INSTITUTO TECNOLÓGICO DE CIUDAD MADERO

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

DOCTORADO EN CIENCIAS DE LA INGENIERÍA

TESIS

MÉTODOS DE PRONÓSTICO DE SISTEMAS DINÁMICOS EN ESPACIO DE ESTADOS

Que para obtener el Grado de
Doctor en Ciencias de la Ingeniería

Presenta
M. C. I. E. José Christian de Jesús Galicia González
D20073015
No. CVU de CONACyT: 551032

Director de Tesis
Dr. Juan Frausto Solís
No. CVU de CONACyT: 31308

Co-director de Tesis
Dr. Juan Javier González Barbosa

Cd. Madero, Tamaulipas octubre 2025

Educación
Secretaría de Educación Pública

TECNOLÓGICO NACIONAL DE MÉXICO

Instituto Tecnológico de Ciudad Madero
Subdirección Académica
División de Estudios de Posgrado e Investigación

Ciudad Madero, Tamaulipas, 01/Septiembre/2025

Oficio No. U10.051/2025
ASUNTO: Autorización de Impresión

C. JOSÉ CHRISTIAN DE JESÚS GALICIA GONZÁLEZ
No. DE CONTROL D20073015
P R E S E N T E

Me es grato comunicarle que después de la revisión realizada por el Jurado designado para su Examen de Grado de Doctor en Ciencias de la Ingeniería, el cual está integrado por los siguientes catedráticos:

PRESIDENTE :	DR.	JUAN FRAUSTO SOLÍS
SECRETARIO:	DRA.	GUADALUPE CASTILLA VALDEZ
VOCAL 1:	DR.	JUAN JAVIER GONZÁLEZ BARBOSA
VOCAL 2:	DR.	RUBÉN SALAS CABRERA
VOCAL 3:	DR.	LUIS FORTINO CISNEROS SINENCIO
SUPLENTE:	DR.	LUCIANO AGUILERA VÁZQUEZ

Se acordó autorizar la impresión de su tesis titulada:

"MÉTODOS DE PRONÓSTICO DE SISTEMAS DINÁMICOS EN ESPACIO DE ESTADOS "

Es muy satisfactorio para esta División compartir con Usted el logro de esta meta. Espero que continúe con éxito su desarrollo profesional y dedique su experiencia e inteligencia en beneficio de México.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Por Mi Patria y Por Mi Bien" ®

RAFAEL CASTILLO GUTIÉRREZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN

EDUCACIÓN
TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE CIUDAD MADERO
DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

ccp. Archivo
RCG/NRV/jar

2025 Año de La Mujer Indígena

100% LIBRE DE PLÁSTICO DE UN SOLO USO

DECLARACIONES DE ORIGINALIDAD, PROPIEDAD INTELECTUAL, CESION DE DERECHOS Y/O CONFIDENCIALIDAD

Declaro que la investigación es original y este documento de tesis son producto de mi directa contribución intelectual ya que sus contenidos no infringen los derechos de terceros, tales como derechos de publicación, derechos de autor y similares.

Todos los datos y las referencias a materiales ya publicados están debidamente identificados con su respectivo crédito e incluidos en las notas bibliográficas y en las citas que se destacan como tal.

Por lo tanto, me hago responsable de cualquier infracción o reclamación relacionada con derechos de propiedad intelectual, exonerando de responsabilidad a mi director de tesis codirectores y al Tecnológico nacional de México, así como a al Instituto Tecnológico de Ciudad de México.

M.C.I.E. JOSÉ CHRISTIAN DE JESÚS GALICIA GONZÁLEZ

DEDICATORIA

Quiero dedicar este trabajo a mi esposa e hija, que, sin su apoyo incondicional, no sería posible alcanzar mis metas.

A mis padres de quienes aprendí esfuerzo y sacrificio.

A todos los miembros de mi comité que me retroalimentaron y estuvieron al pendiente de mi en cada etapa.

A mis directores que con su paciencia y guía me brindaron el soporte necesario para la realización de este trabajo.

Mis compañeros de generación, gracias por hacer amena mi estancia.

AGRADECIMIENTOS

Agradezco profundamente el apoyo económico por parte del gobierno de México a través de del Consejo Nacional de Humanidades, Ciencia y Tecnología (Conahcyt), así como al Instituto Tecnológico de Ciudad Madero (ITCM) y a su división de posgrado e investigación (DEPI).

A mis maestros que participaron directa o indirectamente en este trabajo.

MÉTODOS DE PRONÓSTICO DE SISTEMAS DINÁMICOS EN ESPACIO DE ESTADOS.

JOSÉ CHRISTIAN DE JESÚS GALICIA GONZÁLEZ

Resumen

Los métodos de pronóstico de sistemas dinámicos presentan una disminución en su desempeño conforme aumenta la complejidad de las series de tiempo. En este contexto, las competencias de Makridakis destacan la necesidad de un método general y efectivo.

Los métodos clásicos, como Holt-Winters, ARIMA y SSA, son efectivos para pronósticos a corto plazo, pero su precisión disminuye significativamente a mediano y largo plazo. Según el tamaño de las series de tiempo, se elige entre el análisis de componentes singulares (SSA) o técnicas de aprendizaje profundo, como las redes neuronales LSTM (Long Short Term Memory) y CNN (Convolutional Neural Network), dependiendo de su desempeño en la etapa de validación.

Aunque ambos enfoques son robustos, su desempeño es afectado por el ruido en las series de tiempo (Rosenberg y Maurer, 2008; Othman et al., 2021). Para mitigar este problema, se implementan dos enfoques de filtrado: el Promedio Móvil Simple (SMA, por sus siglas en inglés) y el Filtro de Kalman. Según las características de la señal filtrada, se selecciona el método más adecuado para el pronóstico inicial. Posteriormente, este pronóstico es refinado mediante un análisis de los datos residuales, obtenidos al restar la señal filtrada de la señal real. Los residuales se pronostican utilizando el enfoque más apropiado para sus características, y finalmente, este pronóstico residual se suma al pronóstico de la señal filtrada para generar el pronóstico final.

La generalidad del método se valida mediante pruebas en tres horizontes clave de la pandemia de COVID-19: el inicio, el pico de contagios y un año después del inicio. Estos experimentos se realizan en cuatro países de América: México, Estados Unidos, Brasil y Colombia.

Este enfoque permite dividir el problema en dos etapas gracias al filtrado, que no solo reduce el ruido, sino que también proporciona flexibilidad para seleccionar el método de pronóstico óptimo en cada caso. Como resultado, se logra una mejora en el RMSE (Raíz del Error Cuadrático Medio) del pronóstico en las tres etapas de la pandemia para los cuatro países analizados, en comparación con los métodos clásicos y el estado del arte.

FORECASTING METHODS FOR DYNAMIC SYSTEMS IN STATE SPACE

JOSÉ CHRISTIAN DE JESÚS GALICIA GONZÁLEZ

Abstract

Forecasting methods for dynamic systems experience a decline in performance as the complexity of time series increases. In this context, the Makridakis Competitions highlight the need for a general and effective method.

Classical methods, such as Holt-Winters, ARIMA, and SSA, are effective for short-term forecasts but lose accuracy significantly in medium- and long-term horizons. Depending on the size of the time series, either singular spectrum analysis (SSA) or deep learning techniques, such as LSTM (Long Short Term Memory) and CNN (Convolutional Neural Network), are chosen based on their validation performance.

Although both approaches are robust, their performance is affected by noise in the time series (Rosenberg and Maurer, 2008; Othman et al., 2021). To address this issue, two filtering techniques are implemented: Simple Moving Average (SMA) and the Kalman Filter. Depending on the characteristics of the filtered signal, the most suitable method is selected for the initial forecast. Subsequently, this forecast is refined by analyzing residual data, obtained by subtracting the filtered signal from the actual signal. The residuals are forecasted using the most appropriate method for their characteristics, and finally, this residual forecast is added to the filtered signal forecast to produce the final forecast.

The generality of the method is validated by testing it in three key phases of the COVID-19 pandemic: the onset, the peak of infections, and one year after the start of the pandemic. These experiments are conducted in four American countries: Mexico, the United States, Brazil, and Colombia.

This approach divides the problem into two stages through filtering, which not only reduces noise but also provides flexibility to select the optimal forecasting method in each case. As a result, an improvement in the RMSE (Root Mean Squared Error) of the forecast is achieved

in the three phases of the pandemic for the four analyzed countries, compared to classical methods and the state-of-the-art.

Índice General

Índice Tablas

Índice de Figuras

1 Introducción

El futuro puede entenderse como una consecuencia de una serie de eventos pasados. Bajo esta premisa, realizar predicciones no solo es razonable, sino también factible. Sin embargo, la tarea de pronosticar se vuelve progresivamente más compleja y menos precisa a medida que se incorporan variables adicionales, grados de libertad y no linealidades inherentes a la dinámica de los sistemas [21, 33].

El análisis de series de tiempo, que ha captado el interés de estadísticos, ingenieros, investigadores de operaciones y economistas durante las últimas décadas, se ha consolidado como una herramienta esencial en la generación de pronósticos [22]. Este campo resulta crucial debido a su capacidad para automatizar predicciones, integrando diversas características de los datos y permitiendo así la toma de decisiones informadas en múltiples áreas [36].

Por ejemplo, en la gestión de la cadena de suministro, el análisis de series de tiempo es indispensable para mantener pronósticos actualizados de la demanda. Esto permite planificar niveles óptimos de inventario, garantizando un desempeño que cumpla con los estándares de servicio al cliente. De manera similar, en el suministro de energía eléctrica, no basta con analizar una única variable; es necesario identificar patrones estacionales y otros factores dinámicos para realizar predicciones precisas [26]. En estos casos, los métodos basados en el espacio de estados han emergido como herramientas prometedoras para abordar la complejidad de los sistemas [7, 3].

La precisión en el pronóstico de series de tiempo, sin embargo, se ve comprometida por el ruido, el cual incluye tanto errores de medición como el ruido intrínseco del sistema [Mbaye et al. (2018), De Bézenac et al. (2020)]. Separar este ruido del comportamiento real del sistema es un paso crítico para obtener resultados precisos. Esto se logra mediante el análisis de las características inherentes de la serie, empleando herramientas como la autocorrelación

o métodos espectrales, para posteriormente aplicar técnicas específicas de pronóstico [Golyandina (2020), Kang et al. (2017)].

Los métodos de pronóstico de series de tiempo se utilizan para estimar valores pasados, presentes y futuros, facilitando un control más eficaz de los procesos. Si bien cada método tiene sus ventajas y limitaciones, la práctica ha demostrado que las soluciones híbridas, que combinan diferentes enfoques, suelen ser más efectivas. En este contexto, los métodos basados en espacio de estados han ganado protagonismo recientemente, mostrando un desempeño prometedor en la mejora de la precisión y la robustez de los pronósticos.

1.1 Estado del Arte

Con el objetivo de formular una hipótesis clara y específica, en esta sección se realiza un análisis detallado del estado del arte. Este análisis sirve como base para justificar el planteamiento del problema y contextualizar la contribución de esta investigación.

Las competiciones de Makridakis han sido fundamentales para establecer referencias objetivas sobre la efectividad de los métodos de pronóstico. Estas competencias proporcionan resultados experimentales comparativos entre técnicas actuales, destacando el mejor algoritmo según los requisitos de cada caso [Makridakis et al., 2020].

En la competición M1, se encontró que los métodos simples podían ser tan precisos como los más sofisticados.
La competición M2 demostró que la incorporación de información adicional o el juicio de expertos no necesariamente mejora la precisión de los pronósticos [Makridakis et al., 2018].
En la competición M3, se reafirmaron los hallazgos del M1 y se introdujeron métodos más precisos, mientras que la M4 destacó el valor de los enfoques híbridos, integrando algoritmos de aprendizaje automático para ajustar pesos y lograr una mayor precisión [Makridakis et al., 2020].

1.1.1 Métodos clásicos de pronóstico

Entre los enfoques clásicos destacan los modelos ARMA y ARIMA, reconocidos por su capacidad para manejar ruido en las series de tiempo mediante estimaciones iniciales y algoritmos recursivos [Pankratz, 1983; Baptista et al., 2018]. Por otro lado, los modelos de suavizado exponencial, propuestos por Robert G. Brown, son variantes de ARMA y son particularmente efectivos en series con tendencias y estacionalidades específicas [Svetunkov, 2017; de Livera et al., 2011].

Si bien estos métodos ofrecen mejoras significativas al incorporar elementos estocásticos y cálculos probabilísticos, su desempeño se ve limitado en escenarios con patrones estacionales complejos o múltiples variables, como en la predicción de la demanda eléctrica [Nagbe et al., 2018; López, 2015]. En estos casos, los modelos en espacio de estados han demostrado ser superiores, ya que consideran múltiples interacciones dentro de un sistema y proporcionan estimaciones más precisas [Mbaye et al., 2018; de Bézenac et al., 2020].

1.1.2 Métodos híbridos y combinados

Los métodos híbridos integran las fortalezas de diferentes técnicas, compensando sus debilidades. Por ejemplo, los modelos de suavizado exponencial pueden abordar tendencias, mientras que ARIMA maneja la estacionalidad [Smyl, 2020; Xu and Ren, 2019]. Otros enfoques, como los modelos combinados, no implican interacciones directas entre técnicas individuales, pero sus resultados finales son superiores [Bergmeir et al., 2016; Petropoulos et al., 2018].

En un caso exitoso, Lu et al. [2010] utilizaron modelos conmutados de Markov, basados en espacio de estados y ajustados mediante ARIMA, para analizar la expansión de enfermedades. Aunque efectivo, este enfoque presentó limitaciones frente a series no lineales como las observadas durante la pandemia de COVID-19. Para abordar estas no linealidades, Chimmula y Zhang [2020] aplicaron redes neuronales LSTM, demostrando su capacidad

para identificar patrones a lo largo del tiempo. Sin embargo, su desempeño disminuyó más allá de un horizonte de 17 días, principalmente debido a la limitada cantidad de datos.

Por su parte, Zain y Alturki [2021] emplearon redes neuronales convolucionales (CNN), las cuales reorganizan la dimensión temporal como espacial para identificar patrones. Aunque demostraron ser comparables a LSTM, se observó que CNN es más adecuada para series cortas, mientras que LSTM funciona mejor con datos más extensos [Zeroual et al., 2020].

Un enfoque notable es el de Frausto-Solís et al. [2021], quienes combinaron métodos clásicos y de aprendizaje profundo para un horizonte de 21 días. Este híbrido utilizó componentes semanales para mejorar la precisión en series ruidosas y se probó en México, Estados Unidos, Brasil y Colombia, logrando configuraciones óptimas específicas para cada país.

1.1.3 Enfoques en espacio de estados

Los métodos en espacio de estados permiten una representación más precisa de sistemas dinámicos al modelar múltiples variables simultáneamente. Estos enfoques han mostrado resultados superiores en problemas complejos donde los métodos clásicos o híbridos no logran un desempeño adecuado [Semenoglou et al., 2021]. Además, la combinación de filtros como el de Kalman y técnicas estocásticas ha permitido mitigar el ruido y mejorar la reconstrucción de las series de tiempo, ofreciendo un camino prometedor para el desarrollo de técnicas de pronóstico más robustas.

En síntesis, los métodos híbridos y combinados representan el estado del arte en pronósticos de series de tiempo, destacándose especialmente en sistemas complejos modelados en espacio de estados. Sin embargo, su efectividad depende de la integración adecuada de técnicas y del tratamiento del ruido, elementos clave para avanzar hacia soluciones más precisas y generalizables.

1.2 Hipótesis

Los métodos de pronóstico de sistemas dinámicos para series de tiempo en espacio de estados tienen un desempeño superior a los mejores métodos en el estado del arte.

1.3 Planteamiento del Problema

La mayoría de los sistemas dinámicos conocidos son no lineales, y su comportamiento puede ser tan caótico que predecirlo se convierte en un desafío significativo [Bianchi et al. (2017), Slawek Smyl (2020)]. Este reto se agrava por la inherente falta de fiabilidad en los datos obtenidos de una serie de tiempo, ya que suelen estar contaminados por errores de medición y ruido intrínseco del sistema como se aprecia en la Figura 1.3 [de Bézenac et al. (2020), Alzubaidi et al. (2021)].

Antes de identificar un modelo adecuado para el sistema, es indispensable realizar un análisis exhaustivo de la serie de tiempo. Métodos tradicionales como ARMA, suavizamiento exponencial y Box-Jenkins han demostrado ser útiles para ciertas aplicaciones, pero presentan limitaciones importantes al abordar dinámicas complejas o multivariables.

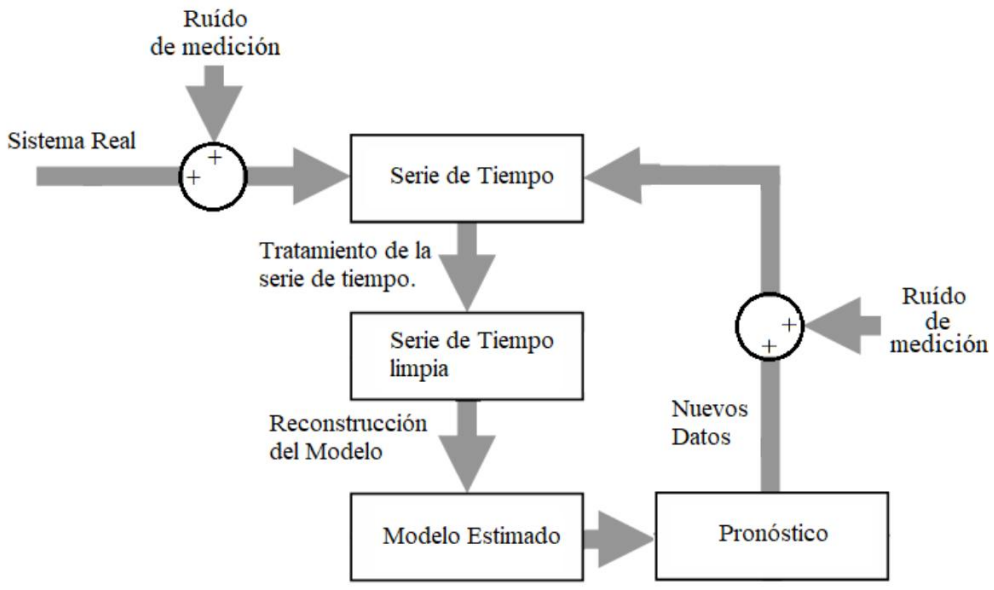

Figura 1.1 Diagrama a Bloques de un Pronóstico con ruido.

En muchos casos, no solo es necesario predecir una variable, sino también modelar la interacción entre múltiples variables. Para este propósito, se propone el Método de Análisis de Series de Tiempo por Espacio de Estados, que extiende las metodologías univariables para permitir predicciones multivariables. Este enfoque no solo realiza predicciones independientes para cada variable, sino que, también valida las interacciones entre ellas, ofreciendo una visión más integral del sistema.

Si bien el análisis en espacio de estados ofrece numerosas ventajas, su implementación requiere manejar un alto costo computacional, especialmente cuando se consideran múltiples variables simultáneamente. Este costo puede ser prohibitivo en aplicaciones que demandan respuestas en línea o en tiempo real, donde la precisión debe equilibrarse cuidadosamente con la velocidad de cálculo.

Para abordar esta problemática, los métodos recursivos como el suavizamiento exponencial y los filtros de Kalman han demostrado ser efectivos, proporcionando un compromiso razonable entre precisión y eficiencia computacional. Sin embargo, estas técnicas todavía enfrentan desafíos al manejar dinámicas complejas y datos ruidosos [**Klein, N., Smith, M. S., & Nott, D. J.** (2020)].

Con el objetivo de superar el desempeño de los modelos de pronóstico tradicionales y reducir los tiempos de estimación, se plantea la posibilidad de hibridar las metodologías clásicas con enfoques modernos de inteligencia artificial. Estas hibridaciones integrarían técnicas de aprendizaje automático para identificar patrones y comportamientos subyacentes en los datos, lo que permitiría enfocar de manera más precisa las estimaciones del filtro de Kalman.

El desarrollo de una metodología híbrida eficaz no solo mejoraría la precisión de las predicciones, sino que también optimizaría los recursos computacionales, garantizando una respuesta en tiempo real que cumpla con los requisitos establecidos. Este enfoque no solo contribuye al avance del campo de los pronósticos de series de tiempo, sino que también abre

nuevas posibilidades para abordar sistemas dinámicos complejos en diversas aplicaciones prácticas.

1.4 Justificación

Es fundamental desarrollar un modelo de pronóstico para series de tiempo que integre múltiples características del sistema analizado, tales como las interacciones entre variables, patrones estacionales y las posibles dinámicas no lineales presentes en los datos. Este enfoque debe ir más allá de la predicción univariante y permitir un análisis multivariable que capture las relaciones complejas entre los distintos componentes del sistema.

Para lograrlo, se requiere un método con la flexibilidad suficiente para manejar datos con alta dimensionalidad y dinámicas complejas sin generar cuellos de botella en términos de costo computacional. Esto es especialmente importante en aplicaciones donde se necesitan respuestas en tiempo real o predicciones a mediano y largo plazo, donde los métodos tradicionales suelen quedarse cortos en precisión y eficiencia.

El estado del arte actual evidencia que la combinación de enfoques clásicos con técnicas modernas, como los métodos de aprendizaje automático, proporciona resultados superiores en comparación con el uso aislado de cualquiera de ellos. Esta combinación permite aprovechar las fortalezas de ambos enfoques: la robustez y simplicidad de los métodos tradicionales y la capacidad de los modelos actuales para manejar grandes cantidades de datos y patrones no lineales.

Por lo tanto, es necesario desarrollar un método híbrido o combinado que no solo pueda abordar problemáticas multivariables, sino que también sea lo suficientemente versátil para adaptarse a diferentes características de las series de tiempo. Este tipo de metodología contribuiría al avance del campo de los pronósticos, permitiendo soluciones más precisas y aplicables a una amplia gama de sistemas dinámicos.

1.5 Objetivos

Objetivo General

Desarrollar un método de pronóstico de series de tiempo por espacio de estados que resuelva problemas donde: se necesite un enfoque multivariable, se necesite información extra sobre la estacionalidad de la serie de tiempo, rango de los datos, volatilidad de los datos, etc., combinando diversas metodologías de pronósticos para obtener un desempeño superior al visto en el estado del arte.

Objetivos específicos

1. Identificación de implementaciones exitosas mediante experimentación utilizando casos de prueba comúnmente utilizadas en la literatura especializada.

2. Diseño de un enfoque multivariable en espacio de estados de las implementaciones más exitosas según el horizonte de pronóstico, cantidad de datos presentes en la serie de tiempo, intervalos de tiempo, entre otras necesidades específicas.

3. Diseñar un algoritmo de pronóstico adaptable al ruido que provea un desempeño igual o superior al estado del arte.

4. Lograr un tratamiento efectivo contra el ruido presente en las series de tiempo.

5. Escritura de artículos para publicación en revistas indexadas.

1.6 Metodología Experimental

La experimentación se realizará en el clúster EHECATL que se encuentra físicamente en el LANTI (Laboratorio Nacional de Tecnologías de la Información) del Campus 3 del Instituto Tecnológico de Ciudad Madero, utilizando software libre como SciLab, Python, o R, entre otros.

Los pasos propuestos por (Montgomery entre otros, 2008) para el proceso de pronóstico de series de tiempo son:

1. Definición del problema: Este paso consisten en entender cómo será utilizando el pronóstico con las expectativas de cliente. Se deben de realizar preguntas tales como el horizonte de proyección o el tiempo de espera, los intervalos de tiempo y el nivel de precisión del pronóstico.

2. Recolección de datos: Consisten en obtener datos históricos relevante de las variables que se desean pronosticar. A menudo es necesario tratar con valores faltantes de algunas variables.

3. Análisis de los datos: Es un paso preliminar para la selección del modelo a utilizar pues es importante identificar en las series de tiempo los componentes de la tendencia, patrones estacionales y cíclicos.

4. Selección de modelo y ajuste: Consiste en seleccionar, uno o más modelos de pronósticos, así como el ajuste del modelo de los datos.

5. Validación del modelo: Consisten la evaluación del modelo de pronósticos para determinar el desempeño para una determinada aplicación.

6. Implementación del modelo de pronóstico: Implica la ejecución del modelo y los resultados del pronóstico usados por el cliente.

7. Monitoreo del desempeño del modelo de pronóstico: Debe ser una actividad que se realice cuando el modelo ha sido utilizado, para asegurarse que se está ejecutando de manera satisfactoria, la naturaleza del pronóstico es que las condiciones cambian sobre el tiempo y el modelo que se ejecutó bien en el pasado podrá alterar su desempeño.

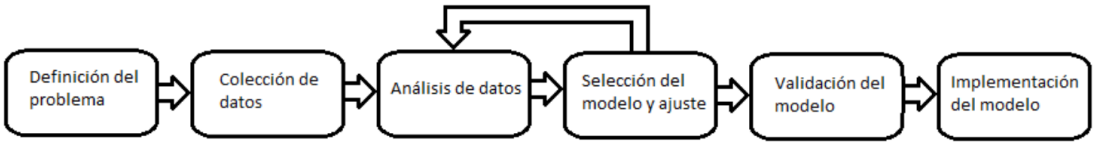

Figura 1.2 El proceso del pronóstico.

1.7 Resultados Esperados

El desarrollo del algoritmo de pronóstico por series de tiempo en espacio de estados utilizando Filtro de Kalman tendrá resultados similares o superiores a los que se reportan en el estado del arte. De esta forma se espera obtener un método novedoso que permita realizar pronósticos de cambio climático, Epidemiológicos, financieros, de cadena de suministros, etc. a corto, mediano y largo plazo. Adicionalmente se espera que, durante el periodo de esta investigación, al menos dos de las publicaciones realizadas y producto de la tesis a desarrollar sean aceptadas para su publicación en revistas indexadas.

1.8 Estructura de la tesis

La presente tesis está compuesta por 6 capítulos ordenados adecuadamente para que el lector pueda comprender los pasos tomados para resolver la hipótesis de esta tesis.

En el Capítulo 2 se presentan los filtros de Simple Media Móvil y de Kalman utilizados como auxiliares en disminución de la complejidad de la serie de tiempo, facilitando la construcción de los métodos presentados en esta tesis.

En el Capítulo 3 se presenta un método de Análisis de Espectro Singular (SSA) automático basado en el desempeño de validación con análisis de residuales.

En el Capítulo 4 se presentan las metodologías más exitosas de aprendizaje profundo: CNN y LSTM, así como sus principales ventajas y desventajas en el uso de pronóstico de series de tiempo.

En el Capítulo 5 se presentan métodos Híbridos desarrollados para este trabajo de tesis: 1. FMFRA: Método de Pronóstico con Filtros y Análisis de Residuales por sus siglas en inglés. 2. H&W con Filtro de Kalman.

En el Capítulo 6 se presentan las conclusiones de esta tesis.

Capítulo 2

2 Filtrado de Series de Tiempo

El pronóstico de series de tiempo caóticas ha despertado un interés creciente en años recientes debido a su alta complejidad, marcada no estacionalidad y considerable volatilidad. Estas características hacen que no exista un método universalmente superior para su pronóstico. En este contexto, los enfoques que combinan o hibridan diferentes técnicas han demostrado un desempeño significativamente mejor que los métodos tradicionales.

De acuerdo con el estado del arte, el desempeño de los métodos de pronóstico está intrínsecamente ligado a la complejidad de la serie de tiempo. Esta complejidad se define como una medida del nivel de dificultad requerido para describir o predecir las propiedades de un sistema [López-Ruiz et al., 1995]. Factores clave que influyen directamente en esta complejidad incluyen:

Frecuencia de muestreo: El espaciamiento temporal de los datos puede variar desde segundos hasta años. Una frecuencia de muestreo más amplia (por ejemplo, anual) puede provocar una pérdida significativa de información, mientras que una frecuencia más alta aumenta la cantidad de datos disponibles, pero también incrementa la complejidad del sistema debido a la mayor densidad de puntos a analizar. Este balance entre pérdida de información y complejidad computacional es crítico en la selección de la frecuencia de muestreo adecuada.

Ruido en la serie de tiempo: El ruido puede manifestarse en forma de datos faltantes, valores atípicos, errores de captura o limitaciones en los métodos de medición que originan la serie

de tiempo. Este ruido introduce incertidumbre en las predicciones y complica la modelización precisa del sistema. La presencia de ruido en una serie de tiempo puede ser modelada matemáticamente mediante la Ecuación 2.1:

$$Y_t = X_t + R_t \qquad \text{Ecuación 2.1}$$

Donde:

Y_t: Es un vector con las observaciones de la serie de tiempo.

X_t: Es un vector con los estados sin ruido de una serie de tiempo.

R_t: Son los ruidos asociados a los estados de la serie de tiempo.

El ruido en una serie de tiempo es indeseable, ya que el propósito principal de un pronóstico es anticipar los estados futuros del sistema de la manera más precisa posible, y no simplemente predecir las observaciones futuras de la serie de tiempo, las cuales incluyen ruido.

Para mitigar los efectos del ruido en las series de tiempo, se emplean diferentes tipos de filtros de señales. La elección del filtro depende de las características del ruido presente y las propiedades de la serie de tiempo. A continuación, se describen las categorías principales:

Filtros Simples

Estos filtros son directos y efectivos para escenarios donde el ruido tiene características bien definidas, como un rango de frecuencias conocido.

SMA (Simple Media Móvil): Calcula el promedio de un número fijo de puntos consecutivos, actuando como un filtro pasa bajas que suaviza las fluctuaciones de corto plazo en la serie.
Suavizado Exponencial: Pone mayor peso en observaciones recientes para capturar tendencias y reducir el impacto del ruido.

En aplicaciones donde los datos provienen de sensores, el ruido de medición generalmente se encuentra dentro de un rango de frecuencias conocidas. Esto permite el uso de filtros simples, como los mencionados, sin necesidad de análisis estadísticos más complejos.

Filtros Estocásticos

Estos filtros realizan un análisis distributivo de los datos para eliminar valores que no pertenecen a la distribución esperada del sistema. Algunos ejemplos destacados son:

Filtro de Box-Cox: Normaliza las series de tiempo y reduce el sesgo de los datos [Bergmeir et al., 2016].

Análisis de Espectro Singular (SSA): Descompone la serie de tiempo en componentes principales para identificar y eliminar ruido [Othman et al., 2021].

Filtro de Kalman: Evalúa estadísticamente las observaciones para determinar cuánto de la señal es ruido y cuánto es real, ofreciendo una estimación óptima del estado del sistema [Alferov y Klimova, 2020; Assimakis et al.].

En este estudio, se analizarán las implicaciones de ambos tipos de filtros con énfasis en Simple Media Móvil (SMA): Actuando como filtro pasa bajas, se utiliza para atenuar las fluctuaciones de alta frecuencia, que suelen representar ruido, y resaltar las tendencias subyacentes en la serie de tiempo. Y Filtro de Kalman: Este método evalúa estadísticamente las observaciones de la serie, diferenciando entre la señal real y el ruido. Su capacidad para ajustarse dinámicamente a nuevas observaciones lo convierte en una herramienta poderosa para el pronóstico de sistemas dinámicos.

2.1 Filtros de Simple Media Móvil

En señales eléctricas donde el ruido está caracterizado y se conoce su frecuencia, se utilizan filtros de frecuencia para eliminar componentes no deseadas. Por ejemplo, si el ruido

tiene una frecuencia alta, se emplea un filtro pasa bajas para discriminar las señales con frecuencias superiores a un umbral especificado.

Estos filtros pasan bajas funcionan mediante la acumulación de datos. En sistemas eléctricos, esta acumulación es lograda mediante un capacitor, mientras que, en series de tiempo, el acumulador puede representarse mediante el Simple Media Móvil (SMA). En casos donde la frecuencia del ruido está bien definida, el Simple Media Móvil Ponderada permite ajustar con mayor precisión la frecuencia de corte deseada.

El Simple Media Móvil consiste en tomar un número definido de datos, n, hacia atrás en la serie, calcular su promedio y asignar este valor como el nuevo dato en la serie filtrada. Este proceso acumula los valores y se resiste a cambios abruptos hasta n pasos atrás, actuando como un filtro pasa bajas que atenúa las fluctuaciones rápidas en la señal.

La representación matemática del filtro SMA está dada por la Ecuación 2.2:

$$SMA_{ni} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \text{Ecuación 2.2}$$

Donde:

SMA_{ni}: Es el ith elemento de la nueva serie de tiempo filtrada

n: Es el número de datos a promediar.

y_i: Es el ith elemento de la serie de tiempo.

A medida que el valor de n aumenta, más armónicos de alta frecuencia son filtrados, lo que permite separar tendencias y componentes estacionales principales.

Un caso práctico del uso del filtro SMA se muestra en la Figura 2.1, donde se analizan series de tiempo correspondientes a los nuevos casos diarios de COVID-19 en cuatro países del continente americano. Los datos fueron filtrados utilizando $n = 3$,5,7, y 100. Los mejores resultados en el pronóstico final se obtuvieron con $n = 3$ y $n =$7, por las siguientes razones:

SMA7: Al tomar el promedio de siete días, captura efectivamente la estacionalidad semanal de la serie de tiempo, suavizando los picos abruptos.
SMA3: Refleja la dinámica de fines de semana, reduciendo los efectos de los picos asociados a días festivos o anomalías.

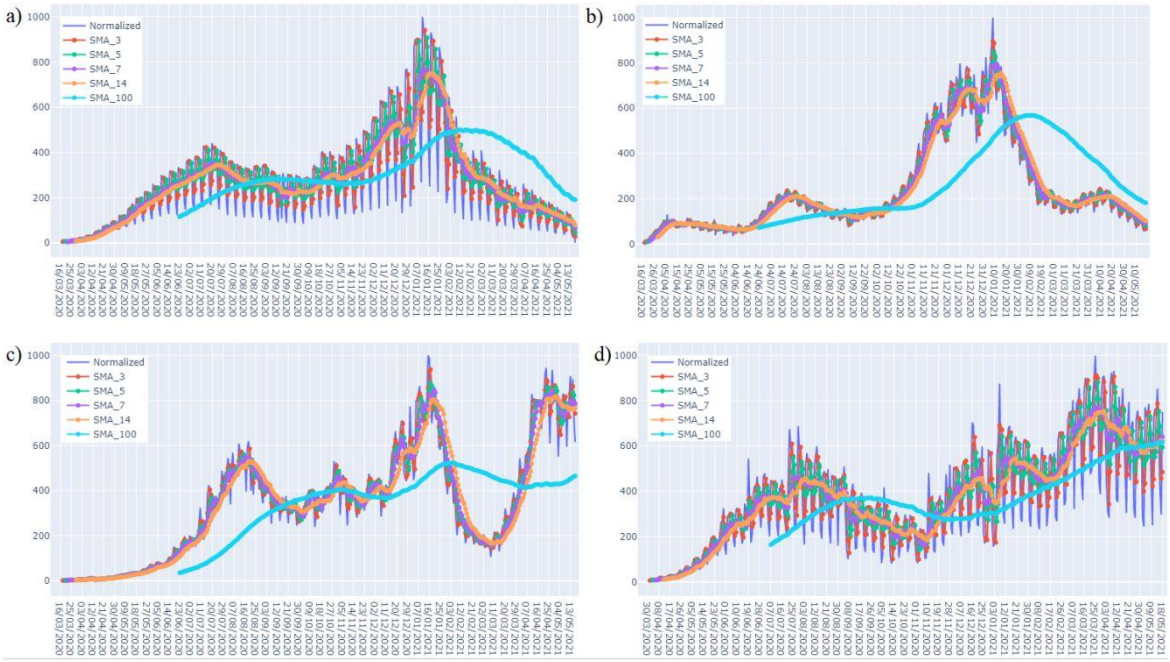

Figura 2.1 Filtros *SMA* de nuevos casos de COVID-19 en: a) México, b) EEUU, c) Colombia, d) Brasil.

En la Figura 2.1, los datos normalizados de cada país se presentan en color azul para ilustrar las dinámicas de los casos diarios entre países con diferentes poblaciones. Las señales filtradas se representan con los siguientes colores: rojo para n=3, verde para n=5, morado para n=7, naranja para n=14 y azul celeste para n=100.

La normalización de los datos permite comparar las dinámicas entre países, mostrando cómo los filtros SMA atenúan el ruido y resaltan las tendencias relevantes.

A medida que el valor de n aumenta, se observa un desfase temporal característico del proceso de acumulación de información en un único punto. Este desfase puede limitar la

utilidad del filtro en ciertas aplicaciones, especialmente en pronósticos en línea donde la información debe ser procesada en tiempo real.

Para abordar el problema de datos faltantes, es factible utilizar SMA centrados, los cuales toman en cuenta tanto datos anteriores como posteriores al punto en análisis. Sin embargo, este enfoque tiene una limitación significativa: requiere información futura, lo que lo hace inviable para aplicaciones en tiempo real donde los datos hacia adelante no están disponibles.

En el análisis de series temporales de países americanos, se observa una relación más estrecha entre países vecinos debido a dinámicas similares en sus patrones sociales, económicos o sanitarios. Por ejemplo, México y Estados Unidos de América presentan series de tiempo con alta similitud, por otro lado, Colombia y Brasil comparten características similares en sus series temporales. Sin embargo, al incrementar el valor de n, las diferencias locales tienden a suavizarse, y las series de países más distantes comienzan a asemejarse. Esto permite a los algoritmos de aprendizaje profundo identificar patrones generales, eliminando fluctuaciones específicas de cada país y favoreciendo la extracción de características globales.

2.2 Filtro de Kalman

El Filtro de Kalman es una herramienta matemática sumamente poderosa para el control de sistemas con ruido significativo en las mediciones. Su capacidad radica en combinar diversos fundamentos matemáticos, como se ilustra en la Figura 2.2. Esta combinación de principios lo hace excepcional para estimar el valor siguiente en una serie temporal, incluso en condiciones de alta incertidumbre.

El Filtro de Kalman no solo filtra el ruido, sino que también ofrece una base matemática robusta para predecir estados futuros. Esto lo convierte en una herramienta clave en el análisis de sistemas dinámicos y series temporales donde los datos están sujetos a ruido de medición o condiciones altamente variables.

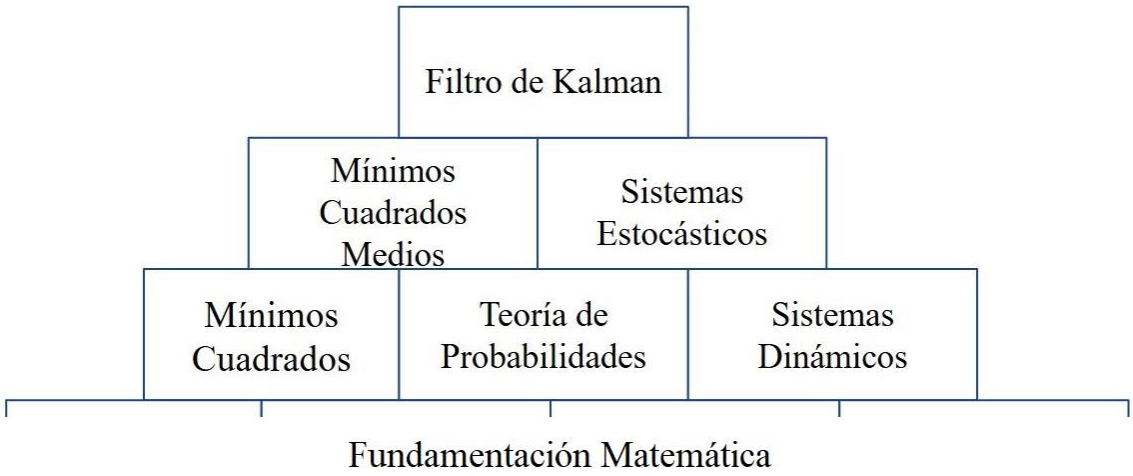

Figura 2.2 Fundamentación del Filtro de Kalman.

El Filtro de Kalman es un algoritmo que permite estimar los estados de un sistema mediante una ponderación estadística entre la predicción proporcionada por el modelo matemático del sistema y las observaciones previas. Este enfoque lo hace particularmente útil en sistemas donde las mediciones están contaminadas por ruido y es necesario obtener estimaciones precisas en tiempo real.

Originalmente diseñado para sistemas discretos, el Filtro de Kalman utiliza un modelo del sistema definido por las Ecuaciones de Estado (2.3) y la Ecuación de Medida (2.4):

$$\alpha_k = T_k\alpha_{k-1} + R_k\eta_k \qquad \text{Ecuación 2.3}$$
$$\eta_k \sim N(0, Q_k)$$

Donde:

α_k: Vector de estados actual.

α_{k-1}: Vector de estados pasado.

T_k: Matriz característica del sistema, describe cómo los estados evolucionan en el tiempo.

R_k: Matriz que determina cómo el ruido afecta los estados.

η_k: Vector de ruido en las entradas, que sigue una distribución normal con media cero y matriz de covarianzas Q_k.

$$Y_k = Z_k \alpha_{k-1} + \varepsilon_k \qquad \text{Ecuación 2.3}$$
$$\varepsilon_k \sim N(0, H_k)$$

Donde:

Y_k: Vector de medidas, que corresponde a los datos observados en la serie de tiempo.

Z_k: Matriz que selecciona los estados relevantes para la salida.

η_k: Vector de ruido de medición, que también sigue una distribución normal con media cero y matriz de covarianzas H_k.

La estimación de los estados se realiza utilizando un observador de orden completo, el cual considera tanto las observaciones disponibles como la dinámica del sistema modelada por las ecuaciones anteriores. Este proceso se ilustra en la Figura 2.3, donde se representa gráficamente la interacción entre las predicciones del modelo y las correcciones basadas en las observaciones.

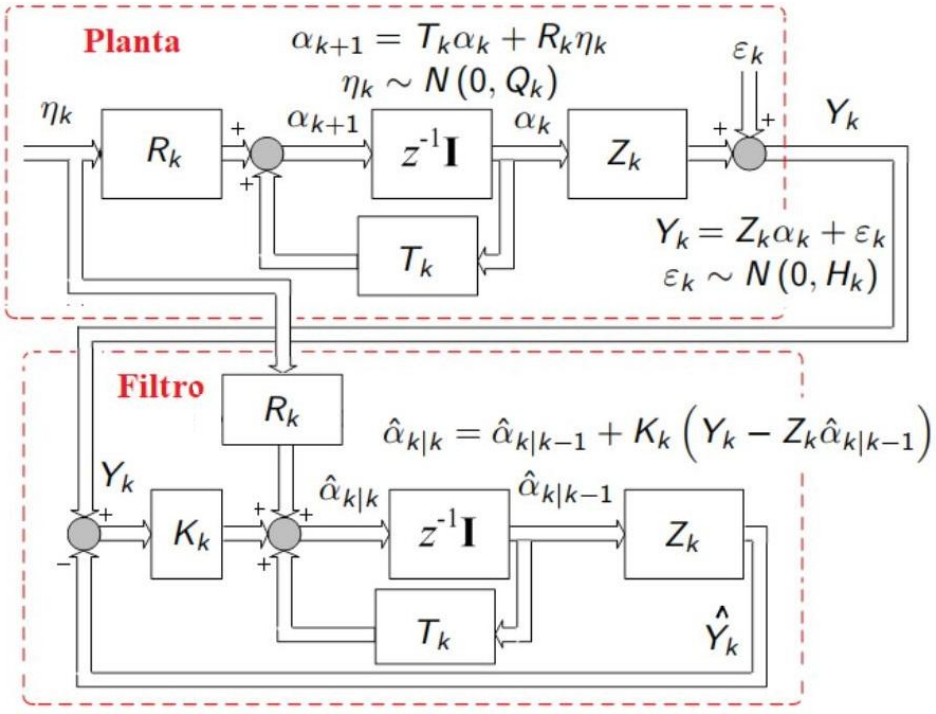

Figura 2.3 Estimador del Filtro de Kalman.

Un observador de orden completo consiste en duplicar el modelo matemático del sistema cuya entrada será el vector de medida Y_k y se expresa mediante la Ecuación 2.5.

$$\hat{\alpha}_{k|k} = \hat{\alpha}_{k|k-1} + K_k(T_k - Z_k\hat{\alpha}_{k|k-1}) \qquad \text{Ecuación 2.5}$$

Donde:

$\hat{\alpha}_{k|k}$: Es la estimación del vector de estados en el tiempo k debido a la observación en el tiempo k.

$\hat{\alpha}_{k|k-1}$: Es la estimación del estado en el tiempo k debido a la observación $k-1$.

K_k: Matriz de ganancias de Kalman.

Y_k: Vector de medidas, que corresponde a los datos observados en la serie de tiempo.

De este modo, el Filtro de Kalman estima el vector de estados actual al actualizar la predicción previa del vector de estados con la medida de la salida. Este proceso se realiza mediante una ponderación conocida como Ganancia de Kalman (K_k), la cual ajusta el balance entre la predicción del modelo y las observaciones disponibles. La Ganancia de Kalman se calcula teniendo en cuenta el ruido presente en las mediciones y las características del modelo del sistema, permitiendo reducir la incertidumbre del vector de estados estimado [de Bézenac et al., 2020].

Tal como se ilustra en la Figura 2.3, este proceso iterativo consiste en:

Predicción: El sistema utiliza las ecuaciones de estado para estimar el vector de estados y la incertidumbre asociada (covarianza).

Actualización: Una vez que se obtiene una nueva medida de la salida, se utiliza la Ganancia de Kalman para ajustar la predicción, disminuyendo la incertidumbre en el estado estimado y mejorando la precisión general.

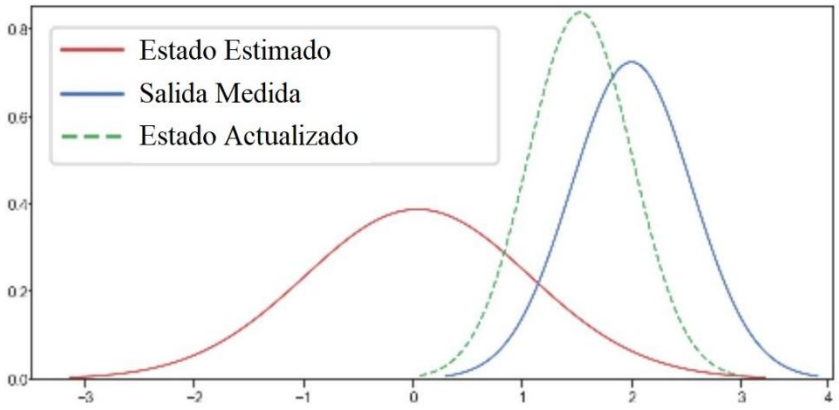

Figura 2.4 Estimador del Filtro de Kalman.

La Figura 2.4 representa gráficamente cómo la incertidumbre asociada a la predicción del estado disminuye tras incorporar la nueva medida de la salida. Este ajuste continuo es el núcleo del algoritmo de Kalman y lo que lo hace eficaz para sistemas dinámicos con ruido e incertidumbre.

Este proceso de actualización y corrección se realiza de manera iterativa, como se muestra en la Figura 2.3, permitiendo una estimación continua y en tiempo real de los estados del sistema.

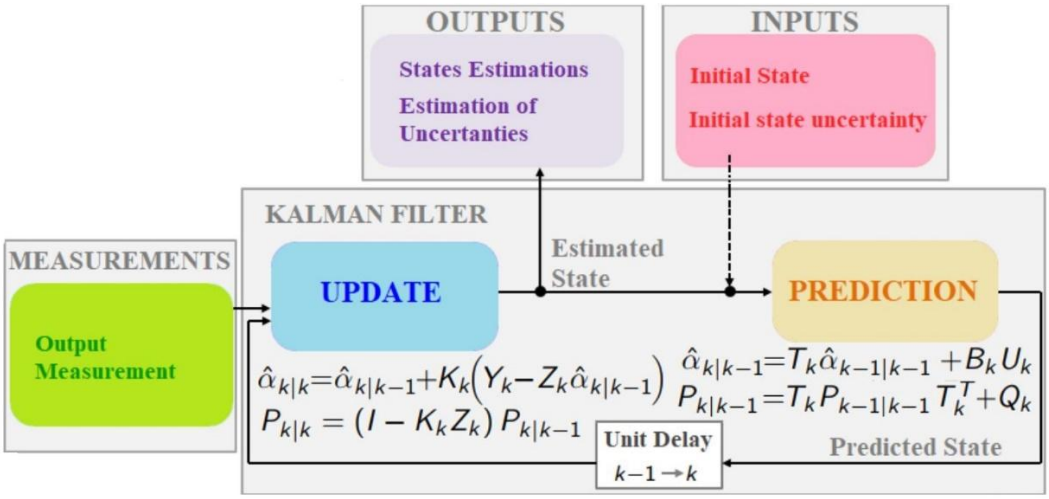

Figura 2.5 Estimador del Filtro de Kalman.

El Filtro de Kalman consta de dos fases fundamentales que se ejecutan de manera iterativa: Predicción y Actualización. Estas fases permiten estimar el estado real de un sistema

dinámico junto con su incertidumbre, minimizando la influencia del ruido de medición (Q) y del ruido intrínseco del sistema (H) en la señal filtrada.

Fase 1: Predicción

Para realizar la predicción inicial, es necesario definir el vector de estados iniciales (α_0) y su matriz de varianzas y covarianzas (P_0). A partir de estas condiciones iniciales, se calcula la predicción del estado en el instante k utilizando la siguiente ecuación:

$$\hat{\alpha}_{k|k-1} = T_k\ \hat{\alpha}_{k-1|k-1} + B_k U_k) \qquad \text{Ecuación 2.6}$$

Donde:

$\hat{\alpha}_{k|k-1}$: Predicción del vector de estados en el instante k basada en el estado previo k-1.

$\hat{\alpha}_{k-1|k-1}$: Es la estimación del estado anterior $k-1$ debido a la observación $k-1$.

T_k: Matriz de transición de estados, que define el comportamiento del sistema.

B_k: Matriz de selección de entradas del sistema.

U_k: Vector de entradas del sistema.

Simultáneamente, se actualiza la matriz de varianzas y covarianzas asociada a la predicción mediante la ecuación:

$$P_{k|k-1} = T_k\ P_{k-1|k-1} T_k^T + Q_k) \qquad \text{Ecuación 2.7}$$

Donde:

$P_{k|k-1}$: Estimación de la matriz de covarianzas en el instante k basada en el instante k-1.

$P_{k-1|k-1}$: Matriz de covarianzas en el instante $k-1$.

Q_k: Matriz de varianzas y covarianzas del ruido del modelo matemático.

Fase 2: Actualización

En la fase de actualización, se incorpora la nueva medición (Y_k) al modelo para ajustar la predicción del estado. La estimación del estado actualizado se calcula con la Ecuación 2.5.

La Ganancia de Kalman (K_k) se calcula con:

$$K_k = P_{k-1|k-1} Z_k^T (Z_k P_{k|k-1} Z_k^T + H_k)^{-1} \quad \text{Ecuación 2.8}$$

Donde:

H_k: Es la matriz de varianzas y covarianzas del ruido de medición.

Finalmente, la matriz de incertidumbre se actualiza según:

$$P_{k|k} = (I - K_k Z_k) P_{k|k-1} \quad \text{Ecuación 2.9}$$

Donde:

$P_{k|k}$: Es la matriz de varianzas y covarianzas actualizada.

I: Es una matriz identidad de las mismas dimensiones que $P_{k|k}$.

El Filtro de Kalman es una herramienta matemática poderosa para estimar estados y pronosticar series de tiempo, siempre que se disponga de un modelo matemático confiable del sistema. Sin embargo, en muchos casos prácticos, como el análisis de pandemias, no se cuenta con un modelo predefinido. En estos escenarios, se puede tratar el modelo matemático como una caminata aleatoria, simplificada en las siguientes ecuaciones:

$$\alpha_{k+1} = T_k \alpha_{k-1} + R_k \eta_k \text{ con } \eta_k \sim N(0, Q_k) \quad \text{Ecuación 2.10}$$

$$y_k = Z_k \alpha_k + \varepsilon_k \text{ con } \varepsilon_k \sim N(0, H_k) \quad \text{Ecuación 2.11}$$

Donde:

α_{k+1}: Vector de estados en el siguiente instante.

R_k: Matriz de selección de entradas.

η_k: Ruido del sistema, con distribución normal.

y_k: Medición del vector de estados.

ε_k: Ruido de medición

Aunque utilizar un modelo simplificado como una caminata aleatoria puede disminuir la precisión del filtro como pronosticador, esta simplificación permite separar la serie de tiempo en componentes de menor complejidad, facilitando el análisis de tendencias y estacionalidades. La Figura 2.6 muestra cómo el filtro de Kalman, aplicado como caminata

aleatoria unitaria, suaviza la serie de tiempo normalizada (en azul), produciendo una señal filtrada (en rojo) que es robusta al ruido.

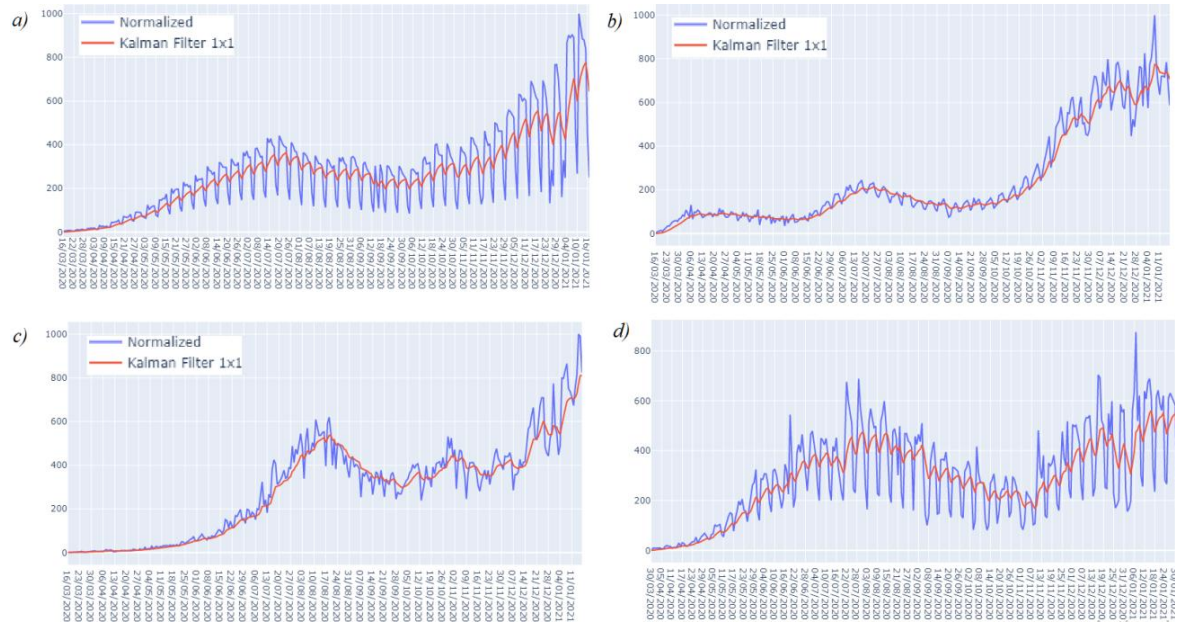

Figura 2.6 Filtro de Kalman para nuevos casos de COVID-19 en a) México, b) Estados Unidos de América, c) Colombia y d) Brasil.

En el caso de estudio de la pandemia de COVID-19, donde no existían precedentes claros, el filtro de Kalman permitió dividir la serie en componentes más manejables, destacando patrones generales y tendencias.

2.3 Conclusiones.

En este capítulo se analizaron los filtros utilizados en esta tesis, cuyo propósito principal es dividir la serie de tiempo en dos componentes: Serie Filtrada y Serie de Residuales. Este enfoque permite una mejor comprensión y manejo de la complejidad inherente a las series temporales.

La Serie Filtrada reduce la complejidad al eliminar ruido y preservar elementos clave como la tendencia y los armónicos principales. Esto facilita la extracción de patrones fundamentales, lo que resulta crucial para mejorar la calidad de los pronósticos en sistemas dinámicos. Por otro lado, al restar la señal filtrada de la señal original, se obtiene la Serie de

Residuales, caracterizada por la ausencia de tendencia y una mínima presencia de estacionalidad. Esta señal residual concentra armónicos menores y ruido que, aunque no son predominantes, pueden contener información útil para ajustar los pronósticos finales.

Este proceso de división ofrece ventajas significativas:

- Pronóstico de señales con poco ruido:

  La señal filtrada, al ser más limpia, permite aplicar métodos de pronóstico clásicos con sintonizaciones específicas para extraer patrones claros y confiables.
- Optimización de los residuales:

  La señal residual, que conserva información más compleja como armónicos pequeños y ruido, se beneficia de métodos especializados en separar componentes sutiles de la serie de tiempo.

Ambos enfoques permiten ajustar las técnicas de pronóstico a las características particulares de cada componente, mejorando así la precisión general del modelo. En capítulos posteriores, específicamente en el Capítulo 4, se analiza en detalle la selección de métodos de pronóstico y las sintonizaciones empleadas para cada una de estas series.

Capítulo 3

3 Análisis de Espectro Singular

El Análisis de Espectro Singular (SSA, por sus siglas en inglés) es una técnica ampliamente utilizada para extraer componentes principales, tanto en imágenes como en series de tiempo. Es, en esencia, un método de filtrado que permite aislar la señal real a partir de los datos medidos. Antes de profundizar en los métodos de pronóstico aplicados a señales filtradas, exploraremos cómo realizar un pronóstico de series de tiempo utilizando SSA.

La aplicación de SSA comienza con la Hankelización de la serie de tiempo, que implica organizar los datos en una matriz de Hankel. Este tipo de matriz, nombrada en honor a Hermann Hankel, es una matriz simétrica donde las diagonales paralelas (de izquierda a derecha) contienen elementos numéricamente iguales, según lo descrito por Golyandina [2020].

3.1 Entrenamiento

El proceso de entrenamiento en SSA consiste en transformar una serie de tiempo de longitud N en una matriz de trayectorias $\boldsymbol{X}$, con L columnas y $K = N - L + 1$ filas. Esto se realiza a través de la Ecuación (3.1):

$$\mathbf{X} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_L \\ x_2 & x_3 & x_4 & \cdots & x_{L+1} \\ x_3 & x_4 & x_5 & \cdots & x_{L+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \cdots & x_N \end{bmatrix} \in \mathbb{R}^{K,L} \qquad \text{Ecuación 3.1}$$

Donde:

$\boldsymbol{X}$: Es llamada matriz trayectoria.

x_i: Es el i-ésimo valor de la serie de tiempo.

N: Es el total de datos en la serie de tiempo.

L: Es el ancho de ventana definido para la serie.

K: Es el número de ventanas calculadas, dado por $K = N - L + 1$.

Dado que la matriz $\boldsymbol{X}$ no es necesariamente cuadrada, no tiene autovalores en su forma original. Sin embargo, cualquier matriz rectangular $\boldsymbol{H} \in \mathbb{R}^{K,L}$ admite una Descomposición en Valores Singulares (SVD, por sus siglas en inglés):

$$\mathbf{X} = \boldsymbol{U\Sigma V^T} \qquad \text{Ecuación 3.2}$$

Donde:

$\boldsymbol{X}$: Es llamada matriz trayectoria.

$\boldsymbol{U} \in \mathbb{R}^{K,K}$: Es una matriz ortonormal de dimensiones $K \times K$.

$\boldsymbol{\Sigma} \in \mathbb{R}^{L,K}$: Es una matriz diagonal que contiene los valores singulares.

$\boldsymbol{V^T} \in \mathbb{R}^{L,L}$: Es una matriz ortonormal de dimensiones $L \times L$.

Los valores singulares, σ, se encuentran en la diagonal principal de la matriz $\boldsymbol{\Sigma}$ y se calculan a partir de los autovalores de la matriz cuadrada $\boldsymbol{X^T X}$ mediante las siguientes ecuaciones:

$$\sigma = \sqrt{\lambda} \qquad \text{Ecuación 3.3}$$

Donde:

$\sigma \in \mathbb{R}^{L}$: Vector de valores singulares, ordenados de mayor a menor.

$\lambda \in \mathbb{R}^{L}$: Vector de autovalores de la matriz cuadrática $\boldsymbol{X^T X}$.

Los autovalores, λ, se obtienen resolviendo la ecuación característica:

$$|\boldsymbol{X^T X} - \lambda \boldsymbol{I}| = 0 \qquad \text{Ecuación 3.4}$$

Donde:

$\boldsymbol{X^T X} \in \mathbb{R}^{L,L}$: Es la forma cuadrática de la matriz trayectoria X.

$\boldsymbol{I} \in \mathbb{R}^{L}$: Matriz identidad.

El tercer elemento, $\boldsymbol{V^T}$, corresponde a la matriz de autovectores unitarios asociados a los autovalores de $\boldsymbol{X^T X}$. Para calcular el primer elemento, $\boldsymbol{U}$, se utiliza la relación ortonormal de las matrices en la descomposición:

$$\boldsymbol{XV = U\Sigma} \qquad \text{Ecuación 3.5}$$

En muchos casos, se tiene que $K > L$. En esta situación, la matriz $\boldsymbol{\Sigma}$ contendrá al menos $K - L$ vectores columna nulos:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_L & \cdots & 0 \end{bmatrix} \qquad \text{Ecuación 3.6}$$

La obtención de la matriz $\boldsymbol{U}$ en la descomposición en valores singulares (SVD) requiere dos fases principales, debido a que algunos vectores asociados a los valores singulares son nulos. Estas fases son:

Fase 1: Cálculo de los primeros L vectores fila

Dado que al menos L valores singulares (σL) son diferentes de cero, es posible calcular los primeros L vectores fila de $\boldsymbol{U}$ usando la siguiente ecuación:

$$u_L = \frac{1}{\sigma_L} \boldsymbol{X} v_L \qquad \text{Ecuación 3.7}$$

Donde:

u_L: Es el L-ésimo vector fila de la matriz $\boldsymbol{U}$.

σ_L: Es el L-ésimo valor singular correspondiente.

Xv_L: Es el L-ésimo vector fila obtenido al multiplicar la matriz $\boldsymbol{X}$ por v_L.

Este cálculo asegura que los vectores fila de $\boldsymbol{U}$ asociados a los valores singulares no nulos están correctamente definidos.

Fase 2: Completar $\boldsymbol{U}$ con vectores ortogonales.

A partir del L-ésimo vector fila, los valores singulares restantes son nulos (σ_L+1,…, σ_K = 0), lo que implica que no hay contribuciones adicionales de $\boldsymbol{\Sigma}$ en esos componentes. Por lo tanto, para completar $\boldsymbol{U}$, se deben encontrar $K-L$ vectores fila adicionales que cumplan con la condición de ortonormalidad:

$$\boldsymbol{W} \in \mathbb{R}^{K,K-L} = span(u_L)^{\perp} \qquad \text{Ecuación 3.8}$$

Donde:

$\boldsymbol{W}$: Representa los $K - L$ vectores fila unitarios que son ortogonales a los vectores definidos en la Fase 1.

$span(u_L)^{\perp}$: Es el espacio ortogonal generado por los vectores u_L.

Con esto, $\boldsymbol{U}$ queda completamente definido, manteniendo su propiedad de ortonormalidad.

Una vez que se ha obtenido la descomposición en valores singulares de $\boldsymbol{X}$, cuyos valores singulares (σ_L) están ponderados de mayor a menor en la diagonal principal de $\boldsymbol{\Sigma}$, es posible reconstruir la matriz $\boldsymbol{X}$ utilizando únicamente un subconjunto de estos valores singulares.

Para reconstruir $\boldsymbol{X}$ con r valores singulares ($r \leq L$), la idea es excluir los valores singulares asociados al ruido y conservar únicamente aquellos que representan las características principales de la serie de tiempo, como la tendencia y la estacionalidad. Esto permite:

1. Reducción de ruido:
   Al omitir los valores singulares más pequeños, se eliminan componentes irrelevantes que pueden estar asociadas a fluctuaciones no significativas en la serie.

2. Conservación de información clave:
   Los valores singulares mayores representan las componentes principales de la señal, asegurando que se preserven los patrones relevantes.

3.2 Realización del Pronóstico

Una vez obtenida la reconstrucción de la serie de tiempo a partir de los valores singulares seleccionados, se procede al pronóstico. Este se realiza comúnmente utilizando un modelo de regresión lineal basado en la matriz trayectoria $\boldsymbol{A}$, el vector pendiente p, y el vector objetivo b, como se define en la Ecuación (3.9):

$$\boldsymbol{A}p = b \qquad \text{Ecuación 3.9}$$

Donde:

$\boldsymbol{A}$: Es la matriz trayectoria $\boldsymbol{X}$**,** truncada eliminando su última columna.

p: Vector columna de pendientes, de dimensiones $L \times 1$.

b: Vector columna correspondiente a los valores observados, truncado de la matriz $\boldsymbol{X}$.

Estas matrices y vectores se definen como:

$$\boldsymbol{A} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_{L-1} \\ x_2 & x_3 & x_4 & \cdots & x_L \\ x_3 & x_4 & x_5 & \cdots & x_{L+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \cdots & x_{N-1} \end{bmatrix}, p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \vdots \\ p_L \end{bmatrix}, b = \begin{bmatrix} x_L \\ x_{L+1} \\ x_{L+2} \\ \vdots \\ x_N \end{bmatrix} \qquad \text{Ecuación 3.10}$$

Para calcular p, se sustituye A por su descomposición en valores singulares (SVD):

$$\boldsymbol{U\Sigma V^T}p = b$$

Premultiplicando ambos lados por la pseudo-inversa de Moore-Penrose ($\boldsymbol{H}^{\dagger}$), definida como $\boldsymbol{H}^{\dagger} = \boldsymbol{V\Sigma^{-1}U^T}$, se obtiene:

$$\boldsymbol{V\Sigma^{-1}U^TU\Sigma V^T}p = \boldsymbol{V\Sigma^{-1}U^T}b$$

Dado que $\boldsymbol{U}$ y $\boldsymbol{V}$ son ortonormales ($\boldsymbol{U^TU = I, \Sigma^{-1}\Sigma = \mathrm{I}, \mathrm{V}V^T = I}$), la solución estimada de p es:

$$p = \boldsymbol{V\Sigma^{-1}U^T}b \qquad \text{Ecuación 3.11}$$

Debido al ancho de ventana L podemos tener dos casos:

1. **Subdeterminado** ($K < L$)**:**

   No hay suficientes observaciones en b para determinar una solución única en p. Esto lleva a soluciones múltiples o aproximadas.

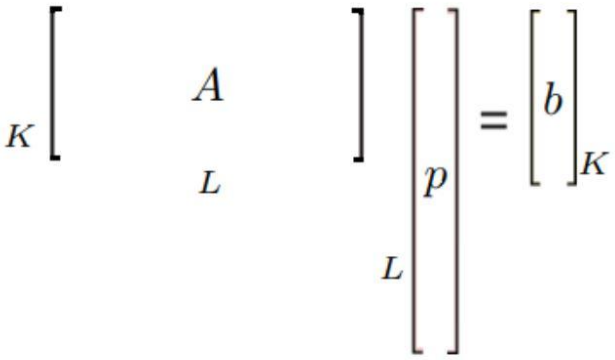

Figura 3.1 Caso subdeterminado de $\boldsymbol{A}p = b$

2. **Sobredeterminado** ($K > L$)**:**

Hay más observaciones que parámetros a ajustar, lo que permite obtener una solución única mediante mínimos cuadrados.

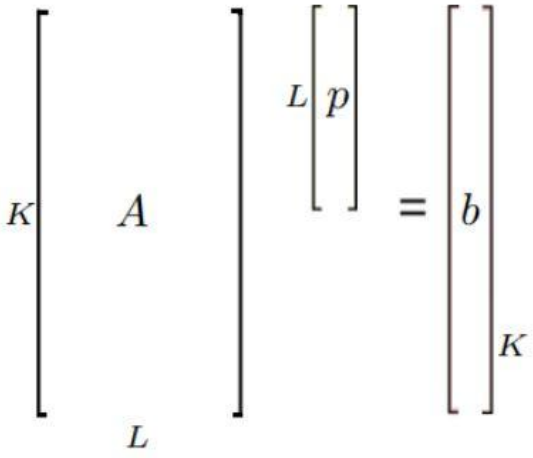

Figura 3.2 Caso sobredeterminado de $\boldsymbol{A}p = b$

Una vez calculado el vector pendiente p, se actualiza la matriz $\boldsymbol{A}$ para formar una nueva matriz $\boldsymbol{A}_{new}$. Esta matriz se construye desplazando las columnas de $\boldsymbol{A}$ hacia la izquierda y colocando el vector de resultados b en la última columna. Entonces se procede con el cálculo del nuevo vector b cuyo único valor nuevo será el último, esto significa que por cada operación se logra pronosticar un paso adelante.

En la Figura 3.3 se muestra el proceso de inicialización de la matriz $\boldsymbol{A}_{new}$, en el inciso a) se muestra la matriz inicial que consiste en devolver el vector de resultados, truncado anteriormente, a la matriz y al realizar la multiplicación con el vector pendiente p se obtiene la primer estimación del vector de resultados b.

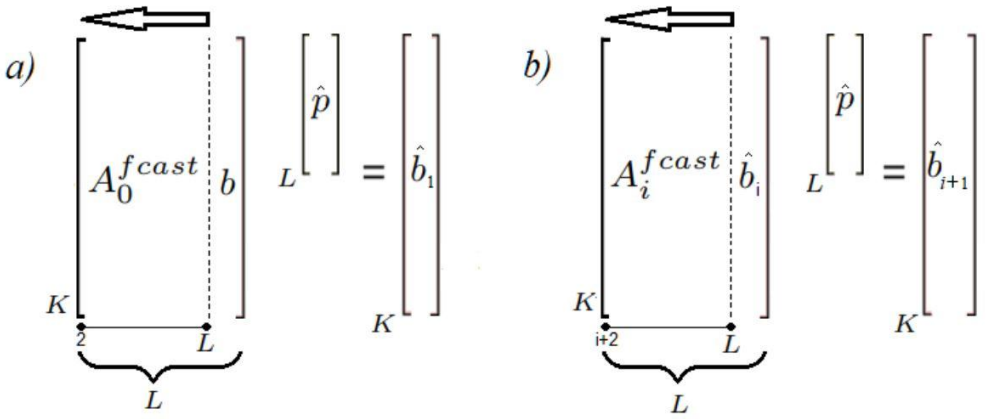

Figura 3.3 Construcción de la matriz $\boldsymbol{A}_0^{fcast}$

De igual forma en la Figura 3.3 en el inciso b) se muestra el procedimiento para extender el horizonte de pronóstico más allá de un paso, es necesario actualizar iterativamente la matriz $\boldsymbol{A}_{new}$. Este proceso consiste en actualizar la última columna de $\boldsymbol{A}_{new}$ del paso anterior con el vector b, pronosticado en el paso anterior, y sustituirlo en la última columna de la matriz $\boldsymbol{A}_{new}$, luego se desplazan el resto de las columnas hacia la izquierda para realizar el siguiente pronóstico:

$$b_i^{fcast} = \boldsymbol{A}_{i-1}^{fcast} p \qquad \text{Ecuación 3.12}$$

Donde:

b_i^{fcast}: Vector pronosticado en la iteración i, donde el último valor representa el pronóstico de un paso adelante en esa iteración.

$\boldsymbol{A}_{i-1}^{fcast}$: Matriz utilizada en la iteración $i-1$ para calcular b_i^{fcast}. Esta matriz se actualiza desplazando las columnas y reemplazando la última con b_{i-1}^{fcast}.

p: Mejor aproximación de la pendiente, calculada previamente mediante la pseudo-inversa de la matriz trayectoria $\boldsymbol{X}$.

3.3 Validación y Prueba

Para evaluar el desempeño del Análisis de Espectro Singular (SSA), se utiliza una serie de tiempo dividida en entrenamiento, validación, y prueba. Los datos de entrenamiento ($x_{train,k}$) se Hankelizan para formar la matriz $\boldsymbol{X_{train}}$, que entrena al modelo. El objetivo es realizar un pronóstico ($\hat{x}_{val,k}$) mediante los siguientes pasos:

1. Separación de los datos en entrenamiento y validación.

   Se elige una separación adecuada de los conjuntos de entrenamiento y validación para medir el desempeño del modelo.

   La Figura 3.4 muestra cómo se realiza esta separación, el propósito de esta configuración es encontrar el vector de pendientes p a partir de los datos de entrenamiento, cuyo desempeño se evaluará con los datos de validación.

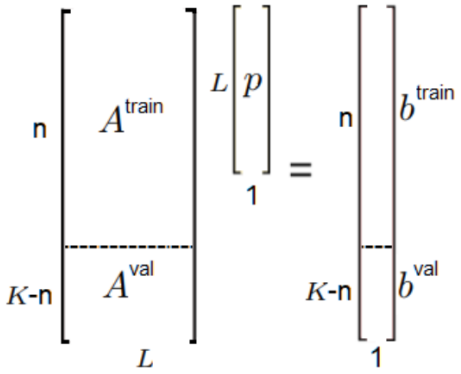

Figura 3.4 Matrices de entrenamiento y validación en formato para regresión lineal.

2. Entrenamiento del Modelo.

   A partir de los datos de entrenamiento, se calcula el vector pendiente siguiendo estos pasos:

- Formulación de $\boldsymbol{A_{train}}$ y b_{train}:

$$\boldsymbol{A}_{train}p = b_{train} \Rightarrow \boldsymbol{U}_{train}\boldsymbol{\Sigma}_{train}{V^T}_{train}x = b_{train} \quad \text{Ecuación 3.13}$$

- Cálculo de p:

El vector pendiente se calcula utilizando la pseudo-inversa de Moore-Penrose, como se muestra en la Ecuación (3.14):

$$p = \boldsymbol{V}_{train}{\boldsymbol{\Sigma}^{-1}}_{train}{U^T}_{train}b_{train} \quad \text{Ecuación 3.14}$$

2. Validación del Modelo

Con p obtenido, se realiza el pronóstico de validación utilizando los datos de validación ($\boldsymbol{A}_{val}$):

- Pronóstico de validación:

$$\boldsymbol{A}_{val}p = \hat{b}_{val} \quad \text{Ecuación 3.15}$$

- El desempeño del modelo se mide calculando el RMSE entre $\hat{b}_{val}$ y los valores reales de validación b_{val}.

3. Optimización de Parámetros

Para optimizar el desempeño del SSA se deben optimizar los parámetros de acuerdo al tamaño del conjunto de datos, en este trabajo utilizaremos el ejemplo de pronóstico de nuevos casos de COVID-19, como los datos con los que se probó estaban limitados a poco más de un año, se contaba con aproximadamente 400 datos, para esta aplicación se prueban diferentes configuraciones de los siguientes parámetros:

- Ancho de ventana L. Se prueba con valores de L desde 7 hasta 40, ya que para $L > 40$ no se observó mejora significativa en el pronóstico.
- Número de valores singulares r. En la descomposición, el rango del vector de valores singulares es:

$$range(\sigma_L) = L \quad \text{Ecuación 3.16}$$

- Para el pronóstico, se utilizan $r \leq L$ valores singulares:

$$range(\sigma_r) \leq range(\sigma_L)$$ Ecuación 3.17

- Se prueban diferentes valores de r, desde 1 hasta L, y se mide el RMSE para identificar el r con mejor desempeño.

4. Selección de los Mejores Parámetros.

- Una vez completado el horizonte de pronóstico para todos los valores de L y r, se selecciona la combinación de L y r que minimiza el RMSE.
- El diagrama de bloques del proceso se muestra en la Figura 3.5.

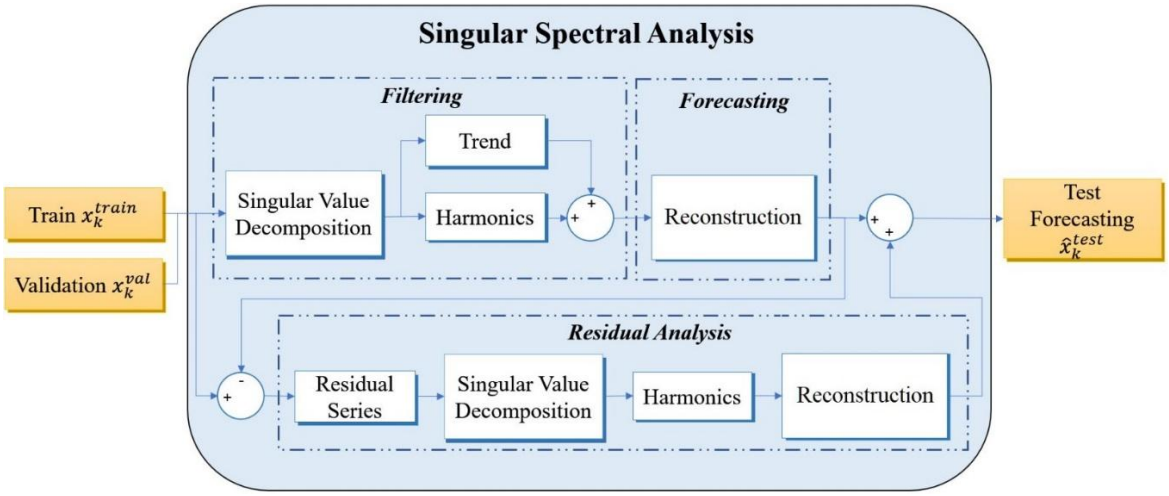

Figura 3.5 Modelo 1: Análisis de Espectro Singular.

Una vez identificado el mejor modelo en la etapa de validación, se procede a realizar el pronóstico de prueba. Este proceso incluye la combinación de los conjuntos de entrenamiento y validación en una única serie de tiempo, la cual se Hankeliza utilizando el ancho de ventana L previamente optimizado durante la etapa de validación.

En esta metodología de Análisis de Espectro Singular (SSA), se realiza un análisis adicional mediante la reconstrucción de la serie de tiempo con los parámetros L y r obtenidos durante

la validación. La serie reconstruida ($\hat{x}_k$) se resta de la serie original (x_k) para generar la serie de residuales (e_k):

$$e_k = x_k - \hat{x}_k \qquad \text{Ecuación 3.18}$$

Donde:

e_k: Serie de tiempo de residuales.

x_k: Concatenación de las series

$\hat{x}_k$: Reconstrucción de la serie de tiempo con los valores L y r optimizados.

La serie de residuales e_k representa los componentes de alta frecuencia (armónicos) y ruido, ya que los elementos de tendencia fueron eliminados durante la reconstrucción. A esta serie se le aplica nuevamente la metodología del SSA para identificar y modelar los armónicos presentes.

1. Valores Singulares de los Residuales:
   Los valores singulares obtenidos en este paso no contienen tendencia; están enfocados únicamente en los armónicos y el ruido.

2. Reconstrucción de los Residuales:
   La reconstrucción incluye únicamente los armónicos más relevantes, descartando el ruido no significativo.

El pronóstico final combina los componentes reconstruidos de tendencia y armónicos de la serie original con los armónicos identificados en el análisis de residuales. Este proceso garantiza que el modelo capte tanto las características principales de la serie como los detalles adicionales encontrados en los residuales.

Para realizar el pronóstico de prueba, se utiliza nuevamente la formulación de regresión lineal con la matriz actualizada y el vector pendiente p, aplicando el modelo sobre los datos concatenados de entrenamiento y validación.

3.4 Conclusiones

El método de Análisis de Espectro Singular (SSA) empleado en este capítulo destaca por su capacidad de realizar un análisis detallado de componentes principales basado en el mejor desempeño obtenido durante la etapa de validación. Este enfoque asegura capturar de manera efectiva las dinámicas más recientes exhibidas por la serie de tiempo, maximizando la utilidad de los valores singulares seleccionados.

Una de las principales ventajas del SSA es su capacidad para descomponer la serie en componentes de tendencia, estacionalidad y ruido. Sin embargo, los valores singulares son discretos, lo que puede dejar fuera ciertas dinámicas relevantes que no se ajustan a la distribución predominante. En este contexto, el análisis de residuales juega un papel fundamental al permitir un ajuste adicional al pronóstico final, incorporando armónicos que no fueron capturados en la primera etapa de reconstrucción.

Aunque el SSA es una herramienta robusta para el análisis y filtrado de series de tiempo, su principal limitación radica en el uso de una regresión lineal simple para el pronóstico. Este enfoque, aunque eficiente, restringe su capacidad para anticiparse a cambios abruptos en la serie. Métodos más avanzados, como los basados en aprendizaje profundo, podrían complementar al SSA para abordar estas dinámicas esporádicas, mejorando así la precisión y robustez del modelo.

Capítulo 4

4 Aprendizaje Profundo

El aprendizaje profundo ha demostrado ser altamente eficaz en el manejo de problemas complejos donde las soluciones evolucionan dinámicamente con el tiempo. En este capítulo, se explorarán dos de los métodos más avanzados basados en Redes Neuronales Recurrentes (RNN, por sus siglas en inglés): las Redes LSTM (Long Short-Term Memory) y las Redes Neuronales Convolucionales (CNN, por sus siglas en inglés). Estos métodos han mostrado resultados sobresalientes en aplicaciones como:

- Predicción del mercado de valores [Moghaddam et al., 2016; López, 2015].
- Pronósticos meteorológicos [Alferov y Klimova, 2020].
- Seguimiento de casos de COVID-19 [Frausto-Solís et al., 2021].

Las Redes LSTM y CNN son modelos poderosos que destacan en un amplio rango de tareas de aprendizaje automático supervisadas y no supervisadas. En particular, son altamente eficaces para problemas donde los datos y las características de las series de tiempo no son directamente interpretables, ya que aprenden representaciones jerárquicas de los datos [Alzubaidi et al., 2021].

La principal ventaja del aprendizaje profundo es su capacidad para modelar sistemas no lineales mediante la inclusión de pequeñas no linealidades en las funciones de activación de cada neurona. Esto permite una mejor aproximación del modelo real del sistema. Con el aumento en el tamaño de las bases de datos y los avances en cómputo paralelo, los métodos

de aprendizaje profundo pueden aprovechar eficientemente los recursos computacionales actuales, superando las limitaciones de métodos más simples, que suelen sufrir de subajuste al no capturar la complejidad de los datos [Bianchi et al., 2017].

4.1 Notación

En un proceso de pronóstico de series de tiempo con RNN, la entrada se define como una secuencia de datos ponderados, donde los valores más recientes tienen mayor relevancia para predecir el siguiente paso. La notación es la siguiente:

- Secuencia de entrada: $x(1),\ x(2), \dots, x(T)$, donde cada x(t) es un vector de valores reales.
- Secuencia de objetivo: $y(1),\ y(2), \dots, y(T)$

Un conjunto de entrenamiento se compone de pares de secuencias de entrada y objetivo. Cuando estas secuencias son finitas, el índice de tiempo máximo se denota como T.

Una red neuronal consiste en un conjunto de neuronas artificiales, también llamadas nodos o unidades, y los enlaces dirigidos entre ellas, que representan sinapsis en una red biológica. Cada entrada de la neurona (j') está ponderada por un peso ($w_{jj'}$). La salida de la neurona es el resultado de aplicar una función de activación ($\ell(.)$) a la suma ponderada de las entradas.

El valor de activación v_j de una neurona j se calcula como:

$$v_j = \ell_j\left(\sum_{j'} w_{jj'} . v_{j'}\right) \qquad \text{Ecuación 4.1}$$

Donde:

v_j: Es el valor de activación de la neurona j.

$v_{j'}$: Es el valor de activación de la neurona de entrada j'.

$w_{jj'}$: Es el peso entre las neuronas j y j'.

ℓ_j: Es la función de activación.

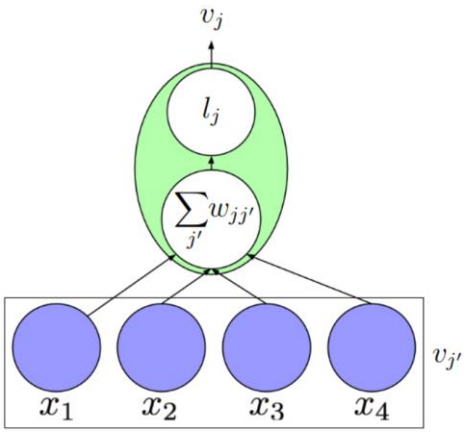

Figura 4.1 Modelo 1: Neurona Artificial.

La Figura 4.1 muestra una representación esquemática de una neurona artificial, destacando las conexiones de entrada, los pesos asociados, y la función de activación que determina la salida. Este modelo es la base para arquitecturas más avanzadas como las LSTM y CNN.

El proceso de aprendizaje en redes neuronales consiste en la actualización iterativa de los pesos para minimizar una función de pérdida $L(\hat{y}, y)$, la cual penaliza la diferencia entre la salida estimada ($\hat{y}$) la salida real (y). Esta optimización se realiza con el objetivo de mejorar la precisión del modelo al ajustar los parámetros de la red.

El algoritmo más ampliamente utilizado para entrenar redes neuronales es la retropropagación. Este método emplea la regla de la cadena para calcular la derivada de la función de pérdida con respecto a los parámetros de la red, permitiendo ajustar los pesos mediante gradiente descendente.

Sin embargo, debido a la naturaleza no convexa de la superficie de la función de pérdida, no hay garantía de alcanzar el mínimo global. En cambio, el proceso puede converger hacia un mínimo local. Este problema, intrínsecamente complejo y clasificado como NP-duro, ha impulsado el desarrollo de numerosos enfoques heurísticos para preentrenamiento y

optimización, los cuales han demostrado resultados impresionantes en diversas tareas de aprendizaje supervisado.

En la actualidad, las redes neuronales son entrenadas comúnmente utilizando el gradiente descendente estocástico (SGD, por sus siglas en inglés) en combinación con mini lotes. Cuando el tamaño del lote es igual a uno, la ecuación de actualización del gradiente estocástico se expresa como:

$$w \leftarrow w - \eta \nabla_w F_i \qquad \text{Ecuación 4.2}$$

Donde:

η: Es la tasa de aprendizaje, que determina el tamaño de los pesos.

$\nabla_w F_i$: Es el gradiente de la función objetivo con respecto a los parámetros w.

El SGD permite un entrenamiento más eficiente al calcular actualizaciones de los pesos basadas en subconjuntos aleatorios de los datos, reduciendo significativamente los costos computacionales y acelerando la convergencia.

4.2 Desvanecimiento del Gradiente

Una de las principales desventajas de las RNN tradicionales es la aparición de dos problemas comunes durante el entrenamiento: el Desvanecimiento del Gradiente y el Gradiente Explosivo (Vanishing Gradient y Exploding Gradient). Estos problemas surgen durante la retropropagación, cuando la derivada de la función de activación alcanza valores extremos, ya sea muy cercanos a cero o demasiado grandes. Esto provoca:

1. Desvanecimiento del Gradiente:
   El error asociado a la neurona es tan pequeño que su contribución al ajuste de los pesos en capas anteriores se reduce significativamente, ralentizando el entrenamiento.

2. Gradiente Explosivo:
   Por el contrario, si el gradiente es muy grande, los ajustes de los pesos se vuelven descontrolados, lo que puede provocar inestabilidad en el modelo.

En la retropropagación, el cálculo del gradiente se realiza utilizando la regla de la cadena. Cuando las derivadas parciales de las funciones de activación, como la tangente hiperbólica (tanh) o el seno hiperbólico (sinh), toman valores cercanos a sus extremos (0 o 1), el entrenamiento puede volverse:

- Demasiado lento: Debido al desvanecimiento del gradiente.
- Errático e inestable: Por el gradiente explosivo.

La Figura 4.2, que muestra cómo la derivada de tanh disminuye drásticamente a medida que los valores se acercan a sus límites.

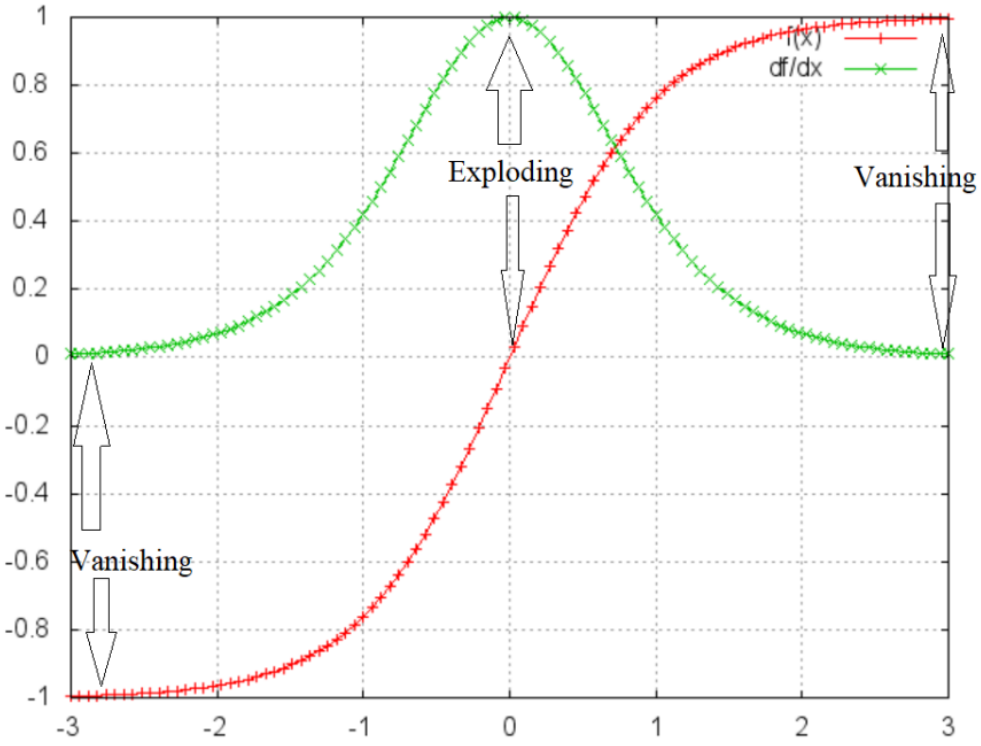

Figura 4.2. Derivada parcial de la función de activación de tangente hiperbólica.

Aunque cambiar la función de activación por una lineal podría mitigar estos problemas, este cambio elimina la capacidad de las RNN para modelar sistemas no lineales, una de sus principales ventajas.

Para abordar estos problemas, surgieron arquitecturas más avanzadas y estables, como las LSTM y las CNN, diseñadas específicamente para manejar el desvanecimiento y explosión del gradiente [Smyl, 2020, Talagala et al., 2018].

- LSTM (Long Short-Term Memory):

  Esta arquitectura resuelve el desvanecimiento del gradiente al mantener un flujo constante de error durante la retropropagación. Lo logra mediante puertas de entrada, salida y olvido, que permiten un control más preciso de la información retenida o descartada [Bianchi et al., 2017].

- CNN (Redes Neuronales Convolucionales):

  Las CNN identifican características relevantes de los datos de entrada mediante convoluciones jerárquicas. Inspiradas en la corteza visual de los gatos, estas redes emulan la capacidad biológica de reconocer patrones complejos en las imágenes, lo que las hace robustas y estables para tareas de percepción [Alzubaidi et al., 2021].

4.3 Larga Memoria de Corto Plazo (LSTM)

Las redes neuronales recurrentes (RNN) son eficaces porque los patrones más recientes tienen un mayor impacto en las predicciones que los patrones más antiguos. Sin embargo, esta ventaja se convierte en una limitación cuando se trata de series de tiempo con doble estacionalidad o patrones de largo plazo. Para superar estas dificultades, se introdujeron los bloques LSTM (Long Short-Term Memory), inventados en 1997 por Hochreiter y Schmidhuber, específicamente para abordar la dependencia a corto plazo y ponderar de manera adecuada los patrones pasados.

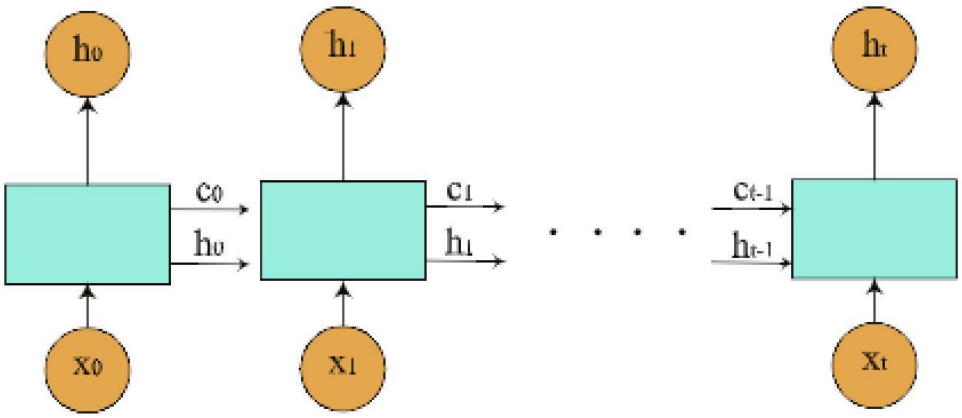

Figura 4.3. Arquitectura de una LSTM compuesta por cuatro capas interactuantes.

Las LSTM reemplazan las neuronas únicas de una RNN tradicional con bloques modulares de cuatro capas interactivas, como se muestra en la Figura 4.3. Estos bloques están compuestos por:

1. Celda de Memoria: Retiene la información relevante a lo largo del tiempo.
2. Puerta de Olvido: Decide qué información debe eliminarse de la celda de memoria.
3. Puerta de Entrada: Determina qué nueva información debe ser almacenada.
4. Puerta de Salida: Decide qué información debe emitirse como salida.

En la Figura 4.4 se muestra la estructura detallada de una celda LSTM. Las celdas LSTM pueden recordar valores independientemente de los intervalos temporales gracias a las tres puertas que regulan el flujo de información entrante y saliente.

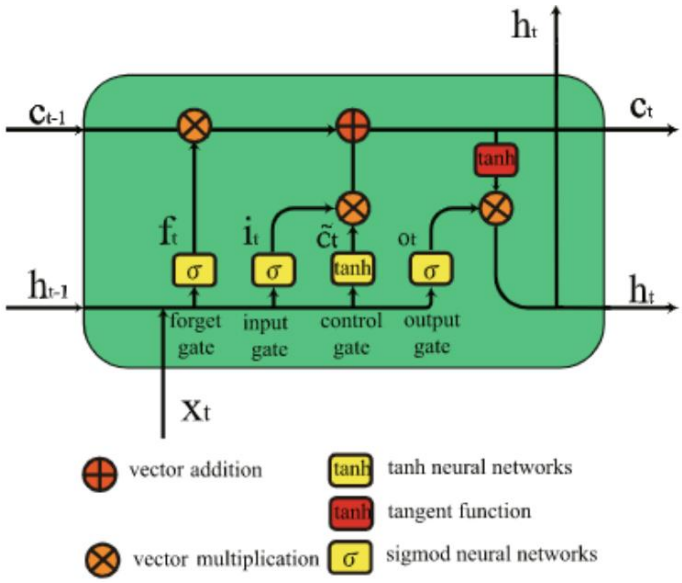

Figura 4.4. Arquitectura de una celda LSTM.

Cada línea en el diagrama comunica un vector desde la salida de un nodo hasta la entrada de otro. Los círculos representan operaciones matemáticas como suma o multiplicación, y las cajas amarillas corresponden a redes neuronales que procesan la información.

Las tres puertas utilizan la función sigmoide (σ) como función de activación. Esta función produce valores entre 0 y 1, lo que indica el grado de bloqueo o paso de la información:

- 0: Bloquea completamente la información.
- 1: Permite pasar toda la información.

Las ecuaciones de las puertas son las siguientes:

$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right) \quad \text{Ecuación 4.3}$$

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \quad \text{Ecuación 4.4}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \quad \text{Ecuación 4.5}$$

Donde:

f_t: Puerta de olvido.

i_t: Puerta de entrada.

o_t: Puerta de salida.

σ: Función sigmoide.

w_x: Pesos asociados a cada puerta.

h_{t-1}: Salida previa.

x_t: Vector de entrada actual.

b_f: Bias asociado a cada puerta.

El funcionamiento interno de una celda LSTM se define por las siguientes ecuaciones:

1. **Celda de Estado Candidata:**

$$\tilde{C}_t = tanh(w_c[h_{t-1}, x_t] + b_c) \quad \text{Ecuación 4.6}$$

2. **Actualización de la Celda de Estado:**

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad \text{Ecuación 4.7}$$

3. **Salida Actual:**

$$h_t = o_t \cdot \tanh(C_t) \quad \text{Ecuación 4.8}$$

Donde:

$\tilde{C}_t$: Celda de estado candidata.

C_t: Celda de estado actual.

C_{t-1}: Celda de estado anterior.

h_t: Salida actual.

$tanh$: Función de tangente hiperbólica.

w_c: Pesos asociados a la celda de estado.

b_c: Bias asociado a la celta de estado.

En la primera iteración, h_{t-1} y C_{t-1} son inicializados en cero. Esto hace que la celda de estado candidata $\tilde{C}_t$ dependa únicamente del vector de entrada procesado por la función tanh.

En las siguientes iteraciones, la celda de memoria C_t se actualiza con los patrones más relevantes y elimina aquellos menos significativos, independientemente del espaciado

temporal. La salida actual h_t siempre está ponderada por la celda de memoria candidata actualizada.

4.4 Redes Neuronales Convolucionales (CNN).

Las redes neuronales convolucionales (CNN, por sus siglas en inglés) son uno de los algoritmos más famosos y comúnmente utilizados en el aprendizaje profundo. Estas redes destacan por su capacidad para identificar automáticamente características relevantes en los datos sin necesidad de supervisión humana directa.

La estructura de una CNN fue inspirada en la compleja secuencia de celdas en la corteza visual del cerebro de los gatos, lo que ha permitido su aplicación en diversos campos como visión computacional, procesamiento del habla, reconocimiento facial, entre otros [Alzubaidi et al., 2021].

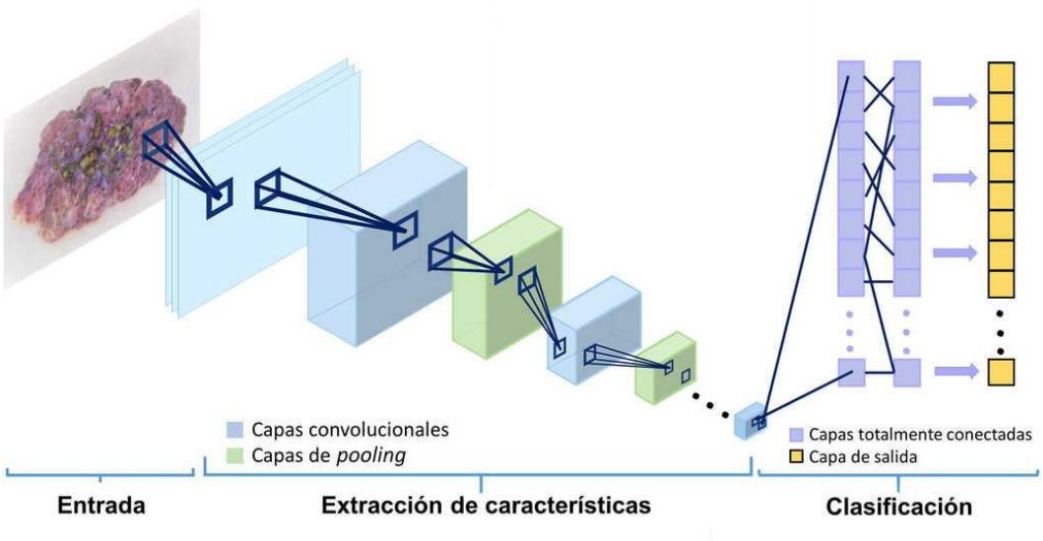

Figura 4.5. Ejemplo de arquitectura CNN para clasificación.

En la Figura 4.5 se presenta un ejemplo de arquitectura típica de una CNN para tareas de clasificación. Esta arquitectura consiste en los siguientes pasos:

1. Capas Convolucionales:

Extraen características relevantes de los datos utilizando filtros o kernels.

2. Capas de Pooling:
Reducen el tamaño de los datos procesados, lo que disminuye la complejidad computacional y ayuda a evitar el sobre-entrenamiento.

3. Capas Adicionales (Opcional):
Se pueden agregar más capas convolucionales y de pooling dependiendo de la arquitectura y la tarea específica.

4. Redes Recurrentes (Opcional):
En algunos casos, los datos procesados por las capas convolucionales se introducen en una RNN para tareas de clasificación o predicción temporal.

5. Capa de Salida:
Proporciona la interpretación final de los datos procesados.

Cada entrada x en un modelo CNN tiene tres dimensiones: Altura (m), es el número de filas en los datos de entrada, ancho (m) que es el número de columnas en los datos de entrada y profundidad (r) que es el número de canales. Por ejemplo:

- En imágenes RGB, cada canal representa un color primario (rojo, verde, azul).
- En series de tiempo, otros canales pueden representar variables adicionales.

La ventana o kernel es una matriz cuadrada de dimensiones $n \times n$ donde $n \ll m$, y su profundidad coincide con el número de canales r.

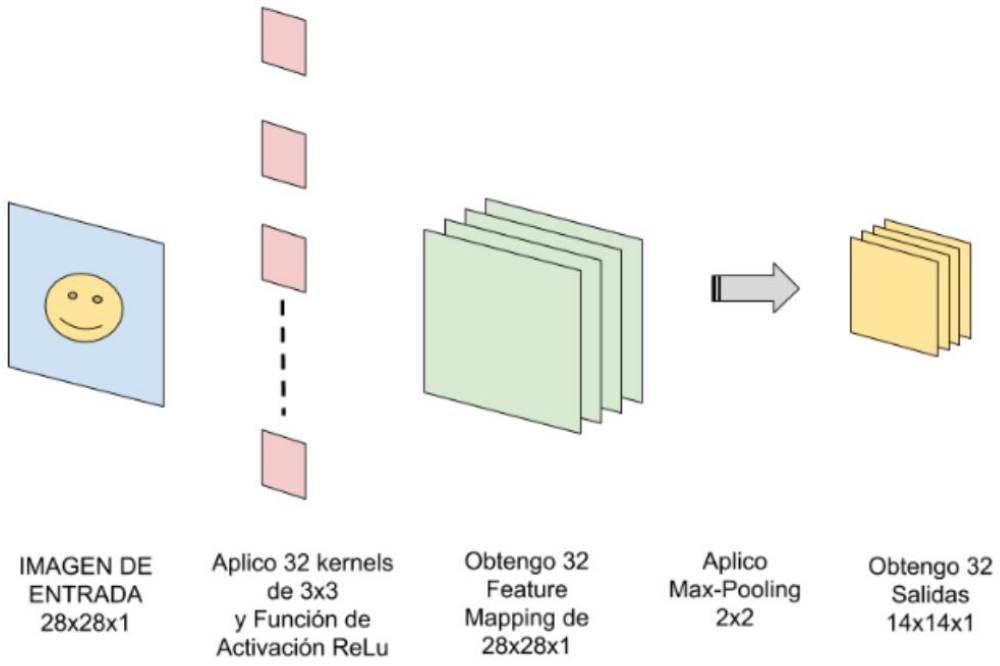

Figura 4.6. Primera Convolución de una CNN.

En la Figura 4.6 se observa un ejemplo de la primera convolución de una CNN. La entrada es una imagen de tamaño 28×28×1 que recibe 32 kernels o filtros, cada uno con una función de activación ReLU. Los kernels (k) tienen dimensiones $n \times n \times r$, donde n es significativamente menor que m.

La salida de la capa convolucional se calcula como el producto punto entre la entrada y los pesos asociados a los filtros, como se describe en la Ecuación 4.9:

$$h_k = f(W_k * x + b_k) \qquad \text{Ecuación 4.9}$$

Donde:

h_k: Es el mapa característico generado por la capa convolucional.

$f(.)$: Es la función de activación, comúnmente ReLU.

W_k: Son los pesos asociados al filtro.

b_k: Es el bias asociado al filtro k.

Los mapas característicos h_k generados por las capas convolucionales tienen un tamaño reducido a $m-n-1$. Este resultado es posteriormente procesado por una capa de pooling con ventanas de tamaño $p \times p$. Las capas de pooling tienen dos objetivos principales:

1. Reducir Dimensionalidad:
   Disminuyen el tamaño de los mapas característicos, lo que acelera el entrenamiento y reduce la carga computacional.

2. Evitar Sobre-entrenamiento:
   Resumen las características más importantes de los datos, mejorando la generalización del modelo.

4.5 Pronóstico

El desempeño de las redes neuronales como LSTM y CNN se evalúa utilizando datos de entrenamiento (x_{train}) transformados en un formato de aprendizaje supervisado. Este proceso entrena el modelo para generar pronósticos ($\hat{x}_{val}$) y seleccionar la mejor metodología entre SSA y redes neuronales, como se ilustra en la Figura 4.7.

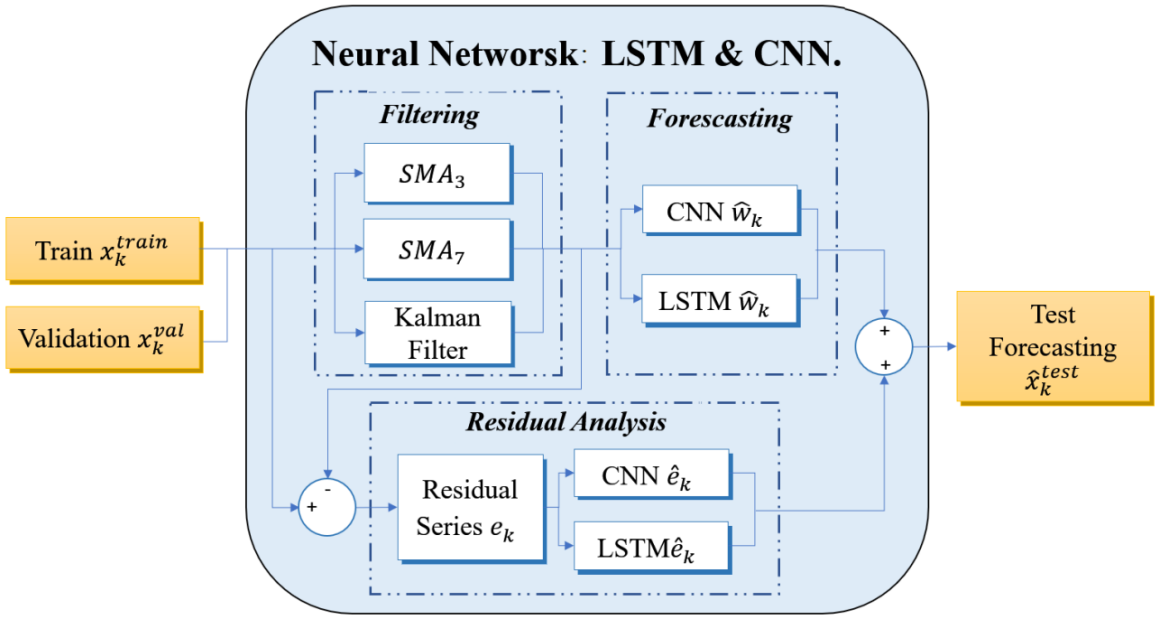

Figura 4.7. Modelo 2: Diagrama a bloques del método de pronóstico por CNN y LSTM.

Para simplificar la serie de tiempo y aislar tendencias o estacionalidades principales del ruido, se aplican los filtros SMA3, SMA7 y el Filtro de Kalman (Capítulo 2). Los parámetros de estos filtros se seleccionaron previamente en función de su desempeño en validación.

El proceso de pronóstico involucra los siguientes pasos:

1. Transformación a Aprendizaje Supervisado:
   La serie filtrada (w_{train}) se transforma en un formato supervisado similar a la hankelización, utilizando un ancho de ventana de 8 días (7 días de contexto más 1 día como objetivo).
2. Entrenamiento del Modelo:
   Los modelos LSTM y CNN se entrenan utilizando las arquitecturas especificadas en la Tabla 4.1:

Tabla 4.1 Arquitecturas para LSTM y CNN para pronóstico de la señal filtrada w_{train}.

Arquitectura	Capa de entrada		Capas ocultas	Capa de salida
LSTM	7 LSTM		42 LSTM	1RNN
CNN	50 CNN	M.P. size:2	50RNN	1RNN

Ambas arquitecturas tienen un batch size de 10, con función de activación ReLU para la capa de entrada, $tanh$ para la capa oculta y ReLU para la capa de salida y son entrenadas durante 50 épocas.

3. Mitigación de Varianza:

Debido a la variabilidad inherente al entrenamiento de RNN (causada por la inicialización aleatoria de los pesos), se entrena cada modelo varias veces (m) y se selecciona el modelo con el mejor desempeño en validación.

En la Figura 4.8 se muestra cómo los errores cuadráticos medios (MSE) varían significativamente con la misma configuración, justificando esta estrategia.

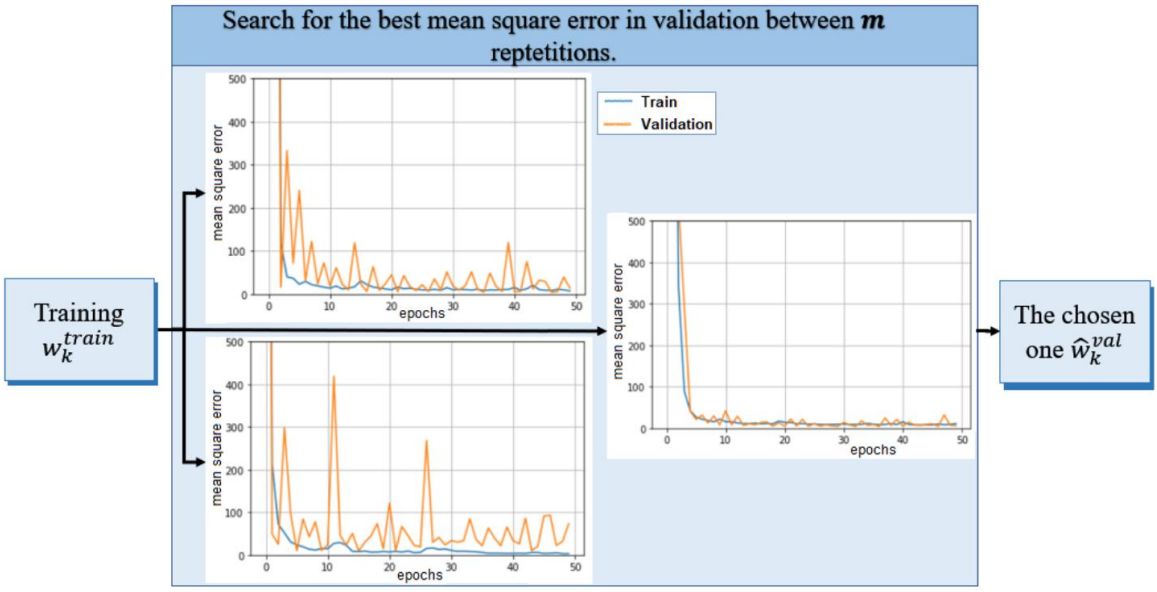

Figura 4.8. Comparación del error cuadrático medio entre pronósticos realizados con la misma configuración.

4. Pronóstico de Residuales:

Se calcula la serie de residuales (e_{train}) a partir de la serie original y su reconstrucción filtrada. Esta serie recibe el mismo tratamiento de transformación y entrenamiento utilizando las configuraciones de la Tabla 4.2:

Tabla 4.2 Arquitecturas para LSTM y CNN para pronóstico de la señal filtrada e_{train}.

Arquitectura	Capa de entrada		Capas ocultas	Capa de salida
LSTM	7 LSTM		42 LSTM	1RNN
CNN	21 CNN	M.P. size:2	50RNN	1RNN

Ambas arquitecturas tienen un batch size de 14, con función de activación ReLU para la capa de entrada, $tanh$ para la capa oculta y ReLU para la capa de salida y son entrenadas durante 60 épocas.

5. Pronóstico de Prueba:

El mejor modelo (LSTM o CNN) seleccionado en validación se utiliza para realizar el pronóstico final de prueba.

La serie de entrenamiento y validación se concatenan y se aplica el filtro identificado en validación antes de entrenar según las configuraciones seleccionadas.

4.6 Seleccionando el Mejor Pronóstico en Validación

Como se mencionó en el paso 3 de la Subsección 4.5, el mejor modelo en validación se selecciona repitiendo la configuración de la Tabla 4.5 un total de m veces. De cada iteración, se elige el modelo que logra el menor MAPE en validación. Los resultados de esta selección se presentan en la Figura 4.9, donde se muestra la desviación estándar de los MAPEs obtenidos.

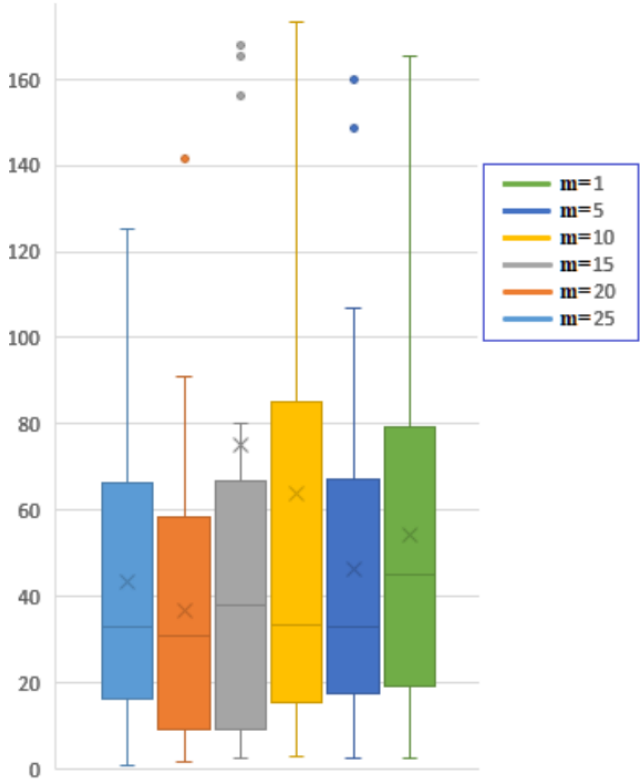

Figura 4.9. Desviación Estándar de los MAPES en Validación al tomar el mejor de m pronósticos.

Para determinar el valor óptimo de m, se realizó una búsqueda exploratoria variando m desde 1 hasta 25. Se midió la desviación estándar de los errores de validación para un horizonte de pronóstico de 21 días. Esta evaluación se repitió 30 veces para cumplir con el Teorema del Límite Central, asegurando resultados estadísticamente significativos.

- Procedimiento:

1. Para cada valor de m, se entrenaron m modelos utilizando los datos de validación.
2. Se seleccionó el modelo con el menor MAPE en validación y se generó un pronóstico.
3. El proceso completo se repitió 30 veces para cada m, generando 30 pronósticos independientes.
4. Se calculó la desviación estándar de los errores para cada m.

- Resultados:

- A medida que aumenta m, la desviación estándar disminuye, lo que indica una mejora en la estabilidad del modelo.
- Más allá de cierto valor de m, los beneficios adicionales son marginales en comparación con el costo computacional.

La metodología asegura que el modelo seleccionado no solo minimice el error de validación, sino que también sea robusto frente a la variabilidad inherente al entrenamiento de redes neuronales.

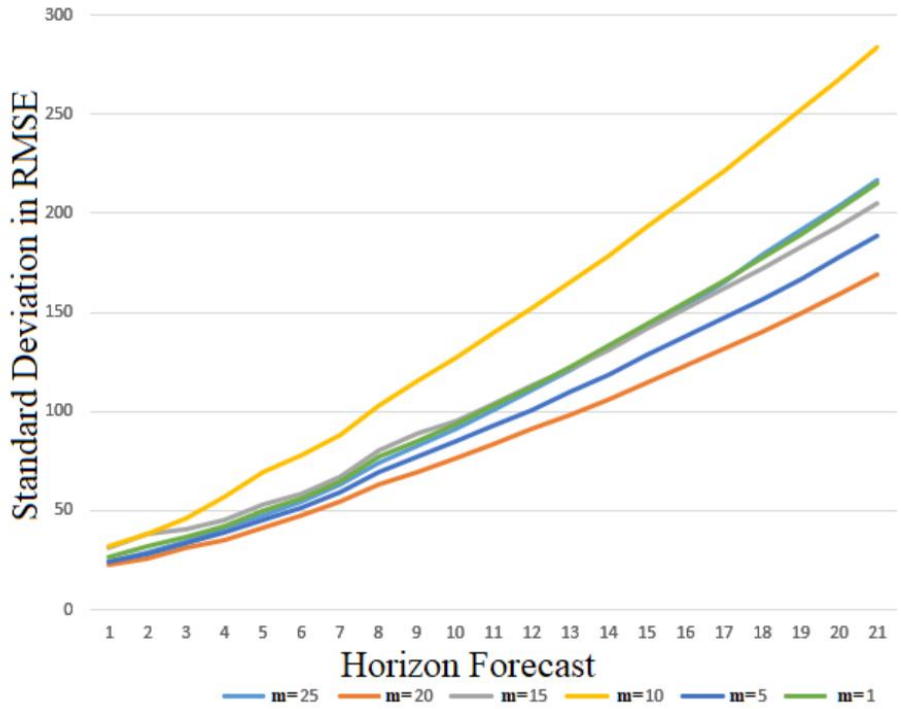

Figura 4.10. Desviación estándar en MAPE de los pronósticos en validación eligiendo diferentes valores de m.

De acuerdo con la Figura 4.10, se observa que seleccionar el mejor modelo de entre $m = 20$ repeticiones produce el pronóstico más preciso en validación. Para valores de $m > 20$, no se detecta una mejora significativa en el desempeño del modelo.

Desde otra perspectiva, al calcular el MAPE de los 30 pronósticos generados para diferentes valores de m y medir su desviación estándar, se observa en la Figura 4.11 que, a medida que aumenta m, los pronósticos atípicos disminuyen. Este comportamiento se traduce en un mejor promedio del MAPE cuando m = 20, destacando además que la mayoría de los pronósticos generados presentan errores cercanos al 0%. Esto respalda que m = 20 es el valor óptimo para garantizar la estabilidad y precisión del modelo.

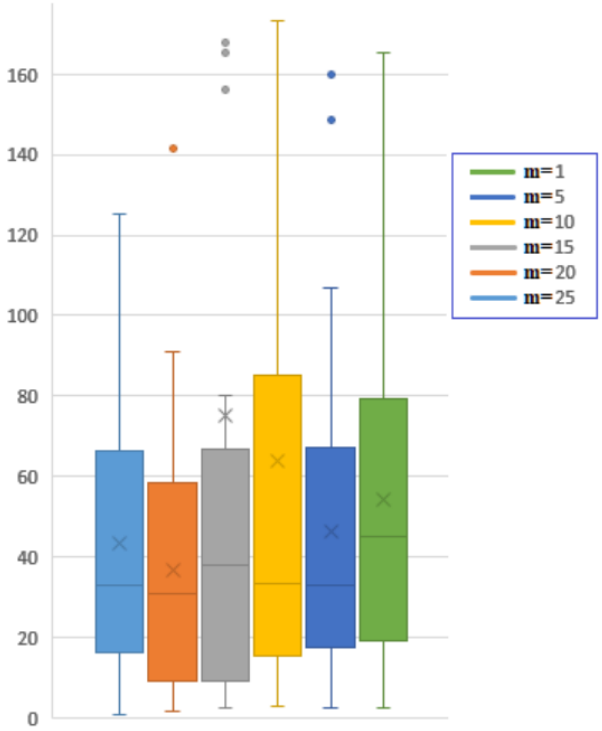

Figura 4.11. Desviación estándar en MAPE de los pronósticos en validación eligiendo diferentes valores de m.

Para el pronóstico de la serie de residuales, se aplica el mismo procedimiento descrito previamente, con el objetivo de maximizar la precisión del resultado final. Este enfoque asegura que tanto la serie filtrada como los residuales sean tratados con configuraciones que minimicen errores y optimicen el rendimiento del modelo.

4.7 Conclusiones

Los métodos de aprendizaje profundo, como LSTM y CNN, han demostrado ser altamente eficaces para el pronóstico de series de tiempo con niveles moderados de complejidad. Sin embargo, ajustar únicamente los parámetros de estos modelos no es suficiente para manejar series de tiempo altamente caóticas y complejas. Para abordar este desafío, se propuso una estrategia basada en la separación de la serie de tiempo en dos componentes principales:

- Serie de tiempo suavizada: Esta contiene la tendencia principal y las estacionalidades más evidentes, lo que permite una sintonización más efectiva de los modelos LSTM y CNN.
- Serie de residuales: Esta retiene el ruido y los armónicos adicionales, eliminando la tendencia y estacionalidades principales, facilitando así un enfoque específico para identificar patrones residuales.

La estrategia de repetir el entrenamiento del modelo m veces y seleccionar el mejor desempeño en validación mostró resultados sobresalientes en la predicción de nuevos casos de COVID-19. No obstante, para otras series de tiempo, es fundamental realizar un análisis exhaustivo para determinar un valor óptimo de m que no sobreajuste el modelo y maximice su rendimiento general.

Por último, el análisis de residuales representa una fase crucial para mejorar el pronóstico, ya que permite incorporar información que no está presente en la serie suavizada. Sin esta etapa, el modelo se limitaría a trabajar únicamente sobre datos filtrados, sin capturar la complejidad de los datos originales. Este enfoque marca una diferencia importante con respecto al SSA, donde el análisis de residuales no siempre garantiza una mejora en el pronóstico final, subrayando así la ventaja de los modelos basados en aprendizaje profundo en escenarios de alta complejidad.

Capítulo 5

5 Hibridaciones de Métodos de Pronóstico

La creciente necesidad de mayor precisión en los pronósticos, impulsada por la especialización de las series de tiempo, ha llevado a la implementación de hibridaciones de métodos especializados. Por ejemplo, en el pronóstico del mercado de valores, es crucial predecir con precisión la tendencia. Sin embargo, los métodos que mejor predicen tendencias son altamente sensibles al ruido, lo que hace necesario integrar técnicas de control de ruido para mejorar el rendimiento del pronóstico final.

Pese a su relevancia, la mayoría de las hibridaciones no han sido ampliamente exploradas. Los estudios recientes se centran principalmente en modelos basados en características, meta-aprendizajes o combinaciones de métodos tradicionales. Solo un número limitado de investigaciones aborda la construcción de auténticas hibridaciones que integren metodologías de manera estructurada y efectiva [Makridakis et al., 2020].

5.1 Método de Pronóstico con Filtros y Análisis Residual (*FMFRA*)

El Método de Pronóstico con Filtros y Análisis Residual (FMFRA) es una técnica híbrida diseñada específicamente para series de tiempo complejas, como las relacionadas con COVID-19. Este método se basa en la separación de series de tiempo mediante filtros, simplificando las tareas de pronóstico para series con grandes volúmenes de datos y altas no linealidades. Para series más cortas, el FMFRA emplea un enfoque automático de SSA con análisis de residuales, logrando un rendimiento superior al SSA tradicional.

No existe un método único para todas las series de tiempo, por lo que cada serie posee características únicas en términos de complejidad y ruido intrínseco. Esto requiere la adaptación de las estrategias de pronóstico a las características particulares de cada serie.

Como se presentó en el Capítulo 4, el éxito de los métodos de aprendizaje profundo depende de su capacidad para identificar múltiples grados de libertad en las series analizadas. En series extremadamente complejas, es necesario incluir más capas intermedias para capturar la dinámica. Sin embargo, esto incrementa el riesgo del problema del gradiente desvaneciente, dificultando la convergencia hacia una solución óptima.

Se identificaron las siguientes estrategias complementarias como las más efectivas para pronóstico de series de tiempo caóticas:

- **Repetición de m pronósticos:** Repetir el proceso de predicción m veces y seleccionar el mejor desempeño en validación.
- **División de la serie en dos:** Separar la serie original en tendencias y estacionalidades principales para ser procesadas con algoritmos de LSTM y CNN.
- **Ajuste de residuales:** Analizados con métodos clasificadores de aprendizaje automático para diferenciar entre ruido y estacionalidades menores.

El FMFRA busca resolver las limitaciones mencionadas al combinar los enfoques de SSA y aprendizaje profundo, siguiendo tres fases principales:

- **Filtrado:** Reducir la complejidad de la serie mediante técnicas específicas (SMA, Filtro de Kalman, entre otros).
- **Pronóstico:** Aplicar algoritmos avanzados (LSTM, CNN o SSA) para predecir las series filtradas.
- **Análisis Residual:** Analizar los residuos para identificar componentes armónicos o ruido residual que puedan mejorar el modelo final.

La Figura 5.1 muestra el flujo completo del método FMFRA, destacando la interacción entre las fases de filtrado, pronóstico y análisis residual.

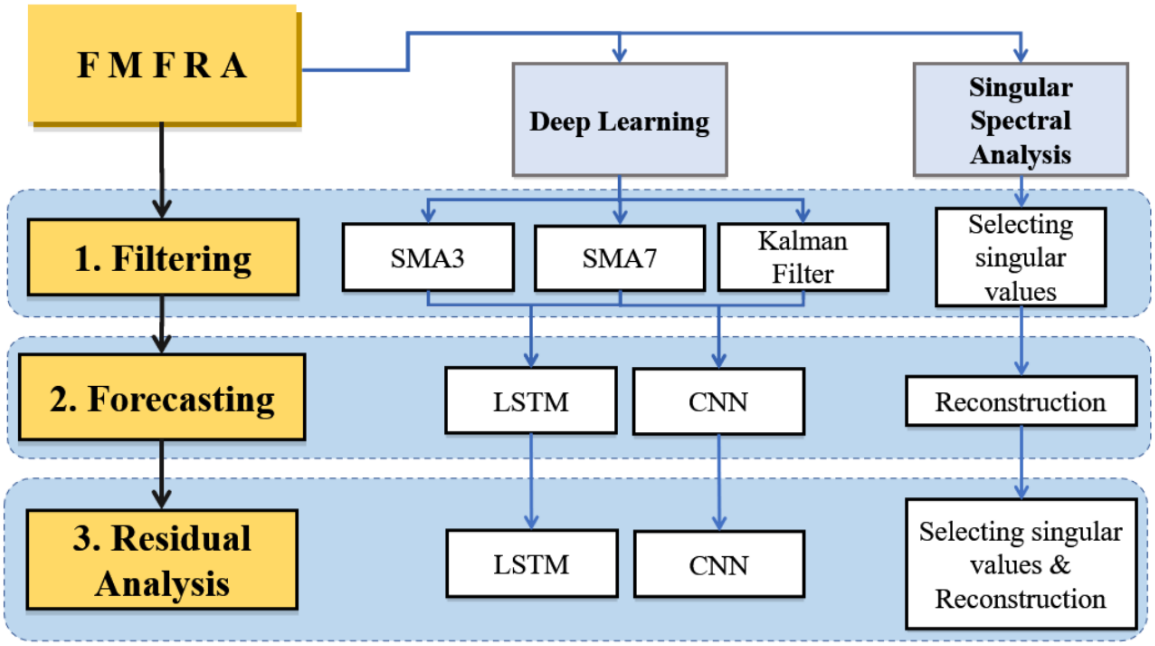

Figura 5.1. Método *FMFRA*.

En el método FMFRA, la serie de tiempo es normalizada para garantizar que el aprendizaje profundo no distinga entre países con poblaciones diferentes, permitiendo que los patrones subyacentes sean analizados sin sesgos debido a magnitudes absolutas. Posteriormente, la serie es dividida en tres conjuntos: entrenamiento, validación y prueba.

El conjunto de entrenamiento se utiliza para calibrar los modelos de aprendizaje profundo y SSA, mientras que el conjunto de validación se emplea para evaluar el desempeño de estos modelos. El método que presenta el mejor desempeño en el conjunto de validación es seleccionado para realizar el pronóstico en el conjunto de prueba, como se ilustra en la Figura 5.1.

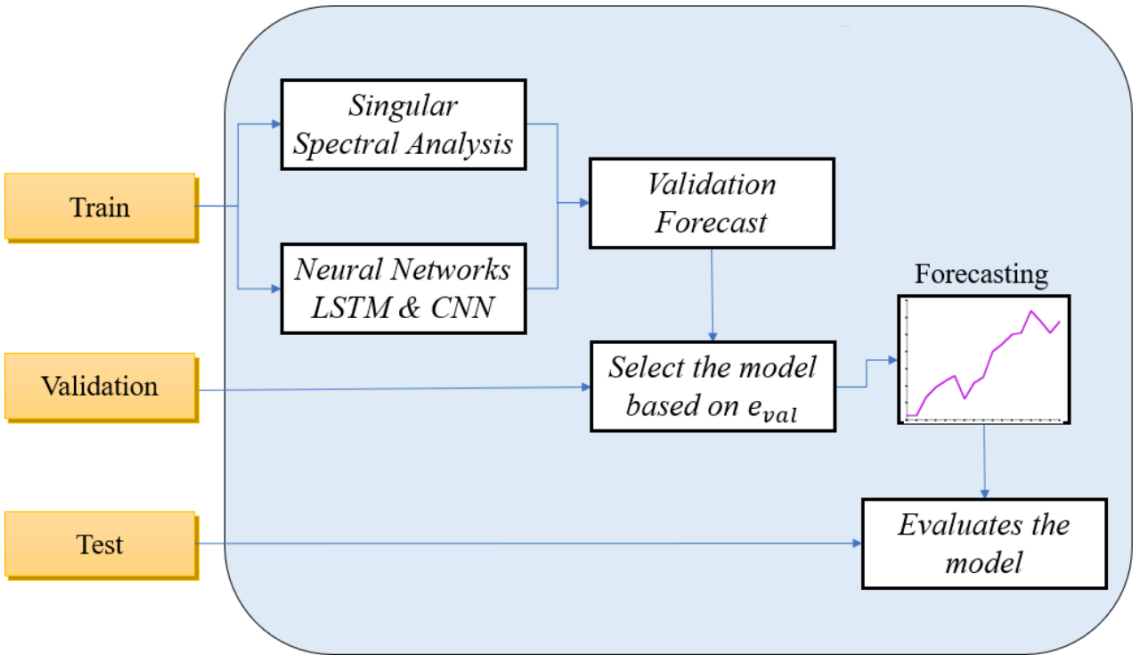

Figura 5.2. Selección del mejor pronóstico por desempeño en validación.

Tal como se explicó en los capítulos correspondientes, tanto el SSA como el Aprendizaje Profundo presentan diferentes especializaciones, sensibilidades al ruido y capacidades para tratar las no linealidades inherentes de las series de tiempo. Mientras que el Aprendizaje Profundo puede almacenar patrones relevantes en su estructura de memoria, el SSA pondera de manera uniforme las dinámicas representadas por sus valores singulares dentro de la matriz trayectoria A. Esta diferencia convierte al Aprendizaje Profundo en una opción más adecuada para series de tiempo con suficientes datos representativos, mientras que el SSA resulta más efectivo en series de tiempo limitadas, ya que puede truncar información irrelevante hasta obtener un pronóstico satisfactorio en validación.

Los métodos de pronóstico que combinan diversas técnicas generalmente se adaptan a las características específicas de las series de tiempo. En contraste, el FMFRA se ajusta dinámicamente a los resultados de validación, permitiendo gran flexibilidad ante cambios abruptos o situaciones inéditas en la serie de tiempo. Esta

capacidad adaptativa le permite superar los métodos del estado del arte en tres instancias clave de la pandemia:

1. **Inicios de la Pandemia**

Durante los primeros meses de la pandemia de COVID-19 en el continente americano, se requería un método de pronóstico eficaz pese a la escasez de datos en las series de tiempo. Esta metodología fue evaluada en los primeros seis meses de la pandemia (del 3 de marzo al 5 de septiembre de 2020). En este período, los métodos basados en Aprendizaje Profundo no lograron resultados satisfactorios debido a la limitada cantidad de datos. En contraste, el enfoque FMFRA con SSA (es decir, el componente de FMFRA basado en Análisis de Espectro Singular) demostró un desempeño superior, como se muestra en la Tabla 5.1, al adaptarse mejor a escenarios con poca información y alta incertidumbre.

Tabla 5.1: Resultados del FMFRA al inicio de la Pandemia.

	EEUU			Brasil			México			Colombia			AVERAGE		
	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE
SSA	15.7	75	6851	21.28	55	6906	21.29	85	1259	15.72	40	1253	18.5	63.75	4067
Kalantari	29.26	75	11973	33.47	**75**	8158	48.55	70	1728	12.52	55	1058	30.95	68.75	5729
XGBoost	11.66	75	5451	32.31	**75**	9497	29.97	80	1481	29.29	55	2156	26.38	**71.25**	4646
RFR	25.5	**85**	8008	47.03	60	10417	33.46	**90**	1647	45.71	40	3273	37.93	68.75	5836
LR	6.97	80	6455	34.08	60	8695	21.67	80	1372	16.77	30	1444	19.87	62.5	4432
FMFRA	**11.05**	75	**4877**	**20.15**	55	**6865**	**17.19**	85	**954**	**12.14**	**60**	**1048**	**12.63**	68.75	**3436**
	$SMA_3 - LSTM$			$SSA - Residuals$			$SMA_3 - LSTM$			$SSA - Residuals$					

La Tabla 5.1 presenta los resultados obtenidos por diferentes modelos durante los primeros seis meses de la pandemia de COVID-19. Puede observarse que el enfoque propuesto FMFRA obtuvo los mejores resultados promedio en términos de MAPE (12.63%), DA (68.75%) y RMSE (3436), superando tanto a los modelos clásicos como a los métodos de aprendizaje automático. Cabe destacar que FMFRA no representa un único modelo, sino una estrategia híbrida cuya configuración se ajusta por país. En particular, el componente de pronóstico que proporcionó mejores resultados para Estados

Unidos de América y México fue $SMA_3 - LSTM$, mientras que para Brasil y Colombia fue $SSA - Residuals$, es decir, un modelo FMFRA basado en análisis de espectro singular y ajuste de residuales.

El FMFRA demuestra un desempeño superior en comparación con los métodos del estado del arte. Aunque técnicas como XGBoost y RFR muestran una aceptable precisión en la dirección de ajuste (DA), no logran resultados satisfactorios en términos de MAPE y RMSE, reafirmando la superioridad del FMFRA.

2. Picos de Contagio

Las series de tiempo asociadas a enfermedades infecciosas presentan un comportamiento más caótico durante los picos de contagios. Este incremento en la volatilidad se debe no solo al aumento acelerado en el número de casos, sino también al ruido en los datos, ocasionado por la incapacidad de los hospitales para registrar adecuadamente cada caso en tiempo real.

En este contexto, el FMFRA destaca por su capacidad de adaptación a estas condiciones de alta complejidad. La Tabla 5.2 muestra el desempeño de esta metodología durante los picos de contagios en América, en el periodo comprendido entre el 3 de marzo y el 18 de noviembre de 2020.

Tabla 5.2: Resultados del FMFRA en picos de contagio en América.

	EEUU			Brasil			México			Colombia			AVERAGE		
	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE
SSA	17.14	**60**	46572	24.08	55	11045	29.65	65	6156	23.58	65	4641	23.61	61.25	17102
XGBoost	20.15	50	53961	34.47	75	15041	31.81	**80**	6192	38.97	45	6970	31.35	62.5	20541
RFR	15.47	40	43731	22.25	60	10202	38.49	65	6660	40.3	50	7316	29.13	53.75	16977
LR	36.7	40	80970	13.69	70	7196	32.25	70	7498	24.84	**70**	4809	26.87	62.5	25118
FMFRA	**9.45**	50	**28864**	**10.34**	**85**	**5442**	**24.29**	70	**6167**	**23.55**	65	**4633**	**16.92**	**67.5**	**11279**
	$SMA_7 - LSTM$			$SMA_7 - LSTM$			$SSA - Residuals$			$SSA - Residuals$					

La mayoría de los métodos de aprendizaje profundo enfrentan dificultades para predecir los picos de contagios de COVID-19 debido a que las pronunciadas tendencias de cambio en las series de tiempo pueden confundirse fácilmente con ruido. Sin embargo, mediante la descomposición de la serie de tiempo con filtros, el FMFRA logra que la tendencia sea correctamente aprendida, ya sea por LSTM o CNN. A pesar de esto, en países con poblaciones muy grandes, como México, la tendencia es más fácil de aprender debido a la mayor cantidad de datos representativos. Por otro lado, en países con poblaciones más pequeñas, como Colombia, los métodos de aprendizaje profundo fueron superados por SSA, que mostró un mejor desempeño al gestionar la menor cantidad de datos y el ruido inherente.

3. Un año de contagios

El desempeño del FMFRA fue evaluado durante un año completo, del 3 de marzo de 2020 al 3 de marzo de 2021, lo que permitió capturar la estacionalidad anual completa. Este periodo permitió a los métodos de aprendizaje profundo, como LSTM y CNN, capturar mejor la tendencia debido a la mayor disponibilidad de datos representativos y patrones consistentes. Sin embargo, dado que la volatilidad fue baja en este periodo debido a los menores niveles de contagios de COVID-19, el SSA demostró un desempeño superior, tal como se muestra en la Tabla 5.3.

Tabla 5.3: Resultados del FMFRA al año de pandemia en América.

	EEUU			Brasil			México			Colombia			AVERAGE		
	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE	MAPE	DA	RMSE
SSA	22.07	85	8043	9.06	**95**	5344	57.33	95	1134	14.67	50	2846	25.16	81.25	4326
XGBoost	20.22	**90**	7253	17.83	80	13784	89.42	80	1525	33	35	6138	40.12	71.25	7175
RFR	42.69	85	14872	14.67	85	9205	68.32	**100**	1269	32.54	50	6021	39.56	80	7034
LR	40.06	85	14510	13.33	80	8384	100.45	**100**	1578	19.12	55	3662	43.24	80	7541
FMFRA	**19.11**	**90**	**6931**	**8.83**	90	**5328**	**54.82**	95	**1071**	**9.15**	**70**	**1734**	**22.98**	**86.25**	**3766**
	$SSA - Residuals$			$SSA - Residuals$			$SSA - Residuals$			$SSA - Residuals$					

La Tabla 5.3 muestra el desempeño del método FMFRA al analizar un año completo de contagios en América. Los resultados evidencian que el FMFRA logra un equilibrio

destacado entre precisión y robustez en todas las métricas, con el mejor desempeño promedio en MAPE (22.98) y RMSE (3766), además de un índice de dirección acertada (DA) competitivo del 86.25%. Esto resalta su capacidad para adaptarse tanto a la tendencia como a las estacionalidades presentes en las series de tiempo.

En comparación, el SSA obtuvo un desempeño similar en métricas globales, especialmente en países con menor volatilidad como Brasil y Colombia, donde la menor complejidad de la serie de tiempo favoreció su enfoque. Sin embargo, métodos como XGBoost, RFR y LR demostraron ser menos efectivos en manejar la combinación de ruido y no linealidades, especialmente en países con series de tiempo más caóticas, como México y Estados Unidos.

El uso del análisis de residuales en el FMFRA permite un ajuste fino en los pronósticos, proporcionando una ventaja significativa frente a los demás métodos al mantener una precisión consistente en distintos escenarios de datos. Esto confirma la flexibilidad y robustez del FMFRA como una herramienta eficaz para abordar la complejidad de los datos pandémicos.

5.2 Filtro de Kalman con Holt and Winters ($FK - H\&W$).

El Filtro de Kalman con Holt and Winters (FK-H&W) combina lo mejor de dos mundos: la robustez y versatilidad de Holt and Winters (H&W) para capturar patrones de nivel, tendencia y estacionalidad, y la capacidad del Filtro de Kalman para responder dinámicamente a cambios abruptos en las series de tiempo.

H&W es una técnica clásica y poderosa, ampliamente reconocida en el ámbito del pronóstico de series de tiempo. Su capacidad para dividir las tareas de estimación en tres componentes principales: nivel, tendencia y estacionalidad, le permite adaptarse eficientemente a patrones complejos. Sin embargo, su debilidad principal radica en su sensibilidad a la volatilidad, es decir, cambios bruscos en la tendencia o el nivel pueden ser subestimados debido a una sintonización conservadora. En estas situaciones, H&W puede

fallar al reflejar adecuadamente la realidad, por esta razón, H&W se ha utilizado como base en metodologías híbridas exitosas, donde la tendencia o estacionalidad es reemplazada o ajustada por métodos más avanzados, como LSTM, CNN, o técnicas de aprendizaje automático.

Por otro lado, el Filtro de Kalman es excepcional cuando se cumplen las siguientes condiciones:

- Modelo Matemático Preciso: Si el modelo describe al menos el 95 % del comportamiento observado para un pronóstico de máximo 7 pasos adelante.
- Ruido Bien Modelado: La identificación adecuada de los diferentes tipos de ruido (medición e intrínseco) permite estimaciones más precisas.
- Estimaciones de máxima verosimilitud para los valores futuros.
- Una medida de incertidumbre estadística ponderada por los datos históricos.

No obstante, en la práctica, rara vez se cuenta con un modelo matemático suficientemente preciso, y el ruido no siempre puede ser modelado con exactitud, lo que limita su efectividad.

La hibridación FK-H&W resuelve estas limitaciones al combinar las fortalezas de ambas técnicas, pues H&W proporciona el modelo matemático base, suficientemente preciso para condiciones normales de operación y con una medida confiable de incertidumbre.

El Filtro de Kalman responde a condiciones anormales, como cambios abruptos en la serie de tiempo. Gracias a su capacidad de medir la incertidumbre del pronóstico, el filtro permite ajustar en tiempo real el modelo de H&W, seleccionando parámetros que mejor se adapten a las nuevas características de la serie.

5.2.1 Ventajas de FK-H&W

- **Adaptabilidad Dinámica:** La capacidad de resintonización en línea del modelo de H&W permite enfrentar escenarios de alta volatilidad sin comprometer la precisión.

- **Mejor Manejo de la Incertidumbre:** Mientras H&W captura patrones de largo plazo, el Filtro de Kalman asegura que las desviaciones significativas sean detectadas y gestionadas oportunamente.
- **Robustez para Series Caóticas:** En condiciones de alta complejidad y ruido, FK-H&W logra combinar la estabilidad de H&W con la precisión adaptativa del Filtro de Kalman, ofreciendo un pronóstico más confiable.

En resumen, FK-H&W se presenta como una metodología híbrida innovadora, capaz de combinar la modelación clásica de series de tiempo con técnicas avanzadas de estimación dinámica, ampliando las posibilidades de pronóstico en escenarios complejos y altamente volátiles.

5.2.2. Metodología.

La Figura 5.3 presenta el diagrama de bloques del Método de Pronóstico Híbrido en Línea Holt & Winter con Filtro de Kalman (H&W − FK). Este modelo combina las fortalezas del suavizado exponencial y el filtro de Kalman, integrando estrategias probadas en competencias de pronóstico [Semenoglou et al., 2021; Makridakis et al., 2020].

Además, permite seleccionar entre al menos ocho configuraciones distintas, optimizando la interacción de estos componentes para mejorar la precisión del pronóstico. Estas configuraciones corresponden a las variaciones del modelo Holt & Winters, donde la tendencia y la estacionalidad pueden ser aditivas o multiplicativas, resultando en las siguientes combinaciones: HW-AA, HW-AM, HW-MA y HW-MM. Adicionalmente, cada una de estas configuraciones puede incluir o no un factor de amortiguación, lo que da lugar a un total de ocho modelos distintos.

Esta flexibilidad permite adaptar el modelo para capturar de manera más precisa las características específicas de los datos, considerando tanto cambios suaves como

fluctuaciones más dinámicas. Con las configuraciones definidas, el modelo se formula en el espacio de estados y se evalúa mediante el RMSE para medir el error de estimación.

Posteriormente, se introduce el Filtro de Kalman, que proporciona capacidades avanzadas, como la identificación de datos atípicos, la evaluación dinámica de incertidumbre y el manejo de valores faltantes. Este filtro permite realizar pronósticos en tiempo real ajustando la incertidumbre según los límites definidos por el usuario.

Si el modelo cumple con los criterios de aceptación establecidos, el pronóstico procede utilizando las configuraciones y ajustes determinados. En caso contrario, se requiere una reevaluación y ajuste del modelo basado en los datos más recientes ingresados al sistema.

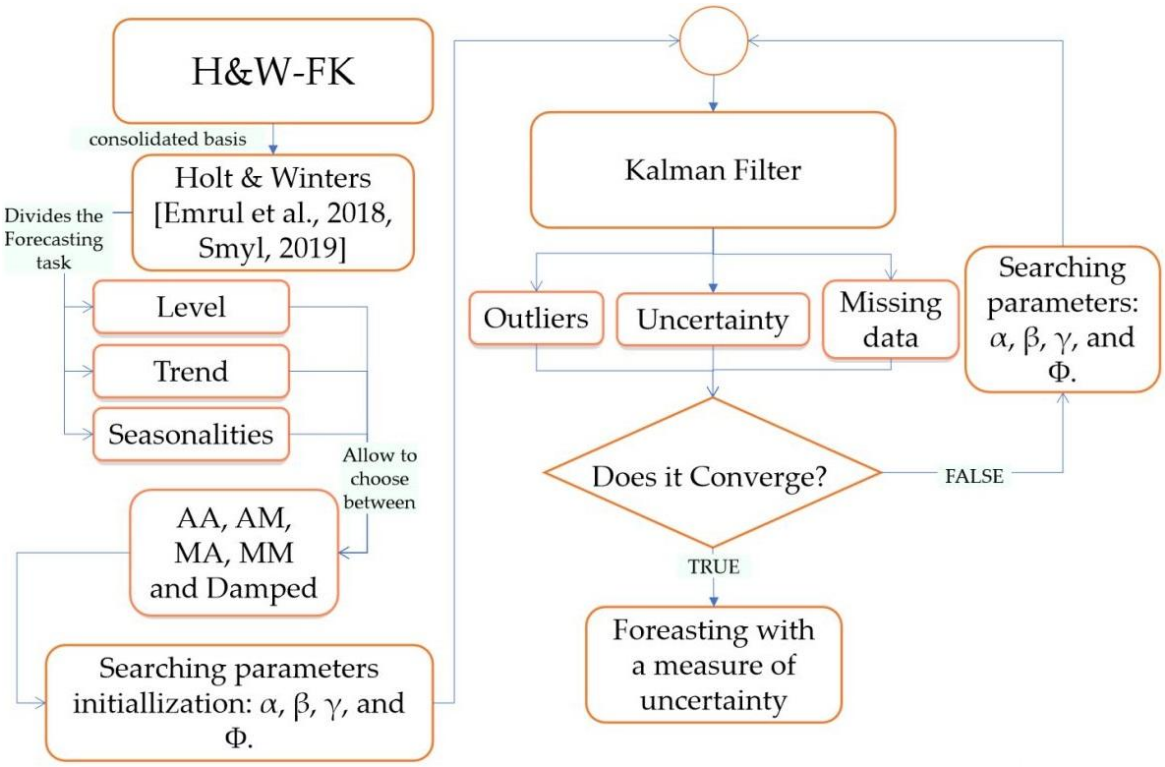

Figura 5.3. Método Híbrido H&W-FK.

Se destacan varias ventajas significativas al realizar un modelo en espacio de estados que se ajusta automáticamente en línea, especialmente para tareas cruciales como el

pronóstico de la demanda en sectores tales como energía, agua, cadena de suministros, y mercados financieros. Estas ventajas incluyen:

- Adaptabilidad en Tiempo Real: Se permite una adaptabilidad instantánea a los cambios en los datos de entrada, lo que es crucial en entornos dinámicos como los mercados de energía o financieros.
- Manejo de Datos No Estacionarios: La capacidad para manejar datos cuyas propiedades estadísticas cambian con el tiempo es proporcionada por estos modelos, lo que permite seguir efectivamente las tendencias y patrones cambiantes sin la necesidad de recalibración manual.
- Incorporación de Múltiples Fuentes de Datos: Se facilita la integración de múltiples fuentes y tipos de datos, como datos históricos, en tiempo real y variables exógenas, lo que resulta valioso en la predicción de demanda donde factores externos pueden influir significativamente.
- Mejor Gestión de la Incertidumbre: La capacidad de estos modelos para estimar y actualizar las incertidumbres asociadas con las predicciones en tiempo real es proporcionada, permitiendo a los usuarios tomar decisiones más informadas y gestionar riesgos de manera efectiva.
- Eficiencia en la Estimación de Componentes Ocultos: La eficacia para estimar componentes ocultos o no observables de los procesos, como tendencias subyacentes y ciclos estacionales, es una ventaja clave de estos modelos.
- Optimización de Recursos: Por medio de estos modelos, se permite una asignación más eficiente y un uso más racional de los recursos, lo que puede reducir costos y mejorar la sostenibilidad en la gestión de energía y agua.
- Resiliencia ante Fallos y Valores Atípicos: Se hace posible la filtración de ruido y el manejo de valores atípicos de forma automática, lo que otorga robustez y confiabilidad a los modelos para la toma de decisiones críticas.

Los modelos en espacio de estados que se sintonizan automáticamente en línea como herramientas poderosas para la predicción en sectores críticos, ofreciendo adaptabilidad, precisión y una eficiencia mejorada en la toma de decisiones basadas en datos complejos y

cambiantes. Para series de tiempo donde la estacionalidad y la tendencia son parsimoniosas, el modelo con tendencia y estacionalidad aditiva H&W(AA) ha mostrado mejores resultados.

5.2.3. Modelo matemático por H&W.

El propósito de definir un modelo en espacio de estados es para poder aplicar las ventajas del álgebra matricial y poder controlar el sistema o estimarlo, definido por la Ecuación 5.1:

$$\mu_t = l_t + b_t + s_{t-h} \qquad \text{Ecuación 5.1}$$

Donde:

μ_t: Es el pronóstico en el instante t.

l_t: Es el nivel en el instante t.

b_t: Es la tendencia en el instante t.

s_{t-h}: Es la estacionalidad en el instante t y h es el tamaño de la estacionalidad.

A su vez, cada uno de los 3 términos de la Ecuación 5.1 que son el nivel, la tendencia y la estacionalidad respectivamente, están definidas por las Ecuaciones 5.2, 5.3y 5.4 que incorporan un término del error de estimación con el propósito de sintonizar los parámetros al minimizar este error:

$$l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t \qquad \text{Ecuación 5.2}$$

$$b_t = \phi b_{t-1} + \alpha\beta\varepsilon_t \qquad \text{Ecuación 5.3}$$

$$s_t = s_{t-h} + \gamma\varepsilon_t \qquad \text{Ecuación 5.4}$$

Donde:

l_{t-1}: Es el nivel del instante anterior.

b_{t-1}: Es la tendencia del instante anterior.

α, β y γ: Son los parámetros de sintonización del triple suavizamiento exponencial.

ε_t: Se define como el error de estimación.

ϕ: Es el amortiguamiento de la tendencia que a su vez es otro parámetro a estimar.

s_{t-h}: Es la estacionalidad en el instante $t - h$.

De esta forma quedan listas las ecuaciones de pronóstico para representarlas en espacio de estados definida por las Ecuaciones 5.5 y 5.6:

$$Y_t = \mathrm{Z}x_t + \varepsilon_t \qquad \text{Ecuación 5.5}$$

$$x_{t-1} = Tx_t + \mathrm{R}\eta_t \qquad \text{Ecuación 5.6}$$

Donde:

Y_t: Es el vector de medidas.

Z: Es la matriz de selección de estados.

x_t: Es el vector de estados en el instante actual cuyos estados están formado por $x_t = [l_t \quad b_t \quad s_t \quad s_{t-1} \quad s_{t-2} \quad s_{t-3} \quad s_{t-4} \quad s_{t-5} \quad s_{t-6}]^T$ para una estacionalidad semanal.

ε_t: Es el término del error.

x_{t-1}: Es el vector de estados del siguiente instante.

T: Es la matriz característica del sistema.

R: Es una matriz que define como se incorpora el ruido de medición en la ecuación, llamada matriz de ruidos o perturbaciones.

La Ecuación 5.6 para una estacionalidad semanal queda definida:

$$\begin{bmatrix} l_t \\ b_t \\ s_t \\ s_{t-1} \\ s_{t-2} \\ s_{t-3} \\ s_{t-4} \\ s_{t-5} \\ s_{t-6} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} l_{t-1} \\ b_{t-1} \\ s_{t-1} \\ s_{t-2} \\ s_{t-3} \\ s_{t-4} \\ s_{t-5} \\ s_{t-6} \\ s_{t-7} \end{bmatrix} + \begin{bmatrix} \alpha \\ \alpha\beta \\ \gamma \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \eta_t \qquad \text{Ecuación 5.7}$$

Con este planteamiento en espacio de estados en función del error de estimación ηt, se puede plantear una búsqueda de parámetros α, β y γ dentro de la matriz R que minimicen el error de estimación a fin de encontrar el mejor modelo de Holt&Winters con tendencia y estacionalidad aditivas (H&W(AA)) para los datos de entrenamiento.

Para un modelo H&W(AM) la estimación sería:

$$\mu_t = (l_{t-1} + b_{t-1})s_{t-h} \quad \text{Ecuación 5.8}$$

$$l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t / s_{t-h} \quad \text{Ecuación 5.9}$$

$$b_t = \phi b_{t-1} + \alpha\varepsilon_t / s_{t-h} \quad \text{Ecuación 5.10}$$

$$s_t = s_{t-h} + \gamma\varepsilon_t / (l_{t-1} + b_{t-1}) \quad \text{Ecuación 5.11}$$

El modelo matemático de Holt & Winters proporciona una sólida base para representar series de tiempo con múltiples componentes dinámicos, como el nivel, la tendencia y la estacionalidad, dentro de un espacio de estados. Esta formulación permite utilizar herramientas avanzadas de álgebra matricial para la estimación y el control del sistema, optimizando los parámetros para minimizar el error de estimación. Sin embargo, en situaciones donde las series de tiempo presentan alta volatilidad o cambios abruptos, como se explorará en la siguiente sección, la capacidad del modelo puede verse limitada. Para abordar esta problemática, se introduce el Filtro de Kalman, que complementa y potencia el enfoque al proporcionar una medida en tiempo real de la incertidumbre, permitiendo ajustes dinámicos de los parámetros y una mayor adaptabilidad a las variaciones en la serie de tiempo.

5.2.4. Filtro de Kalman

En [Smyl, 2020], se descubrió que, al tratar con series de tiempo complejas, el parámetro de la tendencia dejaba de ser válido en las partes más volátiles de la serie de tiempo. Su solución fue introducir un modelo LSTM dilatado y modificado para ajustarse a diferentes estacionalidades, con el propósito exclusivo de pronosticar la tendencia. Sin embargo, esta estrategia, aunque efectiva, depende de validación cruzada con diferentes series de tiempo que describen fenómenos físicos parecidos, entre otras técnicas. En este trabajo de tesis, nos enfocamos en métodos de pronóstico cuya única información disponible es la serie de tiempo en sí misma.

El Filtro de Kalman proporciona una medida de la incertidumbre asociada al nivel, tendencia y estacionalidad, permitiendo la resintonización en línea de los parámetros para ajustarse a nuevos eventos conforme son medidos. Este método utiliza el modelo matemático establecido por las Ecuaciones de medición (5.5) y de estados (5.6) para formar la planta presentada en la Figura 5.3. Procedemos a definir la matriz de varianzas y covarianzas del ruido de observación H_k, descrita en la Ecuación 2.4, que representa la incertidumbre en las observaciones. Si asumimos que las observaciones presentan un ruido independiente y distribuido normalmente, H_k se define como la varianza del error de observación ε_k de la ecuación de observación. En este caso, se inicializa como la varianza de los residuos del modelo de suavizamiento exponencial con respecto a la serie de tiempo original.

Por otro lado, la matriz de varianzas y covarianzas del ruido intrínseco Q_k, representa el ruido inherente en la evolución del proceso (en el vector de estados). Si asumimos que el ruido del proceso es independiente para cada componente del estado y tiene varianza constante, Q_k será una matriz diagonal con estas varianzas. Inicialmente, asumimos una varianza pequeña para el nivel y la tendencia, y una mayor varianza para la estacionalidad, proporcionando al modelo más flexibilidad para realizar estimaciones, como se describe en la Ecuación 5.12:

$$Q_k = \begin{bmatrix} Q_{trend} & 0 \\ 0 & Q_{seasonal} \end{bmatrix} = \begin{bmatrix} \alpha^2 & \alpha\beta & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha\beta & \beta^2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \gamma^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \gamma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \gamma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma^2 \end{bmatrix} \quad \text{Ecuación 5.12}$$

Para las condiciones iniciales, es necesario definir el vector de estados inicial $x0$, el cual contendrá las estimaciones iniciales de los componentes del estado. En este método, estas

estimaciones se basan en los primeros valores conocidos de la serie de tiempo. Además, se debe establecer la matriz de covarianza inicial $P0$, que representará la incertidumbre asociada a las estimaciones iniciales del estado.

5.2.5. Resultados

Mediante una búsqueda de cuadrícula, se determinó el modelo de Holt & Winters (H&W) óptimo para la serie de tiempo de nuevos casos de COVID-19 en Brasil, representado en la Figura 5.4.

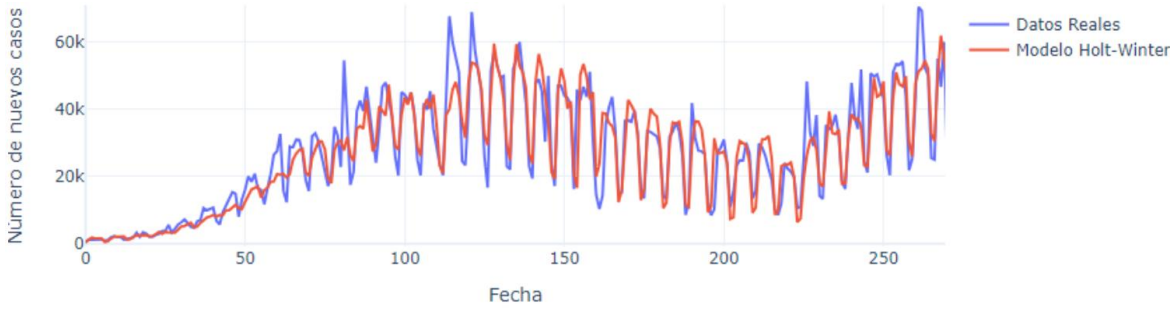

Figura 5.4. Comparación entre los datos reales y las predicciones del modelo $H\&W\,(AA)$

La línea azul muestra los datos reales, mientras que la línea roja representa el ajuste del modelo durante el entrenamiento, con los parámetros de suavizamiento α=0.2171, β=0.0001, y γ=0.2505. Las condiciones iniciales establecidas fueron l_0=955.067, b_0=77.788, y los valores estacionales iniciales s_{t-h} = [−646.555, −48.198, 509.088, 331.081, 278.373, 303.445, −727.234].

Con este modelo matemático, se procedió a implementar el filtro de Kalman, que consta de dos fases principales: la fase de pronóstico, en la que se calcula el estado y la incertidumbre (covarianza) asociada, y la fase de actualización, donde las predicciones son ajustadas con base en nuevas observaciones. Para esta implementación, se utilizaron los parámetros del modelo H&W como valores iniciales para el filtro de Kalman, incluyendo los coeficientes de suavizamiento y los valores iniciales del nivel, tendencia y estacionalidad.

Aunque el filtro de Kalman es sensible a la configuración inicial, este tiende a converger hacia los estados y covarianzas reales conforme se observan más datos. Por lo tanto, los estados iniciales fueron definidos como $x_0 = [l_0, b, s_0, s_1,...,s_6]$, y se asumió una alta incertidumbre inicial representada por $P_0 = I \cdot x_0 \cdot 1000$.

Es crucial destacar que, aunque los estados estimados del modelo abarcan todos los componentes (l_t, b_t, y s_t), solo el nivel (l_t), la tendencia (b_t) y la primera componente estacional (s_0) contribuyen directamente a la salida, como se describe en la Ecuación 5.1. La matriz de selección de salidas se define d0e la siguiente manera en la Ecuación 5.13:

$$Z = [1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0] \qquad \text{Ecuación 5.13}$$

Entonces, la salida estimada (y_{est}) se calcula como: $y_{est} = Z * x_{est}$. Mientras que la incertidumbre (σ) se estima considerando únicamente los componentes l_t, b_t y s_0 según la Ecuación 5.14.

$$\sigma = (Var(l_t) + Var(b_t) + Var(s_0) + 2 * Cov(l_t b_t) + 2 * Cov(l_t s_0) + 2 * Cov(b_t s_0))^{\frac{1}{2}} \qquad \text{Ecuación 5.14}$$

En esta ecuación, Var(lt), Var(bt) y Var($s0$) son las varianzas de los estados l_t, b_t y s_0, respectivamente, obtenidas de los elementos diagonales de la matriz de covarianzas P. Por otro lado, las $Cov(l_t b_t)$, $Cov(l_t s_0)$ y $Cov(b_t s_0)$ representan las covarianzas entre estos estados, extraídas de los elementos fuera de la diagonal de P. Estas covarianzas se multiplican por dos debido a la simetría de la matriz P.

En la Figura 5.2.2 se muestra el desempeño de las estimaciones obtenidas con nuestro modelo en espacio de estados para un H&W(AA), utilizando una matriz Q_k conservadora que prioriza el modelo sobre las mediciones. El análisis comprende las fechas del 30 de marzo al 17 de mayo de 2020, donde la línea azul representa los datos reales de nuevos casos de COVID-19 registrados en Brasil. En contraste, la línea roja muestra las estimaciones del método KF-H&W, mientras que la sombra gris refleja la incertidumbre del pronóstico,

correspondiente a la desviación estándar calculada a partir de la matriz de covarianzas P, como se describe en la Ecuación 5.14.

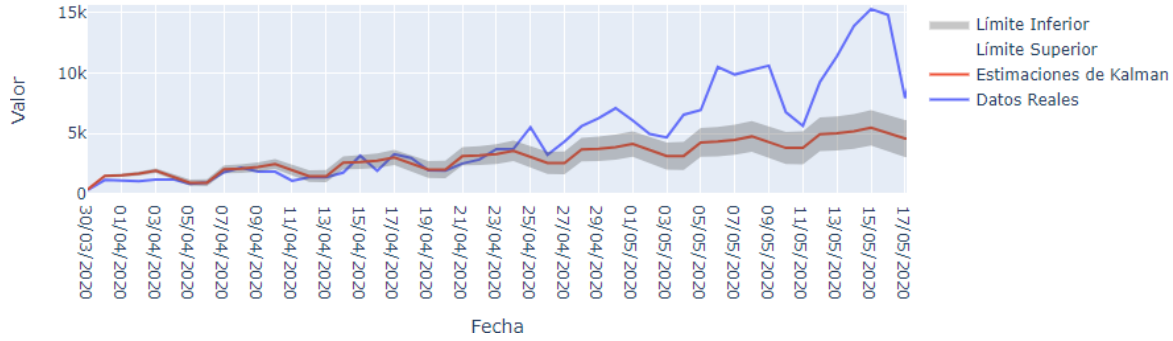

Figura 5.5. Comparación de datos reales contra estimaciones de Kalman con incertidumbre Q_k conservadora.

Dentro de los parámetros de sintonización del Filtro de Kalman, se encuentran la matriz de ruido de medición (R_k) y la matriz de ruido del proceso (Q_k). Estas matrices se ajustan generalmente mediante un análisis de los residuos del modelo, lo que permite determinar si el ruido proviene principalmente del instrumento de medición o si es intrínseco al proceso.

En este caso, la sintonización de Q_k se realiza en función de los residuos del modelo Holt & Winters (H&W(AA)) utilizado para describir el sistema, como se define en la Ecuación 5.15:

$$Q_k = \begin{bmatrix} Q_{trend} & 0 \\ 0 & Q_{seasonal} \end{bmatrix}$$

$$= \begin{bmatrix} var_{res} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & var_{res} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{var_{res}}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{var_{res}}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{var_{res}}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{var_{res}}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{var_{res}}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{var_{res}}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{var_{res}}{2} \end{bmatrix} \quad \text{Ecuación 5.15}$$

Donde:

Q_k: Es la matriz de ruido del proceso.

Q_{trend}: Es la matriz de la tendencia y el nivel.

$Q_{seasonal}$: Es la matriz de ruido de la estacionalidad.

var_{res}: Es la varianza de los residuos calculados al sustraer la serie de tiempo real de la estimada por el modelo simple de H&W(AA).

Para esta configuración, se establece un valor grande para el ruido asociado al nivel y la tendencia (Q_{trend}), debido a la elevada varianza observada en estos componentes. En contraste, el ruido de la estacionalidad ($Q_{seasonal}$) se define como la mitad de dicha varianza. En comparación con configuraciones previas, esta Q_k presenta valores significativamente mayores, lo que implica una mayor ponderación de los datos medidos sobre los datos estimados.

La Figura 5.16 ilustra el comportamiento de las estimaciones de nuestro modelo en espacio de estados utilizando H&W(AA) con una configuración ampliada de Q_k. El periodo analizado abarca las fechas del 30 de marzo al 14 de agosto de 2020. En la figura, la línea azul representa los datos reales de nuevos casos registrados de COVID-19 en Brasil, mientras que la línea roja corresponde a las estimaciones del modelo KF -

H&W. La sombra gris indica la incertidumbre asociada al pronóstico, calculada como la desviación estándar de las predicciones, extraída de la matriz de varianzas y covarianzas P según la Ecuación 5.14.

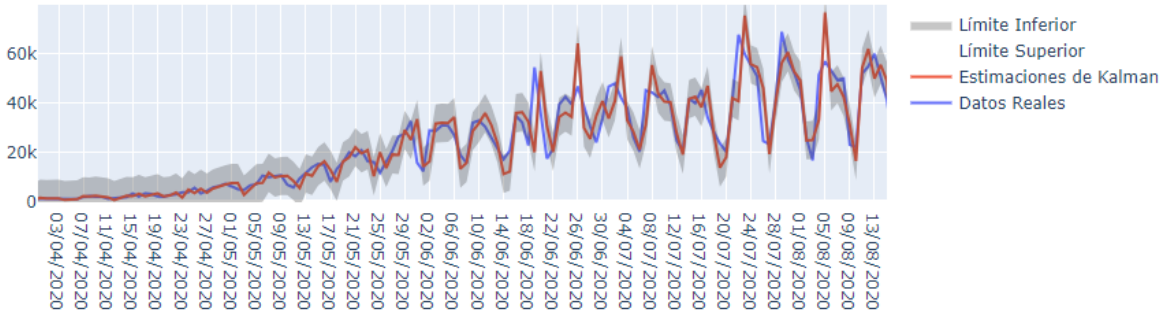

Figura 5.6. Comparación de datos reales contra estimaciones de Kalman con incertidumbre Q_k mayor.

Dado que Q_k es mayor, la incertidumbre inicial en las estimaciones es significativamente superior en comparación con la mostrada en la Figura 5.5, que utiliza una configuración conservadora de Q_k. Sin embargo, esta mayor permite una convergencia más rápida de los estados, reduciendo progresivamente la incertidumbre a medida que se obtienen nuevas observaciones. Esto facilita que el modelo en espacio de estados realice predicciones más precisas para los estados futuros.

Es importante destacar que el modelo matemático H&W(AA) no es válido para toda la serie de tiempo. Por lo tanto, debe ser actualizado con alguna de sus variantes, como H&W(AM), H&W(MA) o H&W(MM), y acompañado de una resintonización de los parámetros α, β, γ y ϕ, según las características de la serie.

Para complementar el análisis visual presentado en la Figura 5.6 y atender las observaciones de los revisores, se calcularon las métricas MAPE, RMSE y DA para comparar el desempeño del modelo con dos configuraciones distintas de Q_k.:

Configuración de Q_k.	**MAPE (%)**	**RMSE**	**DA (%)**
Conservadora (Figura 5.5)	12.83	5487	72.5
Ajustada (Figura 5.6)	10.96	4372	77.5

Tabla 5.4 Comparaciones de sintonización utilizando Q_k.

Estos resultados muestran que incrementar Q_k mejora significativamente la precisión del modelo en términos de MAPE y RMSE, además de aumentar el porcentaje de direcciones acertadas (DA). Esta mejora se atribuye a la mayor adaptabilidad del modelo ante cambios abruptos en la dinámica de la serie de tiempo.
Cabe destacar que, si bien una configuración más grande de Q_k introduce inicialmente una mayor incertidumbre, esta se ve compensada por una convergencia más rápida del filtro, lo cual reduce la incertidumbre conforme se incorporan nuevas observaciones.

Gracias a la formulación del modelo en espacio de estados y en función del error de predicción, la sintonización del Filtro de Kalman puede aprovechar herramientas avanzadas de control moderno (Control Proporcional, Integral y Derivativo, Estabilidad de Lyapunov, Lógica Difusa, entre otras), aprendizaje automático (Descenso del Gradiente, LightGBM, XGBoost, RFR, etc.) y aprendizaje profundo (LSTM, CNN, GRU, Transformers, entre otros).

5.1 Conclusiones

Las series de tiempo, especialmente aquellas asociadas a fenómenos complejos como epidemias, suelen contener un alto nivel de ruido que dificulta el desempeño de los métodos de pronóstico tradicionales. La técnica de filtrado aplicada en este trabajo no solo mitiga este problema, sino que permite descomponer la serie en dos componentes: una serie filtrada y

una serie de residuales. Esta descomposición posibilita el desarrollo de pronósticos especializados y optimizados para cada componente, maximizando la precisión de las predicciones.

El análisis del estado del arte revela que no existe un método general que ofrezca un desempeño consistente bajo diversas condiciones. Esto se hace evidente al observar cómo un mismo método puede fallar al pronosticar en diferentes contextos, como entre países con características epidemiológicas y poblacionales distintas. Para abordar esta limitación, se propuso el método FMFRA, que selecciona dinámicamente el enfoque más adecuado en función de la cantidad de datos disponibles y las características específicas de la serie de tiempo. Por otro lado, el método FK−H&W introduce una estimación en línea personalizada que permite ajustar los parámetros del modelo de manera dinámica, adaptándose a las necesidades actuales del sistema.

Ambos enfoques han demostrado ser soluciones flexibles y robustas frente a las variaciones y desafíos inherentes al pronóstico de series de tiempo caóticas, marcando un avance significativo respecto a los métodos tradicionales.
complejidad.

6 Conclusiones y Trabajos Futuros

6.1 Conclusiones

En esta tesis se propuso el Método de Pronóstico con Filtros y Análisis Residual (FMFRA) como una estrategia híbrida para abordar el pronóstico de series de tiempo complejas, con aplicaciones particulares al estudio del comportamiento de la pandemia de COVID-19. A través de la combinación de técnicas de filtrado (SMA y Filtro de Kalman), análisis espectral (SSA) y modelos avanzados de aprendizaje profundo (LSTM y CNN), el FMFRA logró una separación efectiva entre componentes principales y residuales de las series, lo cual permitió aplicar modelos especializados a cada parte.

Los resultados obtenidos muestran que FMFRA supera a métodos clásicos y modernos como XGBoost, Random Forest, Kalantari y SSA puro, en métricas clave como MAPE, RMSE y porcentaje de direcciones acertadas (DA), especialmente en los primeros meses de la pandemia, cuando la disponibilidad de datos era limitada. Esta superioridad se evidenció tanto en el componente basado en aprendizaje profundo como en su variante basada en SSA, adaptándose según la disponibilidad y calidad de los datos de cada país.

Una solución clave presentada en este trabajo es la repetición del entrenamiento del modelo LSTM o CNN hasta alcanzar configuraciones óptimas, lo que mejora significativamente los pronósticos al seleccionar los pesos más adecuados dentro de un conjunto de soluciones. Sin embargo, este enfoque también tiene limitaciones prácticas, ya que el costo computacional crece rápidamente con el número de repeticiones. Se identificó un número óptimo de 15 repeticiones, considerando que el pronóstico se realiza sobre una señal previamente filtrada, lo que reduce la necesidad de aprender interacciones complejas causadas por el ruido o los datos atípicos.

Los filtros que demostraron mayor eficacia en este trabajo fueron SMA3, SMA7 y un Filtro de Kalman basado en una caminata aleatoria simple. No obstante, se señala que otras técnicas de filtrado, como Holt-Winters y SARIMA, también tienen potencial cuando el análisis de residuales no es necesario. Estos métodos, aunque efectivos en contextos específicos, son menos recomendables cuando se aplica el análisis de residuales, el cual es un punto central de este trabajo.

Se destaca que la integración de métodos de filtrado con análisis de residuales y aprendizaje profundo es particularmente robusta frente a las complejidades de las series de tiempo. Esta metodología puede ser ampliada para abordar problemas aún más complejos y explorar el uso de modelos híbridos que combinen técnicas clásicas y modernas.

6.2 Recomendaciones y Trabajo Futuro

Para trabajos futuros, se propone el desarrollo de un Filtro de Kalman basado en una triple caminata aleatoria, similar al modelo matemático de Holt-Winters. Este enfoque podría combinar las ventajas del suavizado exponencial con la robustez al ruido que ofrece el Filtro de Kalman. Además, se recomienda formular este filtro en función del error, permitiendo su representación en espacio de estados, de forma análoga al modelo SARIMAX, pero con un enfoque más preciso y adaptable al comportamiento dinámico de la serie de tiempo.

Finalmente, el avance hacia métodos de pronóstico más flexibles y adaptativos puede abrir nuevas oportunidades en el análisis de series de tiempo altamente caóticas y de múltiples dimensiones. La combinación de técnicas de aprendizaje profundo, análisis de residuales y modelos estadísticos avanzados puede consolidarse como una herramienta esencial en disciplinas que requieren alta precisión y capacidad de adaptación frente a datos volátiles.

[10]; como ejemplo, la figura 2-1 corresponde al grafico de una serie de tiempo correspondiente al precio de un activo. [10]

Glosario

LanTI.- Laboratorio Nacional de Tecnologías de la Información

CONACYT.- Consejo Nacional de Ciencia y Tecnología.

Spyder IDE.- The Scientific Python Development Environment

Bibliografía

[1] Alferov, Y. V., & Klimova, E. G. (2020). Experience of using the Kalman filter to correct numerical forecasts of surface air temperature. Hydrometeorological Research and Forecasting, 4(378), 28–42. https://doi.org/10.37162/2618-9631-2020-4-28-42

[2] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Springer International Publishing. https://doi.org/10.1186/s40537-021-00444-8

[3] Baptista, M., Sankararaman, S., Medeiros, I. P., Nascimento, C., Prendinger, H., & Henriques, E. M. P. (2018). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. Computers and Industrial Engineering, 115, 41–53. https://doi.org/10.1016/j.cie.2017.10.033

[4] Bézenac, E. de, Rangapuram, S. S., Benidis, K., Bohlke-Schneider, M., Kurle, R., Stella, L., ... & Januschowski, T. (2020). Normalizing Kalman Filters for Multivariate Time Series Analysis. Advances in Neural Information Processing Systems, 33(NeurIPS).

[5] Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation. International Journal of Forecasting, 32(2), 303–312. https://doi.org/10.1016/j.ijforecast.2015.07.002

[6] Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting. Springer. https://doi.org/10.1007/978-3-319-70338-1

[7] Chae, Y. T., Horesh, R., Hwang, Y., & Lee, Y. M. (2016). Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings, 111, 184–194. https://doi.org/10.1016/j.enbuild.2015.11.045

[8] Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons and Fractals, 135. https://doi.org/10.1016/j.chaos.2020.109864

[9] de Livera, A. M., Hyndman, R. J., & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. Journal of the American Statistical Association, 106(496), 1513–1527. https://doi.org/10.1198/jasa.2011.tm09771

[10] Farsadi, M., Mohammadzadeh Shahir, F., & Babaei, E. (2018). Power system states estimations using Kalman filter. 10th International Conference on Electrical and Electronics Engineering (ELECO 2017), 808–812. https://doi.org/10.1109/EPEC.2013.6802979

[11] Frausto-Solís, J., Hernández-González, L. J., González-Barbosa, J. J., Sánchez-Hernández, J. P., & Román-Rangel, E. (2021). Convolutional Neural Network Component Transformation (CNN CT) for Confirmed COVID-19 Cases. Mathematical and Computational Applications, 26(2), 29. https://doi.org/10.3390/mca26020029

[12] Golyandina, N. (2020). Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing. Wiley Interdisciplinary Reviews: Computational Statistics, 12(4). https://doi.org/10.1002/wics.1487

[13] Hanggara, F. D. (2021). Forecasting Car Demand in Indonesia with Moving Average Method. Journal of Engineering Science and Technology Management, 1(1), 1–6. https://jes-tm.org/index.php/jestm/index

[14] Ju, J., Liu, K., & Liu, F. Prediction Of SO_2 Concentration Based On AR-LSTM Neural Network.

[15] Kang, Y., Hyndman, R. J., & Smith-Miles, K. (2017). Visualising forecasting algorithm performance using time series instance spaces. International Journal of Forecasting, 33(2), 345–358. https://doi.org/10.1016/j.ijforecast.2016.09.004

[16] Khairalla, M., Ning, X., & AL-Jallad, N. (2018). Modelling and optimisation of effective hybridisation model for time-series data forecasting. The Journal of Engineering, 2018(2), 117–122. https://doi.org/10.1049/joe.2017.0337

[17] Klein, N., Smith, M. S., & Nott, D. J. (2020). Deep Distributional Time Series Models and the Probabilistic Forecasting of Intraday Electricity Prices. arXiv preprint. https://arxiv.org/abs/2010.01844

[18] López-Ruiz, R., Mancini, H. L., & Calbet, X. (1995). A statistical measure of complexity. Physics Letters A, 209(5–6), 321–326. https://doi.org/10.1016/0375-9601(95)00867-5

[19] Luo, J., Zhang, Z., Fu, Y., & Rao, F. (2021). Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. Results in Physics, 27, 104462. https://doi.org/10.1016/j.rinp.2021.104462

[20] Mbaye, A., Ndong, J., Ndiaye, M. L., Sylla, M., Aidara, M. C., Diaw, M., Ndiaye, M. F., Ndiaye, P. A., & Ndiaye, A. (2018). Kalman filter model, as a tool for short-term

forecasting of solar potential: Case of the Dakar site. E3S Web of Conferences, 57. https://doi.org/10.1051/e3sconf/20185701004

[21] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

[22] Makridakis, S., Assimakopoulos, V., & Spiliotis, E. (2018). Objectivity, reproducibility, and replicability in forecasting research. International Journal of Forecasting, 34(4), 835–838. https://doi.org/10.1016/j.ijforecast.2018.05.001

[23] Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., & Talagala, T. S. (2020). FFORMA: Feature-based forecast model averaging. International Journal of Forecasting, 36(1), 86–92. https://doi.org/10.1016/j.ijforecast.2019.02.011

[24] Moghaddam, A. H., Moghaddam, M. H., & Esfandyari, M. (2016). Predicción del índice del mercado bursátil utilizando una red neuronal artificial. Journal of Economics, Finance and Administrative Science, 21(41), 89–93. https://doi.org/10.1016/j.jefas.2016.07.002

[25] Nagbe, K., Cugliari, J., & Jacques, J. (2018). Short-term electricity demand forecasting using a functional state space model. Energies, 11(5), 1–24. https://doi.org/10.3390/en11051120

[26] Othman, A., Aziz, A. A. A., Ahmad, N. A., Mohd, M. H., & Adam, S. I. M. (2021). Analysing Trends and Forecasting of COVID-19 Pandemic in Malaysia using Singular Spectrum Analysis. Journal of Applied Science and Engineering, 37(3), 121–134.

[27] Pankratz, A. (1983). Forecasting with univariate Box. Wiley. ISBN: 0-471-09023-9.

[28] Petropoulos, F., Hyndman, R. J., & Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work? European Journal of Operational Research, 268(2), 545–554. https://doi.org/10.1016/j.ejor.2018.01.045

[29] Rodrigues Júnior, S. E., & Serra, G. L. de O. (2017). A novel intelligent approach for state space evolving forecasting of seasonal time series. Engineering Applications of Artificial Intelligence, 64(December 2016), 272–285. https://doi.org/10.1016/j.engappai.2017.06.016

[30] Sememoglou, A. A., Spiliotis, E., Makridakis, S., & Assimakopoulos, V. (2021). Investigating the accuracy of cross-learning time series forecasting methods. International Journal of Forecasting. https://doi.org/10.1016/j.ijforecast.2020.11.009

[31] Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting, 36(1), 75–85. https://doi.org/10.1016/j.ijforecast.2019.03.017

[32] Snyder, R. D., Koehler, A. B., & Ord, J. K. (2002). Forecasting for inventory control with exponential smoothing. International Journal of Forecasting, 18(1), 5–18. https://doi.org/10.1016/S0169-2070(01)00109-1

[33] Talagala, T. S., Hyndman, R. J., & Athanasopoulos, G. (2018). Meta-learning how to forecast time series. Monash Econometrics and Business Statistics Working Papers.

[34] Xu, X., & Ren, W. (2019). A hybrid model based on a two-layer decomposition approach and an optimized neural network for chaotic time series prediction. Symmetry, 11(5). https://doi.org/10.3390/sym11050610

[35] Zain, Z. M., & Alturki, N. M. (2021). COVID-19 Pandemic Forecasting Using CNN-LSTM: A Hybrid Approach. Journal of Control Science and Engineering. https://doi.org/10.1155/2021/8785636
[36] Zhong, L., Mu, L., Li, J., Wang, J., Yin, Z., & Liu, D. (2020). Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model. IEEE Access, 8, 51761–51769. https://doi.org/10.1109/ACCESS.2020.2979599