269142 0.000000 0.0 0.0
## hid1 0 0 0.0000000 0.00000000 3.063779 1.0 0.0
## hid2 0 0 0.0000000 0.00000000 -1.625018 0.0 1.0
## out1 0 0 0.0000000 0.00000000 0.000000 0.0 0.0
## con1 0 0 -1.6681117 -0.02226167 0.000000 0.3 0.0
## con2 0 0 1.7424634 1.50031424 0.000000 0.0 0.3
```

```
p <- getNormParameters(y)
```

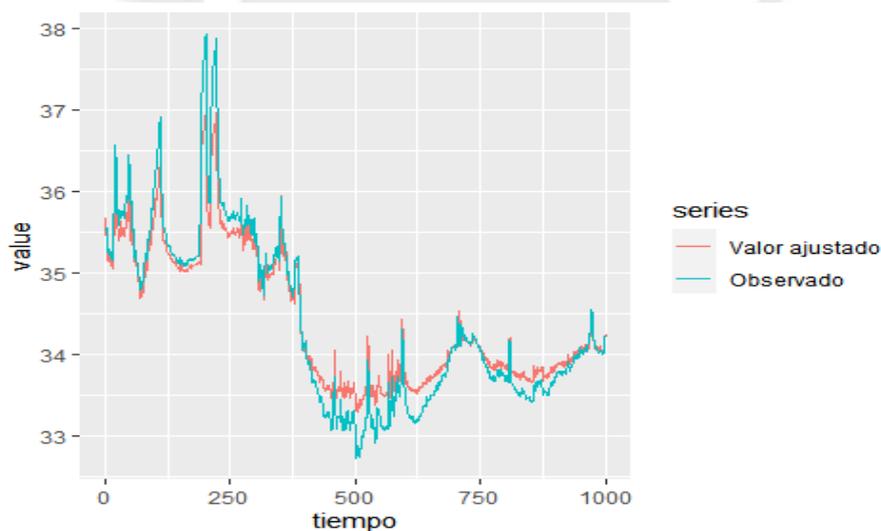
```
train_norm <- denormalizeData(as.ts(fit$fitted.values), p)
```

```
m <- data.frame(cbind(time=1:1000,round(train_norm,2), datos[1:1000]))
```

```
colnames(m)<- c("tiempo", "Valor ajustado", "Observado")
```

```
m<- melt(m,id.vars= 'tiempo', variable.name = 'series')
```

```
ggplot(m, aes(tiempo,value))+geom_line(aes(colour=series))
```



```
a<- accuracy(as.ts(train_norm), as.ts(datos[1:1000]))
```

```
#Predicción a 60 minutos
```

```
bn<-accuracy(as.ts(pred),outputs[1001:1060,1] )
```

```
p <- getNormParameters(y)
```

```
pred_norm <- denormalizeData(as.ts(pred), p)
```

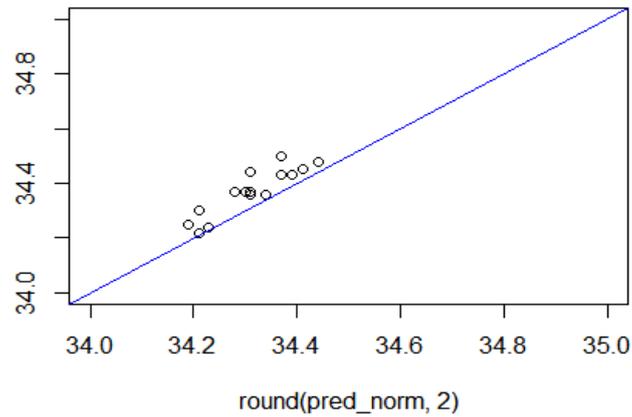
```
comparacion<- cbind(round(pred_norm,2), datos[1001:1060])
```

```
colnames(comparacion) <- c("Predicción", "Observado")
```

```
plot(round(pred_norm,2), datos[1001:1060], xlim=c(34,35), ylim=c(34,35))
```

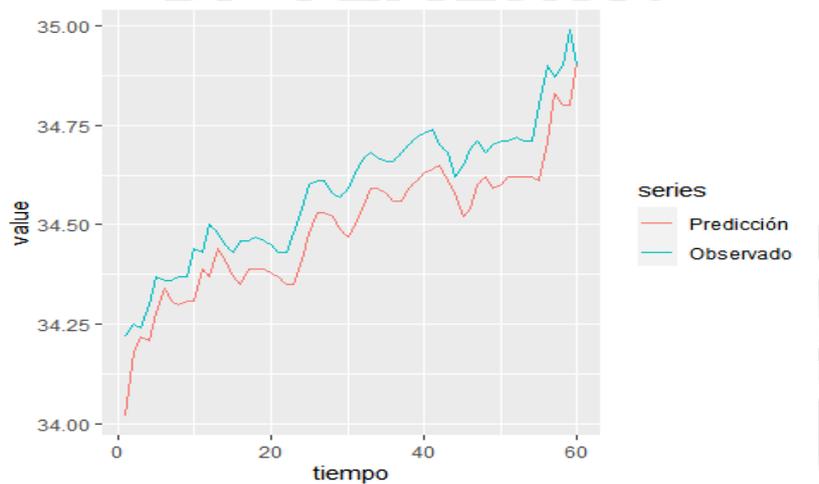
```
b<-accuracy(as.ts(pred_norm),as.ts(datos[1001:1060]))
```

```
abline(coef = c(0,1),col="blue")
```



#Plot

```
m <- data.frame(cbind(time=1:15,round(pred_norm,2), datos[1001:1060]))
```



Fit/ train / real

print(a)

```
##      ME  RMSE  MAE  MPE  MAPE  ACF1
## Test set -0.03344495 0.2589388 0.1971243 -0.1184409 0.5720603 0.8904173
##      Theil's U
## Test set 2.546008
```

#Fit 15

Fit pred/real 15

print(b)

```
##      ME  RMSE  MAE  MPE  MAPE  ACF1
## Test set 0.06021992 0.07042187 0.06021992 0.1750807 0.1750807 -0.2880276
##      Theil's U
## Test set 1.866346
```