MODELADO Y RECUPERACIÓN DE REGISTROS DE SERIES TEMPORALES GPS USANDO LONG SHORT-TERM MEMORY (LSTM)

Axel Nicolás Salamanca Castillo

Informe de Habilitación Profesional presentado al

Departamento de Ciencias Geodésicas y Geomática

Universidad de Concepción, Campus Los Ángeles

En cumplimiento del requisito parcial

Para obtener el título de

Ingeniero Geomático

Escrito bajo la orientación del profesor

Mg. Aharon Cuevas Cordero (Dpto. Ciencias Geodésicas y Geomática)

Aprobado por la comisión

Dr. Henry Montecino Castro (Dpto. Ciencias Geodésicas y Geomática)

Mg. Andreaw Stheven Rifo Pereira (Dpto. Ciencias Geodésicas y Geomática)

Los Ángeles

Marzo, 2025

RESUMEN

El uso de series de tiempo GPS para estudios de diferentes fenómenos, se ve condicionado debido a la presencia de discontinuidades y a la ausencia de datos en los registros, provocados principalmente por fenómenos físicos, tales como, cargas hidrológicas y mareas terrestres, pero también, por causas relacionadas con el sistema de adquisición de datos, tales como, cambios de antenas de los receptores y fallas en los dispositivos de almacenamiento. Debido a la naturaleza estocástica de estos fenómenos, los enfoques tradicionales para el modelamiento de series temporales GPS no logran exitosamente su propósito. Así, en el presente estudio se evalúa el uso del algoritmo Long Short-Term Memory (LSTM) para modelar series temporales GPS, con el propósito de completar datos faltantes de los registros, debido a la capacidad del algoritmo de conservar información relevante a lo largo de la secuencia recibida, y de esta manera permitiendo predicciones a largo plazo. Se diseñaron diferentes estrategias para modelar series y se evaluaron utilizando un conjunto de estaciones de medición continua GPS de Brasil. Para el entrenamiento de los modelos, se utilizaron como entrada datos de carga hidrológica, movimiento del polo, temperatura superficial, presión atmosférica, y efemérides del sol y la luna. Los resultados muestran errores de validación de RMSE=5mm, MAE=4mm y un R^2=0.85 en la componente *up*. Por otra parte, utilizando el mecanismo de atención *Feed Forward Attention Mechanism* (FFAM) como capa complementaria de los modelos LSTM, los rendimientos incrementan en un 0,6% en el RMSE al utilizarse en un modelo regresor, siendo este poco significativo en modelos simples. Es importante destacar que la calidad de los modelos obtenidos está condicionada a la cantidad de registros disponibles

de la serie temporal, como también el comportamiento de esta, donde los saltos y ruido no debidamente tratados incurren en modelos con alto error.

DEDICATORIA

Esta tesis está dedicada a mis familiares que me han apoyado a lo largo de este camino, en especial a mi padre, mi madre y mi hermana por su ayuda, paciencia y dedicación que me brindaron durante mis años de educación.

AGRADECIMIENTOS

Quiero ofrecer mi más sincero agradecimiento a todas aquellas personas que han contribuido de alguna manera en la realización de esta tesis. En primer lugar, agradezco al profesor Aharon Cuevas por su guía experta y su orientación desde el punto de vista de un experto en la materia.

También, agradezco a aquellos servicios que han proporcionado los datos usados en esta investigación, siendo estos Nevada Geodetic Laboratory, National Center of Environmental Prediction, Jet Propulsion Laboratory, International Earth Rotation and Reference Systems Service y Centro Alemán de Geo-investigaciones. Ya que sin los datos proporcionados por estos servicios este proyecto no sería posible.

Además, agradezco a mis familiares y amigos que han aportado su comprensión y aliento durante el desarrollo de esta investigación, en especial a mi madre Bernarda Castillo, cual apoyo fue fundamental para la culminación de este proyecto.

ÍNDICE DE CONTENIDOS

Lista de Tablas

Lista de Figuras

Lista de Símbolos, Nomenclatura o Abreviaciones

AP: Presión Atmosférica

DL: *Deep Learning*

DECs: Declinación del Sol con respecto a la Tierra

DECm: Declinación de la Luna con respecto a la Tierra

DELs: Rango Aparente del Sol con respecto a la Tierra

DELm: Rango Aparente de la Luna con respecto a la Tierra

FFAM: *Feed Forward Attention Mechanism*

GPS: *Global Positioning System*

GNSS: *Global Navigation Satellite Systems*

HYD: Carga Hidrológica

LSTM: L*ong Short-Term Memory*

MM.CC: Mínimos Cuadrados

ML: *Machine Learning*

NGL: Nevada Geodetic Laboratory

Px: Movimiento del Polo en el eje X

Py: Movimiento del Polo en el eje Y

RAs: Ascensión Recta del Sol con respecto a la tierra

RAm: Ascensión Recta de la Luna con respecto a la tierra

RNN: *Recurrent Neural Network*

TEM: Temperatura Superficial

1. Introducción

Las series temporales Global Positioning System (GPS) consisten en registros de datos de posicionamiento recopilados continuamente a lo largo del tiempo mediante receptores GPS. Estas series son utilizadas en la realización de marcos de referencia terrestres, la detección de deformaciones de la corteza terrestre, como también para la investigación del cambio climático y la dinámica de los glaciares (Blewitt, 1997).

Una de las principales dificultades en el modelado de series temporales GPS es el manejo de las variaciones estacionales y no estacionales causadas por factores geofísicos, tales como, mareas terrestres sólidas, cargas atmosféricas, variaciones hidrológicas y expansiones térmicas del lecho rocoso, pero, además, por variación producto de fenómenos locales, entre los cuales se pueden identificar, la actividad tectónica y los terremotos (Nordman, 2010). Otras dificultades están relacionadas con el funcionamiento de los sistemas de adquisición de datos que provocan falta de datos y discontinuidades en la serie temporal, debido principalmente a cambios de antenas de los receptores o mal funcionamiento del equipamiento (*e.g.*, baterías de los receptores).

Tradicionalmente, las series temporales GPS son modeladas utilizando modelos de trayectorias con parámetros estimados mediante mínimos cuadrados (Blewitt & Lavallée, 2002), y más recientemente, utilizando modelos robustos como MIDAS (Blewitt, Hammond, & Kreemer, 2018), empleado, por ejemplo, por el centro de procesamientos Nevada Geodetic Laboratoy (NGL) (http://geodesy.unr.edu/NGLStationPages/GlobalStationList). Sin embargo, este

procedimiento no está exento de errores residuales (Heflin, et al., 2020) y limitaciones al momento de modelar series temporales con ausencia de datos (Blewitt, Hammond, & Kreemer, 2018). Para tratar con estos aspectos, se han empleado abordajes alternativos para modelar series temporales GPS. Por ejemplo, en Luo, Mayor & Heck (2012) se utilizaron los algoritmos *AutoRegressive Moving Average* (ARMA) y *AutoRegressive Integrated Moving Average* (ARIMA) para el modelado y correlación temporal de series temporales GPS, en Ji & Shen (2020) se utilizaron funciones de Wavelet para el modelado de series temporales GNSS con el fin de detectar valores atípicos. En Jiao et al., (2024) se utilizó el filtro de Kalman para el modelado de series temporales *Global Navigation Satellite System* (GNSS) y 5G, aprovechando la densidad de antenas emisoras de esta última, y en (Zhu, et al., 2023) se usaron análisis de componentes principales para interpolar datos faltantes en series temporales de vapor de agua precipitable derivados de series temporales GNSS.

En el último tiempo han surgido múltiples abordajes basados en aprendizaje automático (*machine learning* – ML) y aprendizaje profundo (*deep learning* – DL) para modelar series temporales GNSS, en algunos casos utilizando variables adicionales a la serie de datos. Por ejemplo, en Gao et al. (2022) se entrenaron modelos basados en *Gradient Boosting Decision Tree* (GBDT), *Support Vector Machi*ne (SVM) y *Long Short-Term Memory* (LSTM), para predecir series temporales GNSS de la componente *up*. Además, se utilizaron los parámetros de orientación terrestre, efemérides del sol y la luna, datos de temperatura y presión atmosférica, y datos hidrológicos. Para predicciones de periodos menores a seis meses, el modelo basado en GBDT presentó el mejor rendimiento con un RMSE=3 *mm*, mientras que el modelo basado en LSTM obtuvo el mejor rendimiento para

periodos de seis meses, con un RMSE=7 *mm*, siendo LSTM más preciso que GBDT para predicciones realizadas en periodos de tiempo largo. Por otro lado, en Alevizakou et al. (2018) se diseñaron modelos predictores *uni-step* y *multi-step* basados en *Recurrent Neural Networks* (RNN) para predecir las componentes XYZ de las series temporales GNSS de 1000 estaciones de medición continua de Estados Unidos. Los modelos alcanzaron rendimientos con errores (*mean absolute error* – MAE) de 2 *mm* y 2.6 *mm* para las componentes X, Y y Z, respectivamente.

En estudios relacionados con la recuperación de registros faltantes en series temporales GPS, en Wang et al. (2021) se entrenó un modelo predictor *uni-step* basado en LSTM y se entrenó con los datos de la serie temporal GNSS (*up*) de la estación de medición continua XJSS de China. Utilizando ventanas deslizantes multi-escala de 10, 15 y 20 días, se realizaron predicciones para completar registros faltantes de 1 día, alcanzando un error de predicción de RMSE=3.2 *mm*. Xin et al. (2022), utilizaron registros de series temporales de vibraciones del suelo en zonas cercanas al sensor con datos faltantes. Se evaluaron dos modelos predictivos diferentes: el primero empleó directamente los datos de vibración del suelo de los sensores cercanos, mientras que, en el segundo modelo, estos datos fueron transformados mediante una Transformada de Fourier, el objetivo fue verificar la precisión del modelo predictivo, esperando que ambos enfoques tuvieran una precisión similar. Como resultado, se logró una precisión del 95% en la recuperación de los registros de la serie temporal del sensor objetivo. En Zhang et al. (2019) se recuperaron los registros de series temporales de sensores de calidad del agua a partir de los datos de los sensores en conjunto con datos climatológicos. En dicho trabajo se asumió que los datos faltantes

siguen una distribución *Missing Completely At Random* (MCAR) para interpolar los registros faltantes en las series temporales a modelar, por lo que a partir de estas series con datos interpolados se construyó un modelo *encoder-decoder* con dos capas LSTM, se configuró una ventana deslizante de longitud variable para cada capa paralela del modelo. Usando esta metodología se obtuvo una recuperación de datos de concentración de nitratos en el agua, logrando un desempeño del modelo con errores no significativos.

En el presente estudio, se propone evaluar modelos basados en el algoritmo LSTM para la recuperación de registros de series temporales GPS en la su componente vertical (*up*). Se emplearán las series temporales de estaciones permanentes de Brasil en el periodo de tiempo 2010-2024, debido a que Brasil es un país con una actividad tectónica poco significativa, por lo que el vulcanismo y la sismicidad son factores menos influyentes en los registros de las estaciones GPS.

1.1. Objetivo General

Evaluar modelos basados en *Long Short-Term Memory* (LSTM) para la recuperación de registros de series temporales GPS.

1.2. Objetivos Específicos

- Caracterizar lagunas de datos en los registros de las series temporales GPS.
- Identificar y seleccionar características que proporcionen información relevante para el modelado de series temporales GPS.
- Diseñar e implementar modelos basados en el algoritmo LSTM para el modelado de series temporales GPS.

- Evaluar el rendimiento de los modelos para la recuperación de registros de series temporales GPS.

2. Marco Teórico

2.1. Modelado de series temporales GPS

De acuerdo con Bos et al. (2020) las series temporales geodésicas representan un conjunto de mediciones continuas o discretas. Para el caso de las series de tiempo GPS, corresponden con la posición de una estación GPS, resultado del seguimiento constante de satélites GPS y reflejan cambios en las coordenadas, usualmente geográficas (latitud, longitud, y elevación) o en un sistema topocéntrico (este, norte y up); debido a movimientos tectónicos, deformaciones de la corteza, subsidencia, levantamientos, y otros procesos geodinámicos.

Las series temporales GPS pueden ser descritas mediante dos tipos de modelados, los modelos determinísticos y los modelos estocásticos. En los modelos determinísticos, las series temporales GPS son descritas por modelos cinemáticos teniendo por características, la tendencia, la estacionalidad y saltos (Bevis, Bedford, & Caccamise II, 2020), como se expresa en la ecuación (*1*).

$$x(t) = x_{trend} + x_{jumps} + x_{cycle} \quad (1)$$

Los saltos se describen como combinaciones lineales de funciones escalón de Heaviside en tiempos de salto prescritos t_j. El número de saltos, n_j, puede ser cero, uno o más. Los saltos incluyen saltos cosismicos, que son movimientos reales del terreno, y saltos 'artificiales' asociados con cambios en la antena GPS y/o su rotódomo (también llamado radomo), cambios en el monumento de la antena, entre otros. Casi todas las series temporales de GPS exhiben un ciclo estacional de desplazamiento que puede modelarse

como una serie de Fourier de 4 términos con periodicidades de 1 y 0.5 años (Bevis, Bedford, & Caccamise II, 2020). El modelo de trayectoria más común invoca una tendencia de velocidad constante (Bevis, Bedford, & Caccamise II, 2020):

$$x(t) = x_R + v(t - t_R) + \sum_{j=1}^{n_j} b_j H(t - t_j) + \sum_{k=1}^{n_F} [s_k \sin(\omega_k t) + c_k \cos \omega_k t)] \quad (2)$$

Donde t_R es un tiempo de referencia arbitrario, a menudo establecido como el tiempo promedio de observación, $x_R = x(t_R)$ es la posición de referencia, y v es el vector de velocidad de la estación, que se asume constante. La función H es la función escalón de Heaviside o función unidad, y el vector b_j describe la dirección y magnitud del salto que ocurre en el momento t_j, y n_j es el número de saltos. Cada una de las tres componentes de la serie temporal posee los vectores s_k y c_k, los cuales son los coeficientes de Fourier para resolver la función armónica con frecuencia angular ω_k, y donde n_F es el número de frecuencias distintas. La frecuencia angular $\omega_k = 2\pi/\tau_k$, donde τ_k es el periodo correspondiente. Para modelar ciclos de desplazamiento anual, se elige el periodo fundamental $\tau_1 = 1$ año, y los periodos de las armónicas superiores $\tau_k = \frac{1}{k}$ años. Esto asegura que el ciclo construido a partir de n_F senos y n_F cosenos (y un total de n_F frecuencias o periodos) se repite solo una vez al año. Casi siempre es adecuado establecer $n_F = 2$, especificando así una serie de Fourier de 4 términos (Bevis, Bedford, & Caccamise II, 2020). Por otro lado, el modelo estocástico es fundamental para capturar los efectos no modelados en el modelo determinístico, describiendo las propiedades estadísticas del vector de observaciones mediante una matriz de covarianza. Una correcta especificación de esta matriz permite obtener la mejor estimación lineal insesgada de los parámetros

desconocidos. Para modelar series temporales GPS, la matriz de covarianza de las observaciones es, en muchos casos, solo parcialmente conocida, por lo que se expresa como una combinación lineal de matrices de cofactores conocidas, cuyas varianzas se estiman mediante el método de estimación de componentes de varianza por mínimos cuadrados (LS-VCE). En el caso de las series temporales de posiciones GNSS, la matriz de covarianza suele modelarse como la suma de ruido blanco y ruido coloreado de ley de potencias, como el flicker y el paseo aleatorio. El método LS-VCE permite estimar iterativamente las amplitudes de estos ruidos y tiene propiedades óptimas que garantizan estimaciones insesgadas y de mínima varianza. No obstante, la estimación simultánea de varios componentes de varianza puede dar lugar a valores negativos, problema que se resuelve con una variante no negativa de LS-VCE (NNLS-VCE). Finalmente, aunque la mayoría de los métodos de estimación de varianza asumen distribuciones normales, el LS-VCE proporciona resultados equivalentes al método REML en estos casos, garantizando así estimaciones óptimas y confiables (Amiri-Simkooei, 2020). Ahora, con las series temporales GPS descritas matemáticamente, es necesario conocer cómo se estiman tradicionalmente los parámetros necesarios para modelar este tipo de series temporales, para esta tarea es común el uso de algoritmos tales como, Mínimos Cuadrados, mientras que algoritmos como MIDAS, ARMA, ARIMA, funciones Wavelet, filtro de Kalman, entre otros, los cuales se basan en estadísticos, son métodos alternativos al modelado convencional. Uno de los más ampliamente usados es Mínimos Cuadrados, este algoritmo se basa en el ajuste de una función matemática a los datos observados de una serie temporal, minimizando la suma de los errores cuadrados entre los valores observados y los valores predichos por el modelo.

Los desplazamientos que afectan los sitios GPS, y, en consecuencia, la posición observada, se clasifican en tres grupos: movimientos mareales, movimientos no mareales y desplazamientos que afectan los puntos de referencia internos dentro de los instrumentos de observación.

Entre los movimientos mareales se encuentran las mareas de tierra sólida, la carga de mareas oceánicas, la carga de presión atmosférica, la marea polar y la carga de marea polar oceánica. Estos fenómenos son removidos con modelos específicos para dicha tarea, aunque no están exento de residuales (Petit & Luzum, 2010). Las cargas mareales pueden ser modeladas con información de ascensión recta del sol (RA_s), la declinación del sol (DEC_s) y rango aparente del sol (DEL_s), la ascensión recta de la luna (RA_m), declinación de la luna (DEC_m) y rango aparente de la luna (DEL_m) (Petit & Luzum, 2010). Los efectos polares (marea y carga) pueden ser descritos por medio de series de datos del movimiento del polo (P_x, P_y). Entre los desplazamientos no mareales se pueden encontrar principalmente dos fenómenos, las cargas hidrológicas (Puskas, Meertens, & Phillips, 2017), y la carga atmosférica no mareal, la cual no siempre es considerada en la remoción de efectos sobre series temporales GPS, pero, podrían ser removidas de las series temporales GPS (Petit & Luzum, 2010). Finalmente, los desplazamientos que afectan a los sitios de referencia (deformaciones del lecho rocoso y de los monumentos sobre los que se ubican las estaciones GPS) y los componentes internos de los instrumentos de observación, son consecuencia de la difusión térmica (Yan, et al., 2010).

2.2. Long Short-Term Memory (LSTM)

El aprendizaje de máquinas (*machine learning* – ML) surge de la necesidad de resolver problemas que operan bajo la descripción de listas de reglas formales y matemáticas. En este contexto, el verdadero desafío surge cuando se desea resolver problemas que se basan en problemas cognitivos complejos, en los cuales no siempre existen reglas claras, las cuales se basan en la experiencia, el aprendizaje y el contexto. A partir de esta problemática surge el aprendizaje profundo (*deep learning* – DL), que es una técnica que permite a las maquinas aprender y comprender la información previa. Estos algoritmos, al aprender de datos previos no se indican formalmente las reglas como en algoritmos ML, con lo que la computadora "aprende" conceptos complejos construyéndolos a partir de otros más simples, es decir, desarrolla ecuaciones complejas a partir de datos previos, con las cuales relaciona la información proporcionada con la predicción realizada (Bengio, Goodfellow, & Courville, 2016)

De este último enfoque general se deprenden las redes neuronales recurrentes (*Recurrent Neural Network* – RNN), las cuales están diseñadas para procesar datos secuenciales como texto, audio y series temporales. A diferencia de las redes tradicionales, las RNNs poseen conexiones recurrentes que les permiten mantener información a lo largo del tiempo, creando un estado de memoria que se actualiza en cada paso temporal. Esto es clave para capturar dependencias contextuales en los datos. Las RNNs cuentan con celdas interconectadas que reciben entradas tanto del estado actual como del estado de memoria anterior, permitiendo procesar datos en secuencia (ver Figura 1). Estas celdas comparten los mismos parámetros a lo largo de todos los pasos temporales, lo que simplifica el modelo

y facilita la generalización. La recurrencia en las conexiones permite actualizar el estado de memoria con cada nueva entrada, recordando información clave de pasos previos en la secuencia. En cada paso, la red recibe la nueva entrada junto con el estado oculto anterior, que contiene pesos, sesgos y características aprendidas de la secuencia. Este diseño hace que las RNN sean una herramienta poderosa para tareas como el análisis de texto y la predicción de series temporales. (Lim & Zohren, 2020).

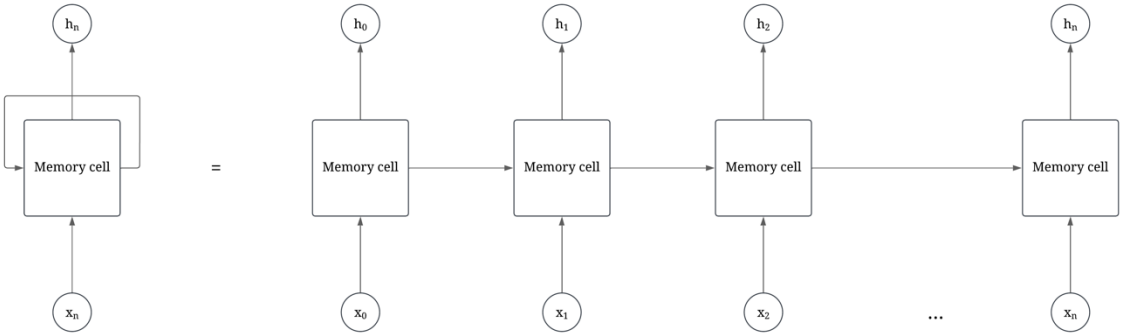


Figura 1. Ejemplo de una Red Neuronal Recurrente
Basado en: (Anonimo, 2015)

Un inconveniente de las RNNs se produce cuando las secuencias de entrada son demasiado grandes, en estos casos, la información inicial se pierde a medida que los pasos de tiempo del modelo se alejan del inicio de la secuencia, lo que ocasiona gradientes explosivos o que se desvanecen, lo que significa que los pesos son demasiado altos o son demasiado pequeños respectivamente, lo que se traduce en predicciones imprecisas de acuerdo con (Lim & Zohren, 2020).

Una red de memoria a corto-largo plazo (*Long Short-Term Memory* – LSTM) es un tipo de RNN para modelar secuencias de datos y resolver problemas de dependencia a largo plazo. Su principal característica es su capacidad para recordar información durante largos períodos, gracias a su estructura interna de "celdas de memoria" que controlan el flujo de

información, permitiendo que la red retenga u olvide datos según sea necesario, que en teoría, soluciona el problema de los gradientes, debido a que este algoritmo mejora el flujo de gradientes dentro de la red, mediante el uso de un estado de celda (c_t), que almacena información a largo plazo (esto quiere decir que la información relevante contenido a lo largo de toda la secuencia es conservada por el algoritmo), modulado a través de una serie de compuertas (Hochreiter & Schmidhuber, 1997).

Las compuertas reciben el nombre de compuerta de entrada, compuerta de olvido y compuerta de salida, donde la celda de memoria (nombre que recibe la unidad básica de un algoritmo LSTM) recuerda valores en cualquier paso de tiempo de la secuencia y las tres compuertas regulan el flujo de información hacia dentro y fuera de la celda. De esta manera, la información útil puede retenerse y la información inútil puede eliminarse (Hochreiter & Schmidhuber, 1997).

En una RNN típica de una capa oculta, para el conjunto de entradas $X = (x_1, x_2, \dots, x_N)$, $H = (h_1, h_2, \dots, h_N)$ es el vector oculto (también conocido como estado oculto) de la celda, las secuencias de salida $\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N)$ son calculadas iterando desde $n = 1$ a N las ecuaciones:

$$h_n = \mathcal{A}(W_{xh}x_n + W_{hh}h_{n-1} + b_h) \quad (3)$$

$$\hat{y}_n = \hat{f}LSTM(x_n) = W_{hy}h_n + b_y \quad (4)$$

Donde $\mathcal{A}$ es la función de la capa oculta, la cual suele ser la función sigmoidal en algoritmos RNN convencionales, W_{xh}, W_{hh} y W_{hy} indican las matrices de pesos entre las entradas y los vectores ocultos, entre diferentes pasos de tiempo de los vectores ocultos, y

entre los vectores ocultos y de salida, respectivamente, y b_h, b_y representan los sesgos correspondientes a W_{hh} y W_{hy}.

Adicionalmente, las ecuaciones de las puertas de entrada, olvido, salida, estado oculto y el estado particular de cada celda para los algoritmos LSTM, están dadas por las siguientes ecuaciones:

$$i_n = \sigma(W_{xi}x_n + W_{hi}h_{n-1} + W_{ci}c_{n-1} + b_i) \quad (5)$$

$$f_n = \sigma(W_{xf}x_n + W_{hf}h_{n-1} + W_{cf}c_{n-1} + b_f) \quad (6)$$

$$o_n = \sigma(W_{xo}x_n + W_{ho}h_{n-1} + W_{co}c_{n-1} + b_o) \quad (7)$$

$$c_n = f_n \odot c_{n-1} + i_n \odot \tanh(W_{xc}x_n + W_{hc}h_{n-1} + b_c) \quad (8)$$

$$h_n = o_n \odot \tanh(c_n) \quad (9)$$

Donde σ es la función sigmoide logística, definida como $\sigma(x) = \frac{1}{1+e^{-x}}$; i_n, f_n, c_n y o_n corresponde a, puerta de entrada (*input gate*), puerta de olvido (*forget gate*), estado de la celda (*cell state*) y puerta de salida (*output gate*), respectivamente; b_i, b_f, b_c y b_o son los correspondientes sesgos y los subíndices de cada matriz de pesos esta dado por la puerta o función a la que están asignados, además, el subíndice n corresponde al número de iteración en el proceso predictivo (Gao, Li, Chen, Jiang, & Feng, 2022), esta estructura puede ser observada en la Figura 2.

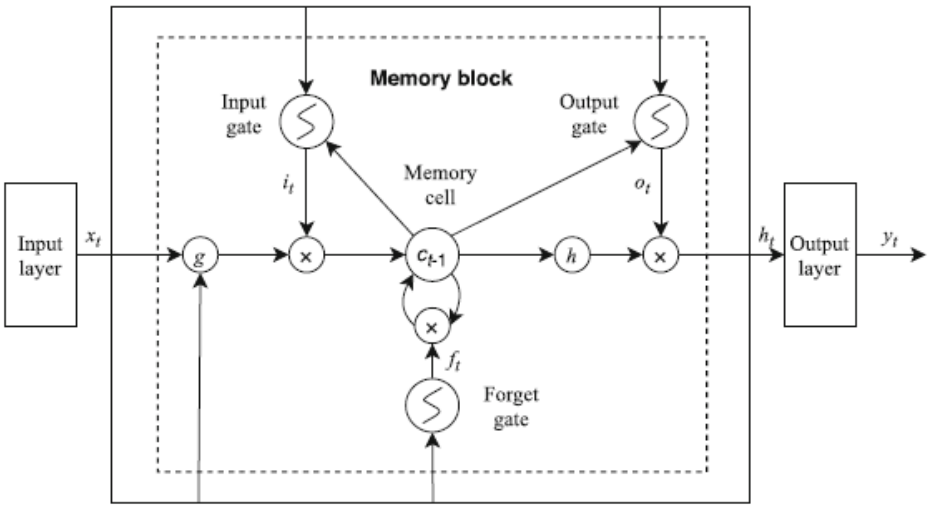


Figura 2. Arquitectura de una red LSTM
Fuente: Gao, Li, Chen, Jiang & Feng (2022)

2.3. Estrategias de ajuste de hiper-parámetros

Los modelos basados en algoritmos del ML o DL utilizan parámetros que no se pueden estimar directamente a partir de los datos, los cuales controlan la precisión de estos, la complejidad, la velocidad de convergencia, el tiempo de entrenamiento o si este tiene o no tiene sobre ajuste. A este tipo de parámetros se les denomina hiper-parámetros porque no existe una fórmula analítica disponible para calcular un valor apropiado (Kuhn & Johnson (2016).

En el caso de modelos basados en LSTM se identifican los siguientes hiper-parámetros: longitud de la entrada, número de neuronas o celdas de estado, número de capas, tasa de abandono (porcentaje de neuronas que se desactivan en un momento aleatorio del entrenamiento, que permite verificar el sobreajuste del modelo), optimizador, tasa de aprendizaje, función de activación, tamaño de lote y número de iteraciones.

En particular, la búsqueda en grilla consiste en evaluar todos los modelos obtenidos a partir de las combinaciones de hiper-parámetros posibles para el universo de hiper-parámetros definido, de esta forma, el modelo con mejor rendimiento definirá los mejores hiper-parámetros. La estrategia conocida como búsqueda aleatoria es una variación de la búsqueda en grilla, la cual busca en distintos puntos aleatorios del espacio de búsqueda de hiper-parámetros, permitiendo reducir el tiempo y los recursos computacionales (Muhammed, 2023). Adicionalmente, para cada entrenamiento/validación se sugiere realizar entrenamiento con validación cruzada, esto, con el propósito de evaluar el desempeño del modelo con todos los datos.

2.4. Feed-Forward Attention Mechanism

Un mecanismo de atención es una técnica utilizada en modelos de aprendizaje profundo para mejorar la capacidad de la red en la selección de información relevante en secuencias de datos, asignando diferentes pesos a distintas partes de la entrada. Su propósito es permitir que el modelo se enfoque en los elementos más importantes para la tarea específica, mejorando la interpretación y eficiencia en el procesamiento de secuencias largas (Zhang, Thorburn, & Xiang, 2019). Existen diversos tipos de mecanismos de atención, entre ellos, *Feed-Forward Attention Mechanism* (FFAM), *Self-Attention*, *Bahdanau Attention* y *Luong Attention*.

Propuesto por Raffel & Ellis (2015), FFAM es un mecanismo de atención utilizado en modelos de aprendizaje profundo, cuya función principal es asignar diferentes pesos a las entradas de la red para mejorar la eficiencia en el procesamiento de secuencias. Este mecanismo busca resolver el problema de la degradación de la información en modelos

secuenciales, específicamente la dificultad que enfrentan las RNNs para mantener información relevante en memoria de largo plazo. A medida que las secuencias se vuelven más largas, la capacidad de la red para recordar datos clave disminuye debido a problemas como el desvanecimiento del gradiente (Hochreiter & Schmidhuber, 1997) y la pérdida de contexto. El FFAM aborda estos desafíos al permitir una mejor propagación de la información a lo largo de la secuencia, optimizando así la retención de dependencias a largo plazo.

El FFAM se implementa como una capa posterior a las capas LSTM (ver Figura 3), toma el estado oculto de la celda (h_n) generado por la capa anterior y, mediante la función de activación interna, (*e.g.*, ReLU o tanh) calcula los puntajes de atención para cada una de sus componentes. Estas funciones de activación se utilizan para introducir no linealidad y garantizar que los puntajes de atención se mantengan dentro de un rango controlado, evitando valores extremos que puedan afectar la estabilidad del modelo. Posteriormente, se aplica una función *Softmax* sobre estos puntajes, transformándolos en una distribución de probabilidad en la que cada componente del estado oculto recibe un peso relativo. La función *Softmax* garantiza que la suma de todos los puntajes normalizados sea igual a 1, permitiendo que el modelo asigne mayor o menor importancia a cada componente del estado oculto según su relevancia en el contexto de la tarea de aprendizaje. De esta manera, FFAM mejora la capacidad del modelo para extraer información significativa y optimizar la generación de la salida.

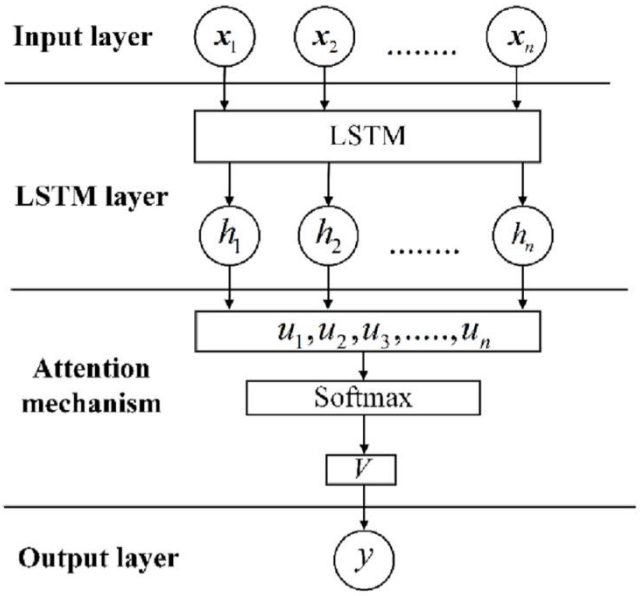


Figura 3. Diagrama de flujo de un modelo combinado de LSTM y un Mecanismo de atención. paraFuente: Kang et al. (2023)

2.5. Selección de características

La selección de características es un proceso esencial en el desarrollo de modelos ML/DL, ya que permite al algoritmo identificar las variables más relevantes, facilitando la construcción de representaciones internas que optimicen el rendimiento del modelo. Este proceso no solo mejora la precisión de las predicciones, sino que también, reduce la complejidad computacional y el riesgo de sobreajuste (Zhao, Tsay, & Kronqvist, 2023). Adicionalmente, el resultado de aplicar la selección de características dependerá del conjunto de datos, es decir, para modelos con las mismas características del dominio del problema, pero diferentes conjuntos de datos podrían no tener las mismas características seleccionadas (Ghimire, et al., 2022)

Para realizar la selección de características existen diversos métodos que pueden ser agrupados en tres grupos (Ghimire, et al., 2022): métodos basados en filtrado, métodos de

clasificación y métodos de envoltura (*wrapper*). Los métodos de filtrado se basan en seleccionar las características en base a aplicar métricas sobre las características, por ejemplo, selección de características usando como criterio la correlación entre las características o información mutua (*mutual information* – MI), donde de este último se desprenden la dependencia entre dos características, incertidumbre de la salida, disminución de la incertidumbre de la salida o la a divergencia de Kullback–Leibler (Ghimire, et al., 2022). Los algoritmos de clasificación, como por ejemplo *Random Forest* (RF), calculan la importancia de cada característica en el entrenamiento del modelo y con esto se seleccionan en base a un umbral mínimo de importancia.

Los métodos *wrapper* utilizan el rendimiento de las predicciones de los modelos para discriminar, y así, seleccionar las características (Chandrashekar & Sahin, 2014). Estos métodos puedes ser aplicado de dos formas: usando algoritmos de selección secuencial de características (*Sequential Feature Selection* – SFS) y algoritmos de búsqueda heurística (*Heuristic Search*). Los algoritmos SFS comienzan con un conjunto vacío (o completo) y agregan características (o eliminan características) hasta que se alcance el mínimo error en el proceso de entrenamiento. Para acelerar la selección, se elige un criterio que aumenta incrementalmente la función objetivo hasta que se alcanza el máximo con el menor número de características. Por otra parte, los algoritmos de búsqueda heurística evalúan diferentes subconjuntos de características para optimizar la función objetivo. Se generan diferentes subconjuntos, ya sea buscando en un espacio de búsqueda definido o, generando soluciones al problema de optimización (Chandrashekar & Sahin, 2014).

2.6. Índices de rendimiento

Para evaluar el desempeño de modelos predictivos, se recurre a un conjunto de métricas estadísticas para proporcionar una evaluación cuantitativa de la capacidad predictiva del modelo, su grado de ajuste a los datos observados y la correlación que existe entre los datos predichos y los datos reales. En Kuhn & Johnson (2016) se proponen diversos índices para la evaluación de modelos predicción y regresión: raíz del error medio cuadrático (*root mean squared error* – RMSE), error medio absoluto (*mean absolute error* – MAE) y coeficiente de determinación R^2.

El rendimiento del modelo es evaluado a partir de una comparación entre las predicciones realizadas por el modelo ($\hat{y}$) y los valores observados (y) sobre el conjunto de prueba o test. El índice más usado para caracterizar la capacidad predictiva del modelo es el RMSE, el cual se puede interpretar como la media de la diferencia al cuadrado entre las predicciones y los valores observados en el modelo.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{10}$$

Considerando que el RMSE tiende a penalizar los errores más grandes que los pequeños, el MAE mide el desempeño de los modelos predictivos, a partir de la diferencia absoluta entre los valores predichos y los observados.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{11}$$

Para conocer la correlación entre los datos predichos y los observados, se emplea el coeficiente de determinación R^2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (12)$$

Donde $\bar{y}$ es la media de los valores observados, $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ es la suma de los cuadrados de los residuos, es decir, la sumatoria del cuadrado de los valores observados menos los valores predichos y $\sum_{i=1}^{n}(y_i - \bar{y})^2$ es la suma de los cuadrados totales, es decir, es la sumatoria del cuadrado de los valores observados menos la media de estos.

3. Datos y métodos

El presente estudio consiste en evaluar diferentes arquitecturas de modelos basados en el uso de LSTM para la recuperación de datos de series temporales GPS. Los modelos son entrenados con datos de temperatura, presión atmosférica, efemérides solares y lunares, y el movimiento del polo, todas estas series de datos con una resolución diaria, al igual que las series temporales GPS. El periodo del estudio está comprendido entre los años 2010 y 2024.

La Figura 4 muestra el diagrama con el flujo de trabajo para el desarrollo del presente estudio. Una vez descargados los datos desde sus respectivas fuentes, son extraídas de estos las variables de interés. Luego, se realizó la selección de características, se agruparon y se generaron los conjuntos de entrenamiento y evaluación (*train*/*test*). A continuación, se realizó el ajuste de hiper-parámetros usando el algoritmo SFS de los métodos *wrapper*, se configuran los modelos en función del objetivo de estos, incluyendo la definición de sus capas, y se finaliza con el entrenamiento de los diferentes modelos (Modelo Predictor, Modelo Interpolador Geográfico, Modelo Regresor) que fueron evaluados con los índices de rendimiento.

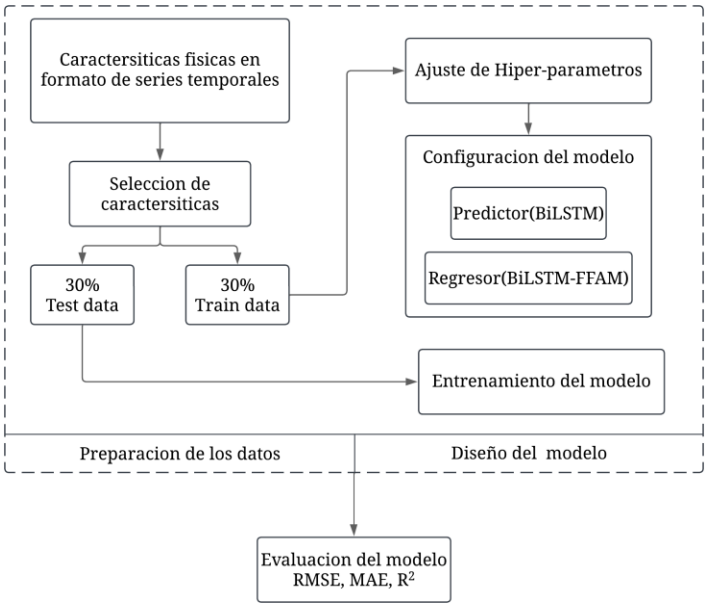


Figura 4. Diagrama de flujo del desarrollo de los modelos LSTM. Basado en: Kang et al. (2023)

3.1. Datos

3.1.1. Series temporales GPS

En este estudio se utilizaron un total de 214 estaciones permanente de la Red Brasileña de Monitoreo Continuo (RBMC), distribuidas por todo el país mostrado en la Figura 5.

Figura 5. Distribución de estaciones GPS (marcas rojas) en Brasil.

Los registros cuentan con coordenadas calculadas con el método *Precise Point Positioning* (PPP), siendo procesadas usando el software GipsyX Version 1.0, se utilizan productos finales diarios Repro3.0 del JPL, integrando órbitas, relojes, parámetros de actitud y correcciones ionosféricas y troposféricas avanzadas. Se emplea un pre-procesamiento riguroso de archivos RINEX, eliminando datos no GPS y aplicando filtros de calidad. La observación principal es la fase portadora libre de ionósfera, con muestreo de 5 minutos y modelado de efectos geofísicos como mareas y cargas oceánicas. Se usa un filtro de Kalman estocástico para la estimación diaria de coordenadas en el sistema de referencia IGS14, con correcciones de relojes y órbitas fijas de JPL (Blewitt, Hammond, & Kreemer, 2018). Las precisiones ofrecidas por las estaciones pertenecientes al IGS14 proporciona posiciones con una precisión del orden de 1-2 mm en el plano horizontal y 3-4 mm en el plano vertical. Estos datos se obtuvieron desde el servicio Nevada Geodetic Laboratory

(NGL) en formato ASCII, disponibles en http://geodesy.unr.edu/NGLStationPages/GlobalStationList.

Las series temporales GPS presentan en promedio una laguna de datos del 44% para el periodo comprendido entre 2010 y 2024 y como se puede observar en el mapa de calor presentado Figura 6, las series temporales ordenadas en sentido Oeste (arriba) a Este (abajo) presenta una mayor cantidad de falta de datos a medida que se acerca a la costa, es decir, a medida que las estaciones GPS se ubican más al Este, sus series temporales presentan una mayor cantidad de lagunas de datos, adicionalmente las series temporales presentan una mayor cantidad de laguna de datos en el primer tercio del periodo abarcado.

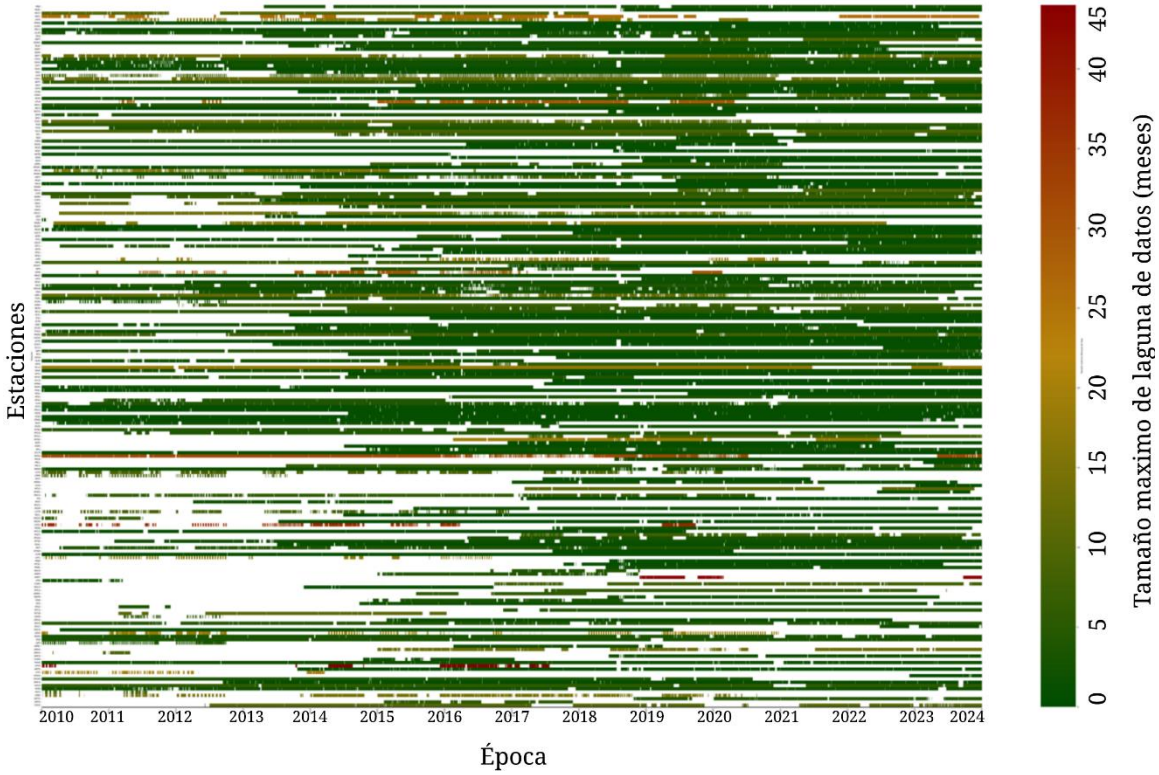


Figura 6. Mapa de calor con los registros de las series temporales GPS de RBMC, ordenados desde arriba hacia abajo, en sentido Oeste a Este.

Otro aspecto relevante que se puede observar corresponde a la longitud de los periodos de las lagunas de datos (épocas adyacentes de espacios). Como se puede ver en el histograma

presentado en la Figura 7, el 50% de las lagunas presentes en las series temporales tienen un periodo de hasta 2 días, mientras el 90% de las lagunas de datos presentes tienen un periodo de hasta 16 días.

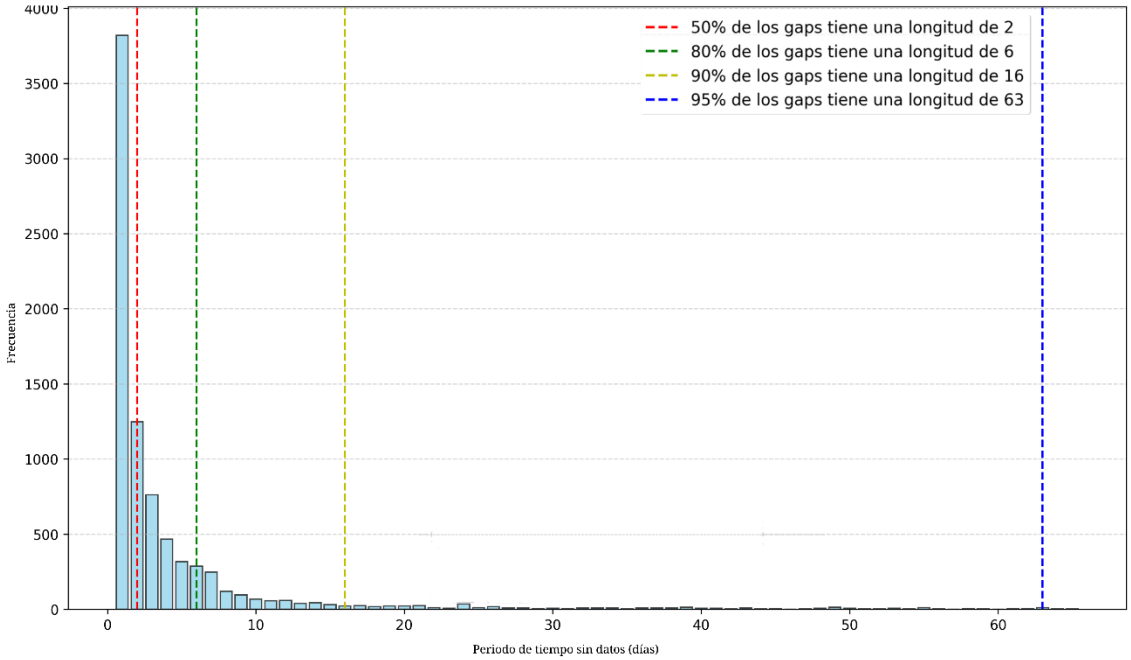


Figura 7. Frecuencia de periodos sin registro de datos.

3.1.2. Carga hidrológica

Considerando que la carga hidrológica local en cada estación tiene efectos significativos sobre las series de tiempo GPS, sobre todo en su componente vertical, provocando desplazamientos de varios milímetros dependiendo de la ubicación de la estación (Herring, et al., 2016), esta información fue considerada en el modelado por medio de los datos obtenidos desde el modelo *Land Surface Discharge Model* (LSDM) version v1.3, el cual simula globalmente el transporte y almacenamiento de agua vertical y horizontal superficial. La física y la parametrización se basan en el *Hydrological Discharge Model (*HDM) y el *Simplified Land Surface Scheme* (SLS). Impulsado por la precipitación, la

evaporación y la temperatura, el modelo hidrológico LSDM captura los procesos más importantes de transporte de masa de agua continental y los compartimentos de almacenamiento (humedad del suelo, nieve, ríos y lagos, escorrentía, drenaje) (Dill, 2008), proporcionados por el Centro Alemán de Geo-investigaciones (GFZ) en forma de grillas 0.5°x0.5° con una resolución temporal de 24 horas. Estos datos fueron descargados como archivos NetCDFs desde el repositorio de productos del ESMGFZ (http://rz-vm115.gfz.de:8080/repository).

3.1.3. Temperatura superficial y presión atmosférica.

La presión atmosférica es un factor que afecta la posición de las estaciones GPS (Petit & Luzum, 2010), debido a que tiene influencia sobre la temperatura superficial, que provoca expansión térmica del lecho rocoso y de los monumentos en los que se instalan los equipos GPS (Yan, et al., 2010). En el presente estudio se utilizaron los datos derivados del modelo NCEP-NCAR, el cual es un modelo de pronóstico y análisis de variables climatológicas. Específicamente utiliza un sistema de análisis/pronóstico para disponibilidad datos de variables climatológicas posteriores a 1948, pero además cuenta con datos registrados de dichas variables a múltiples resoluciones temporales hasta 2024. Estos datos se encuentran disponibles como grillas en archivos NetCDF y pueden ser descargados del servicio web (https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html) del National Center of Environmental Prediction (NCEP). Para el presente estudio, se utilizaron las variables de temperatura del aire (*air temperature*) y presión atmosférica (*pressure*), ambas con resolución espacial de 2.5°x2.5° y resolución temporal diaria.

3.1.4. Series temporales de movimiento del polo

Los parámetros de orientación terrestre proporcionan información de las cargas provocadas por cambios en el potencial centrífugo debido al movimiento del polo (Petit & Luzum, 2010). Son de interés para el presente estudio las coordenadas del movimiento del polo en los ejes X e Y (Px, Py), principalmente, debido a su influencia en la marea polar y la marea polar oceánica, estos efectos se producen influencia de los cambios en la fuerza centrífuga inducidos por el movimiento del polo, siendo la marea polar la deformación producida por estos cambios tanto en la corteza terrestre como en los océanos, mientras que la carga polar oceánica son las deformaciones producidas exclusivamente en los océanos. Estos datos fueron obtenidos desde el International Earth Rotation and Reference System Service (IERS) a través del sitio web https://www.iers.org/IERS/EN/DataProducts/EarthOrientationData/eop.html, donde se encuentra publicados en forma de registros tabulares en archivos de tipo ASCII con resolución temporal diaria.

3.1.5. Series temporales de efemérides solares y lunares

Las coordenadas del sol y la luna en el International Celestial Reference Frame (ICRF) están dadas por ascensión recta del sol (RA_s), declinación del sol(DEC_s) y rango aparente (también denominado rango observado, se refiere a la distancia entre el observador y el cuerpo celeste) del sol (DEL_s), ascensión recta de la luna (RA_m), declinación de la luna(DEC_m) y rango aparente de la luna (DEL_m), las cuales influyen sobre la carga de marea oceánica, las mareas de tierra sólida y la carga de presión atmosférica (Petit & Luzum, 2010). Estos datos de series temporales provienen del modelo JPL Ephemerides

DE440, los cuales se encuentran disponibles en formato ASCII y pueden ser descargados desde el sitio web del Jet Propulsion Laboratory (JPL) California Institute of Technology (https://ssd.jpl.nasa.gov/horizons/app.html#/).

3.2. Selección de características

Para la selección y definición de características para el modelado del problema, se implementa el método *wrapper* SFS. Mediante este método se entrenaron modelos de manera iterativa, partiendo desde una característica por modelo, obteniendo en la primera iteración tantos modelos como números de características existentes, los que posteriormente se compara su rendimiento en base a las métricas RMSE, MAE y R^2, siendo seleccionada la característica perteneciente al modelo con mejor rendimiento, la cual se mantendrá fija como una característica del modelo en las iteraciones sucesivas. De manera similar a la primera iteración, se entrenan tantos modelos como características no fijas existentes, repitiéndose así el proceso hasta que se entrene un modelo con todas las características definidas. Para terminar la selección de características, se evalúan los rendimientos de los mejores modelos resultantes en cada iteración, siendo seleccionadas las características del modelo con mejor rendimiento.

3.3. Diseño de los modelos

El algoritmo LSTM tiene variantes tales como, LSTM Vanilla (o clásico), LSTM Bidireccional (BiLSTM), Stacked LSTM, entre estos. Para el diseño de los modelos desarrollados en el presente estudio se usó la variante BiLSTM, debido a que este presenta mejor desempeño en comparación a las otras variantes (Siami-Namini, Tavakoli, & Siami, 2019). BiLSTM tiene la particularidad de recorrer las entradas en sentido de avance y

retroceso, de esta manera se recorre la serie de datos desde “el pasado al futuro” y del “futuro al pasado”.

Para generar las entradas del entrenamiento y testeo de los modelos, se define una ventana deslizante genérica de tamaño $t - n$, donde n es la cantidad de pasos antes de la época de predicción t (ver Figura 8).

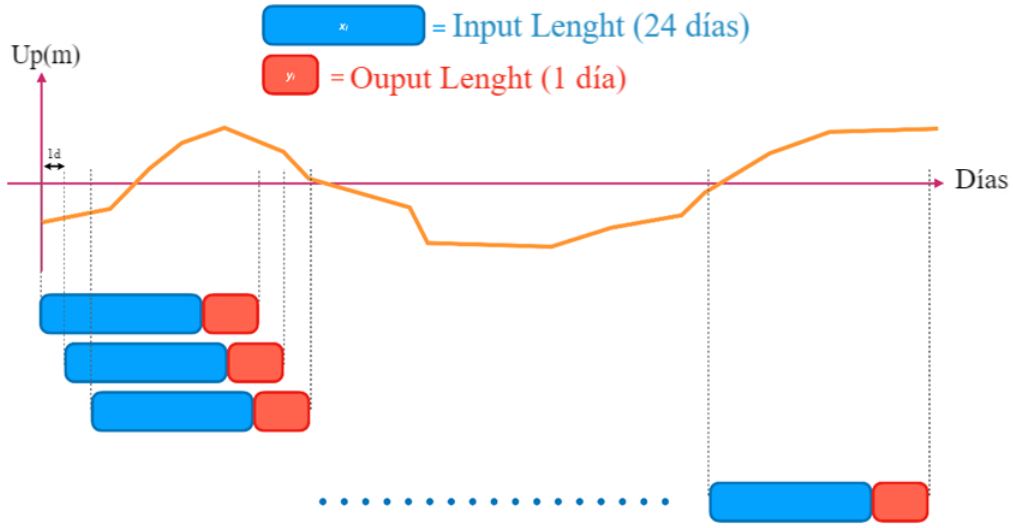


Figura 8. Diagrama una función de ventana deslizante.
Fuente: Elaboración propia.

A partir del algoritmo BiLSTM, se proponen dos estructuras para los modelos a evaluar, la primera de ellas adoptando un enfoque de modelo de predicción, y el segundo, un adoptando un enfoque de modelo de regresión.

El modelo de predicción busca recuperar (predecir) datos de una o más épocas a partir de datos de uno o más épocas del pasado. Para este caso, se propone una configuración de capas, con una capa de entrada, capa(s) BiLSTM, capa Dense y una capa de salida. En la capa Dense, la función de activación será ReLU, debido a su simplicidad y la condición de no linealidad, necesaria para los modelos predictivos (Ramachandran, Zoph, & Le, 2017).

Adoptando un enfoque de modelo de regresión, se busca de recuperar datos de la serie temporal GPS (*up*) de una época a partir de características (o variables) adicionales de la misma época. A partir de este enfoque, se proponen dos diseños de modelos particulares. Por simplicidad, al primero se le denominará Regresor, mientras que, al segundo se le denominará Interpolador Geográfico. El modelo Regresor, utilizará las 12 características de interés (Time, P_x, P_y, HYD, AP, TEM, DEC_s, DEL_S, RA_s, DEC_m, DEL_m y RA_m) para recuperar los registros de la serie temporal GPS (*up*), y su estructura de capas será la siguiente, capa de entrada, capa(s) BiLSTM, capa FFAM, capa Dense y capa de salida (ver Figura 9). Se utilizará la función Linear como función de activación.

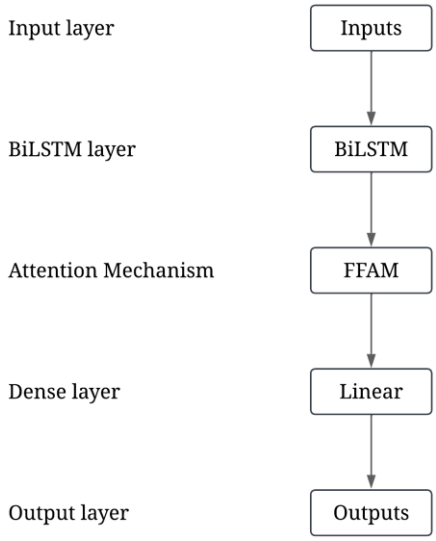


Figura 9. Diagrama de flujo de un modelo combinado LSTM-FFAM.

Por otra parte, el modelo Interpolador Geográfico busca recuperar datos de la serie temporal GPS de una estación, a partir de variables de un conjunto de estaciones cercanas. Se utilizarán como entradas, las 12 variables adicionales de la serie GPS, la componente *up*, y la posición de cada estación (latitud y longitud), esto último, para efectos de establecer

la relación espacial con las estaciones cercanas. Este modelo mantiene la misma estructura de capas del modelo Regresor. La estructura de los modelos diseñados puede ser observada en la Figura 10.

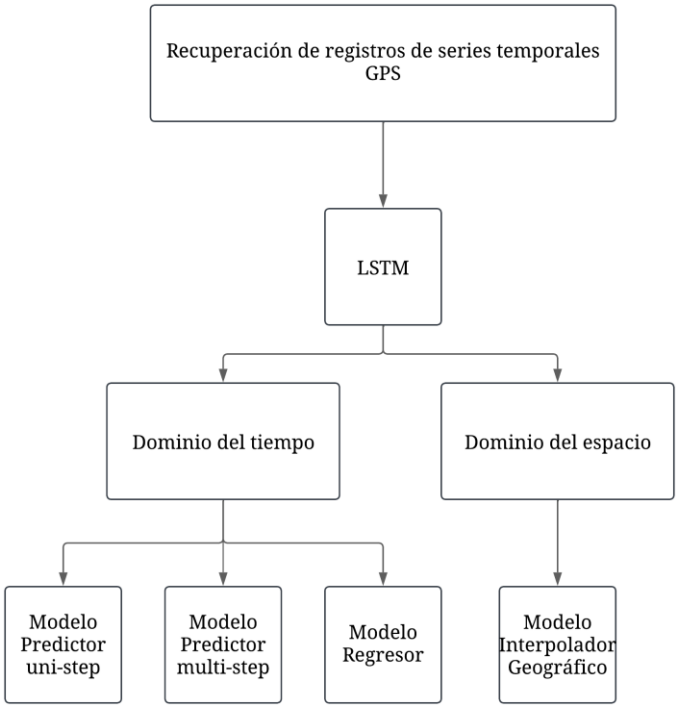


Figura 10. Estructura de los modelos diseñados.

3.4. Entrenamiento de los modelos

En la fase de entrenamiento, las características son escaladas o normalizada, con el propósito de evitar la saturación de la función de activación, estabilizar y acelerar el proceso de optimización, y evitar que algunas características, dada la magnitud de sus valores, predominen en la función de perdida (Lima & Souza, 2023). Los datos serán normalizados usando *Min-max scaling* en el rango [0,1], esto, considerando que se utilizará la función de activación ReLU. Luego, se dividen los datos en una proporción 70/30 para formar los conjuntos de entrenamiento y prueba, respectivamente.

Para el ajuste de hiperparámetros, que busca encontrar el conjunto de parámetros con mejor desempeño, se utilizará el método de búsqueda aleatoria (*random search*) con validación cruzada con $k - folds = 3$, definiendo como límite de la búsqueda el universo de híper-parámetros presentado en la Tabla 1, el cual fue definido en base al espacio de búsqueda de estudios como (Wang, Jiang, Li, & Lu, 2021) y experimentación.

Hiperparámetro	**Rango de Valores**
Número de celdas de estado	[8,128]
Número de capas	[1,5]
Tasa de abandono	[0.0,0.5]
Tasa de aprendizaje	[0.5,0.000005]
Optimizador	Adam, RMSprop, SGD, Nadam
Longitud de entrada	[1,500]
Tamaño de lote	[1, 3875]
Número de Iteraciones	[1,50000]

Tabla 1. Límites del espacio de búsqueda de hiper-parámetros.

Para el caso del modelado de predicción se considera evaluar diferentes longitudes de salidas (ventanas de predicción expresado en días). El universo de ventanas de salidas está definido por los límites [7,133] con intervalos de 7.

Para el modelado tipo Interpolador Geográfico, se evaluará el uso de diferentes cantidades de estaciones cercanas, estableciendo el rango de límites entre [2,90].

3.5. Evaluación del rendimiento de los modelos

El desempeño de los modelos será evaluado usando los índices RMSE, MAE y R^2 para determinar el rendimiento del modelo al realizar las predicciones. Para el conjunto de datos de validación se considera el 30% de los datos.

4. Resultados

4.1. Selección de características

A partir de las 12 variables de interés del presente estudio: tiempo, RA_s, DEL_s, DEC_s, RA_m, DEL_m, DEC_m, P_x, P_y, AP, TEM y HYD, en el proceso de selección de características generó un total de 78 modelos. La Tabla 2 identifica dos modelos con los mejores desempeños del proceso. El modelo A, entrenado con todas las características, y el modelo B, resultante de la selección de características con el método SFS con las variables P_x, RA_s y tiempo.

		Modelo A	Modelo B
Características	**Time**	X	X
	$\boldsymbol{RA_s}$	X	X
	$\boldsymbol{RA_m}$	X	
	$\boldsymbol{DEL_s}$	X	
	$\boldsymbol{DEL_m}$	X	
	$\boldsymbol{DEC_s}$	X	
	$\boldsymbol{DEC_m}$	X	
	$\boldsymbol{P_x}$	X	X
	$\boldsymbol{P_y}$	X	
	TEM	X	
	AP	X	
	HYD	X	
Índices de rendimiento	**RMSE**	0.006	0.005
	MAE	0.005	0.004
	$\boldsymbol{R^2}$	0.70	0.78

Tabla 2. Desempeño de los mejores modelos, resultados del proceso de selección de características.

4.2. Diseño e implementación del modelo

4.2.1. Modelo de predicción

En la Tabla 3 se muestran algunos hiperparámetros y el rendimiento de los modelos con mejor desempeño resultante del proceso de ajuste de hiperparámetros. Considerando que

los rendimientos son similares entre las diferentes configuraciones evaluadas, se prefiere la configuración menos compleja (debido al principio de riesgo estructural), para este caso el Modelo 1.

Modelo	Número de Celdas	Número de Capas	Tasa de abandono	Tasa de aprendizaje	RMSE (m)
1	112	1	0.2	$5*10^{-4}$	0.080
2	48	3	0.1	$5*10^{-3}$	0.077
3	48	5	0.0	$5*10^{-6}$	0.148
4	112	2	0.3	$5*10^{-3}$	0.081

Tabla 3. Mejores hiper parámetros de cada ejecución del ajuste de hiper-parámetros con validación cruzada.

La Tabla 4 muestra los valores de los hiperparámetros obtenidos del proceso de ajuste de hiper-parámetros. Estos valores fueron usados para evaluar modelos con diferentes periodos de predicción (longitud de salida en días). La Figura 11 muestra de forma gráfica la variación del error en función del periodo de predicción. Se puede observar un aumento a medida que se extiende el periodo de predicción. Es importante mencionar que, a pesar de que el error se mantiene en el orden de milímetros, para una ventana de predicción mayor a 105 días el coeficiente de determinación R^2 desciende abruptamente, pasando de 0.80 en un modelo con longitud de salida de 7 días a 0.36 en un modelo con longitud de 112 días. Estos resultados presentan una alta coherencia con el rendimiento esperado para un modelo basado en LSTM de acuerdo con los resultados observados en Gao et al. (2022).

Hiperparámetro	Valor
Longitud de entrada	215
Número de capas	1
Número de celdas	112
Tasa de abandono	0.2
Tasa de aprendizaje	$5*10^{-4}$
Tamaño de lote	322
Número de iteraciones	100
Optimizador	Adam

Función de activación	ReLU
Función de perdida	MSE

Tabla 4. Hiperparámetros definidas para el modelo Predictor.

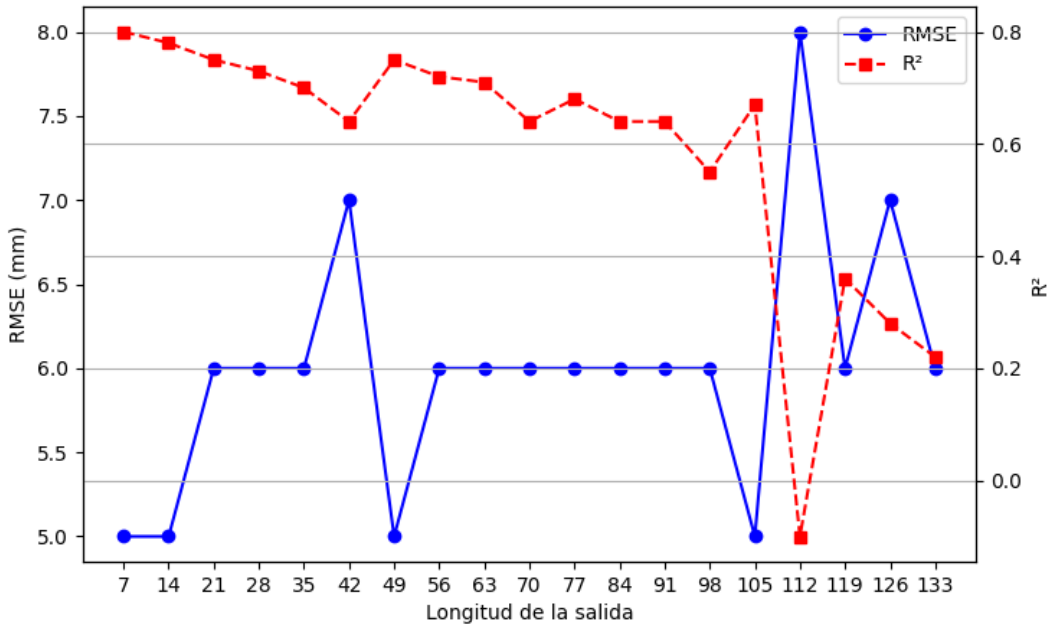


Figura 11. Rendimiento modelo de predicción para longitud de salida variable.

4.2.2. Modelos de regresión

Para los modelos Interpolador Geográfico el proceso de entrenamiento es similar al de los modelos tipo Predictor. En este ajuste, se establecen inicialmente cuatro estaciones cercanas a la estación a recuperar o estación objetivo, la cual se requiere recuperar, con dichas características se realiza el ajuste de hiper-parámetros, resultando en los conjuntos de hiperparámetros mostrados en la Tabla 5, donde puede observarse que el conjunto de la primera ejecución del ajuste es el que presenta el mejor rendimiento, por lo que a pesar de que este posea una mayor cantidad de celdas de estado que el conjunto de hiper-parámetros que le sigue, este es seleccionado al tener un mejor rendimiento en función de su RMSE.

Modelo	Número de Celdas	Número de Capas	Tasa de abandono	Tasa de aprendizaje	RMSE (m)
1	112	1	0.2	$5*10^{-6}$	0.059
2	56	1	0.3	$5*10^{-2}$	0.062
3	24	3	0.3	$5*10^{-2}$	0.063
4	80	2	0.1	$5*10^{-3}$	0.063

Tabla 5. Mejores combinaciones de hiperparámetros ordenadas para los modelos Interpolador Geográfico.

Por lo consiguiente, los hiper-parámetros definidos para los modelos tipo Interpolador geográfico son los mostrados en la tabla 6, los cuales son similares a los definidos para los modelos tipo Predictor. Adicionalmente, el rendimiento del modelo al ser evaluado con las métricas seleccionadas es el siguiente: RMSE=0.006m, MAE=0.005m y R^2=0.52.

Hiper Parámetro	Valor
Longitud de entrada	1
Número de Celdas	112
Número de Capas	1
Tasa de abandono	0.2
Tasa de aprendizaje	$5*10^{-6}$
Optimizador	Adam
Tamaño de lote	322
Número de Iteraciones	100
Función de Activación	Lineal
Función de Perdida	MSE

Tabla 6.Hiper-parámetros definidos para el modelo Interpolador Geográfico.

Usando este conjunto de hiper-parámetros, se entrenan modelos con múltiples estaciones, partiendo con 2 estaciones hasta 90. Los resultados obtenidos son graficados en la Figura 12. Al observar los resultados, se deduce que los mejores resultados se encuentran en los modelos que incluyen hasta 18 estaciones, lo que podría indicar el máximo de estaciones a usar para realizar la recuperación de registros de la estación objetivo, aunque en este caso al poseer resultados similares, es preferible el uso de la menor cantidad de estaciones en el modelo para mantener la mayor simplicidad de este y evitar el riesgo estructural del modelo, además se observa una tendencia descendente en el rendimiento de los modelos a medida que estos tienen un mayor número de estaciones.

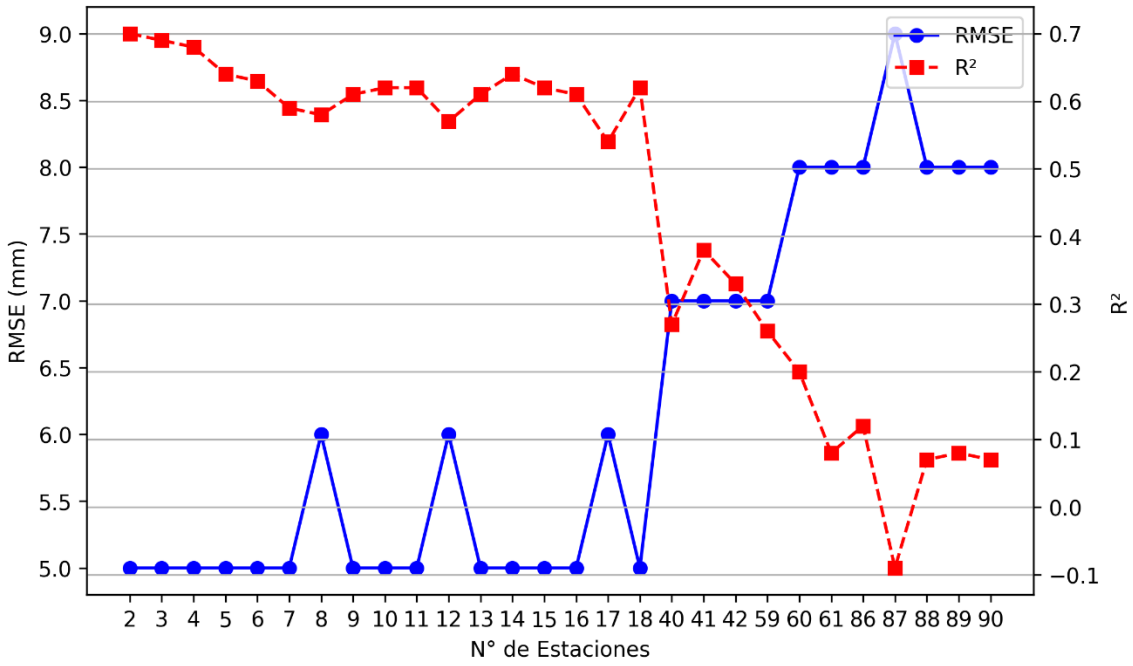


Figura 12. Rendimiento de modelos Interpolador Geográfico con diferentes cantidades de estaciones cercanas.

Pasando al modelo Regresor, los resultados del ajuste de hiper-parámetros son mostrados en la Tabla 7, donde se aprecia que los conjuntos de hiper-parámetros desde el 2 al 5 son aquellos obtenidos mediante el ajuste de hiper-parámetros propuestos, mientras que el primer conjunto es resultante de evaluar el conjunto de hiper-parámetros del modelo Interpolador Geográfico en el modelo Regresor, obteniendo un resultado mejor que el obtenido en el ajuste previamente realizado, en adición de que este modelo es de una sola capa, a diferencia de los otros cuatro conjunto que son de 2 o 3 capas.

Modelo	**Número de Celdas**	**Número de Capas**	**Tasa de abandono**	**Tasa de aprendizaje**	**RMSE (m)**
1	112	1	0.2	$5*10^{-6}$	0.047
2	16	3	0.4	$5*10^{-2}$	0.064
3	120	2	0.0	$5*10^{-2}$	0.065
4	56	3	0.0	$5*10^{-2}$	0.065
5	56	3	0.4	$5*10^{-2}$	0.065

Tabla 7. Mejores conjuntos de hiper-parámetros para el modelo Regresor.

Con lo cual el conjunto de hiper-parámetros definido para el modelo Regresor es idéntico al del modelo Interpolador Geográfico mostrado en la Tabla 6, variando exclusivamente en la cantidad de características que posee cada modelo. Adicionalmente, el rendimiento del modelo al ser evaluado con las métricas seleccionados es: RMSE=0.005m, MAE=0.004m y R^2=0.69.

4.3. Evaluación de los modelos

4.3.1. Modelo Predictor

Con respecto al modelo Predictor, se evalúa el rendimiento del modelo usando los últimos cuatro años de la serie temporal, la comparativa se realiza usando los datos verdaderos de la serie temporal con los valores predichos por el modelo para cada dieciséis épocas de predicción, generando los resultados mostrados en la Tabla 8, estos resultados son relativamente estables a lo largo de las 16 épocas, dando como resultado valores de precisión cercanos para cada una de las épocas predichas, esto es debido a que el algoritmo LSTM como se ha descrito con anterioridad, tiene la capacidad de almacenar información de toda la serie temporal y aprovecharla para predicciones a corto y largo plazo. Ahora con respecto a los valores que muestran las respectivas métricas, estos se encuentran en un rango aceptable para el modelado de series temporales, considerando que el modelado no se realizó con las coordenadas de la serie temporal como características de entrada al modelo.

Época	**RMSE(m)**	**MAE(m)**	**R^2**
1	0.005	0.004	0.81
8	0.005	0.004	0.79
16	0.006	0.005	0.78

Tabla 8. Rendimiento del modelo en el conjunto de validación para la predicción de 16 épocas consecutivas.

4.3.2. Modelos regresores

Evaluando el rendimiento del modelo Interpolador Geográfico, se tienen el rendimiento mostrado para el conjunto de hiper-parámetros es de RMSE=0.006m, MAE=0.005m y R^2=0.52 en la componente *up*, esto siendo aplicado sobre una estación elegida, por lo que es necesario evaluar este modelo en estaciones diferentes a la usada en el entrenamiento.

Para evaluar el rendimiento del modelo en diferentes estaciones, se define un número de 4 estaciones cercanas a la estación objetivo, y se evaluara un total de 4 estaciones objetivo, ubicadas en el norte, sur, este y oeste de Brasil, para evaluar la capacidad de generalización de la estructura de modelo propuesta. Para ello se toman los datos de las estaciones POAL para el norte, BOAV para el sur, SAGA para el este y PBCG para el oeste, de los cuales se obtuvieron los resultados mostrados en la Tabla 9.

Estación	RMSE (m)	MAE (m)	R^2
POAL	0.005	0.004	0.33
BOAV	0.007	0.006	0.59
SAGA	0.009	0.007	0.50
PBCG	0.006	0.005	0.11

Tabla 9. Rendimiento del modelo Interpolador Geográfico en las estaciones POAL, BOAV, SAGA y PBCG.

Como puede observarse en los resultados obtenidos, el RMSE y MAE en el caso de las estaciones PBCG y POAL están en un rango aceptable, pero sus R^2 es bajo, mientras que en el caso de las estaciones BOAV y SAGA los valores de RMSE y MAE son un poco más elevados, pero manteniéndose en un rango aceptable, mientras que sus R^2 están es un rango aceptable (entendiendo por aceptable a valores de rendimiento inferiores al centímetro). Para tratar de explicar estas diferencias, se realiza una caracterización en términos de la

densidad de datos de las estaciones objetivo y de las estaciones cercanas en los cuatro casos (ver Tabla 10).

Estación GPS	Densidad (%)	Densidad promedio estaciones cercanas (%)
POAL	98	56
BOAV	93	64
SAGA	94	75
PBCG	98	72

Tabla 10. Densidades de datos de las estaciones objetivo y estaciones cercanas.

En general, para todos los casos, la estación objetivo mantiene sobre un 90% de densidad de datos, mientras que, la densidad de datos promedio de las estaciones cercanas se encuentra entre ~50% y ~75%.

Para evaluar el efecto en el rendimiento del modelo debido de la densidad de datos de las estaciones cercanas. Se usaron las estaciones POAL y PBCG (de bajos rendimiento) y se evaluaron diferentes escenarios, seleccionando estaciones cercanas en función de la densidad de datos a través de cuartiles de densidades. Como se puede ver en la Tabla 11, no existen diferencias significativas entre los escenarios propuestos. Sin embargo, a medida que la distancia promedio entre las estaciones es mayor, se puede apreciar un descenso en el coeficiente de determinación (R^2).

Estación	Densidad mínima	Densidad máxima	Distancia promedio (km)	RMSE (m)	MAE (m)	R^2
POAL	20%	40%	1174	0.005	0.004	0.12
POAL	40%	60%	1422	0.006	0.004	0.14
POAL	60%	80%	1430	0.006	0.004	0.16
POAL	80%	100%	297	0.005	0.004	0.37
PBCG	20%	40%	2379	0.006	0.005	0.11
PBCG	40%	60%	766	0.006	0.005	0.07
PBCG	60%	80%	1154	0.006	0.005	0.12
PBCG	80%	100%	281	0.006	0.005	0.18

Tabla 11. Rendimiento de los modelos Interpolador Geográfico para diferentes densidades de datos de estaciones cercanas.

Luego, se realizaron evaluaciones de modelos entrenados con diferentes cantidades de estaciones. La evolución del desempeño para las estaciones POAL y PBCG se muestran en la Figura 13.

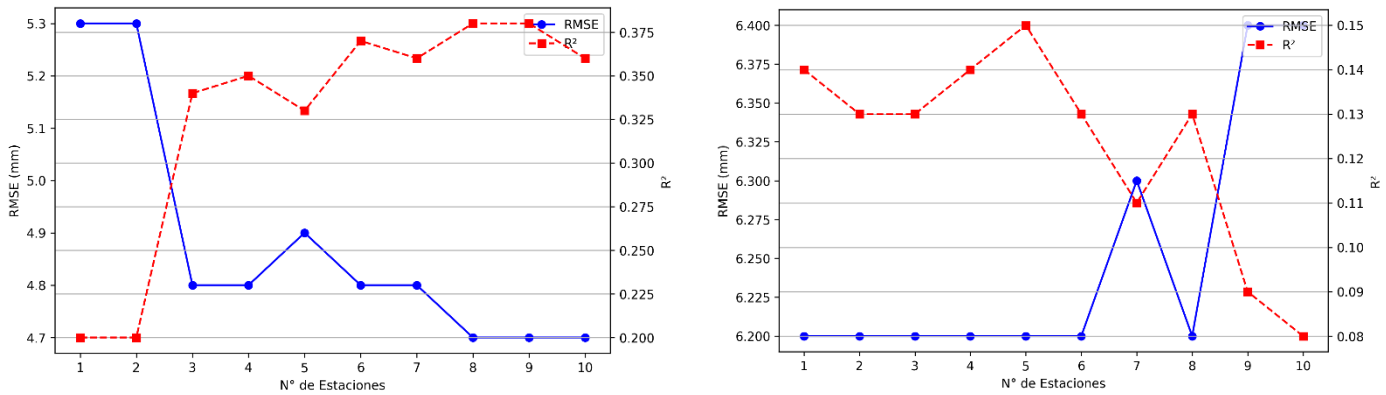
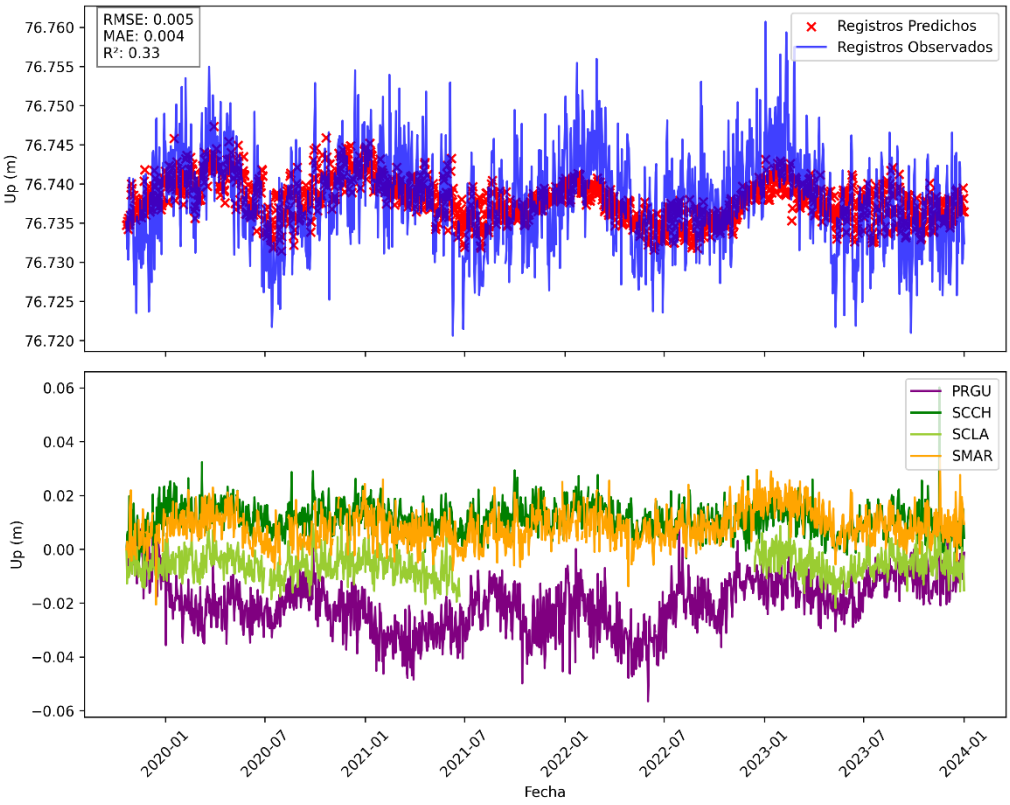
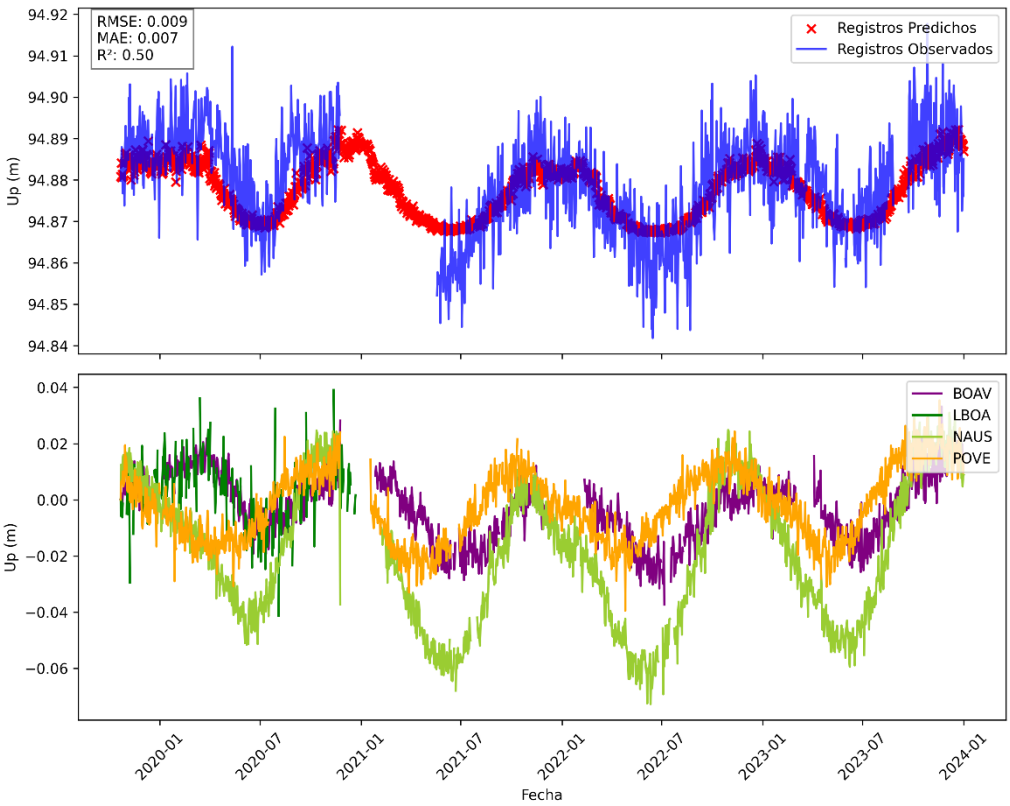


Figura 13. Evolución rendimientos de modelos de regresión, estaciones a) POAL y b) PBCG.

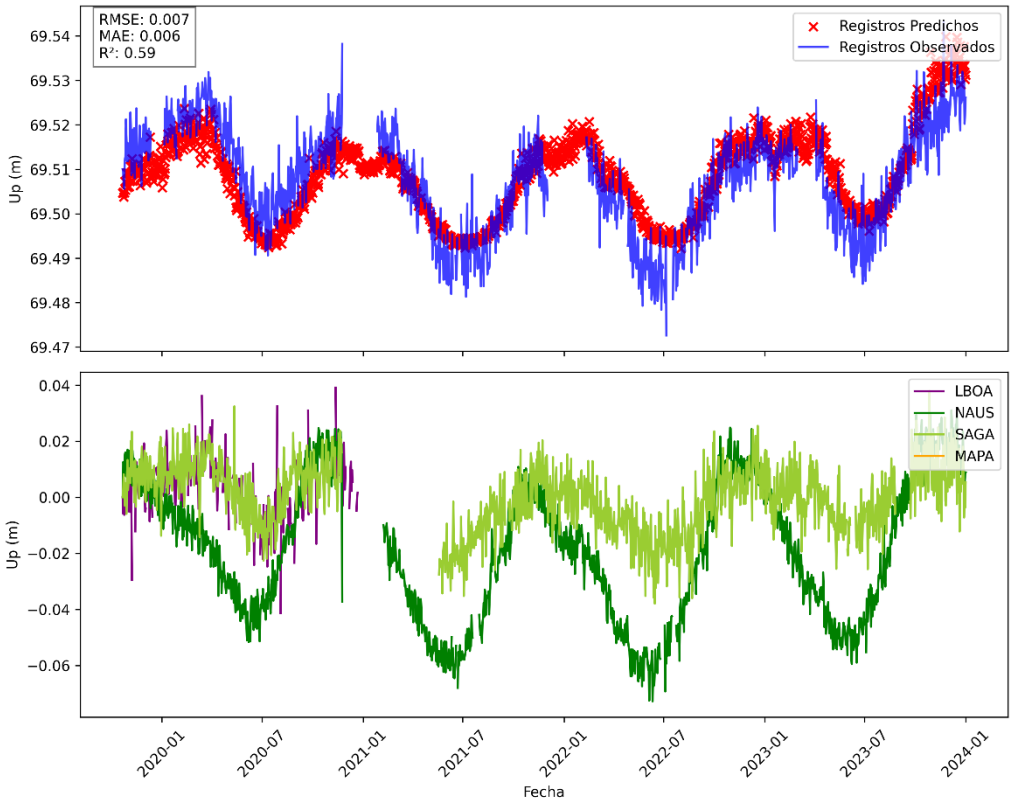
La Figura 14 muestra las series temporales GPS de las estaciones POAL, SAGA, BOAV y PBCG, la serie temporal del modelo, y las series GPS de las estaciones cercanas, empleadas en la recuperación de la respectiva estación objetivo. Como puede observarse en las series temporales (puntos azules) de los modelos con bajo rendimiento, es que estas cuentan con mayor ruido en sus datos, lo que a su vez provoca una diferencia de amplitud en la serie temporal a modela, fenómeno que podría ser el responsable de que los modelos (línea azul) tengan bajo rendimiento.


RMSE: 0.005
MAE: 0.004
R²: 0.33
Registros Predichos
Registros Observados
Up (m)
PRGU
SCCH
SCLA
SMAR
Fecha

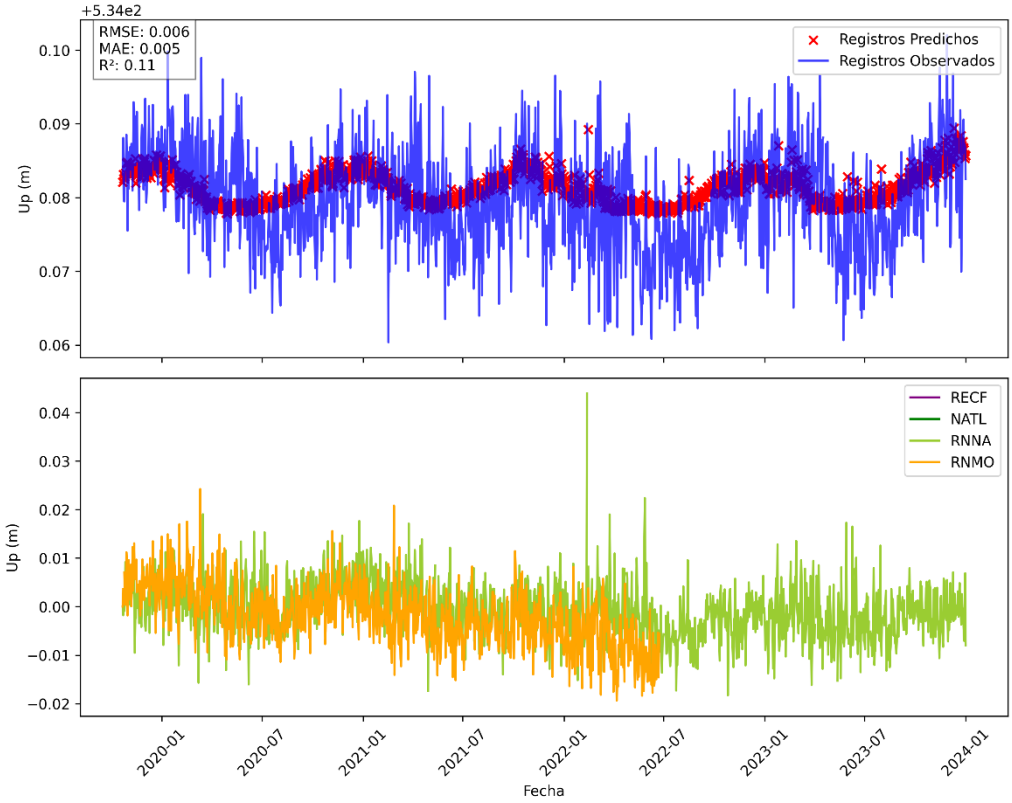
a)


RMSE: 0.009
MAE: 0.007
R²: 0.50
Registros Predichos
Registros Observados
Up (m)
BOAV
LBOA
NAUS
POVE
Fecha

b)


RMSE: 0.007
MAE: 0.006
R²: 0.59
Registros Predichos
Registros Observados
Up (m)
69.54
69.53
69.52
69.51
69.50
69.49
69.48
69.47
LBOA
NAUS
SAGA
MAPA
0.04
0.02
0.00
−0.02
−0.04
−0.06
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
2023-07
2024-01
Fecha

c)



d)

Figura 14. Series temporales (up) de las estaciones objetivo a) POAL, b) SAGA, c) BOAV y d) PBCG.

Pasando a evaluar el conjunto de hiper-parámetros definido para el modelo Regresor, se tiene el siguiente rendimiento: RMSE=0.005m, MAE=0.004m y R^2=0.69. Este rendimiento al igual que pasa en el caso del modelo Interpolador Geográfico debe ser evaluado para otras estaciones.

Para esta tarea se entrenaron modelos para recuperar datos de todas las estaciones de la red RBMC, para esto se ocupó el diseño BiLSTM-FFAM y el diseño con BiLSTM, la intención de esto es comprobar como varía el rendimiento de los modelos a medida que la densidad de los registros de las estaciones va en aumento y la diferencia existente entre ocupar la capa FFAM en el modelo Regresor. Estos resultados se reflejan en las tablas Tabla 12 y Tabla 13.

	R^2			RMSE (mm)		
Cuartil	mín	media	máx	mín	media	máx
Q1	-8.69	-0.64	0.79	0.005	0.011	0.050
Q2	-1.80	0.14	0.83	0.004	0.008	0.035
Q3	-9.60	-0.88	0.82	0.005	0.012	0.043
Q4	-9.02	-0.79	0.75	0.005	0.011	0.048

Tabla 12. Rendimiento en RMSE y R2 de los modelos Regresor agrupados por cuartiles para el modelo BiLSTM.

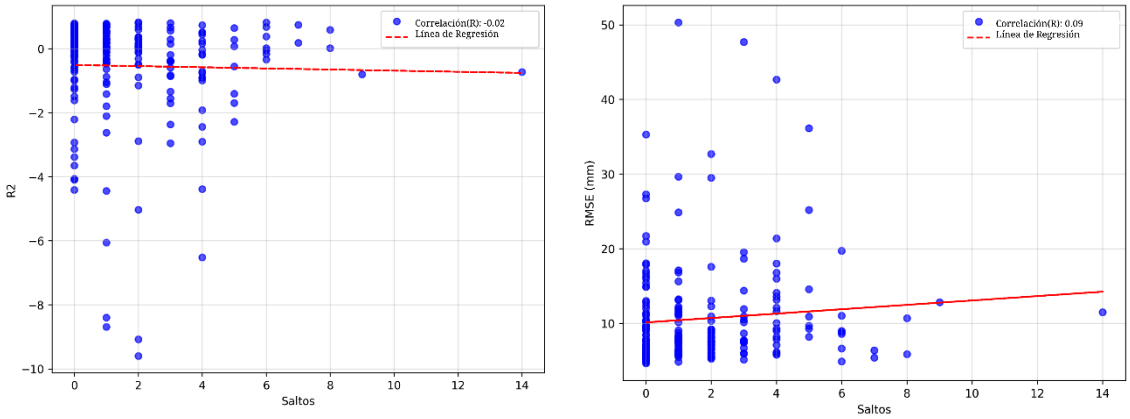
	R^2			RMSE (mm)		
Cuartil	mín	media	máx	mín	media	máx
Q1	-10.95	-0.71	0.78	5	11	5
Q2	-1.81	0.14	0.83	5	8	35
Q3	-8.82	-0.89	0.82	5	12	44
Q4	-9.52	-0.76	0.74	5	11	48

Tabla 13. Rendimiento en RMSE y R2 de los modelos Regresor agrupados por cuartiles para el modelo BiLSTM-FFAM.

Como puede observarse en ambas tablas, los R^2 mínimos superan el -1, pero esto es en situaciones puntuales y no es generalizado en todos los modelos. El RMSE promedio no está condicionado por la densidad de datos, ya que los valores más altos se encuentran en el primer cuartil de estaciones, pero los valores de RMSE más bajos se encuentran en el segundo cuartil, mientras que los valores medios de RMSE del tercer y cuarto cuartil son similares, por lo que esto demuestra que los modelos basados en LSTM no son dependientes de la densidad de datos del fenómeno que se quiera modelar. Comparando los modelos BiLSTM y BiLSTM-FFAM se tiene una diferencia mínima entre ambos donde en promedio los modelos BiLSTM-FFAM son un 0.6% mejores en el RMSE que los modelos BiLSTM, un 0.7% mejores en MAE y 34% peores de media en R^2.

Además, se nota que existen valores altos que incrementan el RMSE promedio de los cuartiles, por lo que teniendo en cuenta las conjeturas realizadas a partir de aplicar el modelo Interpolador geográfico sobre estaciones RBMC se plantea comparar la cantidad de saltos con el RMSE y R^2 de estos, lo que da como resultado lo mostrado en la Figura

15, en la cual se calcula la correlación de Pearson entre el rendimiento de los modelos Regresor entrenados y la cantidad de saltos pertenecientes a las series temporales de dichos modelos.



(a) *(b)*

Figura 15. Correlación rendimiento contra número de saltos. (a)Correlación entre R^2 y la cantidad de saltos. (b) Correlación entre RMSE y la cantidad de saltos.

Puede observarse que no existe una relación directa entre la cantidad de saltos y su RMSE y R^2, existiendo incluso una correlación negativa en esta última, lo que indica que la cantidad de saltos no es un factor determinante para el mal rendimiento de algunas estaciones. Pero llama la atención que existen modelos especialmente malos en esta figura, por lo que se decidió comprobar el comportamiento de las series temporales de tres modelos, uno con un mal rendimiento con la estación AMHU, un modelo con un rendimiento intermedio con la estación MCLA y un modelo con un buen rendimiento con la estación AMCR. Donde se observan saltos producto de actividad tectónica en las dos primeras estaciones mostradas, teniendo un salto más grande la estación AMHU y un salto no tan pronunciado con la estación MCLA, pero está presentando ruido en los datos, mientras que la serie temporal de la estación AMCR no presenta estos comportamientos. Los saltos observados en estas series temporales coinciden con los saltos registrados en el

sitio del NGL. Por lo que se puede inferir que el modelo es sensible al ruido y a la actividad tectónica en las series temporales, lo que una vez más provocaría variaciones de amplitud importante en los datos de la serie temporal a modelar.

Siguiendo con esta línea de análisis, se excluyeron los modelos de las estaciones MTBA, LPIN, LDOU y LPLN, debido a que los resultados presentan bajos rendimientos, alcanzando un RMSE de 1.1 metros y un R^2 negativo, por lo que es necesario conocer el comportamiento de las series temporales correspondientes a estos modelos, esto puede observarse en la Figura 16.

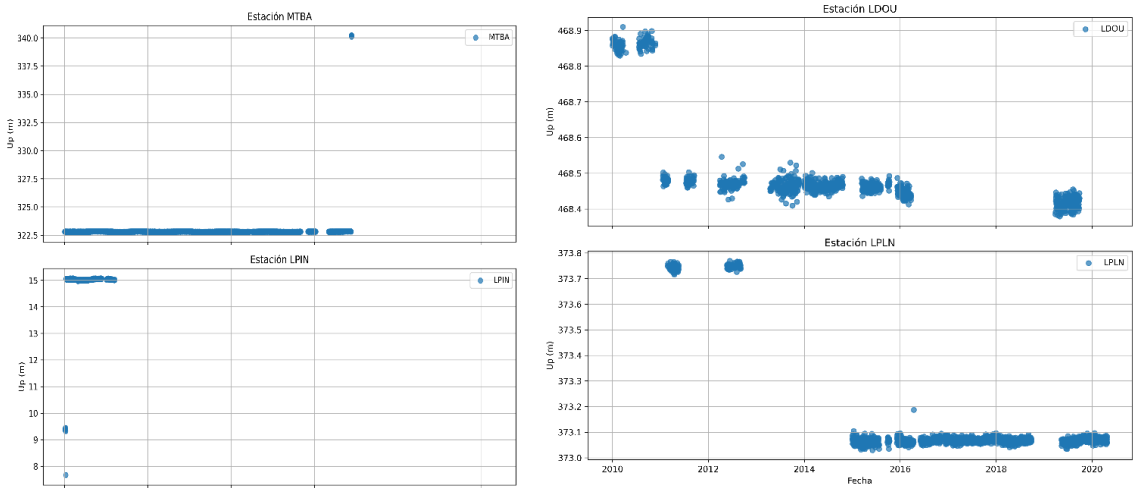


Figura 16. Series temporales de las estaciones MTBA, LPIN, LDOU y LPLN.

Con esta figura, puede interpretarse que la diferencia de valores en la serie temporal es la fuente de error de estos modelos, los cuales son producidos por saltos en las series, donde los saltos en las estaciones MTBA y LDOU son productos de actividad tectónica, mientras que el de las estaciones LPIN no registran causa para los saltos en la información ofrecida por el NGL, además observa cierto ruido en los datos de la estación LDOU. Por lo que, a partir de los análisis realizados se puede interpretar que los modelos basados en BiLSTM son sensibles ante ruido en los datos y una variación de amplitud en los datos, ya que esto

puede introducir al modelo un comportamiento erróneo, el cual con posterioridad intentara replicar, lo que provoca resultados como los comentados con anterioridad, por lo que un tratamiento de los saltos y el ruido de las series temporales previo al modelado de estas es fundamental para este tipo de modelos.

5. Discusión y análisis de resultados

Observando los resultados obtenidos en la etapa de selección de características aplicada sobre el enfoque predictor usando SFS de los métodos *wrapper*, se identifica que esta etapa es importante al momento de diseñar un modelo DL, debido a que al ocupar exclusivamente las características más relevante mejora los resultados y disminuye la complejidad de los modelos, lo que ayuda a la estabilidad de estos. Por otra parte, comparando los resultados del modelo con todas las características (Modelo A) y el modelo con las características más relevantes (Modelo B), se observa una diferencia promedio de 1 mm en RMSE, lo que es una mejora significativa en aplicaciones que requieran de un alta precisión en los registros de series temporales GPS, esta diferencia puede observarse con mas notoriedad en la Figura 17, donde se entrenaron modelos de tipo Predictor con cada configuración de características para un periodo de predicción de 16 días.

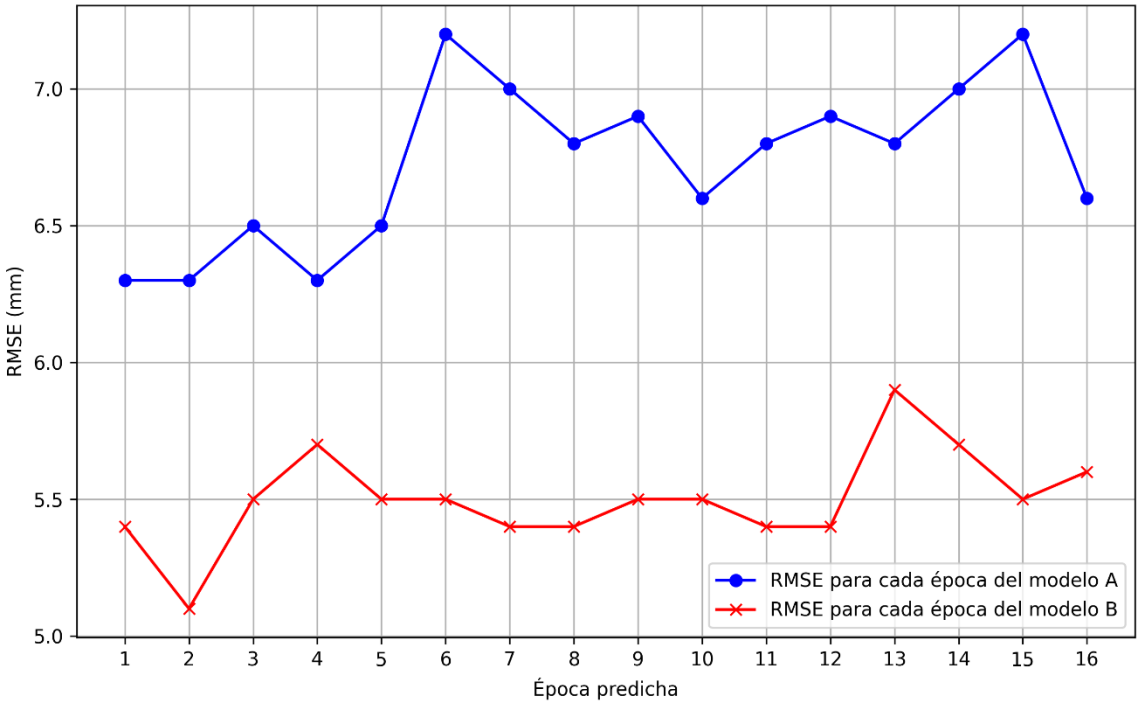


Figura 17.Comparativa de rendimiento en RMSE entre el modelo Predictor A y el modelo Predictor B.

El enfoque predictor permite la obtención de modelos de alta precisión, esto es demostrado por los rendimientos que presentan los modelos de este enfoque, alcanzando en las métricas RMSE= 5 mm, MAE= 4 mm y R^2 a 0.78 en la componente *up* y manteniendo los resultados en un rango similar a medida de que el periodo de predicción aumenta, pero, al coste de una mayor complejidad en el modelo, esto es debido a la longitud de entrada, traduciéndose en un mayor consumo de recursos computacionales y un mayor tiempo de entrenamiento, el cual oscila entre 5 minutos y 8 minutos dependiendo de los hiper-parámetros definidos, lo que vuelve el realizar un ajuste de hiper-parámetros en una tarea demandante.

Pasando al enfoque regresor, se observa que los modelos como Interpolador Geográfico alcanzan rendimiento de hasta RMSE= 6 mm, MAE= 5 mm y R^2 a 0.52 en la componente *up*, lo que lo hace un modelo preciso pero con alto riesgo estructural al añadir los *up* de las estaciones cercanas a la estación objetivo y sus coordenadas LAT y LON, pero al tener una longitud de entrada de 1 día, lo hace un modelo rápido de entrenar, oscilando entre 1 minuto y 1.5 minutos en su entrenamiento, lo que lo hace un buen modelo tanto en resultados como en tiempo de entrenamiento. Por otro lado, los modelos tipo Regresor alcanzan rendimiento de hasta RMSE= 5 mm, MAE= 4 mm y R^2 a 0.69 en la componente *up*, lo que lo hace un modelo con buen rendimiento, sumando el hecho de que este modelo es simple al solo poseer 12 características y longitud de entrada igual a 1 día, lo que permite que sus tiempos de entrenamiento ronden en 1 minuto. Comparando ambos tipos de modelos del enfoque regresor, se observa que en simplicidad y resultados el modelo Regresor es una mejor opción que el modelo Interpolador Geográfico, pero el primero dependiendo de tener una

cantidad de datos adecuada de la serie temporal a recuperar al momento de entrenar un modelo, mientras que el modelo Interpolador Geográfico al depender de los registros de series temporales de estaciones cercanas puede aplicarse en casos en que la serie temporal objetivo tiene pocos datos.

Comparando los mejores modelos de ambos enfoques, se observan rendimientos similares variando en R^2, donde los modelos como Predictor son más precisos en dicha métrica, pero, en términos de tiempos de entrenamiento y simplicidad de los modelos, los modelos tipo Regresor son una mejor opción que los modelos tipo Predictor.

6. Conclusión

En el presente estudio se evaluaron diferentes estrategias para la recuperación de registros de series temporales GPS, por medio del uso del algoritmo LSTM, el cual en situaciones de ausencia de datos en la variable objetivo no presenta problemas de rendimiento significativos al momento de entrenar los modelos. Para la evaluación, se usaron datos series temporales GPS la red de estaciones permanentes de RBCM.

La selección de características se llevó a cabo utilizando el abordaje SFS del método *wrapper*, con el cual se obtuvo que las características con mayor impacto en el modelo son Time, RA_s y , P_x. Los modelos con dichas características comparado con los modelos con todas las características son más precisos al recuperar la componente up de las estaciones ocupadas durante el entrenamiento. Como consideración se debe tener en cuenta que este procedimiento se realzó exclusivamente en el enfoque predictor, por lo que en un enfoque como regresor, las características resultantes al aplicar el procedimiento podrían no ser necesariamente las mismas.

Analizando los tres tipos de modelos diseñados, se determina que el mejor modelo es el Regresor seguido del modelo Interpolador Geográfico, a pesar de que los modelos del enfoque predictor tienen un mejor rendimiento. Esto se debe a que los modelos pertenecientes al enfoque regresor son más simples que los modelos del enfoque predictor, lo que reduce el riesgo estructural de los modelos. Pero estos modelos del enfoque regresor tienen limitantes, en el caso de los modelos tipo Regresor, se debe tener un volumen considerable de datos para entrenar el modelo, mientras que los modelos del tipo

Interpolador Geográfico dependen de la cantidad de datos disponibles de las estaciones cercanas.

Al momento de evaluar los modelos entrenados, se descubrieron condicionantes en el modelado, las cuales son dependientes de las series temporales GPS que van a ser recuperadas, la primera es la presencia de ruido en las series temporales, lo que podría provocar que el modelo tienda al sobre ajuste al aprender este ruido y no las características importantes de la serie temporal, el otro punto que influye en el rendimiento de los modelos son las variaciones de amplitud en los valores de la serie temporal, esto ocasiona que el modelo no logre relacionar correctamente los valores de la serie temporal con las características del modelo, lo que provocaría sobre ajuste en el modelo.

Para finalizar, es recomendable tratar los valores atípicos y variaciones significativas de amplitud presentes en las series temporales GPS que van a ser recuperadas, aunque esto no fue realizado en el presente estudio debido a que se buscaba diseñar modelos basados en LSTM con series temporales con la menor cantidad de alteraciones posibles en ellas. Otro punto por tratar es la inclusión de variables relacionadas con actividad tectónica en los sitios en que se emplazan las estaciones GPS, ya que este es un aspecto no considerado en este estudio, el cual en países con actividad sísmica significativa podría llegar a tener un impacto al momento de entrenar modelos basados en LSTM.

7. Referencias

Alevizakou, E., Siolas, G., & Pantazis, G. (2018). Short-term and long-term forecasting for the 3d point position changing by using artificial neural networks. *ISPRS Int J Geo Inf*. doi:https://doi.org/10.3390/ijgi7030086

Amiri-Simkooei, A. (2020). Least Squares Contribution to Geodetic Time Series Analysis. En J. Montillet, & M. Bos, *Geodetic Time Series Analysis in Earth Sciences* (pág. 185). Neuchatel, Lausanne, Switzerland: Springer Nature. doi:https://doi.org/10.1007/978-3-030-21718-1_6

Anonimo. (27 de Agosto de 2015). *Understanding LSTM Networks*. Obtenido de colah's blog: https://colah.github.io/posts/2015-08-Understanding-LSTMs/

Bengio, Y., Goodfellow, I., & Courville, A. (2016). *Deep Learning.* (T. Ditterich, Ed.) MIT Press.

Bevis, M., Bedford, J., & Caccamise II, D. (2020). The Art and Science of Trajectory. En J. Montillet, & M. Bos (Edits.), *Geodetic Time Series Analysis in Earth Sciences.* Springer Geophysics. doi:https://doi.org/10.1007/978-3-030-21718-1_1

Blewitt. (1997). Basics of the GPS Technique: Observation Equations.

Blewitt, G., & Lavallée, D. (2002). Effect of annual signals on geodetic velocity. *J. Geophys. Res.,*. doi:doi:10.1029/2001JB000570

Blewitt, G., Hammond, W., & Kreemer, C. (2018). Harnessing the GPS Data Explosion for Interdisciplinary Science. *Eos*. doi:https://doi.org/10.1029/2018EO104623

Bos, M., Montillet, J., Williams, S., & Fernandes, R. (2020). Introduction to Geodetic Time Series. En J.-P. Montillet, & M. S. Bos (Edits.), *Geodetic Time Series*

Analysis in Earth Sciences. Springer Geophysics. doi:https://doi.org/10.1007/978-3-030-21718-1_2

Cabello-Solorzano, K. O., Peña, M., Correia, L. J., & Tallón-Ballesteros, A. (2023). The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis. *Springer*. doi:https://doi.org/10.1007/978-3-031-42536-3_33

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 16-28.

Dill, R. (2008). *Hydrological model LSDM for operational Earth rotation and gravity field variations.* Potsdam: GFZ German Research Centre For Geosciences.

Firouzjaee, J., & Khalilian, P. (2024). The Interpretability of LSTM Models for Predicting Oil Company. *International Journal of Energy Research*. doi:https://doi.org/10.1155/2024/5526692

Gao, W., Li, Z., Chen, Q., Jiang, W., & Feng, Y. (2022). Modelling and prediction of GNSS time series using GBDT, LSTM and SVMmachine learning approaches. *Journal of Geodesy*. doi:https://doi.org/10.1007/s00190-022-01662-5

Ghimire, S., Deo, R., Wang, H., Al-Musaylh, M., Casillas-Pérez, D., & Salcedo-Sanz, S. (2022). Stacked LSTM Sequence-to-Sequence Autoencoder with Feature Selection for Daily Solar Radiation Prediction: A Review and New Modeling Results. *Stacked LSTM Sequence-to-Sequence Autoencoder with Feature Selection for Daily Solar Radiation Prediction: A Review and New Modeling Results*. doi:https://doi.org/10.3390/en15031061

Heflin, M., Donnellan, A., Parke, r. J., Lyzenga, G., Moore, A., & Ludwig, L. (2020). Automated estimation and tools. *Earth Sp Sci*.

Hérnandez Orallo, J., Ramírez Quintana, M., & Ferri Ramírez, C. (2004). *Introducción a la minería de datos.* Madrid: Pearson Educación S.A.

Herring, T., Melbourne, T., Murray, M., Floyd, M., Szeliga, W., King, R., . . . Wang, L. (2016). Plate boundary observatory and related networks: GPS data análisis methods and geodetic products. *Rev Geophys*, 759–808.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation, 9*(8), 1735-1780.

Ji, K., & Shen, Y. (2020). A Wavelet-Based Outlier Detection and Noise Component Analysis for GNSS Position Time Series. *International Association of Geodesy Symposia, vol 152*. doi:https://doi.org/10.1007/1345_2020_106

Jiao, H., Tao, X., Chen, L., Zhou, X., & Ju, Z. (2024). GNSS/5G Joint Position Based on Weighted Robust Iterative Kalman Filter. *Remote Sens*. doi:https://doi.org/10.3390/rs16061009

Kang, Q., Chen, E., Li., Z., H., L., & Liu, L. (2023). Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling. *Underground Space*. doi:https://doi.org/10.1016/j.undsp.2023.05.006.

Kuhn, M., & Johnson, K. (2016). *Applied Predictive Modeling.* Springer Nature. doi:DOI 10.1007/978-1-4614-6849-3

Lim, B., & Zohren, S. (2020). Time-series forecasting with Deep Learning: A Survey. *The Royal Soceity Publishing*. doi:https://doi.org/10.1098/rsta.2020.0209

Lima, F. T., & Souza, V. M. (2023). A Large Comparison of Normalization Methods on Time Series. *Big Data Research, 34,* 100407.

Liu, G. (2023). *Machine Learning with Python: Theory and Applications.* Singapore ; Hackensack, NJ :: World Scientific Publishing Co. Pte. Ltd.

Luo, X., Mayer, M., & Heck, B. (2012). Analysing Time Series of GNSS Residuals by Means of AR(I)MA Processes. *Springer*. doi:https://doi.org/10.1007/978-3-642-22078- 4_19

Montgomery, D., Jennigs, C., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting.* New Jersey: WILEYI NTERSCIENCE .

Muhammed, M. (2023). Hyperparameter Optimization of a Parallelized LSTM for Time Series Prediction. *Vietnam Journal of Computer Science*. doi:https://doi.org/10.1142/S2196888823500033

Nordman, M. (2010). *Improving GPS Time Series for Geodynamic Studies(Academic Dissertation, Finnish Geodetic Institute, University of Helsinki).* Helsinki: Finnish Geodetic Institute.

Petit, G., & Luzum, B. (2010). IERS conventions: Technical report. *Bureau International des Poids et mesures sevres (France)*.

Provotar, O., Linder, Y., & Veres, m. (2019). Unsupervised Anomaly Detection in Time Series Using LSTM-Based Autoencoders. *IEEE International Conference on Advanced Trends in Information Theory (ATIT).* doi:DOI: 10.1109/ATIT49449.2019.9030505

Puskas, C., Meertens, C., & Phillips, D. (2017). Hydrologic loading model displacements from the national and global data assimilation systems (NLDAS and GLDAS). *UNAVCO Geodetic Data Service*.

Raffel, C., & Ellis, D. (2015). Feed-Forward Networks with Attention Can Solve Some Long-Term Memory Problems. *arXiv*. doi:https://doi.org/10.48550/arXiv.1512.08756

Ramachandran, P., Zoph, B., & Le, Q. (2017). Searching for Activation Functions. *Google Brain*. doi:https://doi.org/10.48550/arXiv.1710.05941

Siami-Namini, S., Tavakoli, N., & Siami, A. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. *IEEE International Conference on Big Data*. doi:DOI: 10.1109/BigData47090.2019.9005997

SORKUN, M., İNCEL, Ö., & PAOLI, C. (2020). Time series forecasting on multivariate solar radiation data using deep learning (LSTM). *Turkish Journal of Electrical Engineering and Computer Sciences*. doi: https://doi.org/10.3906/elk-1907-218

Wang, J., Jiang, W., Li, Z., & Lu, Y. (2021). A new multi-scale sliding window LSTM framework (MSSW-LSTM): a case study for GNSS time-series prediction. *Remote Sens*. doi:https://doi.org/10.3390/rs13163328

Wang, J., Nie, G., Gao, S., Wu, S., Li, H., & Ren, X. (2021). Landslide Deformation Prediction Based on a GNSS Time Series Analysis and Recurrent Neural Network Model. *Remote Sens*. doi:https://doi.org/10.3390/rs13061055

Xianfeng, T., Huaxiu, Y., Yiwei, S., Charu, A., Prasenjit, M., & Suhang, W. (2020). Joint Modeling of Local and Global Temporal Dynamics for Multivariate Time Series

Forecasting with Missing Values. *AAAI Conference on Artificial Intelligence*. doi:https://doi.org/10.1609/aaai.v34i04.6056

Xin, T., Yang, Y., Zheng, X., Lin, J., Wang, S., & Wang, P. (2022). Time Series Recovery Using Adjacent Channel Data Based on LSTM: A Case Study of Subway Vibrations. *Appl. Sci.* doi:https://doi.org/10.3390/app122211497

Yan, H., Chen, W., Zhu, Y., Zhang, W., Zhong, M., & Liu, G. (2010). Thermal effects on vertical displacement of GPS stations in China. *Chin J Geophys*, 252–260. doi:https://doi.org/10.1002/cjg2.1492

Zhang, S., Gong, L., Zeng, Q., Li, W., Xiao, F., & Lei, J. (2021). Imputation of GPS Coordinate Time Series Using missForest. *Remote Sens*. doi:https://doi.org/10.3390/rs13122312

Zhang, Y., Thorburn, P., & Xiang, W. F. (2019). SSIM—A Deep Learning Approach for Recovering Missing Time Series Sensor Data. *IEEE Internet of Things Journal*. doi:10.1109/JIOT.2019.2909038

Zhao, S., Tsay, C., & Kronqvist, J. (2023). Model-based feature selection for neural networks: A mixed-integer programming approach. *Springer*. doi: https://doi.org/10.1007/978-3-031-44505-7_16

Zhu, D., Zhong, Z., Zhang, M., Wu, S., Zhang, K., L, i. Z., . . . Liu, J. (2023). An Improved Principal Component Analysis Method for the Interpolation of Missing Data in GNSS-Derived PWV Time Series. *Remote Sensing*. doi:https://doi.org/10.3390/rs15215153

8. Anexos

Archivos comprimidos de:

- Datos características empleadas por los modelos.
- Códigos usados para selección y descarga de estaciones de Brasil.
- Códigos de funciones de análisis exploratorio, entrenamiento de los modelos y generación de gráficos y mapas.
- Códigos del proceso del trabajo.