

PRONÓSTICO DIARIO DE ÍNDICE UV BASADO EN MEDICIONES IN SITU EN EL ÁREA METROPOLITANA DE MONTERREY

T E S I S

Que para obtener el grado de

Maestro en Ciencias

con Orientación en

**Ciencias de la Computación y
Matemáticas Industriales**

Presenta

Giovanni Gamaliel López Padilla

Director de Tesis:

Dr. Oscar Susano Dalmau Cedeño

Co-director de Tesis:

Dra. Adriana Ipiña Hernández

Autorización de la versión final

Datos del jurado

Alumno:

Giovanni Gamaliel López Padilla
Centro de Investigación en Matemáticas, CONACyT
Maestría en Ciencias con Especialidad en Computación y Matemáticas Industriales

Director de tesis:

Dr. Oscar Susano Dalmau Cedeño
Centro de Investigación en Matemáticas, CONACyT

Co-directora de tesis:

Dra. Adriana Ipiña Hernández
Instituto de Física de Rosario, CONICET-UNR

Sinodal 1:

Dra. Graciela María de los Dolores González Farías
Centro de Investigación en Matemáticas, CONACyT

Sinodal 2:

Dr. Adrián Pastor López Monroy
Centro de Investigación en Matemáticas, CONACyT

Agradecimientos

A la Dra. Adriana Ipiña, deseo expresar mi más sincero agradecimiento por brindarme la invaluable oportunidad de colaborar en diversos proyectos a su lado. Su guía y apoyo a lo largo de los años que hemos compartido han sido esenciales en mi crecimiento académico y profesional. Su confianza en mí y su liderazgo inspirador han sido un factor determinante en mi desarrollo. Gracias a esta convivencia, he adquirido experiencia, conocimientos y perspectivas que han enriquecido mi persona, ya sea desde el aspecto personal y académico.

Al Dr. Oscar Dalmau, quiero expresar mi agradecimiento por su apoyo y orientación a lo largo de esta travesía académica. Durante su clase, su dedicación y conocimiento enriquecieron mi perspectiva y me inspiraron a adentrarme en un campo completamente nuevo. Además, agradezco profundamente su disposición para asumir el rol de director de mi tesis, a pesar de que el tema era completamente desconocido para él. Su voluntad de embarcarse en este desafío y su compromiso con mi crecimiento académico fueron evidentes desde el principio.

A mis padres, Irene Padilla y Francisco López, quiero expresar mi más profundo agradecimiento por su apoyo incondicional desde el momento en que tomé la decisión de estudiar fuera de la ciudad. Su presencia y aliento han sido un faro de luz en mi camino académico. Este logro que hoy celebro es, en gran medida, un reflejo del inmenso esfuerzo y el apoyo que siempre me han brindado. Estaré eternamente agradecido por todo lo que han sacrificado y ofrecido para que yo pudiera alcanzar mis metas. Su amor y respaldo han sido invaluables en mi vida, y este éxito es también suyo.

A mis queridos amigos, María García, Edgar López, Iancarlo Espinosa y Natalia Huitzil, deseo expresar mi más profundo agradecimiento por su apoyo y amistad, ya que su presencia ha sido un pilar tanto en el ámbito personal como en el académico a lo largo de la maestría.

A mi querida amiga, Laura Aramburo quiero agradecerle por haber sido testigo de mi crecimiento y desarrollo desde una perspectiva distinta a la de mi entorno cotidiano. Tu amistad y tu capacidad para brindar una visión fresca y única son un tesoro que valoro profundamente. Gracias por estar a mi lado y por compartir este viaje de autodescubrimiento.

A mis queridas amigas, Mayra y Marcela Medellín, deseo manifestar mi más profundo agradecimiento por el apoyo que me brindaron desde el proceso de admisión hasta la actualidad. Su amistad fue un consuelo en los momentos de desafío y tensión, ya que con sus chistes y risas me permitieron distenderme y olvidar temporalmente las preocupaciones que pude enfrentar. Compartir buenos momentos con su compañía fue un regalo invaluable que atesoraré siempre. Gracias por ser parte esencial de esta etapa de mi vida académica y personal.

A mi querida amiga, Paulina Vázquez, quiero expresar mi sincero agradecimiento por su apoyo y consejo invaluables en diversos aspectos de mi vida. Sin su presencia e intervenciones, mi trayectoria actual sería significativamente distinta. Paulina ha sido una influencia esencial que ha enriquecido mi vida en múltiples niveles, guiándome y ofreciéndome valiosas perspectivas que han moldeado mi presente de manera fundamental. Su amistad y sabios consejos han dejado una huella imborrable en mi camino.

Finalmente, al Consejo Nacional de Ciencia y Tecnología (CONACyT), por haberme apoyado económicamente a través de su programa de becarios, el cual me permitió realizar mis estudios de maestría en el Centro de Investigación en Matemáticas.

Resumen

La exposición a la radiación solar ultravioleta (UV) por un tiempo prolongado puede ocasionar diversos efectos nocivos al ser humano, como el eritema cutáneo (enrojecimiento), daño al ADN, cáncer de piel y cataratas. Conocer la intensidad solar es necesario para evitar estos daños y proteger la salud de la población. El índice UV (IUV) se emplea como medida del riesgo de sufrir quemadura solar o eritema relacionado con la sobreexposición a la radiación UV. Los modelos de pronósticos del índice UV desempeñan un papel fundamental para proveer información sobre la protección solar. En esta Tesis se proponen diversos modelos basados en métodos de descomposición de señales y redes neuronales para elaborar el pronóstico horario del IUV a lo largo del día, basado en mediciones de la radiación solar global in situ del Sistema Integral de Monitoreo Ambiental (SIMA) y la fracción de nubes estimada satelitalmente por el instrumento OMI/NASA, en el Área Metropolitana de Monterrey. Con el fin de extraer patrones espaciales, los modelos propuestos realizan una descomposición de la radiación solar global por medio del método Empirical Wavelet Transform (EWT) o el Variational Mode Decomposition (VMD), las cuales son ingresadas a la red neuronal. Para determinar el rendimiento de los modelos propuestos, se calcularon la Raíz del Error Cuadrático Medio (RMSE), el Error Absoluto Medio (MAE) y el coeficiente de determinación (R^2). Los resultados de los experimentos realizados muestran que el modelo basado en el codificador del Transformers (Arclight) y el modelo perceptrón multicapa (MLP), los cuales emplean la descomposición EWT, obtienen un desempeño semejante o mejor que los demás modelos propuestos para el pronóstico horario del IUV para el siguiente día y el séptimo día consecutivo. Sin embargo, no existe una diferencia significativa en el desempeño de los modelos propuestos para el pronóstico del IUV para el siguiente día. Una ventaja computacional del modelo Arclight es su bajo número de parámetros, por lo que estos modelos pueden ser ejecutadas en dispositivos electrónicos con recursos computacionales bajos como teléfonos móviles.

Palabras clave: radiación solar global, índice UV, descomposición de señales, mediciones satelitales, redes neuronales, Transformers, pronóstico

Dedicatoria

A mis padres, hermanos y amigos. Los quiero.

¿Dónde estarás ahora? ¿Estarás escribiendo un artículo en un laboratorio de investigación en Estados Unidos? ¿Estarás trabajando toda la noche en un experimento? Es posible que estés pensando en tu padre. No te acuerdas de las tres semanas que pasamos juntos. Pero eso está bien. En este momento, estás respirando, hablando, pensando en algo muy complicado. Eso es suficiente. Estás viendo el mismo mundo que yo, este mundo con un futuro desconocido.

Okabe Rintarou. Steins;Gate.

Lista de iniciales

SMARTS	Simple Model of the Atmospheric Radiative Transfer of Sunshine
SIMA	Sistema Integral de Monitoreo Ambiental
GHI₀	Irradiancia solar global extraterrestre
GHI	Irradiancia solar global horizontal
OMS	Organización Mundial de la Salud
VMD	Variational Mode Decomposition
IMF	Funciones modales intrínsecas
RNN	Redes neuronales recurrentes
OMI	Ozone Monitoring Instrument
S_{er}	Espectro de acción eritémica
EWT	Empirical Wavelet Transform
LSTM	Long-Short Term Memory
CNN	Red neuronal convolucional
ANN	Red neuronal artificial
MLP	Perceptrón multicapa
seq2seq	Secuencia a secuencia
K_T	Índice de claridad
UV	Ultravioleta
IUV	Índice UV

Índice general

Datos del jurado	I
Agradecimientos	II
Resumen	IV
Dedicatoria	V
Lista de iniciales	VI
1. Introducción	1
1.1. Objetivos	2
1.2. Contenido de la tesis	3
2. Antecedentes y trabajos relacionados	4
3. Marco teórico	6
3.1. Radiación solar	6
3.1.1. Posición del observador	7
3.1.2. Posición solar	8
3.1.3. Irradiancia Global Horizontal Extraterrestre	9
3.1.4. Índice de claridad	10
3.1.5. Índice UV	11
3.1.6. Modelo SMARTS	12
3.2. Métodos de descomposición	12
3.2.1. Empirical Wavelet Transform	13
3.2.2. Variational Mode Decomposition	14
3.3. Métodos de interpolación	16
3.4. Modelos basados en redes neuronales	17
3.4.1. Perceptrón	17
3.4.2. Perceptrón Multicapa	18
3.4.3. Redes convolucionales	18
3.4.4. Long-Short Term Memory	20
3.4.5. Aprendizaje secuencia a secuencia	22
3.4.6. Redes Neuronales Residuales	22
3.4.7. Mecanismos de atención	22
3.4.8. Transformer	24
4. Metodología	28
4.1. Datos	28
4.1.1. Mediciones <i>in situ</i>	28
4.1.2. Datos satelitales	32

4.1.3. Estimación del IUV	33
4.1.4. Aumentación de datos	35
4.2. Entrenamiento de la red neuronal	36
4.2.1. Segmentación de los conjuntos de datos	37
4.2.2. Datos de la entrada y salida esperada	37
4.2.3. Función de pérdida	39
4.2.4. Problemas de sobreajuste	40
4.3. Propuesta del modelo	41
4.4. Modelos implementados	42
4.4.1. Modelos básicos	43
4.4.2. Modelos híbridos	44
4.4.3. Modelos residuales	45
4.4.4. Modelos basados en Transformers	46
5. Resultados	50
5.1. Métricas de rendimiento	50
5.1.1. Error Absoluto Medio	50
5.1.2. Error Cuadrático Medio	51
5.1.3. Coeficiente de determinación	51
5.2. Número de parámetros	51
5.3. Tiempo de entrenamiento y ejecución	52
5.4. Predicción diaria	54
5.4.1. EWT	54
5.4.2. VMD	58
5.4.3. Comparación EWT y VMD	62
5.4.4. Comparación con mediciones <i>in situ</i> UVA+UVB	63
5.5. Predicción semanal	64
5.5.1. EWT	64
5.5.2. VMD	68
5.5.3. Comparación EWT y VMD	71
6. Conclusiones	73
Referencias	75
A. Anexo	83

Introducción

La radiación solar que llega a la Tierra tiene un impacto significativo en la salud humana, ya que la exposición moderada a la misma es beneficiosa al ser humano debido a que desencadena la síntesis de vitamina D en la piel. Esta hormona ayuda a la prevención del raquitismo, la osteoporosis y favorecer la circulación sanguínea (Jiménez, Moya, y Jiménez, 2015; Masip, Morell, y Serra, 2008; Neale y cols., 2023; Webb y Holick, 1988). Sin embargo, la exposición excesiva puede producir efectos nocivos como el cáncer de piel, cataratas y deterioro gradual e irreversible del colágeno, la elastina y otras fibras que dan estructura a la piel (Gilaberte y cols., 2011; Modenese y cols., 2016). El uso regular de sustancias realizan una protección frente a los efectos de la radiación solar antes de los 18 años puede reducir la incidencia del cáncer de piel en la capa más interior de la epidermis hasta 78% (Honeyman, 2002; Lara del Valle, 2010; Rosso y cols., 1996). Antes de los 18 años se ha acumulado 70% del daño total que recibirá el ADN en la piel y que se manifestará en la edad adulta, por lo que es importante iniciar las medidas de protección desde la infancia (Jurado-Santa Cruz, Medina-Bojórquez, Gutiérrez-Vidrio, y Ruiz-Rosillo, 2011; Stern, 1986). El cambio climático ha provocado un aumento en la temperatura en la superficie terrestre, en especial en Centroamérica debido a que se encuentra en la franja latitudinal con mayor intensidad solar (Pounds, Fogden, y Campbell, 1999; Still, Foster, y Schneider, 1999). En 1992 se definió el Índice UV (IUV) debido a que hubo un aumento en la radiación UV en consecuencia del agotamiento en la capa de ozono (Heckman, Liang, y Riley, 2019). La Organización Mundial de la Salud (OMS) adoptó el IUV como indicador estándar para representar la intensidad de la radiación UV solar en la superficie terrestre. El índice se expresa como un valor mayor a cero, y cuanto más alto, la probabilidad de lesiones en los ojos y cutáneas es mayor, además estas lesiones pueden producirse en un menor tiempo. El desarrollo de un modelo de pronóstico del IUV es esencial para obtener estimaciones e informar el riesgo que se presentará en una región, con el fin de proteger a la población de manera anticipada sobre los niveles de la radiación UV. Además, facilita la planificación y protección en entornos al aire libre, mejorando la seguridad y el bienestar de los individuos en su vida diaria y actividades laborales. En esta Tesis se desarrollarán modelos basados en redes neuronales para el pronóstico del IUV en el Área Metropolitana de Monterrey debido a que se encuentra en la franja latitudinal con mayor intensidad del IUV. La figura 1.1 se muestra el promedio mensual del IUV en el mes de agosto del 2010 estimado por el instrumento satelital Ozone Monitoring Instrument (OMI) de la NASA.

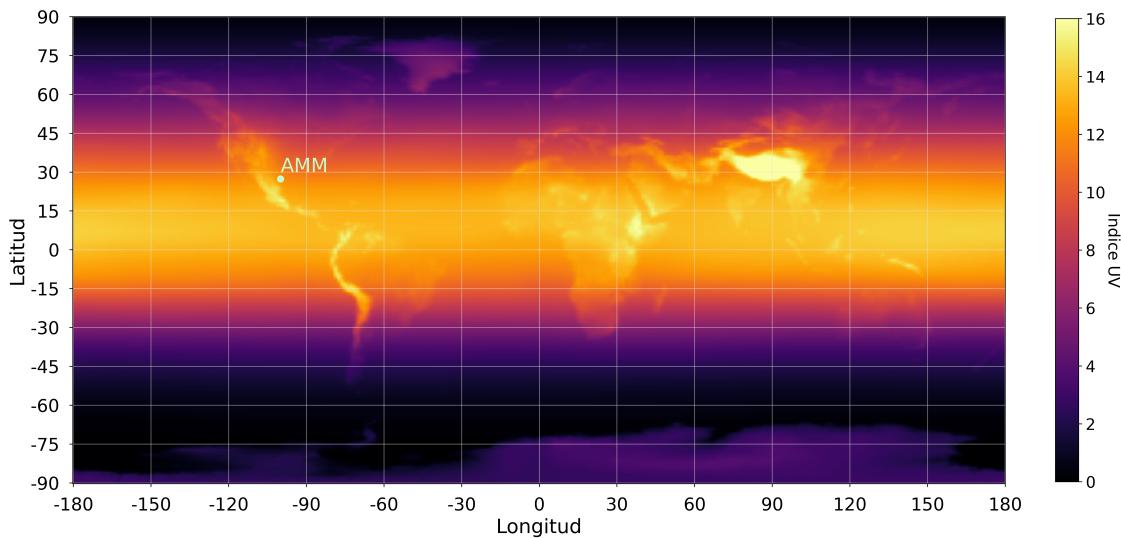


Figura 1.1: Promedio mensual del IUV para el mes de agosto del 2010 estimado por el instrumento OMI de la NASA. El punto indica la posición geográfica del Área Metropolitana de Monterrey.

1.1. Objetivos

El objetivo general de esta Tesis es analizar y evaluar el rendimiento de diversos modelos basados en redes neuronales y métodos de descomposición de señales para obtener un pronóstico horario del IUV para el siguiente día. Se busca obtener una comprensión profunda de cómo estos modelos pueden abordar eficazmente el desafío del pronóstico del IUV con precisión. Para lograrlo, se plantean los siguientes objetivos específicos:

- **Filtrar y reconstruir datos:** Filtrar datos históricos medidos por el Sistema Integral de Monitoreo Ambiental del Área Metropolitana de Monterrey. Esto implica el análisis de la calidad de los datos, la corrección de posibles errores y la normalización de las variables relevantes.
- **Diseñar los modelos basados en redes neuronales:** Desarrollar diferentes arquitecturas de redes neuronales, como redes neuronales convolucionales, recurrentes, redes neuronales basadas en atención y Transformer. Entrenar modelos utilizando los datos preparados y ajustar hiperparámetros para optimizar el rendimiento y la precisión de las predicciones diarias a lo largo de las horas del día.
- **Evaluar y comparar el rendimiento de los modelos:** Analizar la diferencia del rendimiento para el pronóstico del IUV a lo largo del día para el día siguiente y una semana después obtenidos por los modelos propuestos al realizar pronósticos

con base en la descomposición de señales utilizando el método Empirical Wavelet Transform y Variational Mode Decomposition.

1.2. Contenido de la tesis

A continuación, se describe de forma breve el contenido de los siguientes capítulos.

- Con el fin de brindar un contexto de las diferentes técnicas y modelos que se han propuesto para resolver problemas de pronóstico, en el capítulo 2 se mencionan los trabajos previos relacionados con el pronóstico del IUV y radiación solar en diferentes escalas temporales.
- En el capítulo 3 se introducen los conceptos básicos de la radiación solar e IUV y se presenta el modelo de transferencia radiativa SMARTS. Además se definen los métodos de descomposición (Empirical Wavelet Transform y Variational Mode Decomposition) y el método de interpolación. De la misma forma se describen las diferentes arquitecturas de redes neuronales implementadas en este trabajo.
- En el capítulo 4 se muestra la metodología empleada para realizar el filtrado y reconstrucción de la base de datos histórica de la radiación solar global del Sistema Integral de Monitoreo Ambiental y de la columna total de ozono y fracción de nubes estimadas por el instrumento satelital OMI. Se detalla la arquitectura de los modelos propuestos y las técnicas empleadas durante el entrenamiento de las redes neuronales.
- En el capítulo 5 se analizan los resultados de las diferentes propuestas en este trabajo para generar el pronóstico del IUV horario, para el día siguiente y para una semana después.
- Por último en el Capítulo 6 se desglosan las conclusiones y trabajos a futuro.

Antecedentes y trabajos relacionados

En 1993, el estudio de la radiación solar en México inició con una prueba piloto organizada por el primer programa de información pública para hispanohablantes a través de la red de televisión mexicana Televisa ([TrentUniv, Televisa, y Instituto Tecnológico y de Estudios Superiores de Monterrey, 1993](#)). Esta prueba tenía el objetivo de monitorear la radiación solar entre los rangos de 280 a 315 nm (UVB). En 1996, se diseñó la primera red de monitoreo de radiación UV en México, la cual se encuentra en la Ciudad de México ([Acosta, 1997](#)). Actualmente este sistema sigue comunicando la radiación UV al público en general ([Secretaría del Medio Ambiente de la Ciudad de México, 2023](#)). Gracias a esta red se han desarrollado investigaciones basadas en mediciones de radiación solar, variables atmosféricas y meteorológicas ([Retama, Ramos-Cerón, Rivera-Hernández, Allen, y Velasco, 2022; Viatte y cols., 2022](#)). Ipiña, López-Padilla, Retama, Piacentini, y Madronich ([2021](#)) demostraron que existe un aumento en la magnitud del IUV debido a una disminución en la contaminación atmosférica. Por ende, la radiación solar debe seguir siendo monitoreada para informar y realizar estudios con el fin de disminuir el riesgo de cáncer de piel en la población. En años recientes los avances de los modelos basados en redes neuronales han propiciado un aumento en la exploración y aplicación de estos modelos para pronosticar variables climatológicas, incluyendo la radiación solar. La radiación solar a nivel de suelo contiene componentes determinísticos y aleatorios, ya que existe una periodicidad en diferentes frecuencias de tiempo (día, mes y año) ([Box, Jenkins, Reinsel, y Ljung, 2015](#)) y es afectada por componentes atmosféricos como las nubes ([Boland, 2008](#)). Es por ello, que la precisión obtenida por las redes neuronales para el pronóstico de la radiación solar se encuentra relacionada con la habilidad del modelo de introducir cambios inducidos por las nubes ([Inman, Pedro, y Coimbra, 2013](#)). Existen diferentes aproximaciones para realizar un pronóstico de radiación solar en diferentes intervalos de tiempo y diferentes ubicaciones utilizando modelos basados en aprendizaje máquina como: árboles regresores con potenciadores de gradiente (GBRT) y levantamiento de gradiente extremo (XGBoost) ([Huang y cols., 2021](#)), y métodos basados en redes neuronales constituidas de capas recurrentes y convolucionales ([Alzahrani, Shamsi, Dagli, y Ferdowsi, 2017](#)). Las redes neuronales, los métodos basados en aprendizaje máquina y sus modelos híbridos han demostrado obtener de manera consistente resultados en el modelado de procesos físicos con naturaleza estocástica, debido a que son capaces de compensar los errores sistemáticos de las mediciones y las variaciones debidas a las condiciones del cielo ([Paulescu, Paulescu, Gravila, y Badescu, 2013](#)). Las redes neuronales son modelos estadísticos que

establecen una relación entre los datos de entrada y salida durante el entrenamiento, esta relación está conformada por varias capas que forman conexiones entre los nodos de la entrada, los nodos internos y la salida de la red neuronal. Las redes neuronales basadas en capas recurrentes (RNN), en especial la capa Long Short-Term Memory (LSTM) es empleada para el pronóstico relacionado con energías renovables debido a que puede obtener patrones temporales de los datos (Husein y Chung, 2019; Pang, Niu, y O'Neill, 2020). Además estas redes usualmente son acompañadas de capas convolucionales, debido a que estas redes son consideradas como extractores de características relevantes de los datos (Zang y cols., 2020). En años recientes se han utilizado técnicas de descomposición de señales para suplementar o reemplazar a la capa convolucional. Para ello las técnicas de descomposición empleadas han sido el método modal variacional (Cannizzaro y cols., 2021; X. Wang, Wang, Yuan, Wang, y Cheng, 2021), transformaciones basadas en ondículas empíricas (Y. Wang, Guo, Ma, y Liu, 2022) y descomposición local de media robusta (Anh, Ravinesh, Mumtaz, Shahab, y Nawin, 2021). A su vez, varios estudios han propuesto a la metodología de ensambles de modelos para pronosticar la radiación solar, ya que al tener diferentes modelos realizando la misma tarea se pueden aprovechar las virtudes de cada uno y así obtener resultados conscientes con mejor precisión (Cornaro, Pierro, y Bucci, 2015; Phy, Byun, y Park, 2022; Solano, Dehghanian, y Affonso, 2022). Tras el surgimiento del modelo Transformer se han propuesto diversos modelos basados en mecanismos de atención debido a que extraen de manera más efectiva patrones temporales de los datos, una ventaja de estos modelos es que la complejidad computacional es menor en comparación en modelos basados en RNN (Sharda, Singh, y Sharma, 2021; Zhou, Li, Wang, y Liu, 2023). En esta Tesis se diseñaron modelos basados en redes MLP, CNN, RNN y Transformers para el pronóstico del IUV a lo largo del día dada la radiación solar medida in situ y el factor de nubes estimado por satélites del anterior. Estos modelos realizan una descomposición de señales de la radiación solar empleando los métodos Empirical Wavelet Transform y Variational Mode Decomposition.

Marco teórico

Este capítulo aborda conceptos fundamentales relacionados con la radiación solar, que son importantes para entender cómo la radiación solar afecta a la atmósfera y a los seres vivos. Además, se introducirá el modelo SMARTS, el cual es una herramienta para estimar la radiación solar en diferentes lugares del mundo. Este modelo utiliza datos atmosféricos, meteorológicos y geográficos para calcular la intensidad de radiación solar que llega a nivel del suelo. También se discutirán los métodos de descomposición de señales, las cuales son técnicas matemáticas empleadas para separar una señal en componentes más simples. Estos métodos son útiles para analizar las diferentes tendencias y patrones presentes en los datos. Posteriormente, se presentará el método *spline* cúbico, que es una técnica de interpolación aplicada para aproximar curvas suaves a través de un conjunto de puntos. Este método es útil para construir curvas que sean suaves y continuas y se utiliza a menudo en la modelización así como en el análisis de datos. Por último, se discutirán las bases de los modelos basados en redes neuronales, los cuales consisten en una clase de modelos de aprendizaje automático que imitan el funcionamiento del cerebro humano. Estos modelos han demostrado ser muy efectivos en una amplia variedad de problemas de predicción y clasificación, y han sido utilizados en muchas aplicaciones prácticas.

3.1. Radiación solar

La radiación solar son ondas electromagnéticas emitidas por el Sol que se propaga en todas las direcciones. Este tipo de ondas se propagan a través del vacío, por lo que pueden atravesar el espacio interplanetario e interestelar y llegar a la Tierra. A cada onda le corresponde una longitud, la cual tiene una cantidad específica de energía. Al conjunto de longitudes de onda se le conoce como espectro solar. Conforme la radiación solar atraviesa la atmósfera terrestre, esta es atenuada, dispersada y absorbida por los diferentes gases y partículas presentes en la atmósfera terrestre. La radiación solar que llega a la superficie de la Tierra se encuentra en el rangopectral infrarrojo, visible y ultravioleta. Estos rangos abarcan las longitudes de onda de 700 nm a 1 mm para el infrarrojo, entre 400 y 700 nm en el visible y de 280 a 400 nm en el ultravioleta (D’Orazio, Jarrett, Amaro-Ortiz, y Scott, 2013). La figura 3.1 muestra los rangos de las longitudes de onda de la radiación solar que llega a la superficie terrestre.

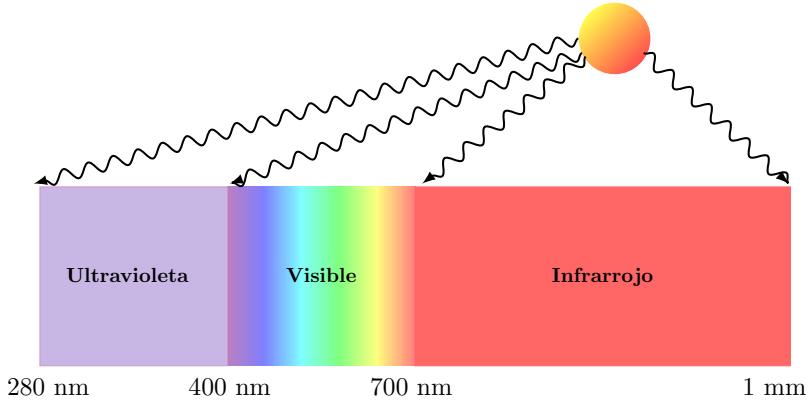


Figura 3.1: Rangos del espectro de la radiación solar que llegan a la superficie terrestre.

La figura 3.2 muestra la comparación entre los espectros de emisión de un cuerpo negro a 5250 K , de la radiación solar medida en el espacio exterior y a nivel de suelo.

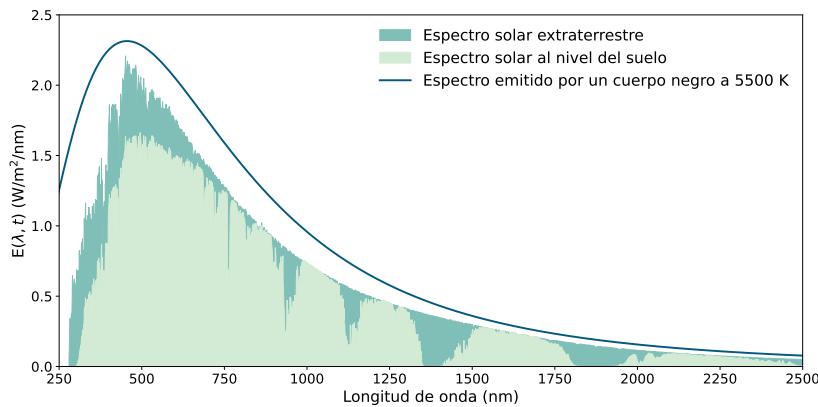


Figura 3.2: Espectro de emisión de un cuerpo negro a 5250 K , irradiado por el Sol en el espacio exterior y espectro solar medido a nivel del suelo.

La irradiancia solar es la energía por unidad de superficie para un rangopectral, la cual se define como:

$$I_s(t) = \int_{\lambda_1}^{\lambda_f} E(\lambda, t) d\lambda \quad (3.1)$$

donde λ_i y λ_f definen el rangopectral considerado, $E(\lambda, t)$ y $I_s(t)$ son el espectro e irradiancia solar para un tiempo t . La irradiancia solar se mide en vatios por metro cuadrado (W/m^2) en el Sistema Internacional de Unidades (SI).

3.1.1. Posición del observador

La posición geográfica de un observador es un aspecto fundamental que influye en una amplia gama de fenómenos y procesos relacionados con el entorno natural y humano.

Para caracterizar la posición geográfica del observador debemos definir los meridianos y paralelos de la Tierra. Los meridianos son semicírculos imaginarios que pasan por los polos norte y sur y los paralelos son semicírculos imaginarios paralelos al eje de la Tierra. La figura 3.3 se muestra una representación de los paralelos y meridianos de la Tierra.

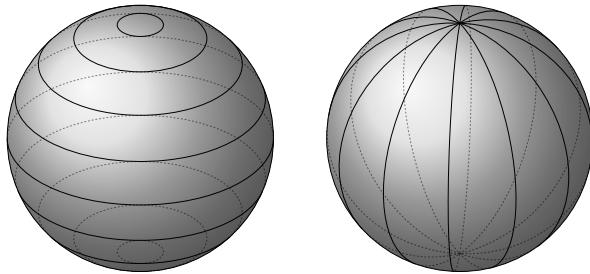


Figura 3.3: Representación gráfica de los paralelos y meridianos de la Tierra respectivamente.

Con base en los meridianos y paralelos se definen la longitud, latitud y franja horaria para un observador. La longitud es la medida del arco entre el meridiano de Greenwich y el meridiano donde se encuentra el observador. Este ángulo se encuentra en el rango de -180° a 180°. La latitud es la distancia entre el paralelo del observador y la línea del Ecuador, la cual es representada en grados y puede tener valores entre -90° y 90°. Las franjas horarias son regiones limitadas por meridianos espaciados por 15 grados empezando por el meridiano de Greenwich, en estas franjas se suele regir por una misma hora oficial. Debido a que la Tierra gira con una dirección de oeste a este, cada que se atraviesa una franja en esta dirección, se suma una hora, en dirección contraria se resta una hora. Sobre meridiano de 180° Se marca la línea internacional de cambio de fecha. Simplificando lo anterior, la hora en cada franja horaria se calcula de la siguiente manera:

$$h_{tz} = \left[\frac{\text{meridiano}}{15} \right] \quad (3.2)$$

donde h_{tz} da como resultado un valor entero.

3.1.2. Posición solar

La intensidad de la radiación solar depende de las coordenadas geográficas, la composición atmosférica y de la posición del Sol en la bóveda celeste. Para determinar su trayectoria se utilizan variables geométricas en función del tiempo. La declinación solar (δ) es el ángulo entre el ecuador y una línea trazada desde el centro de la Tierra al Sol. La declinación

solar varía durante el año y depende del día consecutivo (c), definida como:

$$\delta = 23.45 \sin(\gamma) \quad (3.3)$$

$$\text{donde } \gamma = \frac{360}{365}(c - 81) \quad (3.4)$$

donde γ es el ángulo que define la variación de la declinación a lo largo del año (Allen y cols., 2005). En el caso de los años bisiestos, el denominador es 366. Para considerar la variación del movimiento aparente del Sol sobre el horizonte del lugar debido a la franja horaria y la precesión de la Tierra, se emplea una corrección longitudinal y una perturbación anual debido a la velocidad de la Tierra (EoT). La corrección longitudinal es de 4 minutos por grado entre la longitud local y la longitud correspondiente a la franja horaria de la región. Por otro lado, la EoT se calcula como:

$$EoT = 9.87 \sin(2\gamma) - 7.53 \cos(\gamma) - 1.5 \sin(\gamma) \quad (3.5)$$

Con el fin de considerar estas variaciones al tiempo local, se define el tiempo solar (Iqbal, 1983), el cual es medido en horas y es estimado como:

$$\text{Tiempo solar} = \text{Tiempo local} + \frac{4(\text{longitud} - 15h_{tz}) + EoT}{60} \quad (3.6)$$

Por otro lado, el ángulo horario es la representación del arco que existe entre el plano que contiene la posición del sol y el plano de intersección con el meridiano local, el cual es calculado por:

$$\omega = 15(\text{Tiempo solar} - 12) \quad (3.7)$$

Asimismo, el ángulo cenital es el ángulo entre la línea que apunta al Sol desde el observador y la línea perpendicular hacia el cenit del lugar. Este ángulo es definido como:

$$z = \cos^{-1}(\cos(\phi) \cos(\delta) \cos(\omega) + \sin(\phi) \sin(\delta)) \quad (3.8)$$

donde ϕ es la latitud del sitio donde se encuentra el observador. La figura 3.4 muestra la representación gráfica del ángulo horario y el ángulo cenital.

3.1.3. Irradiancia Global Horizontal Extraterrestre

La Irradiancia solar global extraterrestre (GHI_0) es la radiación total que llega a la primera capa de la atmósfera terrestre (Iqbal, 1983). La GHI_0 varía anualmente debido

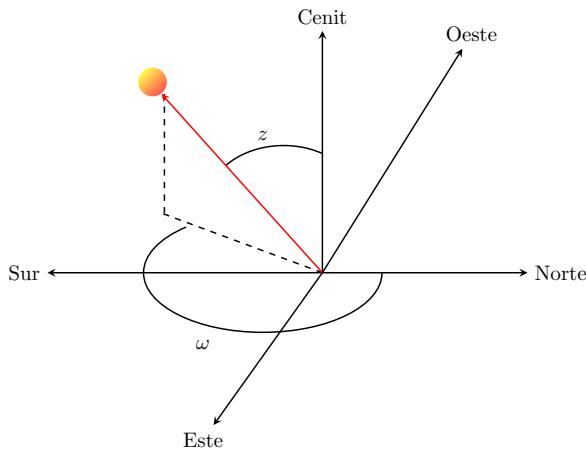


Figura 3.4: Representación gráfica del ángulo cenital (z) y el ángulo horario (ω).

a la trayectoria de la Tierra alrededor del Sol, expresada como:

$$GHI_0 = ISC \left[1 + 0.033 \cos \left(\frac{360c}{365} \right) \right] \cos(z) \quad (3.9)$$

donde ISC es la constante solar con un valor de 1367 W/m^2 (Li, Lian, Wang, Ma, y Zhao, 2011).

3.1.4. Índice de claridad

El índice de claridad (K_T) es definido como la razón de la irradiancia solar global (GHI) a nivel del suelo y la GHI_0 . Este índice fue descrito por Black, Bonython, y Prescott (1954) para estimar la radiación solar en condiciones de cielo despejado basado en mediciones afectadas por nubes. El K_T está definido como:

$$K_T = \begin{cases} \frac{GHI}{GHI_0} & \text{si } GHI_0 \neq 0 \\ 0 & \text{si } GHI_0 = 0 \end{cases} \quad (3.10)$$

El K_T es una representación de la fracción afectada por la atenuación, absorción y dispersión de la GHI_0 luego de pasar la atmósfera. Este valor estima las variaciones de la intensidad solar en presencia de nubes. El rango se encuentra entre 0 y 1. Los valores cercanos a cero indican atenuación por presencia de nubes o altas concentraciones de contaminantes, mientras que aquellos mayores o iguales a 0.8 representan condiciones de cielo despejado o prácticamente despejado.

3.1.5. Índice UV

La radiación UV se divide en tres rangos: UVA que abarca longitudes de onda de 315 a 400 nm, UVB entre 280 y 315 nm y UVC son las longitudes de onda que van de 100 a 280 nm. En el recorrido a través de la atmósfera terrestre, la radiación UVC y aproximadamente el 90 % de la radiación UVB es absorbida debido a que esas longitudes de onda tienen la suficiente energía para desencadenar el rompimiento de moléculas como el ozono, dióxido de carbono, peróxido de hidrógeno y ácido nítrico (McCullough, 1977). La radiación UVA es la menos afectada, es por ello que la radiación UV que impacta a la superficie se compone del UVA y una pequeña porción del UVB (WHO, 2016). La radiación UV genera efectos negativos como quemaduras, envejecimiento prematuro y cáncer en la piel, así como daño al ADN y a los ojos. Para cuantificar el efecto de enrojecimiento (eritema) en un tipo de piel caucásica, debido a la incidencia de la radiación UV, se definió el espectro de acción eritémica (S_{er}) (CIE, 2014). El espectro de acción eritémica es una función que representa la sensibilidad de la piel humana del tipo caucásico al interactuar con la radiación UV. La figura 3.5 muestra el espectro de acción eritémica para cada longitud de onda de la radiación UV.

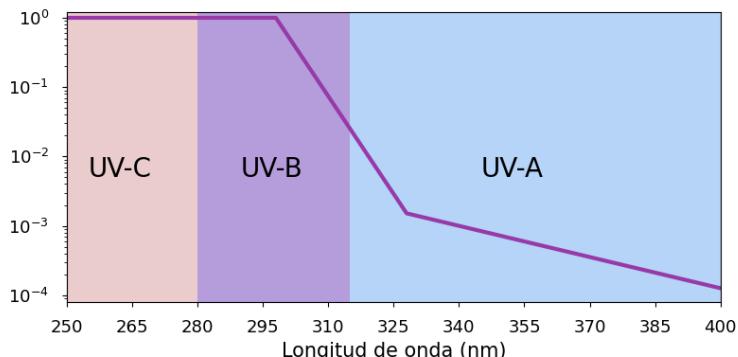


Figura 3.5: Espectro de acción eritémica para cada longitud de onda en la radiación UV.

Con el propósito de dar información de manera más simple al público general sobre los riesgos de sufrir daño en la piel por exposición a la radiación UV, se definió el Índice Ultravioleta (IUV) como (WHO, 2016):

$$\text{IUV}(t) = 40 \int_{280\text{nm}}^{400\text{nm}} E(\lambda, t) S_{er}(\lambda) d\lambda \quad (3.11)$$

donde $E(\lambda, t)$ es el espectro solar medido en $\text{W/m}^2/\text{nm}$ y S_{er} es el espectro de acción eritémica. La Organización Mundial de la Salud (OMS) define una escala para el Índice UV que va de 1 a 11 (o más). Para valores de 11 o superiores, se considera un riesgo

importante para el ser humano debido a que puede sufrir daños en la piel en un periodo de tiempo breve (WHO, WMO, UNEP, y ICNIRP, 2002). La figura 3.6 se representan los niveles de riesgo que contempla la WHO para cada rango del IUV.

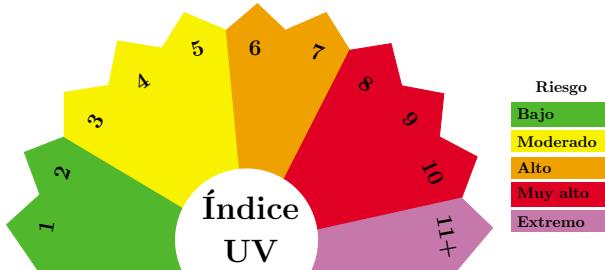


Figura 3.6: Niveles de riesgo de acuerdo al rango de valores del IUV.

3.1.6. Modelo SMARTS

El Simple Model of the Atmospheric Radiative Transfer of Sunshine (SMARTS por sus siglas en inglés) es un modelo de transferencia radiativa desarrollado en fortran, el cual estima la irradiancia directa, difusa y global que incide en la superficie de la Tierra (Gueymard, 2009). El SMARTS calcula la radiación solar en el rango de 280 a 4000 nm, es decir incluye los rangos UVA, UVB, visible e infrarrojo. El SMARTS estima la irradiancia espectral en cualquier lugar de la superficie terrestre, hora del día y día del año. Mediante instrumentos que registran la radiación solar como: radiómetros, espectroradiómetros, biómetros, piranómetros y pirhelíometros, es posible determinar la variación de su intensidad, y con ello, derivar la composición atmosférica. Con base en los cálculos de los espectros de irradiancia solar empleando el SMARTS, también es posible determinar el IUV. Este modelo es de código abierto y para obtenerlo es necesario realizar una petición al National Renewable Energy Laboratory para su acceso (National Renewable Energy Laboratory, 2006).

3.2. Métodos de descomposición

Los métodos de descomposición son técnicas matemáticas que como su nombre lo indica, permiten descomponer una señal o serie temporal en un número finito de componentes más simples, bajo ciertas condiciones y restricciones. Estos métodos son útiles en diversos campos como: la economía, la física, la ingeniería y la ciencia de los datos o en cualquier área que demande un análisis de comportamiento de las señales y series temporales. El objetivo principal de los métodos de descomposición es separar una señal o serie temporal en sus componentes subyacentes, que a menudo corresponden a diferentes patrones o

tendencias en los datos. La descomposición puede variar dependiendo de la naturaleza de los datos y del problema específico. Los métodos de descomposición más comunes incluyen la descomposición clásica de series temporales, el análisis de componentes principales (PCA), el análisis de componentes independientes (ICA) y la transformada de Fourier.

3.2.1. Empirical Wavelet Transform

Las transformaciones basadas en funciones wavelet son una de las técnicas usadas para el procesamiento de señales. El análisis basado en funciones wavelet tiene el propósito de descomponer una señal como una suma finita de funciones modales intrínsecas (IMF por su abreviación en inglés). Asimismo una dada señal, se representa como:

$$f(t) = \sum_{k=0}^N f_k(t) \quad (3.12)$$

donde f_k son las IMF de la señal. Cada IMF es una función dependiente de una amplitud (F_k) y una frecuencia (φ_k) tal que:

$$f_k(t) = F_k(t) \cos(\varphi_k(t)) \quad (3.13)$$

donde se requiere que la función $F_k(t)$ y la derivada de $\varphi_k(t)$ sean no decrecientes. Las transformaciones basadas en funciones de wavelet sobre una señal son calculadas a través del producto interior de la señal y una función wavelet ($\phi_{u,s}$) donde s y u son parámetros que describen la dilatación y traslación de la función. Esta transformación se escribe de la siguiente forma:

$$W(u, s) = \int f(\tau) \phi_{u,s} d\tau \quad (3.14)$$

Por lo tanto, las IMF de la señal son descritas de la siguiente forma:

$$f_k(t) = W_k(u, s) \phi_{u,s} \quad (3.15)$$

Para aplicar el método de transformación basado en funciones de wavelet empíricas (EWT por sus siglas en inglés), se utilizan las funciones de escalado, wavelet empírico (Gilles, 2013), Littlewood-Paley y Meyer para cada segmento definido (Daubechies, 1992). La implementación del método EWT se encuentra disponible en el lenguaje MATLAB, la cual fue desarrollada por Gilles (2023) e implementada en el lenguaje Python por Carvalho, Moraes, Braga, y Mendes (2020). La figura 3.7 muestra las 3 IMF obtenidas con el método

EWT de las funciones descritas en las ecuaciones 3.16 y 3.17.

$$f_1(x) = 0.05x + 0.2 \cos(2\pi x) + 0.1e^{-x^2+3x} \quad (3.16)$$

$$f_2(x) = 0.3 \sin(2.6\pi x) - 0.2 \cos(1.2\pi x) + 0.1e^{-0.1x^2} \quad (3.17)$$

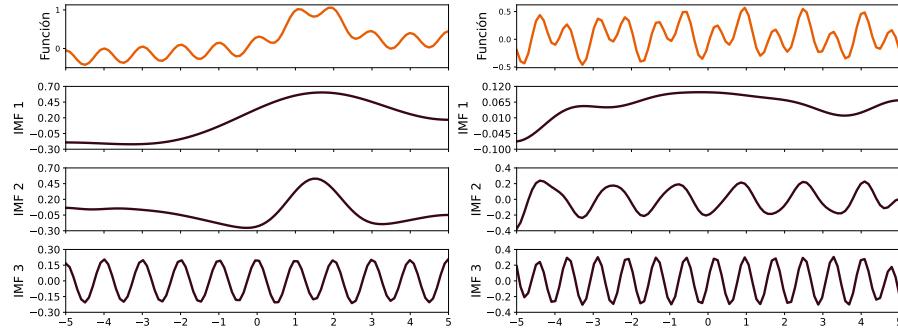


Figura 3.7: IMF obtenidas por el método EWT de las funciones descritas en las ecuaciones 3.16 (izquierda) y 3.17 (derecha).

3.2.2. Variational Mode Decomposition

El método de Variational Mode Decomposition (VMD) tiene aplicaciones en una variedad de campos, los cuales abarcan la ciencia de los materiales, la bioinformática y la geofísica, entre otros. Se ha utilizado para analizar una amplia gama de señales, como señales de audio, señales de radar y señales de electroencefalograma (EEG). El VMD propuesto por [Dragomiretskiy y Zosso \(2014\)](#) es una técnica de análisis de señales no lineales que permite descomponerlas si existen al menos K señales modales con una única frecuencia central (ω_k), en un rango de frecuencias limitado. Esto quiere decir que la señal f puede ser representada como una suma ponderada de K señales modales tal que:

$$f(t) = \sum_{k=1}^K u_k(t) \text{ donde } u_k = A_k(t) \cos(\phi_k(t)) \quad (3.18)$$

donde $u_k(t)$, $A_k(t)$ y $\phi_k(t)$ son las funciones de la señal modal, amplitud y fase en el tiempo t , respectivamente tal que $\phi'_k(t)$ es un función no decreciente y $A_k(t)$ es una función no negativa. El objetivo del método VMD es descomponer una señal en K modos donde cada uno tiene una banda ancha alrededor de una frecuencia ω_k a estimar. [Dragomiretskiy y Zosso \(2014\)](#) propusieron el siguiente esquema para obtener la descomposición:

1. Para cada modo u_k se calcula su señal analítica por medio de la transformada de Hilbert con el fin de obtener una señal con frecuencia no negativas.

2. Cada nodo se trasladara hacia la frecuencia ω_k , el cual será su frecuencia central.
3. El ancho de la banda ancha de la señal se calcula a través del gradiente de la raíz de la norma L^2 .

A este esquema se le conoce como “el problema del rango de onda restringido”, el cual se representa como:

$$\begin{aligned} \min_{\{u_k\}, \{\omega_k\}} & \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_k(t) \right] e^{-i\omega_k t} \right\|_2 \right\} \\ \text{s.t. } & f(t) = \sum_k u_k(t) \end{aligned} \quad (3.19)$$

donde ∂_t es la derivada parcial sobre el tiempo, $*$ es la operación de convolución y $\{u_k\}, \{\omega_k\}$ son los conjuntos de los modos y frecuencias respectivamente. Para resolver el problema descrito en la ecuación 3.19, se utiliza una penalización cuadrática y multiplicadores de Lagrange (λ) para incluir el problema sin restricciones. La combinación de estos dos métodos beneficia en las propiedades de convergencia. Es por ello, que la ecuación 3.19 puede ser escrita como:

$$\begin{aligned} L(u_k, \omega_k, \lambda) = & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{i}{\pi t} \right) * u_k(t) \right] e^{-i\omega_k t} \right\|_2 \\ & + \left\| f(t) - \sum_k u_k(t) \right\|_2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle \end{aligned} \quad (3.20)$$

donde α es un multiplicador de Lagrange que regula el desempeño de la reconstrucción de la señal con base en las señales modales encontradas. Para optimizar la función L , los parámetros $\{u_k\}$, $\{\omega_k\}$ y λ son optimizados siguiendo el método de multiplicadores en direcciones alternadas (ADMM por sus siglas en inglés). La implementación del método VMD se encuentra disponible en el lenguaje MATLAB, la cual fue desarrollada por [Zosso \(2023\)](#) y fue implementada en el lenguaje Python por [Carvalho y cols. \(2020\)](#). La figura 3.8 muestra las 3 IMF obtenidas con el método EWT de las funciones descritas en las ecuaciones 3.16 y 3.17.

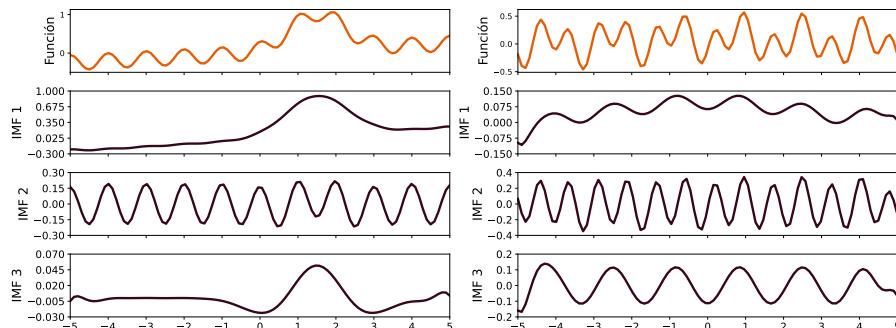


Figura 3.8: IMF obtenidas por el método VMD de las funciones descritas en las ecuaciones 3.16 (izquierda) y 3.17 (derecha).

3.3. Métodos de interpolación

Los métodos de interpolación son utilizados para estimar valores intermedios en una serie de datos. Los métodos de interpolación permiten encontrar una función f mediante un conjunto de coordenadas (x_i, y_i) que cumplan una condición tal que $f(x_i) = y_i$ para cada par ordenado en el conjunto. Existen diversos métodos de interpolación, los más conocidos son aquellos que utilizan funciones lineales o polinomiales. En los últimos años, los métodos que aplican funciones definidas por tramos (*splines*) han cobrado una gran popularidad. La interpolación polinomial tiene como inconveniente que su resultado puede contener una derivada no diferenciable. Para la mayoría de las aplicaciones, la función de la interpolación debe ser diferenciable sobre todo su dominio. El modelo más sencillo de la interpolación lineal por tramos consiste en líneas rectas que unen de manera consecutiva cada punto con la característica de tener una derivada discontinua. En el presente estudio, se emplearon funciones polinomiales de orden p denominadas *spline*. Estas funciones se definen en una serie de $m + 1$ puntos con un conjunto de n funciones S_k definidas en el intervalo $[x_k, x_{k+1}]$ que satisfacen:

$$\begin{aligned} S_k(x_k) &= S_{k+1}(x_k) \\ S'_k(x_k) &= S'_{k+1}(x_k) \\ S''_k(x_k) &= S''_{k+1}(x_k) \\ &\vdots \\ S_k^{(p-1)}(x_k) &= S_{k+1}^{(p-1)}(x_k) \end{aligned} \tag{3.21}$$

para las condiciones de frontera:

$$S_k(x_i) = y_i \tag{3.22}$$

$$S_i(x_{i+1}) = y_{i+1} \tag{3.23}$$

tal que S y sus derivadas sean continuas. Los splines de grado 3 (cúbicos) requieren poco poder de cómputo y sus soluciones tienen una segunda derivada continua. Los splines cúbicos se definen como:

$$S_i(x) = d_i + c_i(x - x_i) + b_i(x - x_i)^2 + a_i(x - x_i)^3 \tag{3.24}$$

donde a_i, b_i, c_i, d_i son coeficientes a estimar, por lo que esta ecuación representa un sistema de ecuaciones con $4n$ incógnitas y $4n$ condiciones. Al tener el mismo número de incógnitas y condiciones, se afirma que los coeficientes existen y son únicos. Aplicando las condiciones de continuidad a la ecuación 3.24 y la primer y segunda derivada se obtiene que la solución

de los coeficientes es:

$$d_i = \frac{1}{3h_i}(c_{i+1} - c_i) \quad (3.25)$$

$$b_i = \frac{1}{h_1}(y_{i+1} - y_i) - \frac{h_i}{3}(2c_i + c_{i+1}) \quad (3.26)$$

$$h_{i-1}c_{i-1} + 2(h_{i-1} + h_i)c_i + h_ic_{i+1} = \frac{3}{h_i}(y_{i+1} - y_o) - \frac{3}{h_{i-1}}(y_i - y_{i-1}) \quad (3.27)$$

donde $h_i = x_{i+1} - x_i$. La ecuación 3.27 es un sistema de ecuaciones tridiagonal donde el coeficiente a determinar es c_i (Ahlberg, Nilson, y Walsh, 2016).

3.4. Modelos basados en redes neuronales

La inteligencia artificial es una rama de las ciencias computacionales encargada de estudiar modelos de cómputo capaces de realizar actividades como predicción y clasificación. Los métodos basados en la inteligencia artificial son un conjunto de algoritmos que por medio de datos obtienen patrones estadísticamente significativos. Una red neuronal artificial (ANN, por sus siglas en inglés) es un modelo computacional con un conjunto de unidades denominadas neuronas, donde cada una de ellas se conecta con otras a través de enlaces. La información de entrada atraviesa la red neuronal realizando una serie de operaciones con el valor contenido en cada nodo (pesos), produciendo un resultado. Puede existir una función activación, que modifica el resultado o impone un límite antes de propagarse a otra neurona. Esta función se conoce como función de activación. Para entrenar a la red neuronal en un problema específico, se usa la función de pérdida, que mide la eficiencia de la red en el trabajo que está ejecutando (Christoffersen y Jacobs, 2004). El entrenamiento supervisado es el proceso de optimizar parámetros de un modelo, utilizando una función de pérdida, conociendo el resultado previo a cada entrada que se evalúe (Cunningham, Cord, y Delany, 2008). En seguida, se expondrán los diferentes tipos de redes neuronales.

3.4.1. Perceptrón

El modelo perceptrón es un modelo donde cada nodo se encuentra conectado entre sí (Rosenblatt, 1958). Para realizar esta conexión entre nodos se incorpora una capa, llamada capa oculta, donde se ejecutan las operaciones entre nodos. Esta capa también se le conoce como capa lineal. Cada nodo en la red neuronal tiene una función de activación. El perceptrón puede ser expresado matemáticamente como:

$$y = f(W^T x) \quad (3.28)$$

donde x son los valores de entrada, W son los pesos de cada nodo, f es la función de activación y y es la salida del perceptrón. La figura 3.9 muestra la arquitectura del modelo perceptrón.

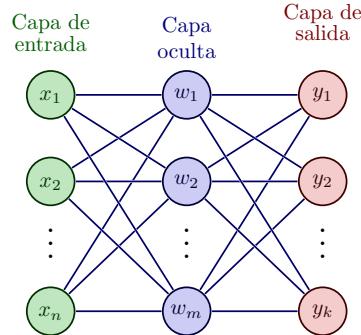


Figura 3.9: Arquitectura del modelo perceptrón.

3.4.2. Perceptrón Multicapa

El perceptrón multicapa usa como base al modelo de perceptrón. Para poder realizar operaciones no lineales, el perceptrón multicapa incorpora más capas ocultas. El perceptrón multicapa está compuesto por una capa de entrada, una capa de salida y n capas ocultas entremedias. El perceptrón multicapa se caracteriza por tener salidas distintas pero relacionadas entre sí, de tal manera que la salida de una neurona es la entrada de la siguiente. En el perceptrón multicapa se pueden diferenciar una 2 fases:

- Propagación en la que se calcula el resultado de salida de la red desde los valores de entrada hacia delante.
- Aprendizaje en la que los errores obtenidos a la salida del perceptrón se van propagando hacia atrás (*backpropagation*) con el objetivo de modificar los pesos de las conexiones para que el valor estimado de la red sea semejante al valor esperado. Esta modificación está basada en el gradiente de la función objetivo.

La figura 3.10 se presenta la arquitectura del perceptrón multicapa.

3.4.3. Redes convolucional

Las redes neuronales que realizan operaciones de convolución en su interior son llamadas redes neuronales convolucionales (CNN, por sus siglas en inglés). Las CNN tienen una arquitectura como las ANN y sus aplicaciones son el reconocimiento de imágenes y

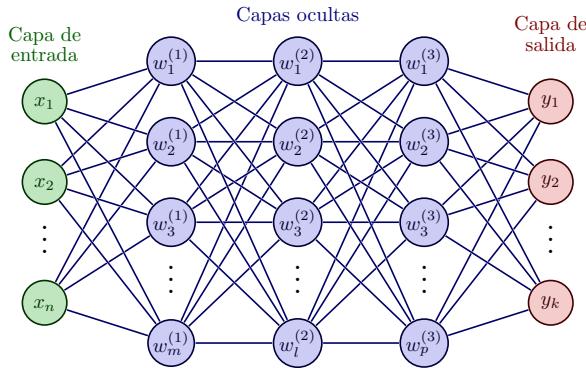


Figura 3.10: Arquitectura del modelo perceptrón multicapa con 3 capas ocultas de dimensiones m , l y p respectivamente.

videos, sistemas de recomendación (van den Oord, Dieleman, y Schrauwen, 2013), clasificación y, segmentación de imágenes, análisis de imágenes médicas, procesamiento de lenguaje natural (Collobert y Weston, 2008), interfaces entre cerebro-computadora (Avilov, Rimbert, Popov, y Bougrain, 2020) y series de tiempo (Tsanidakis y cols., 2017). La arquitectura de las CNN es conocida como Shift Invariant or Space Invariant Artificial Neural Networks (SIANN). Las CNN comparten los pesos de cada capa entre cada filtro de convolución. Los resultados de cada operación de convolución son llamados mapa de características. En las CNN, el método de *backpropagation* es aplicado a cada valor de los filtros. Las CNN son versiones regularizadas del perceptrón multicapa debido a que reconocen patrones más simples y pequeños en sus filtros. La arquitectura de las CNN se muestra en la figura 3.11.

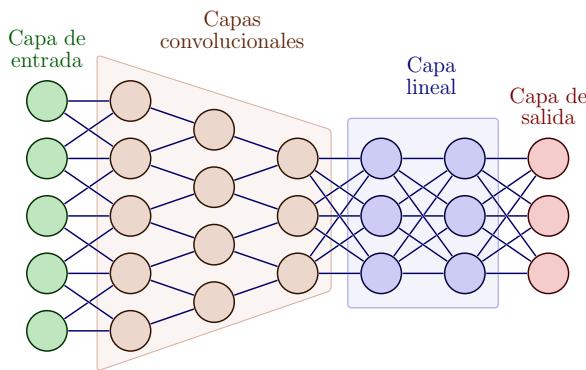


Figura 3.11: Arquitectura de las redes CNN.

3.4.4. Long-Short Term Memory

Las redes neuronales recurrentes (RNN) son una arquitectura de las ANN. Los nodos de las RNN tienen conexiones entre sí creando un ciclo, el cual ayuda a procesar datos secuenciales como texto, audio y video (Sherstinsky, 2020). Las RNN usan un espacio interno, que procesa las secuencias de los datos de entrada. Esta operación puede ejecutarse de forma infinita, en contraste con las CNN que su operación puede realizarse de forma única (Miljanovic, 2012). La figura 3.12 muestra el transporte y operación de la información que entra a las RNN, donde x_t representa el dato x en el tiempo t , h_t , el cual es el estado oculto para el tiempo t y A es la matriz de pesos de la red. Debido a la

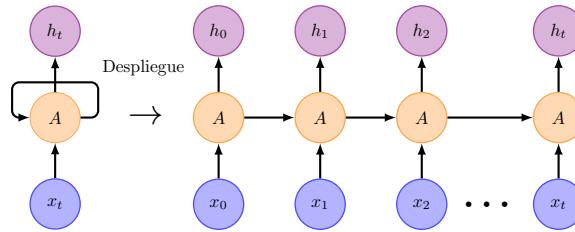


Figura 3.12: Representación del flujo de información en las RNN.

naturaleza de las conexiones recurrentes, las RNN tienen un problema conocido como el "problema del gradiente que se desvanece o explota. Este surge cuando se entrena una red neuronal con datos secuenciales muy largos, haciendo que los gradientes que se propagan hacia atrás sean muy pequeños o muy grandes, lo que dificulta en el proceso de optimización. Para disminuir este inconveniente se emplea la red Long Short-Term Memory (LSTM), la cual actualiza la información de acuerdo a su relevancia, en función del tiempo (Bengio, Simard, y Frasconi, 1994; Hochreiter y Schmidhuber, 1997). A estas operaciones se le denominan compuertas de las LSTM. Las diferentes compuertas se describen en lo sucesivo.

Compuerta de olvido Determina la proporción de información que se mantiene en el estado interno de la LSTM a lo largo del tiempo. Esta proporción es calculada como:

$$f_t = \sigma(U_f x_t + W_f h_{t-1}) \quad (3.29)$$

donde U_f y W_f son los pesos de la operación, x_t son los datos de entrada en el tiempo t y h_{t-1} es el estado oculto de la LSTM para el tiempo $t - 1$.

Compuerta de entrada Es la compuerta que realiza una unión entre la información del estado de la LSTM y los datos de entrada. El estado interno de la LSTM se computa

como:

$$\begin{aligned} i_{t1} &= \sigma(U_i x_t + W_i h_{t-1}) \\ i_{t2} &= \tanh(U_g x_t + W_g h_{t-1}) \\ i_t &= i_{t1} * i_{t2} \end{aligned} \quad (3.30)$$

donde U_i , U_g , W_i y W_g son los pesos para cada compuerta, x_t son los datos de entrada en el tiempo t , h_{t-1} es el estado oculto de la LSTM para el tiempo $t - 1$ y i_t es el estado interno de la LSTM para el tiempo t . Con base en el estado actual de la LSTM, se actualiza el estado de la celda, el cual es definido como:

$$c_t = \sigma(f^t c^{t-1} + i_t) \quad (3.31)$$

donde c_{t-1} es el estado anterior de la celda, f_t es la salida de la compueta de olvido y c_t es el estado actual de la celda.

Compuerta de salida Determina la proporción de la información del estado interno, mientras que el estado oculto es transferido a la siguiente iteración. La salida de la compueta está determinado por:

$$o_t = \sigma(U_o x_t + W_o h_{t-1}) \quad (3.32)$$

donde U_o y W_o son los pesos de la compuerta. El estado oculto es actualizado siguiendo ecuación:

$$h_t = \tanh(c_t) * o_t \quad (3.33)$$

Al conjunto de todos los componentes de la LSTM se denomina celda. La figura 3.13 presenta los componentes y el flujo de la información en una celda de LSTM.

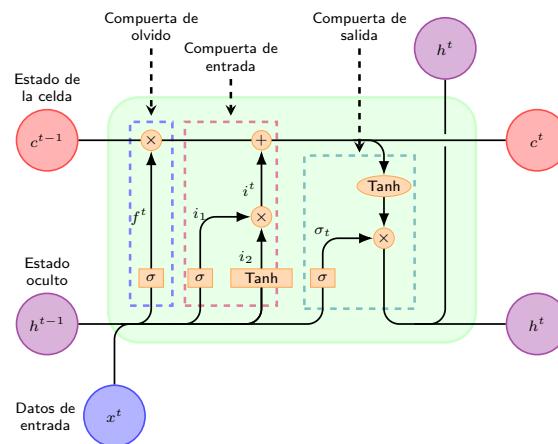


Figura 3.13: Representación visual de una celda de LSTM.

3.4.5. Aprendizaje secuencia a secuencia

Los modelos secuencia a secuencia (seq2seq) son una clase de ANN que son comúnmente utilizados para las tareas de traducción automática, reconocimiento de voz y chatbots (Sutskever, Vinyals, y Le, 2014). Estos modelos consisten en realizar un mapeo de los datos de entrada a un espacio de menor dimensión y finalmente ejecuta una operación para obtener el resultado de la red neuronal. Los modelos seq2seq se componen de tres partes: la capa de codificación, el espacio latente y la capa de decodificación. Las capas de codificación y de decodificación son dos redes neuronales del tipo LSTM, cada una con sus pesos independientes. La capa de codificación obtiene los datos de entrada y los transforma para llevarlos a un espacio de menor dimensión, denominado espacio latente. En seguida, la capa de decodificación lee los datos contenidos en el espacio latente y los transforma para obtener el resultado esperado. La figura 3.14 muestra la representación de la arquitectura de los modelos seq2seq.

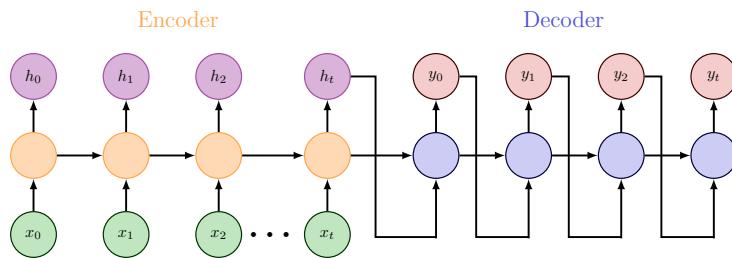


Figura 3.14: Arquitectura de los modelos basados en la estructura seq2seq.

3.4.6. Redes Neuronales Residuales

Un modelo de redes neuronales residuales (ResNet) es una arquitectura de redes neuronales donde los pesos de cada capa son entrenados para llevar a cabo una función residual. Una ResNet se caracteriza por tener conexiones que se bifurcan antes de transferir la información a un bloque de capas neuronales y que finalmente se suma a la información antes y después de entrar al bloque de capas neuronales. Esas conexiones se les denominan conexiones residuales. Inicialmente, esta arquitectura se empleó por He, Zhang, Ren, y Sun (2015) en la competencia de ImageNet(Deng y cols., 2009) del 2015. La figura 3.15 muestra la arquitectura básica de una ResNet.

3.4.7. Mecanismos de atención

Las ANN consisten en un conjunto de capas que realizan transformaciones no lineales. Las interacciones entre capas dificultan el control en la distribución de la información

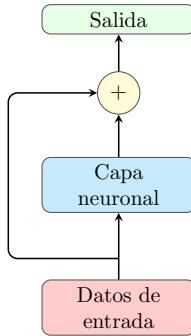


Figura 3.15: Arquitectura básica de los modelos ResNet.

cuando el conjunto de datos tiene una gran dimensión. Las redes generan espacios de gran concentración de información conforme su flujo se incrementa. Los mecanismos de atención brindan una alternativa (Ba, Kiros, y Hinton, 2016). Una ANN con un mecanismo de atención, escala de manera óptima, la representación en los estados ocultos de la red. Los mecanismos de atención infieren para obtener una selección sobre las representaciones en los estados ocultos. Esta selección se encuentra condicionada por el resultado final de la red. Los mecanismos de atención permiten a la red neuronal mantener una relación entre el tamaño de los datos y la memoria utilizada. Además muestra un rendimiento óptimo para diversas tareas como predicción y clasificación. Existen diferentes tipos de mecanismos de atención, cada uno con diversas funciones y maneras de estimar una selección para los datos.

Scaled Dot-Product Attention

El mecanismo de atención *Scaled Dot-Product Attention* consiste en tres matrices: dos matrices de dimensión d_k (denominadas *query* y *key*) y una de dimensión d_v (denominada *value*) (Vaswani y cols., 2017a) tal que:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (3.34)$$

donde Q , K y V son las matrices *query*, *key* y *value*, y x son los datos de entrada del mecanismo. Este mecanismo de atención usualmente está presente en arquitecturas de modelos tipo seq2seq. La figura 3.16 muestra una representación en capas de este mecanismo de atención.

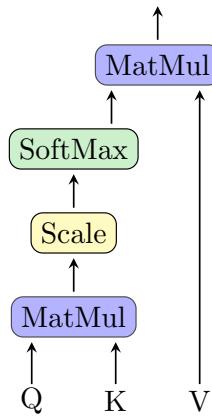


Figura 3.16: Representación del *Scaled Dot-Product Attention*.

Multi-Head Attention

El *Multi-Head Attention*, se basa en el *Scaled Dot-Product Attention* y realiza múltiples operaciones en proyecciones aleatorias. Este mecanismo de atención ha obtenido un buen desempeño reconociendo patrones dentro de datos temporales e imágenes (Vaswani y cols., 2017a). En cada proyección realizada se aplica el *Scaled Dot-Product Attention*, en seguida los resultados concatenados. El *Multi-Head Attention* permite a la red neuronal analizar la información obtenida basada en diferentes proyecciones de los datos. El mecanismo de *Multi-Head Attention* se describe como:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_n)W^o \quad (3.35)$$

donde $\text{head}_i = \text{Attention}(W_i^Q, W_i^K, W_i^V)$

donde las proyecciones de cada *Scaled Dot-Product Attention* son funciones parametrizadas por las matrices W_i^Q, W_i^K, W_i^V . La figura 3.17 muestra el mecanismo Multi-Head Attention.

3.4.8. Transformer

Los Transformers son una arquitectura de red neuronal basada en los modelos seq2seq. Este tipo de modelos fueron propuestos por Vaswani y cols. (2017b) en el área de lenguaje natural. Los modelos basados en la arquitectura de los Transformers han demostrado tener una gran eficiencia y precisión en diversos campos como análisis de imágenes, texto, audio y series de tiempo. La idea detrás de los Transformers es eliminar el uso de las RNN y en su lugar, utilizar una arquitectura completamente basada en mecanismos de atención. Una de las ventajas clave de los Transformers es que pueden procesar secuencias más largas que

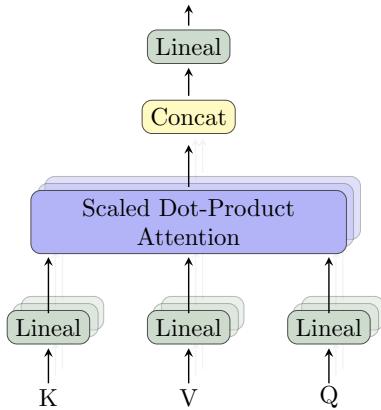


Figura 3.17: Representación del mecanismo *Multi-Head Attention*.

las RNN (Karita y cols., 2019). Además, los Transformers son altamente paralelizables y se pueden entrenar de manera más eficiente que las RNN. La figura 3.18 muestra la arquitectura típica de los Transformers. La arquitectura de los Transformers consta de dos partes principales: el codificador y el decodificador.

Capa Feed Forward

La capa Feed Forward consiste en un conjunto de dos capas lineales, entre estas capas se añade una capa que aplica la función rectificadora (ReLU). El objetivo de la capa Feed Forward es ayudar a agregar información contextual y no lineal a los resultados de las capas con mecanismos de atención.

Enmascaramiento del Multi-Head Attention

El enmascaramiento del *Multi-Head Attention* es una técnica aplicada en los modelos Transformer que se utiliza para controlar la información disponible durante el cálculo del mecanismo de atención. Esta técnica permite que el transformer se enfoque solo en la información relevante y evita que se acceda a ciertos elementos en la secuencia durante el proceso de atención.

Positional encoding

El Transformer por se no contiene capas recurrentes o convolucionales en su interior. Para incluir una idea de la temporalidad de los datos, el Transformer incluye una capa llamada positional encoding. El objetivo de esta capa es incluir información acerca de la posición de cada secuencia dada. Existen diversas maneras para generar esta información, ya sea que la misma red aprenda la posición o esté sea fija (Gehring, Auli, Grangier, Yarats, y

Dauphin, 2017). El Transformer utiliza un positional encoding fijo descrito por:

$$\begin{aligned} PE_{pos,2i} &= \sin(pos/1000^{2i/d_{model}}) \\ PE_{pos,2i+1} &= \cos(pos/1000^{2i/d_{model}}) \end{aligned} \quad (3.36)$$

donde pos es la posición de la secuencia, e i es la dimensión de los datos. Esta información es añadida antes de transmitir la información entrante al codificador y decodificador.

Codificador

El codificador está compuesto de N_C bloques idénticos, cada bloque contiene una capa *Multi-Head Attention* y una capa fully-connected. La capa de *Multi-Head Attention* es aplicada a la información que recibe el codificador, luego se aplica una transformación lineal por medio de la capa fully-connected para mapear la información con la misma dimensión de los datos de entrada. Para algunos procesamientos se añade una capa de normalización entre capas del bloque del codificador (He y cols., 2015). Esto con el fin de tener una arquitectura tipo residual.

Decodificador

El decodificador está compuesto de N_D bloques idénticos, cada uno contiene dos capas de *Multi-Head Attention* y una capa de feed forward. La información que recibe el decodificador es transferida a la capa *Multi-Head Attention*. En este proceso, los pesos correspondientes a las entradas futuras se enmascaran con ceros para garantizar que el modelo no tenga acceso a la información futura. Posteriormente, se combina la información del codificador y del decodificador en la segunda capa de *Multi-Head Attention*. El resultado de este proceso es transferido a la capa feed-forward para mapearlo en un espacio con dimensión igual al de la dimensión recibida por el decodificador. Para controlar la distribución de la información en los espacios entre cada capa se añade un proceso de normalización.

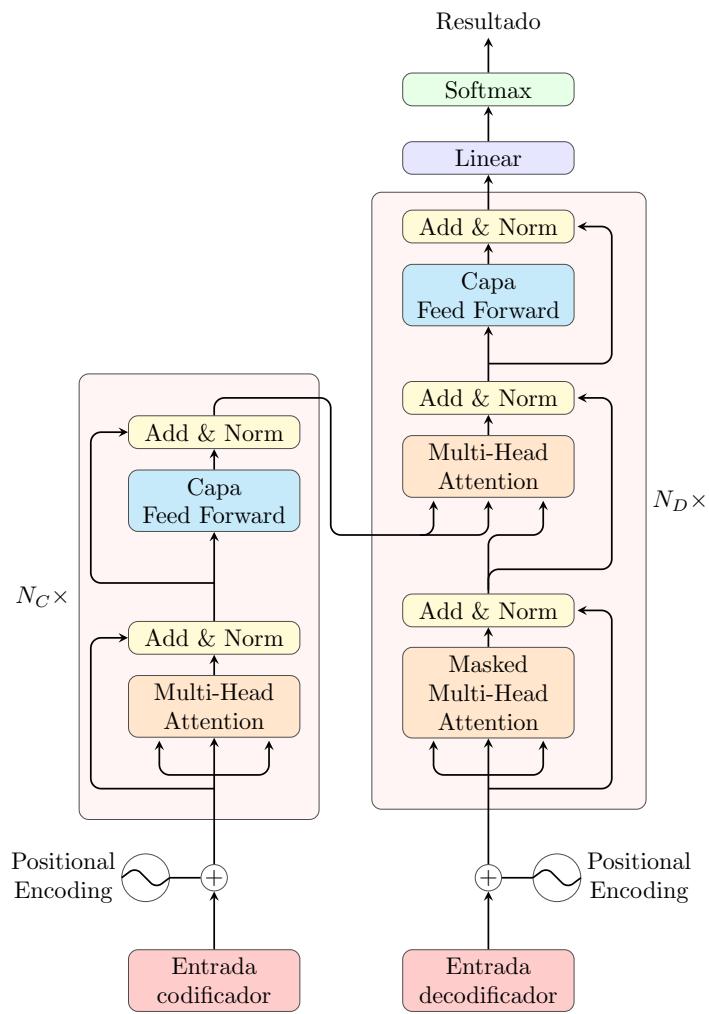


Figura 3.18: Arquitectura de los modelos Transformer.

Metodología

En este capítulo aborda la descripción, filtración y reconstrucción de los datos registrados por el Sistema Integral de Monitoreo Ambiental. Además se detalla el proceso para interpolar los datos horarios de la columna total de ozono y la fracción de nubes estimados de mediciones hechas por el instrumento satelital OMI. A su vez, se describe el tratamiento de los datos para reconstruir los datos faltantes con una resolución horaria. Adicionalmente, se discutirá la técnica de aumentación de datos que fue aplicada a la *CHI* medida, para incrementar la variedad de los datos y el número de muestras. Posteriormente se introducirán las consideraciones para el entrenamiento de las redes neuronales. Estas consideraciones abarcan la segmentación de los conjuntos de entrenamiento, validación y prueba, así como la obtención de los datos de entrada y de salida esperados de las redes neuronales a utilizar, la función a optimizar y las técnicas para evitar el sobreajuste. Finalmente, se expondrá la propuesta del modelo híbrido y las diferentes arquitecturas de redes neuronales para establecer el pronóstico de IUV diario, dadas características recabadas el día anterior.

4.1. Datos

En esta sección, se brindará una descripción detallada sobre el origen y el procesamiento de los datos empleados del presente estudio. Asimismo se enumeran los pasos de preparación y el acondicionamiento que se llevaron a cabo para garantizar la precisión y aplicabilidad en esta investigación.

4.1.1. Mediciones *in situ*

El Área Metropolitana de Monterrey (AMM) es la segunda zona urbana más grande en México, la cual tiene alrededor de 5 millones de habitantes ([CONAPO-INEGI-SEDATU, 2018; INEGI, 2021](#)). En el AMM existen diversas empresas dedicadas a la extracción de materiales para la construcción, actividades industriales y tiene una gran cantidad de vehículos. Desde 1993, el Sistema Integral de Monitoreo Ambiental (SIMA) ha sido el encargado de medir la calidad del aire en el AMM. Actualmente, el SIMA tiene en funcionamiento 15 estaciones de monitoreo distribuidas sobre el AMM. La localización de cada estación se describe en la tabla 4.1. El SIMA mide variables meteorológicas y las concentraciones de contaminantes atmosféricos. Los instrumentos de medición son

Estación	Abreviación	Ciudad	Elevación (m.s.n.m.)	Latitud (°N)	Longitud (°O)
Centro	C	Monterrey	560	25.67	100.34
Sur	S	Monterrey	630	25.57	100.25
Este	E	Apodaca	333	25.79	100.07
Norte	N	Escobedo	528	25.80	100.34
Noreste 2	NE2	Apodaca	432	25.78	100.19
Noroeste	NO	Monterrey	571	25.76	100.37
Noroeste 2	NO2	Garcia	716	25.78	100.59
Noroeste 3	NO3	Monterrey	592	25.78	100.46
Sureste	SE	Guadalupe	492	25.67	100.25
Sureste 2	SE2	Juarez	387	25.65	100.10
Suroeste	SO	Santa Catarina	694	25.68	100.46
Noreste	NE	San Nicolas de los Garza	476	25.75	100.26
Norte 2	N2	San Nicolas de los Garza	520	25.73	100.31
Sureste 3	SE3	Cadereyta de Jimenez	340	25.58	99.99
Suroeste 2	SO2	San Pedro Garza Garcia	636	25.66	100.41

Tabla 4.1: Localización de las estaciones de monitoreo del SIMA en el AMM. La elevación tiene medidas de metros sobre el nivel del mar (m.s.n.m), la latitud y la longitud se encuentran descritas en grados con dirección hacia el norte (°N) y oeste (°O) respectivamente.

verificados regularmente a fin de cumplir la norma NOM-035-SEMARNAT-1993 y NOM-156-SEMARNAT-212 ([Sistema Integral de Monitoreo Ambiental, 2015](#)). Los valores son registrados cada cinco minutos y almacenados en una base de datos, los cuales son promediados cada hora. Para seleccionar las estaciones de monitoreo con una mayor cantidad de mediciones se consideraron las siguientes condiciones:

- Al menos 20 promedios horarios al día, para considerarse una medición diaria representativa.
- Al menos 24 días de medición representativos al mes, para considerarse un mes representativo.

La figura 4.1 se muestra el número de meses representativos por cada año entre el 2015 y 2022, para las estaciones de monitoreo del SIMA. Las estaciones de monitoreo que cumplieron los criterios para el número de meses representativos fueron: Noreste, Noroeste, Sureste y Suroeste. La figura 4.2 se muestra la geolocalización de las estaciones de monitoreo del SIMA. Las mediciones empleadas en la presente Tesis, corresponden a las estaciones indicadas en color rojo y en azul las estaciones no seleccionadas.

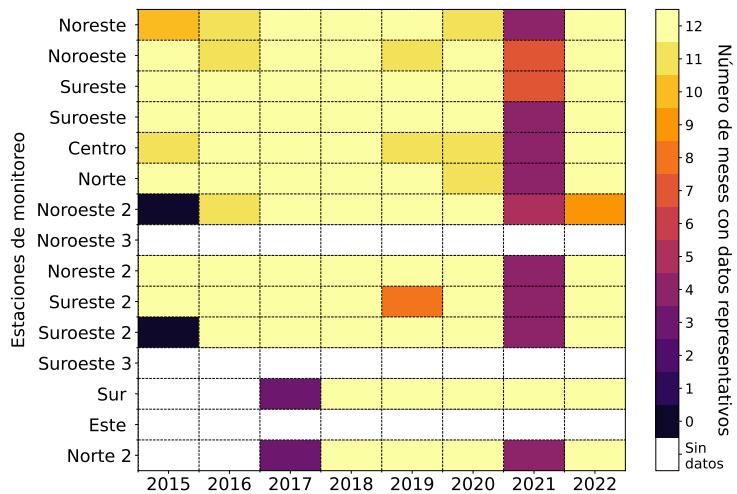


Figura 4.1: Número de meses representativos para cada estación de monitoreo del SIMA dentro del periodo 2015-2022.

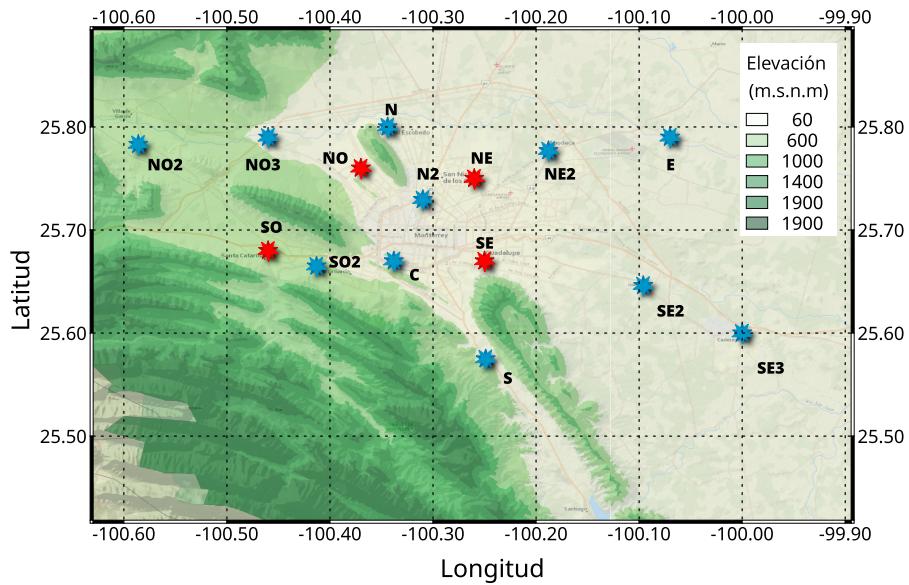


Figura 4.2: Geolocalización de las estaciones de monitoreo del SIMA. En rojo las estaciones seleccionadas y en azul las estaciones no seleccionadas.

En las estaciones de monitoreo seleccionadas, las mediciones de *GHI* son realizadas por piranómetros modelo 095 de la compañía Met One. El sensor del instrumento tiene una respuesta a la radiación entre 285 y 280 nm. Adicionalmente se realizaron una serie de mediciones *in situ* cerca del Palacio de Justicia de Nuevo León (25.75°N , 100.34°O , 541 m.s.n.m), el cual se encuentra a 5 km de la estación Noroeste, con un radiómetro UVA+UVB portátil Solarmeter ®modelo 5.0 propiedad del Instituto de Física Rosario

(CONICET-UNR, Argentina). El sensor tiene una precisión de $\pm 5\%$, esta precisión fue certificada por el instituto nacional de estándares y tecnologías (NIST por sus siglas en inglés) bajo el sistema de referencia estándar de materiales. Este sensor mide en un rango entre 280 y 400 nm (Solarmeter, 2023). Las mediciones realizadas con este sensor fueron obtenidas durante algunos días de cielo despejado de los meses de junio, julio y agosto de 2022, con el propósito de verificar los niveles del Índice UV derivados de los modelos y las mediciones de la GHI del SIMA. La figura 4.3 muestra un piranómetro modelo 095 de la compañía Met One ubicado en la estación Centro y el radiómetro utilizado para obtener las mediciones UVA+UVB. La base de datos del SIMA tiene datos atípicos y



Figura 4.3: Piranómetro de la estación Centro del SIMA (izquierda) y radiómetro UVA+UVB (derecha).

datos faltantes. Para obtener un entrenamiento eficiente, se realizó un preprocesamiento que consistió en un filtrado de datos atípicos y una reconstrucción basada en una métrica de similitud. El proceso de filtración removió aquellos datos que presentan un índice de claridad (K_T) mayor a 0.9. Debido a que los valores entre 0.9 y 1 son poco probables y los valores mayores o iguales a 1 son inviables desde una perspectiva física. Con esto, la base de datos presenta únicamente datos faltantes. La figura 4.4 muestra a la izquierda, la medición adquirida (curva azul) por las estaciones de monitoreo y la curva roja es la GHI_0 obtenida a partir de la ecuación 3.9, mientras que a la derecha se observa la medición diaria (curva verde), ya sin los datos atípicos. El proceso de reconstrucción de datos consiste en reemplazar los valores faltantes con el promedio horario correspondiente. El promedio horario es calculado con los 15 días con mayor similitud considerando los datos diarios medidos un mes antes y después centrados en la fecha que se requiere reconstruir. La

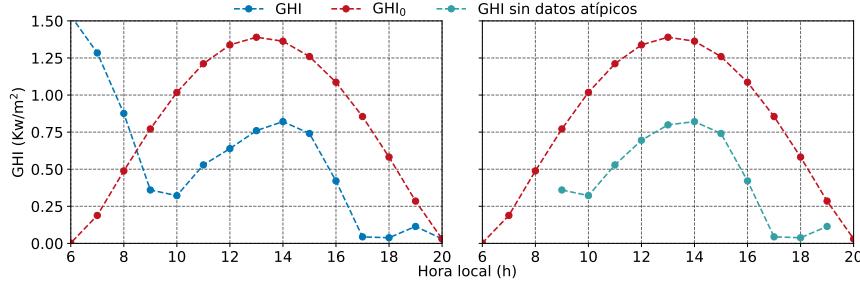


Figura 4.4: Proceso de filtrado del GHI para el 2 de junio de 2021 en la estación Noroeste.

similitud entre mediciones diarias se obtuvo por medio de la expresión:

$$\text{sim}(a, b) = \frac{a^T b}{\|a\| \|b\|} \quad (4.1)$$

donde a representa la irradiancia solar a lo largo del día con datos horarios faltantes y b es la medición diaria completa, dentro del periodo de comparación. Durante este cálculo, los datos horarios faltantes son enmascarados con ceros. La figura 4.5 muestra la reconstrucción de la irradiancia solar global el día 2 de junio de 2021 en la estación Noroeste. La curva verde muestra la irradiancia diaria con datos faltantes, la curva morada es la reconstrucción y la curva roja es la GHI_0 .

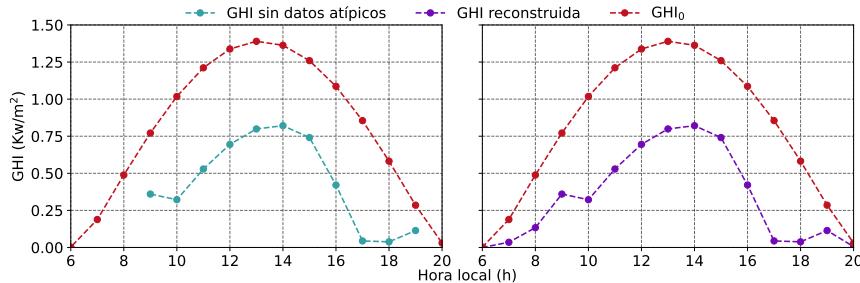


Figura 4.5: Proceso de reconstrucción para el 2 de junio de 2021 en la estación Noroeste.

4.1.2. Datos satelitales

El satélite Aura fue lanzado por la NASA en 2004 como parte del programa Earth Observing System (EOS). A bordo del satélite se encuentra el instrumento Ozone Monitoring Instrument (OMI), el cual estima la Columna Total de Ozono (TOC) en la atmósfera por medio del algoritmo TOMS version-8 ([NASA EOS/Aura Validation Data Center \(AVDC\) - Correlative data, Field of View Predictions, Data Subsets, GEOMS, DCIO, 2023](#)). Este algoritmo deriva la TOC a través de la radiación solar absorbida entre 317.5 y 331.2 nm.

El OMI realiza correcciones debido a la presencia de las nubes y aerosoles. El Level-2 Total Column Ozone Data Product (OMTO3) proporciona estimaciones de la TOC, índice de SO₂, IUV, fracción de nubes y presión a nivel de la superficie. Para el AMM, el satélite sobrevuela diariamente entre las 13 : 00 y las 15 : 00 hora local. En este estudio, se utilizó un spline cúbico para interpolar las mediciones satelitales de la TOC y la fracción de nubes a fin de obtener estos datos con resolución horaria.

4.1.3. Estimación del IUV

Actualmente el SIMA únicamente registra la radiación GHI. Para derivar el componente UV de la radiación solar se empleó el modelo SMARTS y las mediciones de GHI. Con el fin de caracterizar al modelo SMARTS para el AMM, se emplearon las condiciones atmosféricas y de superficie basadas en la [Atmosphere US Standard \(1976\)](#) y el modelo de perfil en altura de los aerosoles de [Shettle y Fenn \(1979\)](#) para zonas urbanas. Adicionalmente se introducen valores adicionales de concentración de los gases acordes al nivel de la tropósfera y un albedo correspondientes al caracterizado para el concreto. El valor de concentración adicional para cada gas se muestra en la tabla 4.2.

Gas	Concentración adicional (ppmv)
CH ₂	0.007
CH ₄	0.3
CO	0.35
HNO ₂	0.002
HNO ₃	0.005
NO	0.2
NO ₂	0.02
NO ₃	5×10^{-5}
O ₃	0.053
SO ₂	0.05

Tabla 4.2: Concentraciones adicionales en partes por millón por volumen (ppmv) en cada gas en la atmósfera terrestre con el fin de adaptar las estimaciones del modelo SMARTS al AMM.

La estimación de la radiación UV consta de dos fases, en la primera se varía el parámetro de la profundidad óptica de aerosol a 550 nm (AOD_{550}) hasta ajustar las estimaciones del GHI obtenidas por el modelo SMARTS con las mediciones *in situ* obtenidas en las estaciones de monitoreo. La caracterización del AOD_{550} es realizada debido a que este componente atmosférico uno de los contaminantes que más afectan a la reducción de la radiación solar UV a nivel del suelo ([Krzyścin y Puchalski, 1998](#)) y existen pocos sitios en

el mundo donde se registra superficialmente (Holben y cols., 1998). En la segunda fase, se realiza una simulación de la radiación UV utilizando el parámetro del AOD_{550} obtenido en la primera fase. En seguida, se detallan los procesos empleados en cada fase para la estimación de la radiación UV. Para la primera fase del método de la caracterización del AOD_{550} se aplica un algoritmo de búsqueda binaria. Este algoritmo varía el AOD_{550} hasta obtener una diferencia relativa (RD) entre 5 % y 6 % entre la GHI estimada por el modelo SMARTS y la GHI medida en las estaciones de monitoreo. Este valor de la RD es una diferencia mínima y razonable de acuerdo a la estimación del modelo. La diferencia relativa es calculada como:

$$RD = 100 \left(\frac{GHI - \widehat{GHI}}{GHI} \right) \quad (4.2)$$

donde \widehat{GHI} es estimación del GHI realizada por el modelo SMARTS. Esta metodología se describe en el algoritmo 1. Con base al AOD_{550} encontrado por el algoritmo 1

Algoritmo 1 Caracterización del AOD_{550nm}

```

 $AOD_i \leftarrow 0$ 
 $AOD_f \leftarrow 1$ 
 $RD \leftarrow \infty$ 
while |  $RD - 5 | > 0.5$  do
     $AOD_{550} \leftarrow \frac{AOD_i + AOD_f}{2}$ 
     $\widehat{GHI} \leftarrow \text{SMARTS}(AOD_{550}; \theta)$ 
     $RD \leftarrow 100 \left( \frac{GHI - \widehat{GHI}}{GHI} \right)$ 
    if  $RD > 6$  then
         $AOD_f \leftarrow AOD_{550}$ 
    else
         $AOD_i \leftarrow AOD_{550}$ 
    end if
end while

```

y las condiciones atmosféricas y de superficie caracterizadas para el AMM, se estima el espectro solar UV a nivel del suelo, utilizando el modelo SMARTS. Mediante la ecuación 3.11 se calcula el IUV correspondiente a las condiciones atmosféricas. Para verificar que las estimaciones obtenidas por el modelo SMARTS tienen un acuerdo razonable con las mediciones *in situ*, se compararon las mediciones UVA+UVB realizadas cerca del Palacio de Justicia de Nuevo León y la radiación UVA+UVB para la estación Noroeste obtenida con el modelo SMARTS. Se consideran estos dos sitios debido a que son cercanos entre sí. La figura 4.6 se muestran las mediciones *in situ* UVA+UVB (color verde) y las estimaciones del modelo SMARTS (color azul) de la radiación solar UVA+UVB para cuatro días

diferentes. En esta figura se muestra que los resultados del modelo SMARTS por medio del algoritmo 3.11 tienen una buena aproximación a las mediciones *in situ*. Por ello, el modelo SMARTS puede aplicarse con un buen acuerdo, utilizando como datos de entrada las condiciones atmosféricas obtenidas a partir de mediciones de *GHI* e información geográfica del lugar.

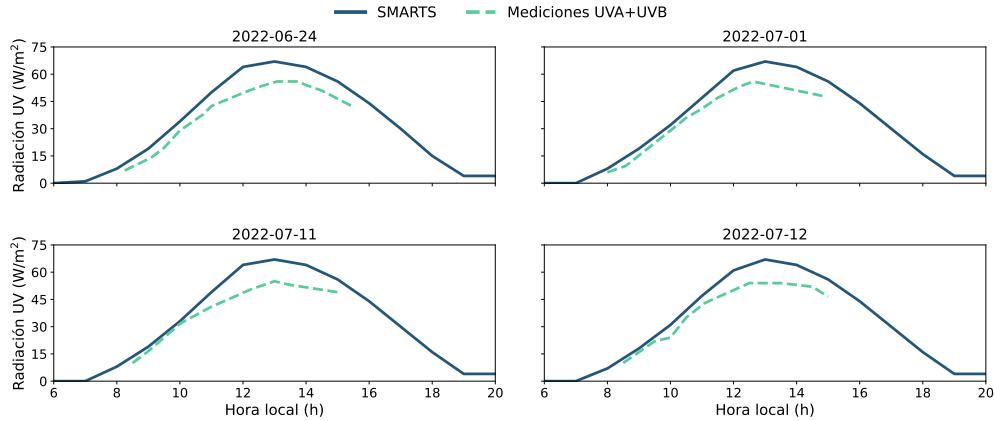


Figura 4.6: Estimación de la radiación UVA+UVB obtenida por el modelo SMARTS con el AOD_{550} obtenido a partir de mediciones de *GHI* empleando el algoritmo 1 (azul) y mediciones de radiación UVA+UVB *in situ* realizadas con un instrumento externo al SIMA en el Palacio de Justicia de Nuevo León.

4.1.4. Aumentación de datos

Los métodos basados en redes neuronales han demostrado obtener buenos rendimientos en diferentes áreas como visión por computadora, procesamiento de lenguaje natural y reconocimiento del habla. Estos métodos han sido aplicados a tareas relacionadas con series de tiempo como clasificación (Han, Zhao, Leung, Ma, y Wang, 2021) y predicción (Gamboa, 2017). El éxito de las redes neuronales se encuentra en la gran cantidad de datos que procesan optimizando los parámetros de la red. En el caso de las series de tiempo, en algunas ocasiones no se cuenta con una gran variedad y cantidad de datos, debido a las características del problema a tratar. Por lo tanto, la red neuronal no consigue obtener un buen rendimiento respecto a los métodos estadísticos y probabilísticos. La aumentación de datos es una herramienta que amplía el conjunto de datos de entrenamiento utilizando los patrones que se encuentren en ellos. El objetivo de la aumentación de datos, es que la red generalice sus resultados durante el proceso de entrenamiento. Esta metodología produce datos sintéticos que abarquen valores intermedios utilizando los datos reales, manteniendo sus propiedades y comportamiento original. En este trabajo, se aplicó una aumentación basada en la suma de variables aleatorias con una distribución gaussiana

sobre las mediciones de *GHI*. La variable aleatoria fue obtenida a partir de una distribución gaussiana, con media cero y una varianza en función de la hora del día y el mes del año. La varianza tiene esta dependencia temporal debido a la estacionalidad de la *GHI*. La figura 4.7 muestra la estacionalidad de la varianza, en función de la hora y el mes del año, implícitamente relacionadas con las mediciones del *GHI*. La aumentación de datos

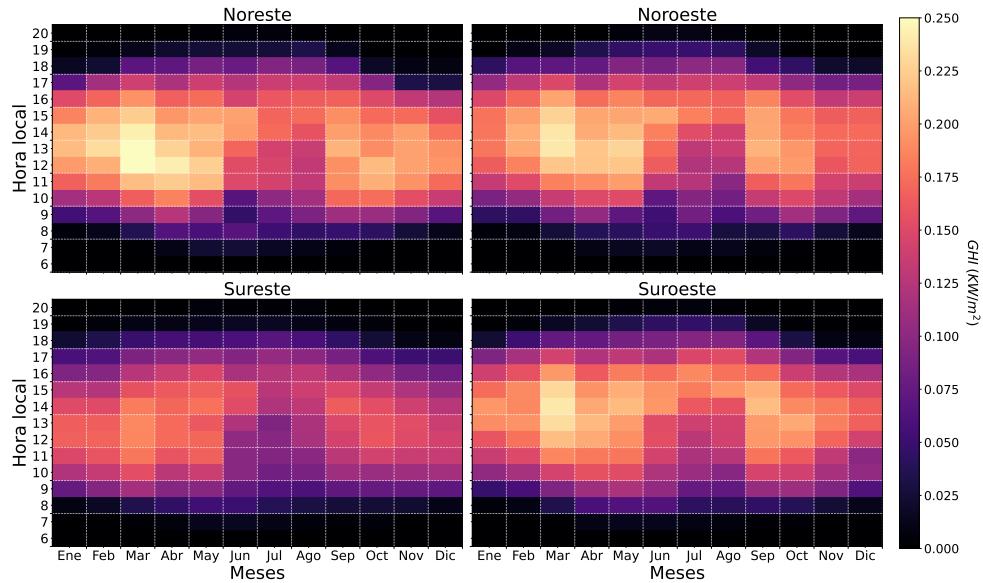


Figura 4.7: Desviación estándar del GHI horario a lo largo de los meses en el periodo 2015-2019 para la estaciones de monitoreo del SIMA.

es descrita como:

$$f'_{y,m,d,h} = f_{y,m,d,h} + 0.2\Phi(0, \sigma_{m,h}) \quad (4.3)$$

donde $f_{y,m,d,h}$ denota la medición de GHI para un año y , mes m , dia d y hora h y $f'_{y,m,d,h}$ es el dato sintético correspondiente para la misma fecha que $f_{y,m,d,h}$.

4.2. Entrenamiento de la red neuronal

La optimización de los parámetros de las redes neuronales es un proceso fundamental en el campo de la inteligencia artificial y el aprendizaje automático, el cual se le conoce como entrenamiento de la red neuronal. El entrenamiento de un modelo neuronal implica alimentar al modelo con un conjunto de datos, con el fin de obtener un mínimo de una función objetivo. Es por ello que la red neuronal ajusta sus parámetros para lograr este propósito. Este proceso se lleva a cabo mediante algoritmos de optimización que actualizan iterativamente los parámetros de la red neuronal. El entrenamiento de modelos neuronales puede ser un proceso intensivo en términos de recursos computacionales y tiempo,

especialmente cuando se utilizan grandes conjuntos de datos o modelos complejos. Sin embargo, la selección adecuada de algoritmos sobre los parámetros en concordancia con los procesos físicos y atmosféricos pueden llevar a modelos altamente precisos, así como útiles para una variedad de aplicaciones. En esta sección abordaremos la segmentación de los conjuntos de datos para el entrenamiento, validación y prueba de los modelos. La información que recibirá las redes neuronales, así como el valor a predecir, la función objetivo a optimizar y las técnicas para evitar el sobreajuste de la red.

4.2.1. Segmentación de los conjuntos de datos

La segmentación de conjuntos de entrenamiento, validación y prueba son pasos en la construcción de modelos de redes neuronales para el análisis de series de tiempo (Hastie, Tibshirani, Friedman, y Friedman, 2009). Cabe destacar que en la creación de subconjuntos de datos en series de tiempo, es necesario respetar el orden cronológico de los datos donde el conjunto de entrenamiento debe contener los datos más antiguos mientras que el conjunto de prueba contiene datos más recientes. El conjunto de datos se divide en subconjuntos de ellos para entrenar el modelo, ajustar sus parámetros y posteriormente se evalúa la capacidad predictiva. Esta metodología es esencial para evitar que el modelo aprenda características específicas del conjunto de entrenamiento. Este esquema es fundamental para garantizar que el modelo sea capaz de realizar mejores predicciones y más precisas a futuro. Para este trabajo los períodos de tiempo para cada subconjunto de datos se muestran en la tabla 4.3.

Conjunto de datos	Periodo
Entrenamiento	2015-2019
Validación	2020
Prueba	2021-2022

Tabla 4.3: Periodos de tiempo definidos para los conjuntos de datos de entrenamiento, validación y prueba de la red neuronal.

La figura 4.8 muestra la representación gráfica de los conjuntos de entrenamiento, validación y prueba para las estaciones de monitoreo seleccionadas del SIMA.

4.2.2. Datos de la entrada y salida esperada

La obtención de los datos de entrada en un modelo de series de tiempo, implica la selección de las características relevantes y la preparación de los datos para su ingreso en un modelo de red neuronal. A continuación se describen los pasos más comunes para la división de datos de entrada y salida esperada en un modelo de series de tiempo.

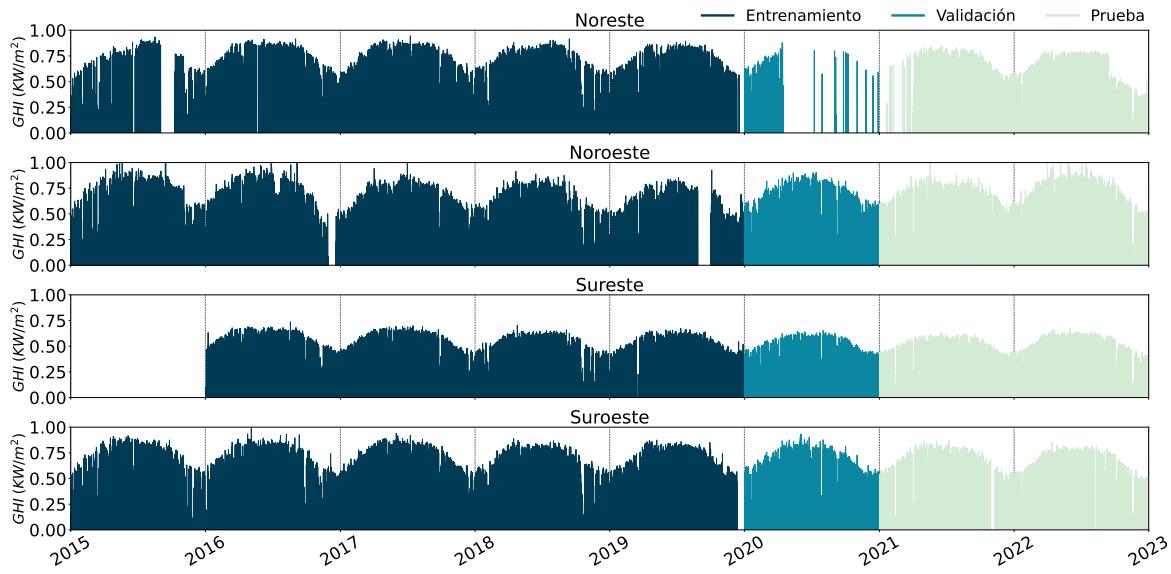


Figura 4.8: Segmentación del conjunto de datos de GHI recolectados por el SIMA para el proceso de entrenamiento, validación y prueba.

Selección de las características relevantes

Las series de tiempo pueden incluir múltiples variables en el mismo periodo, por lo que es importante seleccionar las características relevantes para el problema específico que se está tratando de resolver. En nuestro caso, las características seleccionadas para los datos de entrada y de salida tienen una resolución horaria. Los datos de entrada consideran al GHI , GHI_0 , fracción de nubes estimadas satelitalmente y horas del día, mientras que el resultado de salida es el IUV.

Normalización de los datos

La normalización de los datos permite operar en un único rango de valores que la red neuronal aprende de manera más efectiva. Para este trabajo se normalizaron los parámetros con respecto a su máximo valor físicamente posible. Esto con el fin de que todas las variables se encontrarán entre 0 y 1. La tabla 4.4 contiene los valores máximos para cada variable utilizada.

Ventanas de tiempo

Las ventanas de tiempo recorren la serie temporal dividiendo en subconjuntos de entrada y salida esperada de la red neuronal, estos datos son denominados X y Y . La X denota la matriz o conjunto de matrices que contienen las características relevantes, mientras que Y representa un vector o conjunto de vectores que contienen las salidas esperadas. Las

Característica	Valor máximo
GHI	1
GHI_0	1.5
IUV	15
Hora del día	24
Fracción de nubes	1

Tabla 4.4: Valores usados para normalizar los datos de entrada y salida de las redes neuronales.

ventanas de tiempo pueden definirse como deslizantes o fijas (Kapoor y Bedi, 2013). Los pares XY se utilizan para entrenar y evaluar la red neuronal. La figura 4.9 muestra la metodología empleada para crear estos pares de datos. En esta Tesis, las ventanas tuvieron

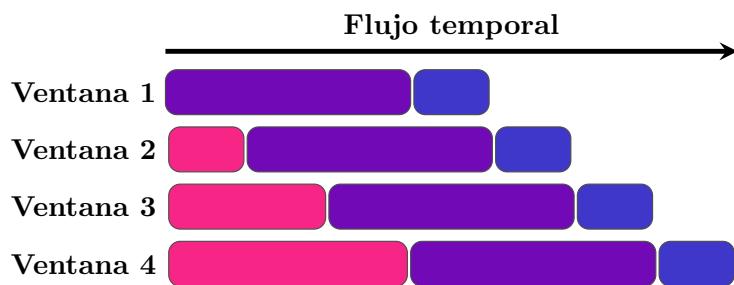


Figura 4.9: Creación de los pares XY , los datos de entrada son representados con el color morado y el color azul ejemplifica a los datos de salida esperada. El cuadro rosa representa los datos pasados que son descartados en cada iteración.

tamaños fijos para la creación de los pares XY . La X considera únicamente los valores obtenidos el día anterior al día pronosticado y Y agrupa las mediciones realizadas durante el día a pronosticar. El horario de los datos XY fue restringido a un horario de 6 a 20 horas, esto debido a que el inicio y final del horario son la hora del amanecer y anochecer a lo largo del año en el AMM.

4.2.3. Función de pérdida

Las funciones de pérdida juegan un papel fundamental en la evaluación del rendimiento de los modelos y su elección adecuada permite generar predicciones precisas (Hastie y cols., 2009). En el contexto de predicción en series de tiempo, las funciones del error cuadrático medio (MSE) y el error absoluto medio (MAE) son las funciones de pérdida más utilizadas. El MSE es sensible a los datos atípicos, esto se debe a que la diferencia entre la predicción y el valor esperado es grande. El MAE es una función no diferenciable en todo su dominio, por lo perjudica el cálculo del gradiente para la retropropagación durante el entrenamiento. El MAE y el MSE comparten la característica de un gradiente

decreciente, causando que la red neuronal no diverja durante el entrenamiento. La función de Huber es una alternativa a las funciones *MSE* y *MAE*, esto debido a que combina las características de estas funciones de pérdida, resultando en una función menos sensible a los valores atípicos que el *MSE* y diferenciable en todo su dominio. La función de Huber contempla un parámetro que controla el punto de transición entre la *MSE* y *MAE*. Si el error es menor al parámetro, se utiliza la *MSE*, mientras que si el error es mayor al parámetro, se utiliza la *MAE*. Por consiguiente, la función de Huber es descrita como:

$$f(x) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{si } |y - \hat{y}| < \alpha \\ \alpha(|y - \hat{y}| - \frac{\alpha}{2}) & \text{si } |y - \hat{y}| \geq \alpha \end{cases} \quad (4.4)$$

donde y es el valor a predecir, \hat{y} es el valor predicho y α es el parámetro a elegir. Las secuencias de IUV contienen cambios críticos en un corto periodo de tiempo, los cuales pueden ser causados por la interacción de la radiación solar y las nubes. En algunos casos, este cambio crítico puede considerarse un valor atípico, es por ello que para este trabajo se emplea la función de Huber con $\alpha = 1$ como función a optimizar en las redes neuronales.

4.2.4. Problemas de sobreajuste

El sobreajuste, también conocido como sobreentrenamiento, es un problema común en el aprendizaje automático, incluyendo las redes neuronales. El sobreajuste es cuando el modelo se ajusta demasiado a los datos de entrenamiento y pierde su capacidad para generalizar (Ying, 2019). Esto se produce cuando el modelo contiene muchos parámetros y poca variabilidad en los datos de entrenamiento. Parte del problema es que el modelo puede comenzar a memorizar los datos de entrenamiento en lugar de aprender patrones generales que puedan generar datos diferentes. El sobreajuste se puede detectar mediante la comparación del rendimiento del modelo en el conjunto de entrenamiento y en el conjunto de validación. Si el rendimiento del modelo es mucho mejor en el conjunto de entrenamiento que en el conjunto de validación, entonces el modelo está sobreajustado a los datos de entrenamiento.

Normalización del gradiente

La normalización del gradiente se utiliza en el entrenamiento de redes neuronales durante el proceso de retropropagación (Zhang, He, Sra, y Jadbabaie, 2020). Cuando los gradientes son demasiado grandes, pueden hacer que los parámetros de la red neuronal cambien drásticamente y causar problemas de estabilidad en el entrenamiento. Por otro lado, si los gradientes son muy pequeños, el entrenamiento puede volverse lento y requerir muchas

iteraciones para converger a una solución. La normalización del gradiente limita la magnitud del gradiente a un valor máximo predefinido, lo que ayuda a mantener su estabilidad durante el entrenamiento y por lo tanto resulta en una red neuronal más robusta. En este trabajo se limitó a 1 al valor de la norma del gradiente.

Detección temprana

La detección temprana es una técnica basada en el rendimiento de la red sobre el conjunto de validación durante el entrenamiento (Prechelt, 1998). Esta técnica es empleada para evitar el sobreajuste y mejorar el rendimiento de la red. La detección temprana consiste en detener el entrenamiento del modelo si el rendimiento bajo el conjunto de validación es de peor calidad a uno obtenido en alguna iteración anterior. Este comportamiento en el rendimiento indica que el modelo empieza a sobreajustarse a los datos del conjunto de entrenamiento y pierde su capacidad para generalizar a nuevos datos. Para esta Tesis se estableció como condición que después de 20 iteraciones sin lograr un mejor rendimiento del modelo en el conjunto de validación, el entrenamiento del modelo se detenga. Los parámetros del modelo que hayan sido empleados para obtener el mejor rendimiento sobre el conjunto de validación serán utilizados en el cálculo del rendimiento de la red con el conjunto de prueba. La figura 4.10 representa el problema del sobreajuste y la detección temprana durante el entrenamiento de la red neuronal.

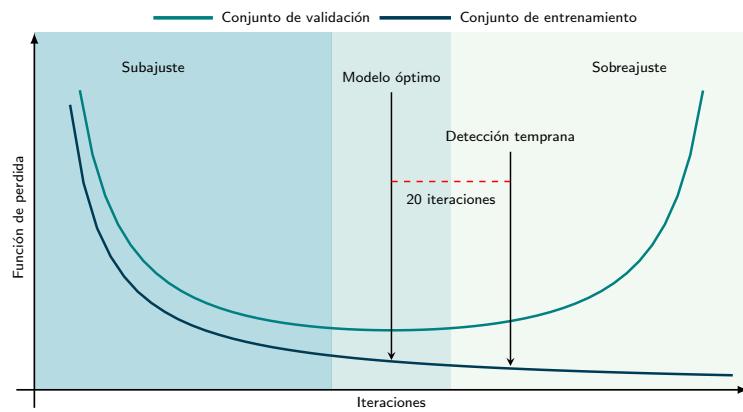


Figura 4.10: Representación gráfica del problema de sobreajuste y obtención de modelo óptimo.

4.3. Propuesta del modelo

Los modelos basados en redes neuronales han demostrado su rendimiento en el pronóstico al capturar patrones complejos y aprender representaciones significativas de los datos. En esta Tesis se proponen diferentes arquitecturas de modelos híbridos para realizar el

pronóstico de IUV por hora, a lo largo del día, dado un conjunto de características del día anterior. El modelo híbrido es descrito como:

$$IUV = f(x, \theta) \quad (4.5)$$

donde x son las características del día anterior (GHI, GHI_0 , fracción de nubes y las horas del día) y θ son los parámetros de la red neuronal. El modelo propuesto aplicará un método descomposición a la GHI para obtener una cantidad fija de IMF. Los modelos híbridos realizan la descomposición por medio del método VMD o el EWT. El método VMD obtendrá 10 IMF, en cambio el método EWT calculará 6 IMF. La figura 4.11 representa el flujo de información del modelo híbrido propuesto.

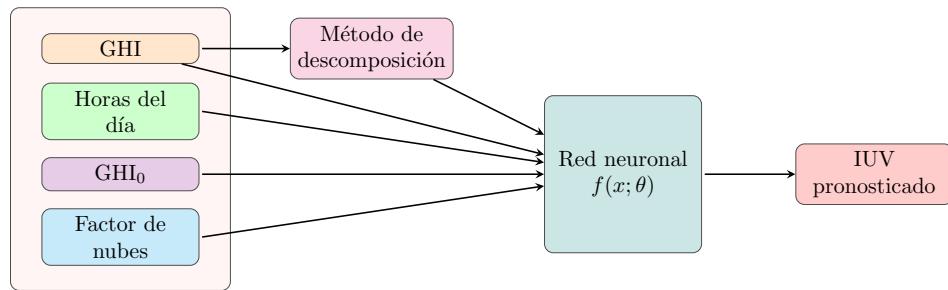


Figura 4.11: Flujo de la información utilizada en los modelos propuestos.

4.4. Modelos implementados

Con el fin de explorar el rendimiento de los modelos basados en redes neuronales para el pronóstico diario del IUV, se realizó un estudio de ablación que considera diversas arquitecturas de redes neuronales. La tabla 4.5 muestra las capas y funciones que incorpora cada modelo propuesto en esta Tesis.

Modelo	CNN	Seq2seq	Mecanismo de atención	Transformer		MLP	ELU
				Codificador	Decodificador		
MLP	X	X	X	X	X	✓	✓
Attention	X	X	✓	X	X	✓	✓
Arclight	X	X	X	✓	X	✓	✓
Transformer	X	X	X	✓	✓	✓	✓
LSTM	X	✓	X	X	X	✓	X
CNN LSTM	✓	✓	X	X	X	✓	X
Residual	X	X	X	X	✓	✓	✓
Attention							
Residual	X	X	X	X	X	✓	✓
MLP							
LSTM	X	✓	✓	X	X	✓	X
Attention							
CNN LSTM	✓	✓	✓	X	X	✓	X
Attention							

Tabla 4.5: Estudio de ablación realizado en esta Tesis. La cabecera de cada columna muestra el nombre de la capa de red neuronal correspondiente y el nombre de la fila contiene el nombre del modelo implementado. La marca de verificación (✓) indica si el modelo contiene la capa de la columna correspondiente y la equis (X) denota que el modelo no hace uno de la capa.

A continuación, se presentan las arquitecturas e hiperparámetros de los modelos implementados en esta Tesis.

4.4.1. Modelos básicos

Perceptrón Multicapa

Se implementó un modelo con arquitectura perceptrón multicapa (MLP) para obtener soluciones no lineales para el pronóstico del IUV. Los hiperparámetros de la capa MLP son los siguientes: 4 capas lineales conectadas consecutivamente donde los tamaños de salida de 128, 64, 32 y 15 respectivamente. Entre cada capa lineal no se aplica una función de activación entre capas. Con el fin de regular los valores cercanos a 0, se agregó una función ELU con $\alpha = 0.001$ a la salida del modelo.

LSTM

Con el fin de extraer patrones temporales en los datos de entrada, se implementó un modelo con arquitectura seq2seq con celdas LSTM. El tamaño de los estados ocultos y vectores de contexto son de 32 en la celda del codificador y decodificador. La información obtenida de la capa seq2seq es dada a una capa MLP. Esta capa MLP contiene dos capas lineales de 1 y 15 tamaños de salida respectivamente.

Attention

Con el objetivo de crear un modelo con un bajo número de parámetros, se implementó una red neuronal basada en el mecanismo de atención *Scaled Dot-Product Attention*, ya que este mecanismo de atención aprende un peso que se aplica a la información de entrada. Este modelo contiene una capa con mecanismo de atención, la cual crea espacio latente de tamaño 8 ($d_k = 8$) y una capa MLP. Esta capa MLP contempla dos capas lineales con 1 y 15 tamaños de salida respectivamente. Para regular el resultado de la red se añadió la función ELU con $\alpha = 0.001$.

4.4.2. Modelos híbridos

El objetivo de las redes híbridas es crear un modelo que contemple las propiedades de las redes neuronales que la componen, por lo que se aprovechan las ventajas de cada una para abordar desafíos complejos. La red híbrida puede mejorar la precisión y la capacidad de generalización del modelo en comparación con el uso de un solo tipo de red. En este trabajo se implementaron las siguientes redes híbridas descritas a continuación.

LSTM Attention

El modelo LSTM Attention es un modelo híbrido, el cual se compone de dos capas LSTM con una arquitectura seq2seq, una capa de atención y una capa MLP. Las capas LSTM aprenden los patrones temporales de los datos de entrada mientras que la capa de atención realiza un peso de la importancia de cada tiempo por medio del mecanismo *Scaled Dot-Product Attention*. Los estados ocultos de las capas LSTM tienen un espacio latente de 32, mientras que la capa de atención crea un espacio latente de 8 ($d_k = 8$). En seguida, la información pasa por la capa MLP, la cual contiene dos capas lineales con tamaños de salida de 1 y 15. La figura 4.12 describe la arquitectura de la red híbrida propuesta.

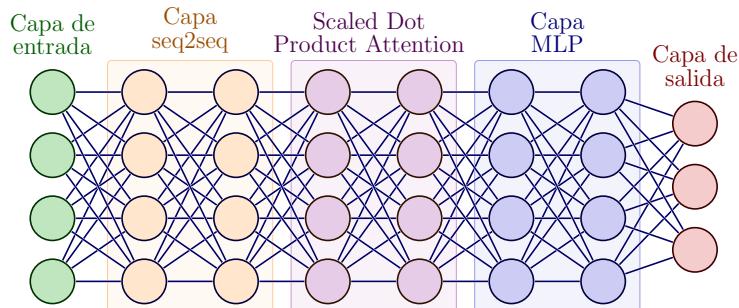


Figura 4.12: Arquitectura de la red híbrida LSTM Attention.

CNN LSTM

Con el fin de crear un modelo que utilice patrones temporales y espaciales, se implementó un modelo híbrido con las capas CNN y LSTM. La capa CNN realiza la operación de convolución, las cuales extraen patrones espaciales. La capa CNN contiene 64 filtros, cada uno con tamaño de 4. A continuación, la información resultante es transferida a una capa con arquitectura seq2seq compuesta por dos celdas LSTM, las cuales extraen los patrones temporales. El espacio latente creado por los estados ocultos de las celdas LSTM tienen un tamaño de 32. En seguida, la información se envía a una capa MLP que contiene dos capas lineales con 1 y 15 de tamaños de salida.

CNN LSTM Attention

Con el propósito de extraer patrones espaciales y temporales e integrarlos por medio de un mecanismo de atención, se propone un modelo híbrido compuesto por capas CNN, LSTM y *Scaled Dot-Product Attention*. Este modelo se compone de 4 capas. La primera capa de este modelo es una capa CNN que contiene 64 filtros de tamaño 4. La segunda capa tiene arquitectura seq2seq que consiste en dos LSTM con un espacio latente de 32 cada una. La tercera capa es una *Scaled Dot-Product Attention* con espacio interno de 8 ($d_k = 8$) y la cuarta capa es un MLP con dos capas lineales, las cuales tienen 1 y 15 de tamaño de salida respectivamente. La figura 4.13 muestra la arquitectura del modelo híbrido CNN LSTM Attention.

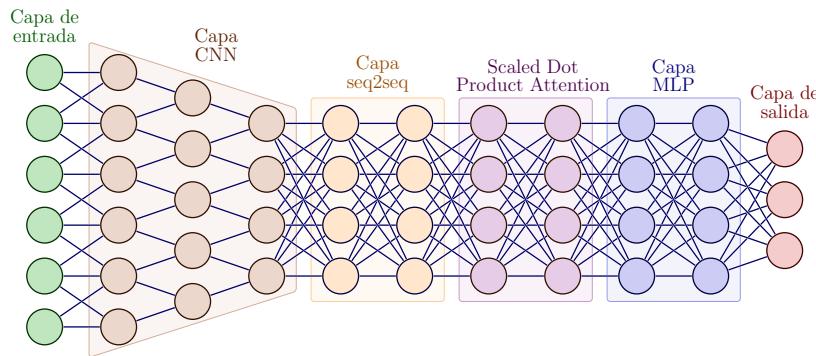


Figura 4.13: Arquitectura del modelo híbrido CNN LSTM Attention.

4.4.3. Modelos residuales

Con el fin de evitar el problema del decaimiento del rendimiento en un modelo varias capas como los modelos híbridos, se propuso una red residual.

Residual MLP

El modelo residual propuesto se compone de un bloque que contiene dos capas lineales y una función de adición y normalización. Este bloque está inspirado en las celdas de las RNN, ya que la ruta de la información crea un ciclo. La primera capa tiene un espacio de salida de 32 y la segunda tiene un espacio de salida de 10 o 14, esto debido a que depende de la descomposición que se realice. Si el modelo realiza la descomposición EWT, el tamaño de salida es 10, en cambio, si la descomposición empleada es VMD entonces el tamaño de salida es 15. La información resultante de la segunda capa lineal es sumada elemento a elemento con la información entrante del ciclo. Para evitar valores muy altos o bajos, se añade una normalización al resultado de la suma. La información resultante del ciclo es dado a una capa MLP que adapta esta información a un vector de tamaño de 15. Esta capa MLP contiene dos capas lineales de tamaños de 1 y 15. Con el fin de regular el resultado de la red, se añadió una función ELU con $\alpha = 0.001$. La figura 4.14 representa la arquitectura de la red residual MLP.

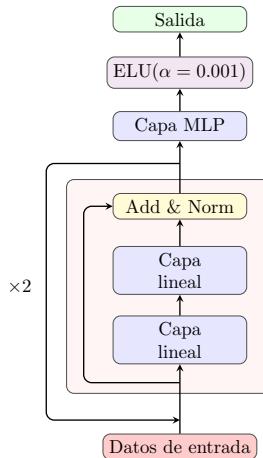


Figura 4.14: Arquitectura del modelo residual MLP.

4.4.4. Modelos basados en Transformers

La característica distintiva de los modelos Transformers es su capacidad para capturar las dependencias de largo alcance en las secuencias de entrada. A diferencia de las arquitecturas recurrentes tradicionales, como las redes RNN, los modelos Transformers no dependen de la propagación secuencial de la información debido a que utilizan mecanismos de atención, la cual relaciona a cualquier posición en la secuencia directamente con cualquier otra. Esto facilita el procesamiento paralelo y la captura de relaciones complejas

en el texto. En esta Tesis incorporamos un conjunto de modelos basados en los bloques de codificación, decodificación y la unión de estos dos.

Transformer

Se implementó un modelo basado en la arquitectura Transformer expuesto por Vaswani y cols. (2017b). A diferencia del modelo Transformer original, nuestra implementación utiliza el *Scaled Dot-Product Attention* como mecanismo de atención en vez del Multi-Head Attention. En seguida del mecanismo de atención se añade una capa lineal con el propósito de obtener una matriz con el mismo tamaño de los datos de entrada. Además el modelo propuesto no utiliza la capa de positional encoding en el codificador y decodificador. El codificador del modelo recibe la información del día anterior que haya sucedido entre las 6 y 15 horas, por otro lado, el decodificador utiliza la información recabada entre las 11 y 20 horas. El tamaño del espacio latente creado por el Scaled Dot-Product Attention en el codificador es de 64, por otro lado, el mecanismo de atención en el decodificador crea un espacio latente de 16. La salida del decodificador del modelo es enviada a una capa MLP para procesar su información. Esta capa MLP contiene dos capas lineales con tamaños de salida 1 y 15. La información resultante de esta red es dada como entrada a una función ELU con $\alpha = 0.001$, esto con el objetivo de controlar los valores cercanos a 0. La figura 4.15 representa la arquitectura del modelo propuesto basado en el Transformer.

Arclight

Con el fin de reducir el costo computacional y el sobreajuste que produce el modelo Transformer, se propuso un estudio de ablación basado en los componentes del Transformer. En este estudio de ablación se consideraron dos modelos que se componen únicamente del codificador o el decodificador. El modelo que contempla el codificador del Transformer fue nombrado como *Arclight*. El tamaño del espacio latente creado por la capa *Scaled Dot-Product Attention* es de 16. La información que resulta del proceso que realiza el decodificador es dada a una capa MLP que contiene dos capas lineales. Estas capas lineales tienen 1 y 15 de tamaños de salida respectivamente. Con el fin de regular la salida del modelo, se aplica una función ELU con $\alpha = 0.001$. La figura 4.16 representa la arquitectura del modelo Arclight.

Residual Attention

El modelo que utiliza el decodificador del Transformer fue llamado Residual Attention. El mecanismo de atención empleado es el *Scaled Dot-Product Attention*. La característica principal de este modelo es que la segunda capa del mecanismo de atención recibe la

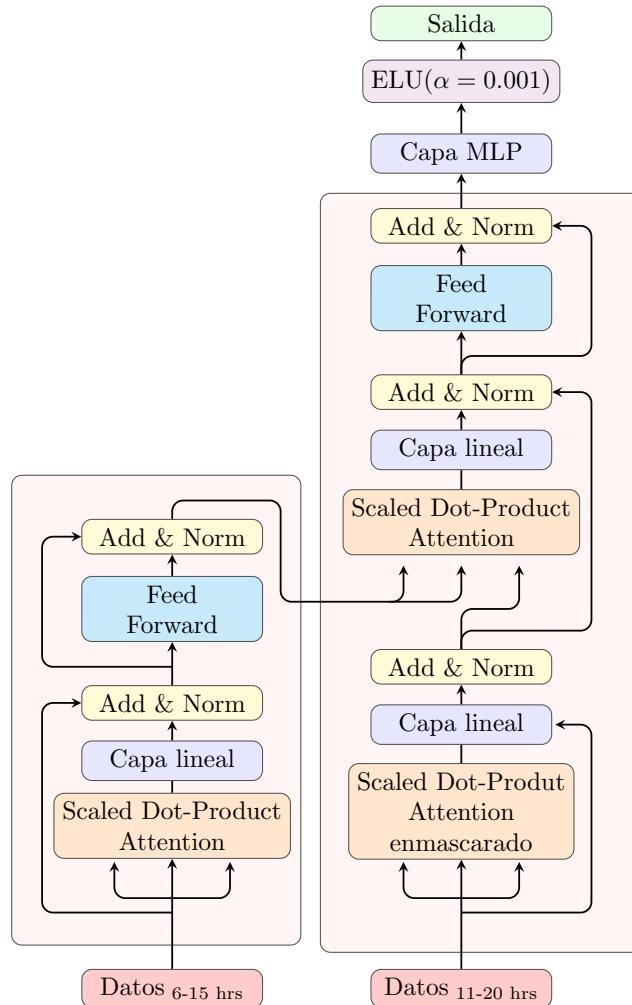


Figura 4.15: Arquitectura del modelo propuesto basado en Transformers.

información de salida de la primera capa de mecanismo de atención como *query* y la información de la entrada es utilizada como *key* y *value*. El tamaño del espacio latente creado por los dos mecanismos de atención es de 16. La salida del decodificador es enviada a una capa MLP, la cual contiene dos capas lineales con tamaños de salida de 1 y 15 respectivamente. La figura 4.17 representa la arquitectura del modelo Residual Attention.

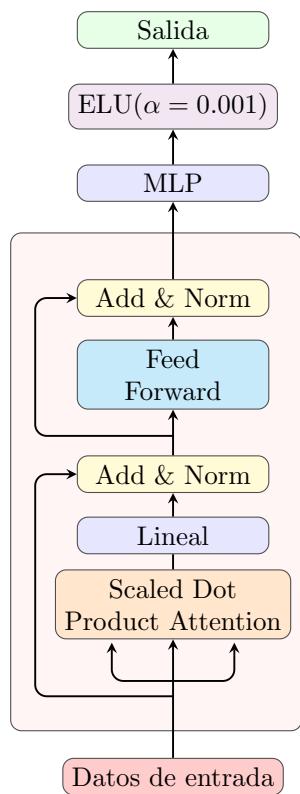


Figura 4.16: Arquitectura del modelo *Arclight*.

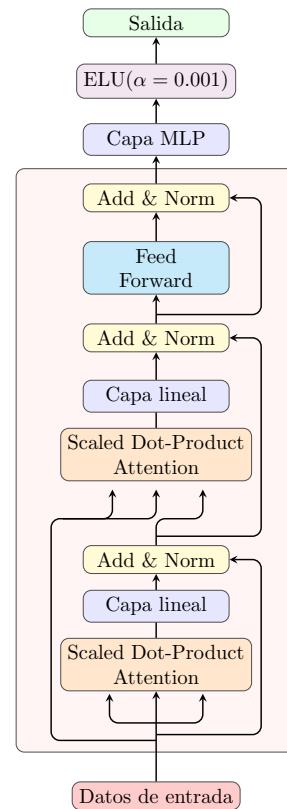


Figura 4.17: Arquitectura del modelo Residual Attention.

Resultados

En esta sección se expondrán y analizarán las métricas obtenidas con los modelos de redes neuronales implementados que realizan la descomposición EWT o VMD con base al conjunto de prueba para cada estación de monitoreo atmosférico. Además se mostrará el número de parámetros que emplea cada modelo en función de la descomposición y el tamaño de datos de entrada que utilicen. Para un análisis más detallado en el rendimiento y variabilidad de los resultados se calcularon las distribuciones de las métricas diarias obtenidas en la predicción del IUV al siguiente día y para el séptimo día consecutivo. Con el propósito de analizar el comportamiento del rendimiento de los modelos a lo largo de una semana, se calculó el coeficiente de determinación para cada día de manera independiente.

5.1. Métricas de rendimiento

Uno de los conceptos importantes en la creación de un modelo, es la medición del rendimiento del mismo basado en datos que no fueron utilizados durante el proceso de entrenamiento y validación. Las funciones encargadas de obtener el rendimiento de un modelo son llamadas métricas (Hastie y cols., 2009). Con el fin de analizar el rendimiento de los modelos desde una perspectiva basada en las diferencias entre los pronósticos obtenidos y los datos observados, se utilizará el Error Absoluto Medio y la Raíz del Error Cuadrático Medio. Debido a que la distribución de las mediciones del IUV varía según las condiciones del cielo, se utilizará el coeficiente de determinación como la métrica principal para evaluar el rendimiento de los modelos. Esta elección se basa en que dicha métrica considera la distribución de los datos en el conjunto de pruebas (Chicco, Warrens, y Jurman, 2021; Yu, Cao, y Zhu, 2019). Para esta sección N es el número de valores observados, \hat{y}_i y y_i es el valor i-ésimo estimado y observado respectivamente.

5.1.1. Error Absoluto Medio

El Error Absoluto Medio (MAE por sus siglas en inglés) es una función que promedia las diferencias entre los datos estimados y observados con una proporción lineal, esto quiere decir que las diferencias obtenidas en cada observación tienen el mismo peso (Willmott y

Matsuura, 2006). El MAE está definido como:

$$MAE = \frac{1}{N} \sum_{i=1}^N | \hat{y}_i - y_i | \quad (5.1)$$

5.1.2. Error Cuadrático Medio

El Error Cuadrático Medio (MSE por sus siglas en inglés) es una métrica que promedia las diferencias entre el valor estimado y observado al cuadrado. El MSE da un mayor peso a aquellas diferencias que obtengan un mayor valor. Por otro lado, las diferencias que obtengan un menor valor tendrán una menor aportación a la métrica. Es por ello que MSE es usado cuando se quieren detectar valores atípicos. Para obtener una métrica con la misma unidad de medida que los datos, se obtiene la Raíz del Error Cuadrático Medio (RMSE por sus siglas en inglés). El RMSE se define como:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (5.2)$$

5.1.3. Coeficiente de determinación

El coeficiente de correlación (R^2) está basado en la razón del error cuadrático medio y la varianza del valor observado. Cuando R^2 es igual a cero, se puede interpretar que el modelo tiene el mismo rendimiento que utilizar el promedio de los datos históricos (Chicco y cols., 2021). El R^2 está definido como:

$$R^2 = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (5.3)$$

donde \bar{y} es la media de los valores observados.

5.2. Número de parámetros

El número de parámetros en cada modelo implementado depende del tamaño de los datos de entrada, el número de capas ocultas y el tipo de arquitectura. En la tabla 5.1 se muestra el número de parámetros que necesita cada modelo implementado en esta Tesis dependiendo del método de descomposición que realice al *GHI*. Se calculó la razón entre el número de parámetros necesarios los modelos basados en la descomposición VMD y los

modelos que usan la descomposición EWT. El incremento en el número de parámetros se debe a que el método VMD obtiene 10 IMF, en cambio, el método EWT obtiene 6 IMF.

Modelo	EWT	VMD	Razón
MLP	30159	37839	1.255
Attention	489	585	1.196
Arclight	1903	3003	1.578
Residual Attention	2823	4739	1.679
Residual MLP	953	1225	1.285
Transformer	6416	10196	1.589
LSTM	14353	14865	1.036
CNN LSTM	23844	24868	1.043
LSTM Attention	15097	15609	1.034
CNN LSTM Attention	24588	25612	1.042

Tabla 5.1: Número de parámetros de los modelos implementados para cada método de descomposición (EWT y VMD) y la razón del número de parámetros de los modelos que realizan la descomposición VMD y los modelos que aplican la descomposición EWT.

5.3. Tiempo de entrenamiento y ejecución

El tiempo de entrenamiento y ejecución de los modelos es un elemento importante a tomar en cuenta cuando se tienen recursos computacionales limitados. En esta Tesis se midió el tiempo transcurrido durante el entrenamiento y la ejecución de los modelos en una computadora con una tarjeta de video y en una computadora con bajo poder computacional. El tiempo requerido para el entrenamiento de los modelos puede variar considerablemente debido a la técnica de detección temprana implementada. Con el fin de obtener un resultado robusto, se ejecutó 20 veces el proceso de entrenamiento y ejecución. La computadora con tarjeta de video se encuentra en el Laboratorio de Supercómputo del Bajío. La tarjeta de video utilizada es la Titan RTX de la marca Nvidia, la cual cuenta con 24 gigabytes de memoria. La computadora con bajo poder computacional utilizada en este experimento tiene instalado una unidad central de procesamiento (CPU) modelo Celeron N4000 de la marca Intel, la cual tiene una frecuencia de 1.10 GHz. Esta computadora cuenta con una memoria de acceso aleatorio (RAM) de 2 gigabytes. La tabla 5.2 contiene el tiempo promedio de entrenamiento y ejecución de los modelos medidos en la computadora con tarjeta de video, la cual muestra que el modelo Residual MLP es el más rápido en el entrenamiento cuando se utiliza el método EWT y el modelo MLP es el más rápido en el entrenamiento cuando se emplea el método VMD. Por otro lado, el modelo MLP es el modelo que tiene la ejecución más rápida.

Modelo	Entrenamiento (min)		Ejecución (ms)	
	EWT	VMD	EWT	VMD
MLP	3	3	2	2
Attention	7	4	2	2
Arclight	9	8	3	3
Residual Attention	9	10	4	4
Residual MLP	3	3	2	2
Transformer	14	11	5	5
LSTM	8	7	5	5
CNN LSTM	12	7	5	5
LSTM Attention	6	5	6	5
CNN LSTM Attention	8	8	5	6

Tabla 5.2: Tiempo promedio empleado en el entrenamiento y la ejecución de los modelos propuestos para cada método de descomposición en la computadora con tarjeta de video. El tiempo de entrenamiento y ejecución se expresan en minutos (min) y milisegundos (ms) respectivamente.

La tabla 5.3 contiene el tiempo utilizado para realizar el entrenamiento y ejecución de los modelos medidos en la computadora con bajo poder computacional. En esta tabla muestra que los modelos Residual MLP y MLP son los más rápidos cuando se realiza el entrenamiento para la descomposición EWT y VMD respectivamente. Por otro lado, el modelo MLP es el modelo con menor tiempo de ejecución para los dos métodos de descomposición.

Modelo	Entrenamiento (min)		Ejecución (ms)	
	EWT	VMD	EWT	VMD
MLP	4	3	4	4
Attention	5	4	5	4
Arclight	13	11	7	6
Residual Attention	18	17	8	8
Residual MLP	4	4	5	5
Transformer	20	17	10	10
LSTM	16	12	17	17
CNN LSTM	21	15	17	16
LSTM Attention	12	12	18	18
CNN LSTM Attention	18	18	16	18

Tabla 5.3: Tiempo promedio requerido para el entrenamiento y ejecución de los modelos propuestos que realizan la descomposición EWT y VMD en la computadora con bajo poder computacional. Los tiempos de entrenamiento y ejecución de los modelos se encuentran en minutos (min) y milisegundos (ms) respectivamente.

Comparando los tiempos requeridos para realizar el entrenamiento y la ejecución de los modelos en la computadora con tarjeta de video y la computadora con bajo poder computacional, se obtiene que en promedio la computadora con bajo poder computacional tarda 1.64 y 2.59 veces más que la computadora con tarjeta de video en realizar el entrenamiento y la ejecución de los modelos respectivamente.

5.4. Predicción diaria

En esta sección se analizaran el rendimiento de los modelos descritos en la sección 4.4 para el pronóstico de IUV con base al conjunto de datos de prueba en cada estación de monitoreo del SIMA. Se calcularon las métricas diarias con el fin de obtener un resultado más detallado de los datos en comparación de las métricas basadas en todo el conjunto de prueba. Para identificar a los cuatro mejores modelos se seleccionaron aquellos que tengan una mayor mediana en las métricas de R^2 diaria. Además, se comparará de manera cualitativa los pronósticos producidos por el mejor modelo y el IUV estimado a partir de las mediciones *in situ* UVA+UVB.

5.4.1. EWT

En esta sección se expondrán los rendimientos de los modelos que realizan la descomposición EWT al pronosticar el IUV para el día siguiente en cada estación de monitoreo.

La tabla 5.4 muestra la mediana de las métricas diarias de los modelos propuestos, en la que puede observar que los modelos MLP, Residual MLP, Residual Attention y Arclight obtienen mejores rendimientos en comparación a los modelos basados en RNN y CNN. Sin embargo, no existe una diferencia significativa entre modelos.

Para realizar un análisis en la variabilidad del rendimiento diario de los modelos, se procedió a obtener las distribuciones de las métricas diarias. La figura 5.1 muestra la distribución de la métrica R^2 de los cuatro modelos seleccionados para cada estación de monitoreo. El modelo con mayor mediana se encuentra del lado izquierdo y de lado derecho se encuentra el modelo con menor mediana de los cuatro mejores.

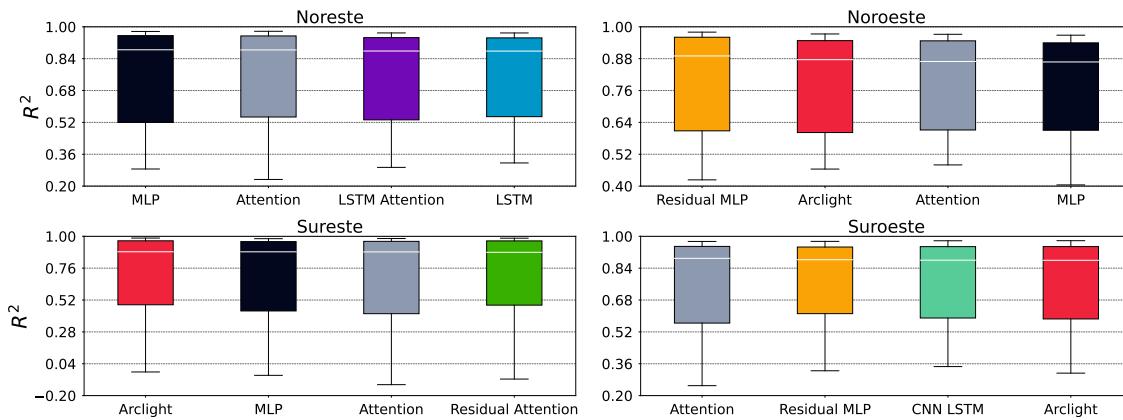


Figura 5.1: Distribuciones de la métrica R^2 diaria de los modelos seleccionados. El modelo con el mayor valor en la mediana se encuentra de lado izquierdo y a la derecha es el modelo con menor mediana de los cuatro modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

De igual manera se calcularon las distribuciones correspondientes a las métricas MAE y $RMSE$. Las figuras 5.2 y 5.3 muestran las distribuciones de las métricas diarias de la MAE y $RMSE$ respectivamente.

Estación	Modelo	R2	MAE	RMSE
Noreste	MLP	0.885	0.060	0.096
	Attention	0.884	0.060	0.098
	Arclight	0.882	0.060	0.096
	Residual Attention	0.878	0.064	0.100
	Residual MLP	0.878	0.069	0.106
	Transformer	0.874	0.064	0.098
	LSTM	0.878	0.066	0.101
	CNN LSTM	0.873	0.066	0.101
	LSTM Attention	0.878	0.064	0.100
Noroeste	CNN LSTM Attention	0.863	0.072	0.107
	MLP	0.868	0.075	0.112
	Attention	0.869	0.068	0.107
	Arclight	0.876	0.067	0.105
	Residual Attention	0.863	0.069	0.106
	Residual MLP	0.890	0.063	0.099
	Transformer	0.861	0.070	0.107
	LSTM	0.856	0.076	0.115
	CNN LSTM	0.863	0.068	0.103
Sureste	LSTM Attention	0.855	0.075	0.111
	CNN LSTM Attention	0.864	0.073	0.108
	MLP	0.883	0.040	0.059
	Attention	0.882	0.039	0.060
	Arclight	0.883	0.038	0.059
	Residual Attention	0.879	0.038	0.059
	Residual MLP	0.879	0.041	0.061
	Transformer	0.867	0.042	0.065
	LSTM	0.878	0.042	0.061
Suroeste	CNN LSTM	0.879	0.041	0.060
	LSTM Attention	0.875	0.043	0.063
	CNN LSTM Attention	0.873	0.045	0.065
	MLP	0.876	0.065	0.100
	Attention	0.889	0.062	0.098
	Arclight	0.880	0.058	0.093
	Residual Attention	0.866	0.065	0.100
	Residual MLP	0.882	0.062	0.098
	Transformer	0.862	0.065	0.099

Tabla 5.4: Mediana de las métricas diarias obtenidas para el pronóstico del día siguiente de los modelos implementados que realizan la descomposición EWT. En negritas se encuentra resaltada el mejor rendimiento para cada métrica en las estaciones de monitoreo.

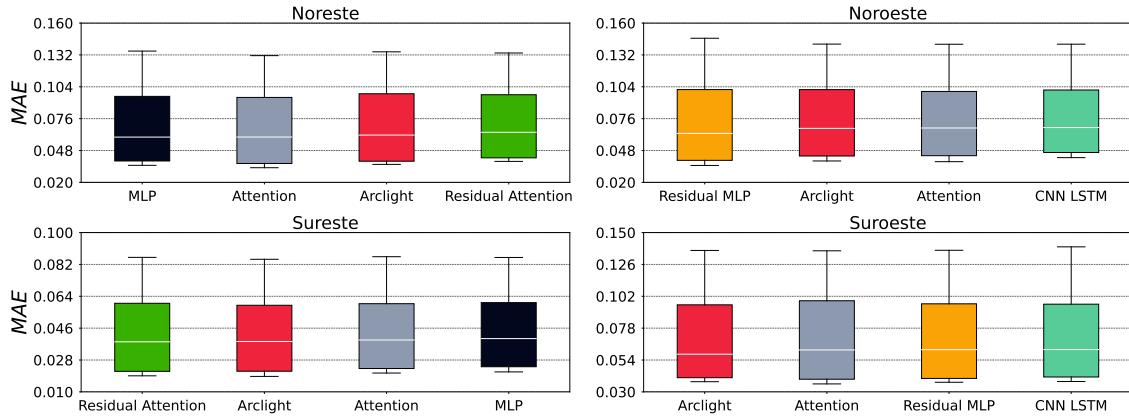


Figura 5.2: Distribuciones de la métrica MAE diaria para los cuatro modelos con mayor mediana en esta métrica. El modelo con el menor valor en la mediana se encuentra de lado izquierdo y a la derecha es el modelo con mayor mediana de los cuatro modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

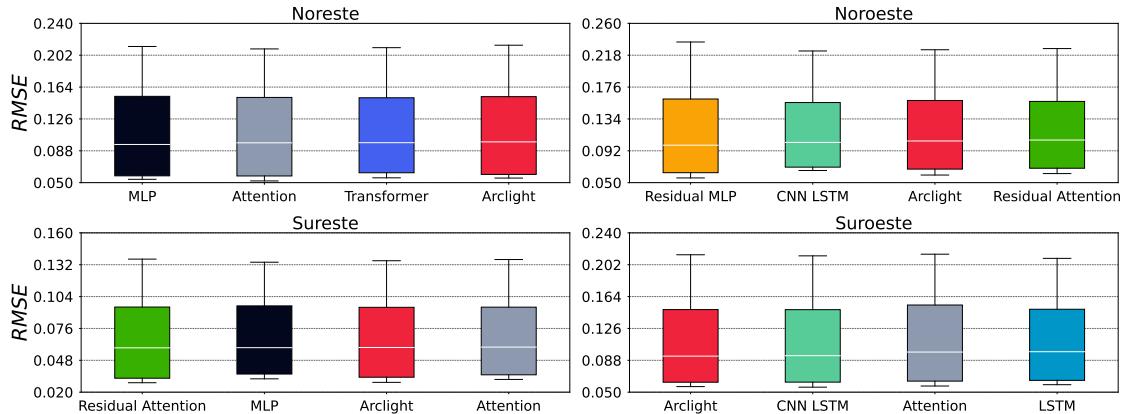


Figura 5.3: Distribuciones de la métrica $RMSE$ diaria para los cuatro modelos con mayor mediana en esta métrica. El modelo con el menor valor en la mediana se encuentra de lado izquierdo y a la derecha es el modelo con mayor mediana de los cuatro modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

Con base en la tabla 5.4 y las figuras 5.1, 5.2 y 5.3, los cuatro modelos seleccionados tienen un desempeño semejante al pronosticar el IUV para el día siguiente. Esto debido a que las distribuciones de las métricas de R^2 , MAE y $RMSE$ tienen distribuciones similares. La figura 5.4 muestra el pronóstico del IUV obtenida con los cuatro modelos seleccionados en cada estación de monitoreo y el IUV obtenido a partir de las mediciones de GHI . Se observa que los modelos tienden a subestimar el IUV en comparación al IUV estimado a partir del GHI . Además, se visualiza que los modelos no son capaces de pronosticar las disminuciones del IUV que son producidas por la presencia de nubes.

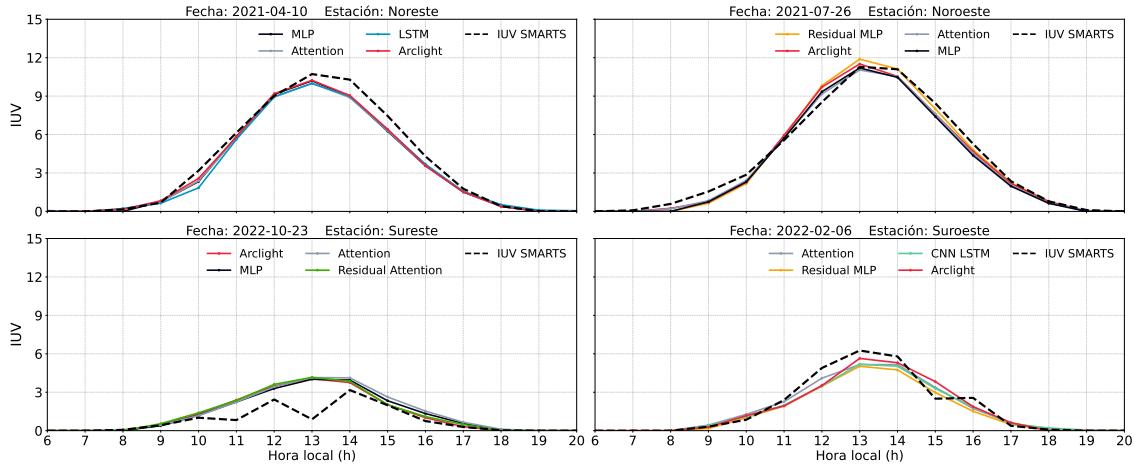


Figura 5.4: Pronóstico del IUV para el siguiente día obtenido con los cuatro modelos con mayor mediana en la métrica R^2 utilizando la descomposición EWT en cada estación de monitoreo. La línea punteada es la estimación del IUV obtenida del modelo SMARTS a partir de mediciones de GHI .

5.4.2. VMD

De igual manera en cómo se realizó en la sección anterior, se analizó el rendimiento de los modelos que emplean la descomposición VMD al pronosticar el IUV para cada estación de monitoreo. La tabla 5.5 muestra la mediana de las métricas diarias por cada modelo. En este caso, no existe una diferencia significativa entre los rendimientos de los modelos analizados. Con el fin de examinar la variabilidad en el rendimiento diario de los modelos, se llevó a cabo un análisis de las distribuciones de las métricas diarias. La representación gráfica de estas distribuciones se muestra en la Figura 5.1, donde se visualiza la distribución de la métrica R^2 para los cuatro modelos seleccionados en cada estación de monitoreo. En esta figura, el modelo con la mediana más alta se encuentra en el extremo izquierdo, mientras que en el extremo derecho se ubica el modelo con la mediana más baja entre los cuatro modelos seleccionados.

Estación	Modelo	R2	MAE	RMSE
Noreste	MLP	0.881	0.061	0.099
	Attention	0.885	0.061	0.098
	Arclight	0.885	0.062	0.097
	Residual Attention	0.877	0.060	0.098
	Residual MLP	0.879	0.066	0.104
	Transformer	0.880	0.060	0.096
	LSTM	0.886	0.064	0.101
	CNN LSTM	0.875	0.061	0.097
	LSTM Attention	0.876	0.065	0.103
Noroeste	CNN LSTM Attention	0.873	0.060	0.097
	MLP	0.867	0.072	0.112
	Attention	0.864	0.069	0.108
	Arclight	0.863	0.070	0.107
	Residual Attention	0.866	0.068	0.106
	Residual MLP	0.882	0.073	0.107
	Transformer	0.859	0.070	0.108
	LSTM	0.867	0.072	0.110
	CNN LSTM	0.866	0.071	0.107
Sureste	LSTM Attention	0.863	0.073	0.112
	CNN LSTM Attention	0.859	0.071	0.110
	MLP	0.884	0.040	0.060
	Attention	0.882	0.040	0.060
	Arclight	0.883	0.039	0.059
	Residual Attention	0.877	0.040	0.061
	Residual MLP	0.881	0.040	0.061
	Transformer	0.875	0.039	0.060
	LSTM	0.883	0.038	0.059
Suroeste	CNN LSTM	0.870	0.039	0.060
	LSTM Attention	0.877	0.041	0.062
	CNN LSTM Attention	0.880	0.042	0.061
	MLP	0.871	0.067	0.104
	Attention	0.878	0.061	0.098
	Arclight	0.865	0.060	0.095
	Residual Attention	0.875	0.059	0.094
	Residual MLP	0.881	0.065	0.100
	Transformer	0.858	0.060	0.097

Tabla 5.5: Mediana de las métricas diarias obtenidas para el pronóstico del día siguiente de los modelos implementados que realizan la descomposición VMD. En negritas se encuentra resaltada el mejor rendimiento para cada métrica en las estaciones de monitoreo.

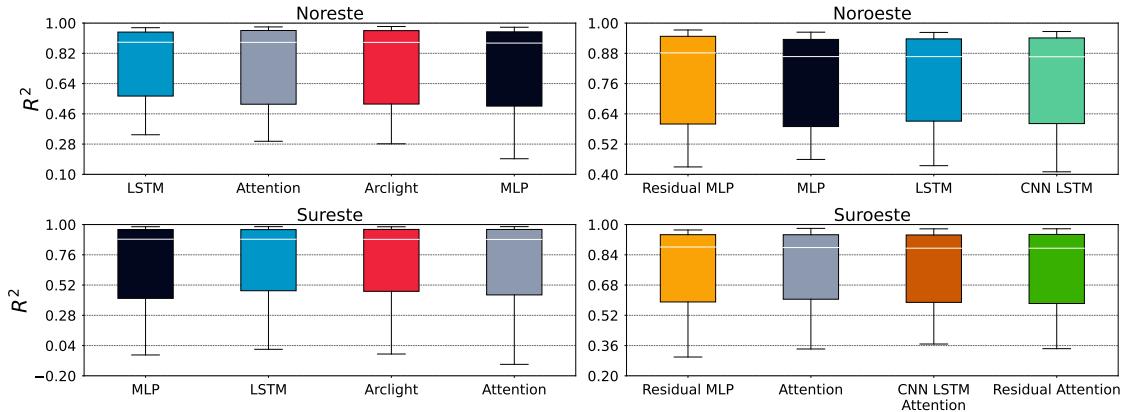


Figura 5.5: Distribuciones de la métrica R^2 diaria calculadas para los modelos seleccionados. El modelo con el mayor valor en la mediana diaria se encuentra de lado izquierdo y a la derecha es el modelo con menor mediana de los modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

De igual manera se calcularon las distribuciones correspondientes a las métricas MAE y $RMSE$. Las figuras 5.6 y 5.7 muestran las distribuciones de las métricas diarias de la MAE y $RMSE$ respectivamente.

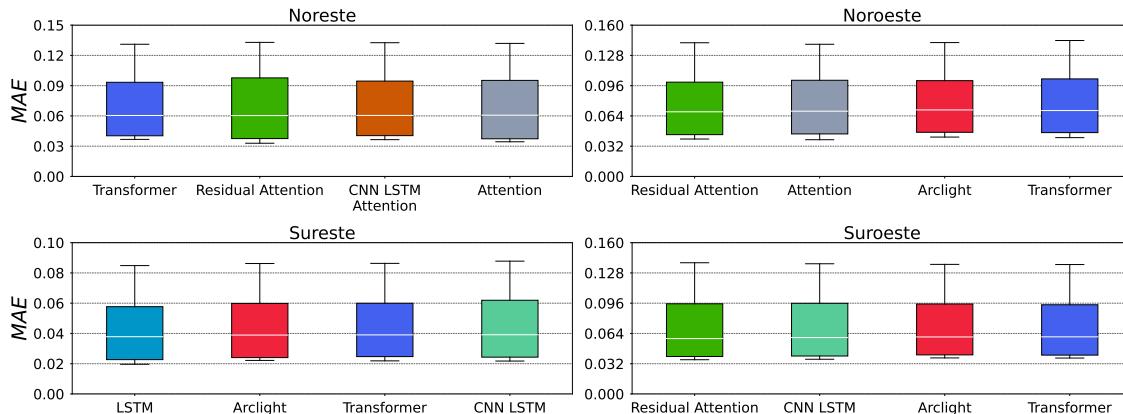


Figura 5.6: Distribuciones de la métrica MAE diaria calculadas para los cuatro modelos basados en la descomposición VMD con mayor mediana. El modelo con el mayor valor en la mediana diaria se encuentra de lado izquierdo y a la derecha es el modelo con menor mediana de los modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

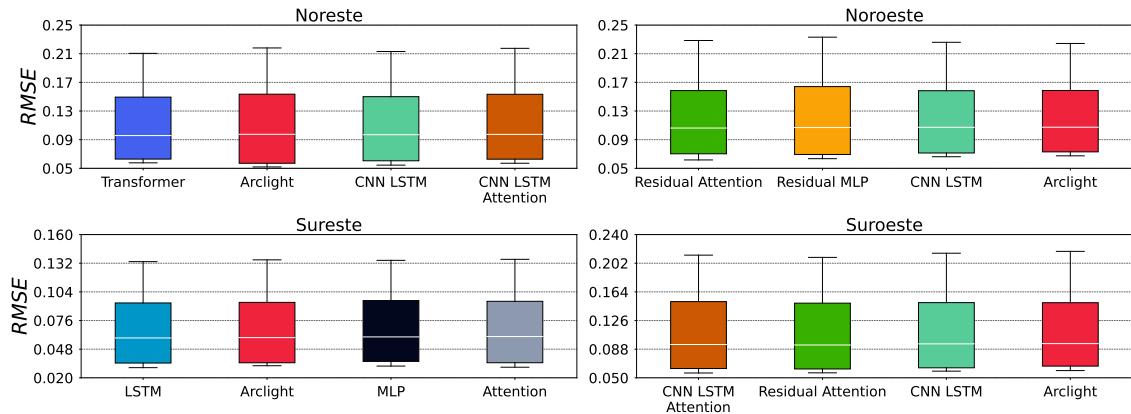


Figura 5.7: Distribuciones de la métrica $RMSE$ diaria calculadas para los cuatro modelos basados en la descomposición VMD con mayor mediana. El modelo con el mayor valor en la mediana diaria se encuentra de lado izquierdo y a la derecha es el modelo con menor mediana de los modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

Con base en la tabla 5.5 y las figuras 5.5, 5.6 y 5.7, podemos asegurar que los cuatro mejores modelos tienen un desempeño semejante al pronosticar el IUV para el día siguiente. Esto debido a que las distribuciones de las métricas de R^2 , MAE y $RMSE$ tienen valores similares. La figura 5.8 muestra el pronóstico del IUV obtenida con los mejores cuatro modelos de cada estación de monitoreo seleccionada y el IUV obtenido a partir de las mediciones de GHI . Se visualiza que los modelos tienden a subestimar el IUV en contraste con la estimación del IUV a partir del GHI . Adicionalmente, se observa que los modelos no son capaces de pronosticar las disminuciones del IUV que son producidas por la presencia de nubes.

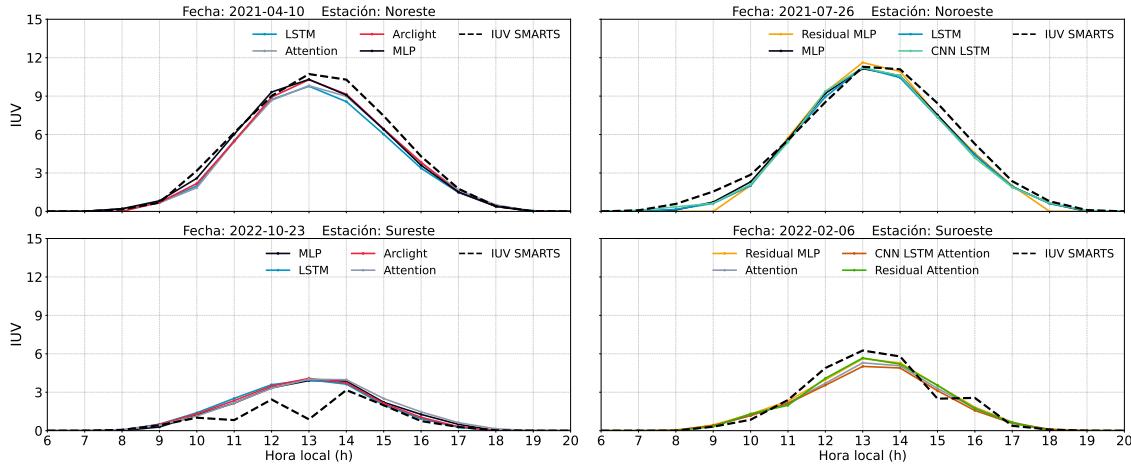


Figura 5.8: Pronóstico del IUV obtenido de los cuatro mejores modelos empleando la descomposición VMD de cada estación de monitoreo seleccionada. La línea punteada es la estimación del IUV obtenida a partir del modelo SMARTS.

5.4.3. Comparación EWT y VMD

Las diferencias en el rendimiento de los modelos que computan la descomposición EWT y VMD en el pronóstico del IUV para el día siguiente no es significativa. Esto debido a que las distribuciones de las métricas diarias y sus medianas obtienen valores similares. Sin embargo, los modelos que utilizan la descomposición VMD necesitan un número mayor de parámetros para realizar el pronóstico. Esta diferencia en el número de parámetros va desde 96 (Attention) hasta 7680 (MLP). El aumento en el número de parámetros se debe a que la descomposición VMD obtiene 10 IMF, en cambio, la descomposición EWT obtiene 6 IMF. El modelo LSTM Attention es el menos afectado debido a que el número de parámetros aumenta un 3.4 % con respecto a la versión que usa la descomposición EWT. El modelo que recibe la mayor aumentación en el número de parámetros es el Residual Attention, ya que aumenta un 67.9 % con respecto a la versión que usa la descomposición EWT. Por otro lado, no existe una diferencia significativa en el tiempo necesario para realizar el entrenamiento y la ejecución de los modelos entre los modelos que utilizan la descomposición EWT y VMD. La tabla 5.6 muestra la mediana de las métricas diarias en la tarea del pronóstico horario del IUV para el siguiente día a partir de los mejores modelos en cada método de descomposición en las estaciones de monitoreo seleccionadas.

Estación	Descomposición	Modelo	R ²	MAE	RMSSE
Noreste	EWT	MLP	0.885	0.060	0.096
	EWT	Attention	0.884	0.060	0.098
	EWT	Arclight	0.882	0.060	0.096
	VMD	Residual Attention	0.877	0.060	0.098
	VMD	Transformer	0.880	0.060	0.096
	VMD	LSTM	0.886	0.064	0.101
Noroeste	EWT	Residual MLP	0.890	0.063	0.099
	VMD	Residual Attention	0.866	0.068	0.106
	VMD	Residual MLP	0.882	0.073	0.107
Sureste	EWT	MLP	0.883	0.040	0.059
	EWT	Arclight	0.883	0.038	0.059
	EWT	Residual Attention	0.879	0.038	0.059
	VMD	MLP	0.884	0.040	0.060
	VMD	Arclight	0.883	0.039	0.059
	VMD	LSTM	0.883	0.038	0.059
Suroeste	EWT	Attention	0.889	0.062	0.098
	EWT	Arclight	0.880	0.058	0.093
	VMD	Residual Attention	0.875	0.059	0.094
	VMD	Residual MLP	0.881	0.065	0.100
	VMD	CNN LSTM Attention	0.875	0.061	0.094

Tabla 5.6: Mediana de las métricas diarias obtenidas por los modelos con mejor rendimiento para cada método de descomposición para la tarea del pronóstico horario del IUV a lo largo del siguiente día. En negritas se resalta la mejor métrica para la estación correspondiente.

5.4.4. Comparación con mediciones *in situ* UVA+UVB

Con base en las mediciones *in situ* de radiación UVA+UVB realizadas por con el radiómetro solarmeter, se estimó el IUV. Para realizar esta estimación se siguió la metodología mencionada en la sección 4.1.3, debido a que la estación Noroeste es la estación de monitoreo más cercana a donde se realizaron las mediciones *in situ* con el radiómetro externo al SIMA. La figura 5.9 se muestra el IUV pronosticado por el modelo con mejores métricas para la estación Noroese, el IUV estimado a partir en mediciones de *GHI* realizadas en la estación Noroeste del SIMA y el IUV estimado a partir de mediciones UVA+UVB obtenidas frente al Palacio de Justicia de Nuevo León. En esa figura se aprecia que el modelo de pronóstico obtiene valores semejantes al IUV basado en mediciones UVA+UVB, teniendo la mayor diferencia cuando se presenta el máximo solar. En cualquier caso, el

pronóstico del IUV realizado tiene menores diferencias en comparación al IUV obtenido con el modelo SMARTS.

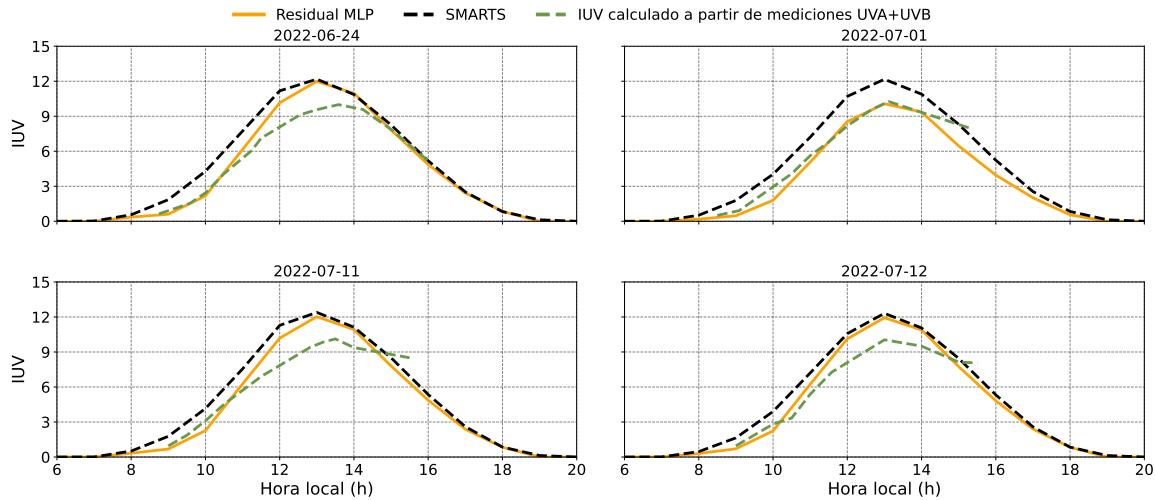


Figura 5.9: IUV pronosticado por el modelo Residual MLP para diferentes días en la estación Noroeste (naranja), IUV estimado a partir de mediciones de *GHI* realizadas en la estación Noroeste (negro) y IUV calculado con base en las mediciones *in situ* UVA+UVB obtenidas frente al Palacio de Justicia de Nuevo León (verde).

5.5. Predicción semanal

En esta sección se analizará el rendimiento obtenido por los modelos descritos en la sección 4.4 para el pronóstico de IUV para el séptimo día consecutivo y a lo largo de la semana con base en el conjunto de datos de prueba en cada estación de monitoreo seleccionada. Se calcularon las métricas para cada día de la semana, con el fin de analizar el comportamiento del rendimiento en función del número de días pronosticados. Para identificar a los cuatro mejores modelos se seleccionaron aquellos con mayor promedio de la mediana diaria de la métrica R^2 .

5.5.1. EWT

En esta sección se analizará el rendimiento de los modelos que realizan la descomposición EWT al pronosticar el IUV para la siguiente semana. La tabla 5.7 muestra la mediana de las métricas diarias para el séptimo día consecutivo. Los modelos MLP, Residual Attention, Arclight y CNN LSTM Attention obtienen un mejor rendimiento en comparación a los modelos basados en CNN y LSTM sin mecanismos de atención. El modelo con el peor rendimiento en todas las estaciones de monitoreo es Attention, esto puede deberse

a que el modelo tiene una poca cantidad de parámetros. Por lo cual, es complicado que el modelo presente un buen rendimiento para el pronóstico con una ventana grande de tiempo. Por otro lado, el modelo Arclight es consistente en las estaciones de monitoreo, esto debido a que obtiene las mejores métricas con valores semejantes. La figura 5.10 muestra el decaimiento de la mediana diaria de métrica R^2 a lo largo de la semana en los modelos seleccionados en cada estación de monitoreo. En esta figura se observa que el decaimiento tiende a seguir un comportamiento lineal, por lo que con ventanas de tiempo más grandes, el rendimiento de los modelos será menor. En general, la métrica R^2 decae entre 11.6 % y 22.8 % para el séptimo día con respecto al primer día de pronóstico.

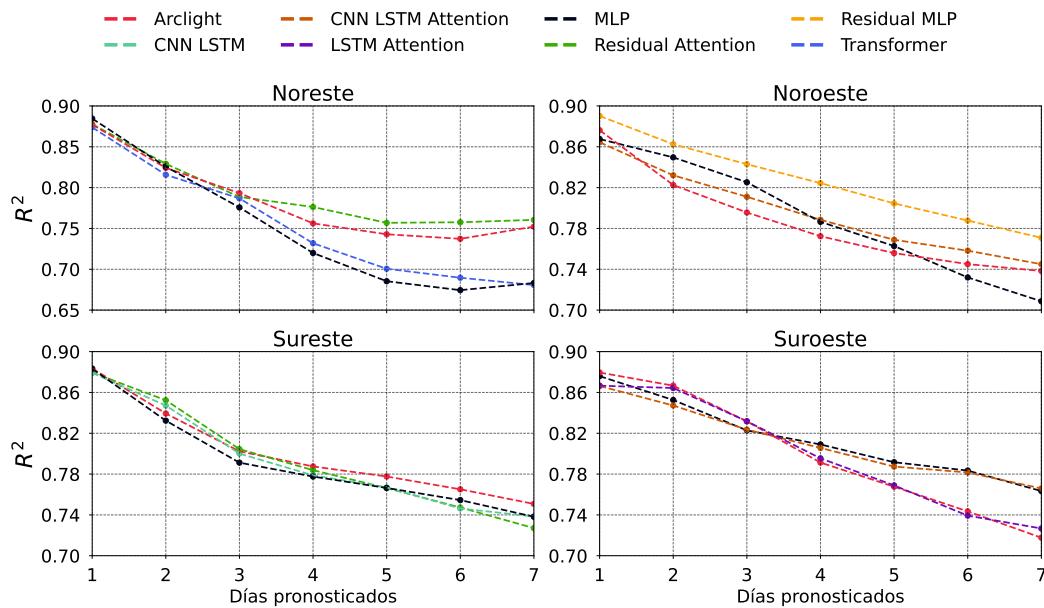


Figura 5.10: Decaimiento de la mediana diaria de la métrica R^2 a lo largo de la semana obtenida por los cuatro modelos con mayor promedio semanal de esta métrica. Las métricas corresponden a los modelos implementados que realizan la descomposición EWT.

Estación	Modelo	R2	MAE	RMSE
Noreste	MLP	0.683	0.079	0.146
	Attention	0.123	0.149	0.266
	Arclight	0.760	0.075	0.134
	Residual Attention	0.760	0.082	0.132
	Residual MLP	0.422	0.125	0.213
	Transformer	0.681	0.082	0.148
	LSTM	0.204	0.167	0.261
	CNN LSTM	0.607	0.099	0.168
	LSTM Attention	0.188	0.166	0.270
Noroeste	CNN LSTM Attention	0.304	0.163	0.237
	MLP	0.709	0.081	0.135
	Attention	0.459	0.102	0.195
	Arclight	0.738	0.084	0.137
	Residual Attention	0.440	0.138	0.236
	Residual MLP	0.771	0.076	0.129
	Transformer	0.545	0.104	0.186
	LSTM	0.526	0.115	0.184
	CNN LSTM	0.695	0.107	0.166
Sureste	LSTM Attention	0.683	0.082	0.148
	CNN LSTM Attention	0.745	0.092	0.151
	MLP	0.738	0.048	0.079
	Attention	0.629	0.053	0.096
	Arclight	0.751	0.047	0.080
	Residual Attention	0.727	0.045	0.081
	Residual MLP	0.614	0.052	0.098
	Transformer	0.705	0.055	0.090
	LSTM	0.688	0.051	0.088
Suroeste	CNN LSTM	0.738	0.046	0.079
	LSTM Attention	0.646	0.058	0.094
	CNN LSTM Attention	0.723	0.055	0.088
	MLP	0.763	0.072	0.125
	Attention	-0.215	0.185	0.308
	Arclight	0.718	0.071	0.138
	Residual Attention	0.612	0.094	0.168
	Residual MLP	0.201	0.148	0.245
	Transformer	0.546	0.105	0.183

Tabla 5.7: Mediana de las métricas de los modelos con descomposición EWT en el pronóstico del IUV para el séptimo día consecutivo. En negritas se encuentra resaltada el mejor rendimiento para cada métrica en cada estación.

Para realizar un análisis de la variabilidad del rendimiento de los modelos para el pronóstico del IUV al séptimo día consecutivo, se estimó la distribución de la métrica R^2 diaria. La figura 5.11 muestra las distribuciones de los cuatro modelos seleccionados para cada estación de monitoreo. El modelo con mayor promedio se encuentra de lado izquierdo y de lado derecho se encuentra el modelo con menor promedio de los cuatro modelos seleccionados.

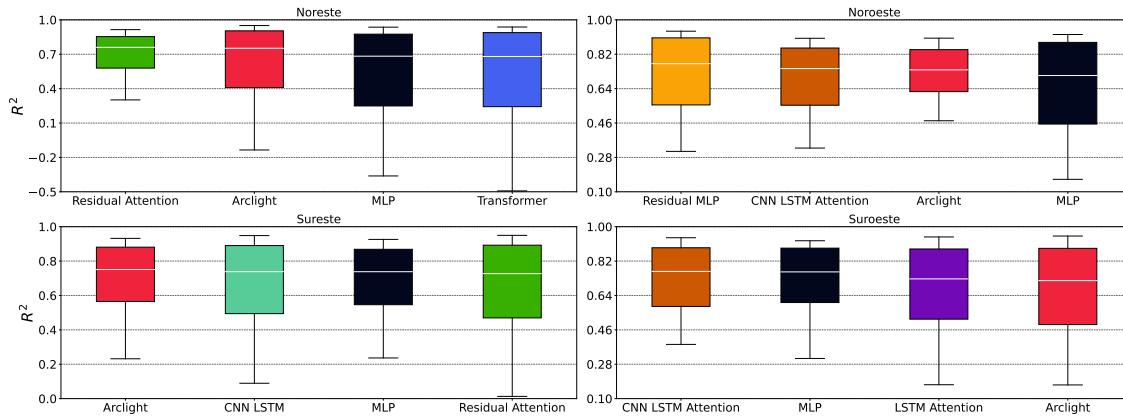


Figura 5.11: Distribuciones de la métrica R^2 en el séptimo día para los cuatro modelos que realizan la descomposición EWT con mayor promedio. El modelo con la mayor promedio encuentra de lado izquierdo y a la derecha el modelo con menor promedio de los cuatro modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

Con esta figura se observa que no existe una consistencia de las distribuciones de la métrica R^2 por cada modelo en las estaciones de monitoreo, ya que un mayor promedio no corresponde a una distribución con un rango de valores más pequeña. Sin embargo, la mediana de las distribuciones en cada modelo es similar entre las estaciones de monitoreo. La figura 5.12 muestra el pronóstico para una semana completa del IUV obtenida con los cuatro modelos seleccionados para cada estación de monitoreo. Se observa que los modelos tienden a subestimar el IUV en comparación al IUV estimado a partir del *GHI*. Además, se visualiza que los modelos no son capaces de pronosticar las disminuciones del IUV que son producidas por la presencia de nubes. Los modelos basados en el Transformer obtienen un buen pronóstico del IUV en todas las estaciones de monitoreo para ventanas de tiempo mayores a un día. En especial el modelo Arclight, el cual se aproxima a la estimación del IUV en el conjunto de prueba.

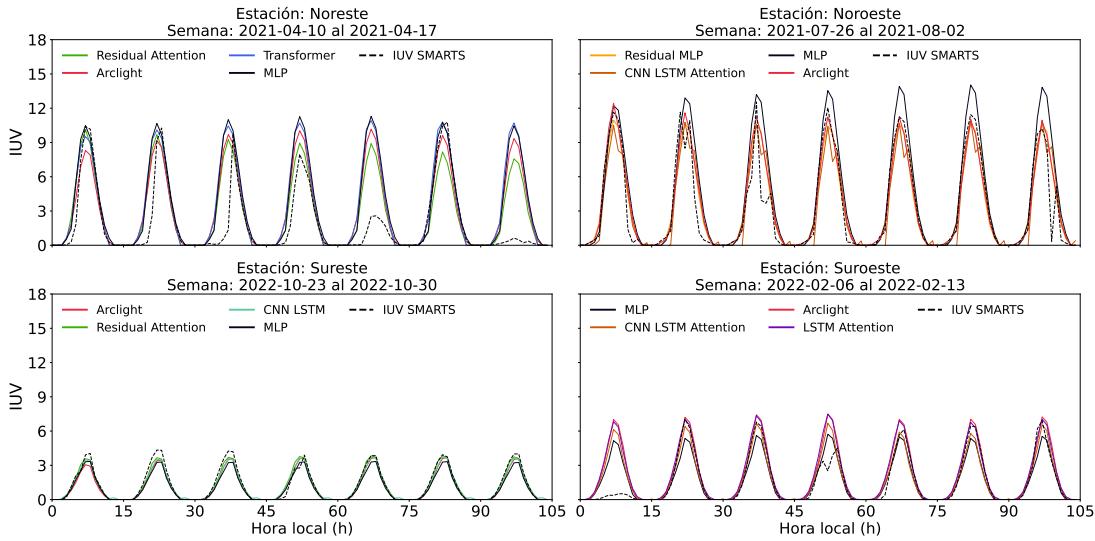


Figura 5.12: Pronóstico semanal del IUV obtenida con los cuatro modelos con mayor promedio semanal de la mediana diaria de la métrica R^2 utilizando la descomposición EWT de cada estación de monitoreo seleccionada. La línea punteada es la estimación del IUV obtenida a partir del modelo SMARTS.

5.5.2. VMD

En esta sección se analizó el rendimiento de los modelos que realizan la descomposición VMD al pronosticar el IUV para la siguiente semana en cada estación de monitoreo seleccionada. La tabla 5.8 muestra la mediana de las métricas diarias en el séptimo día consecutivo para cada modelo. Los modelos MLP, residual Attention, residual MLP y Arclight obtienen los mejores rendimientos en comparación a los demás modelos. Para este caso, el modelo residual MLP, Arclight, LSTM ,LSTM Attention y CNN LSTM Attention presentan rendimientos insuficientes, ya que obtienen valores negativos o cercanos a cero en la métrica R^2 , lo cual indica que es más recomendable realizar pronósticos empleado la media de los datos históricos del IUV que emplear el modelo con el que se calculó la métrica. La figura 5.13 muestra el decaimiento de la métrica R^2 en el tiempo obtenido con los cuatro modelos seleccionados para cada estación de monitoreo. En esta figura se observa con mayor detalle que el modelo MLP supera en rendimiento a los demás modelos, a excepción de los modelos LSTM, LSTM Attention y CNN LSTM en la estación Sureste. Considerando únicamente a los modelos seleccionados, la mediana de la métrica R^2 decae entre 13.5 % y 69.1 % para el séptimo día consecutivo con respecto al primer día de pronóstico.

Estación	Modelo	R2	MAE	RMSE
Noreste	MLP	0.740	0.076	0.128
	Attention	0.245	0.146	0.255
	Arclight	0.238	0.246	0.407
	Residual Attention	0.576	0.095	0.174
	Residual MLP	0.569	0.102	0.179
	Transformer	0.503	0.107	0.195
	LSTM	0.086	0.166	0.290
	CNN LSTM	0.492	0.092	0.162
	LSTM Attention	0.208	0.162	0.267
Noroeste	CNN LSTM Attention	-0.757	0.242	0.403
	MLP	0.712	0.081	0.138
	Attention	0.717	0.076	0.135
	Arclight	0.771	0.100	0.180
	Residual Attention	0.776	0.072	0.127
	Residual MLP	0.647	0.083	0.154
	Transformer	0.737	0.075	0.136
	LSTM	0.584	0.100	0.170
	CNN LSTM	0.605	0.094	0.167
Sureste	LSTM Attention	0.543	0.107	0.180
	CNN LSTM Attention	0.757	0.083	0.142
	MLP	0.712	0.046	0.082
	Attention	0.623	0.054	0.097
	Arclight	0.652	0.047	0.078
	Residual Attention	0.620	0.051	0.097
	Residual MLP	0.738	0.044	0.080
	Transformer	0.736	0.045	0.081
	LSTM	0.730	0.046	0.083
Suroeste	CNN LSTM	0.632	0.058	0.098
	LSTM Attention	0.581	0.061	0.106
	CNN LSTM Attention	0.677	0.05	0.090
	MLP	0.746	0.073	0.130
	Attention	0.582	0.099	0.181
	Arclight	0.696	0.121	0.205
	Residual Attention	0.497	0.107	0.191
	Residual MLP	0.153	0.167	0.277
	Transformer	0.469	0.109	0.195

Tabla 5.8: Medianas de las métricas diarias de los modelos con descomposición VMD en el pronóstico del IUV para el séptimo día consecutivo. En negritas se encuentra resaltada el mejor rendimiento para cada métrica en cada estación.

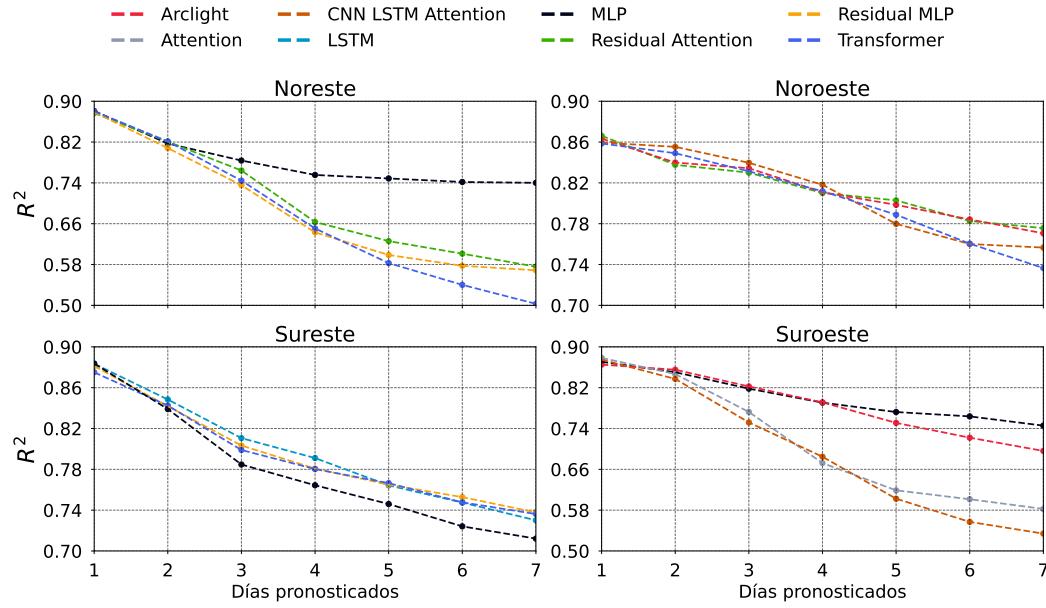


Figura 5.13: Decaimiento de la mediana diaria de la métrica R^2 a lo largo de la semana obtenida de los cuatro modelos que realizan la descomposición VMD con mayor promedio semanal.

Para realizar un análisis acerca de la variabilidad del rendimiento de los modelos seleccionados, se obtuvieron las distribuciones de la métrica R^2 para el séptimo día consecutivo. La figura 5.14 muestra las distribuciones de los cuatro modelos seleccionados.

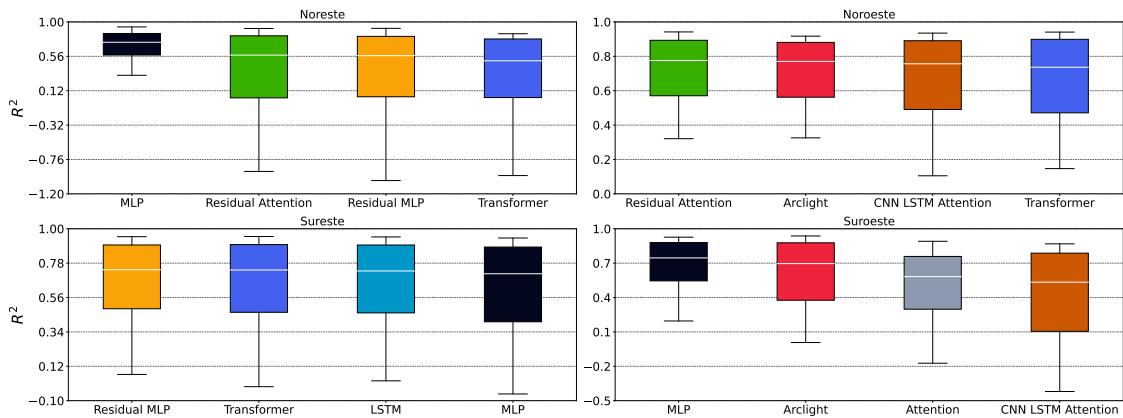


Figura 5.14: Distribuciones de la métrica R^2 obtenidas en el séptimo día para los cuatro modelos que realizan la descomposición VMD con mayor promedio semanal de esta métrica. El modelo con la mayor promedio semanal encuentra de lado izquierdo y a la derecha el modelo con menor promedio de los cuatro modelos seleccionados. La línea blanca representa la mediana, el límite inferior y superior de las cajas son el primer y tercer cuartil, por otro lado, el límite inferior y superior son el 10 y 90 percentiles respectivamente.

El modelo con mayor promedio se encuentra de lado izquierdo y de lado derecho se encuentra el modelo con menor promedio de los cuatro modelos seleccionados. La figura 5.15 muestra el pronóstico del IUV para la semana siguiente usando los cuatro modelos seleccionados en cada estación de monitoreo. Se observa que los modelos tienden a subestimar el IUV en comparación al IUV estimado a partir del *GHI*. Además, se visualiza que los modelos no son capaces de pronosticar las disminuciones del IUV que son producidas por la presencia de nubes. En cambio, cuando se presentan días con cielo despejado de manera consecutiva, el pronóstico de los modelos son cercanos al IUV de los datos de prueba.

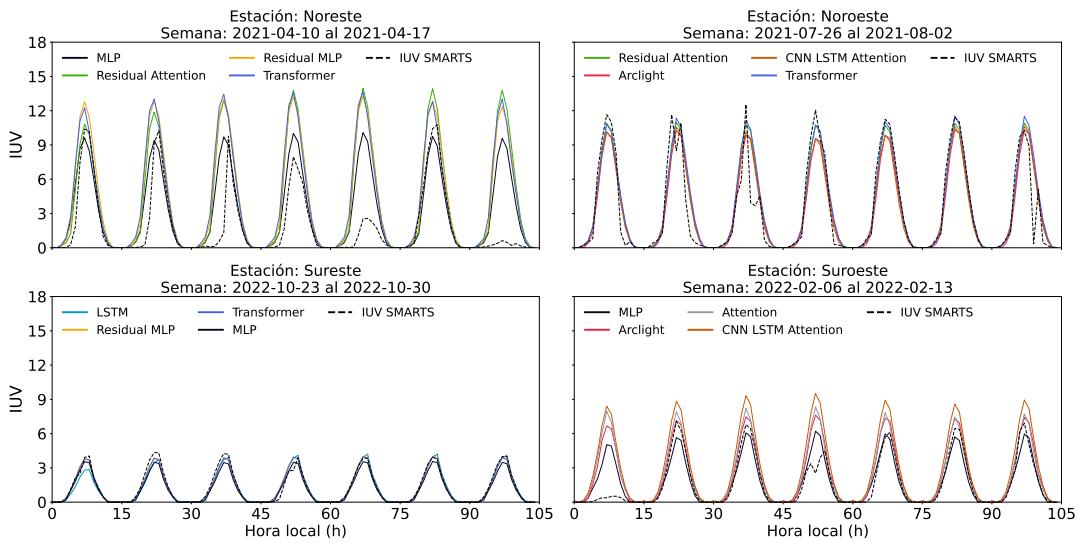


Figura 5.15: Pronóstico semanal del IUV obtenida con los cuatro modelos con mayor promedio semanal de la mediana diaria de la métrica R^2 utilizando la descomposición VMD de cada estación de monitoreo seleccionada. La línea punteada es la estimación del IUV a partir de las mediciones de *GHI*.

5.5.3. Comparación EWT y VMD

Al comparar detalladamente las tablas 5.7 y 5.8, se observa una diferencia en el desempeño de los modelos que utilizan la descomposición EWT en comparación con los modelos que emplean la descomposición VMD al realizar pronósticos de IUV para la semana siguiente. Esto se evidencia a través de valores más altos en las métricas de evaluación, como el *RMSE* y *MAE* y valores más bajos en la métrica R^2 . Estos resultados indican que los modelos basados en la descomposición EWT logran una mejor capacidad de predicción del IUV para la semana siguiente en comparación con los modelos basados en la descomposición VMD. Además, es importante destacar que los modelos basados en la descomposición EWT requieren un número menor de parámetros en comparación con los modelos que emplean la descomposición VMD. Esta diferencia en la complejidad del

modelo puede ser una consideración relevante tanto desde una perspectiva computacional como el uso de recursos. La tabla 5.9 muestra la mediana de las métricas diarias obtenidas en el séptimo día a partir de los mejores modelos para cada método de descomposición en las estaciones de monitoreo.

Estación	Descomposición	Modelo	R ²	MAE	RMSE
Noreste	EWT	Arclight	0.760	0.075	0.134
	EWT	Residual Attention	0.760	0.082	0.132
	VMD	MLP	0.740	0.076	0.128
Noroeste	EWT	Residual MLP	0.771	0.076	0.129
	VMD	Residual Attention	0.776	0.072	0.127
Sureste	EWT	MLP	0.738	0.048	0.079
	EWT	Arclight	0.751	0.047	0.080
	EWT	Residual Attention	0.727	0.045	0.081
	EWT	CNN LSTM	0.738	0.046	0.079
	VMD	Arclight	0.652	0.047	0.078
	VMD	Residual MLP	0.738	0.044	0.080
Suroeste	EWT	MLP	0.763	0.072	0.125
	EWT	Arclight	0.718	0.071	0.138
	EWT	CNN LSTM Atention	0.766	0.074	0.126
	VMD	MLP	0.746	0.073	0.130

Tabla 5.9: Mediana de las métricas diarias obtenidas en la tarea del pronóstico horario del IUV para el séptimo día con los modelos con mejor rendimiento en cada método de descomposición en las estaciones de monitoreo. En negritas se resalta la mejor métrica para la estación correspondiente.

Conclusiones

En esta Tesis se presentó una propuesta basada en redes neuronales y métodos de descomposición de señales para el pronóstico del IUV horario en el siguiente día. Los modelos propuestos emplean arquitecturas de redes recurrentes, redes convolucionales, operaciones residuales y mecanismos de atención. Estos modelos tienen como datos de entrada el *GHI*, *GHI₀* y el factor de nubes del día previo al pronosticado. Posteriormente se realizan una descomposición del *GHI* aplicando el método EWT o VMD. Para evaluar el rendimiento de los modelos se obtuvieron las métricas del Error Absoluto Medio (*MAE*), la Raíz del Error Cuadrático Medio (*RMSE*) y el coeficiente de determinación (*R²*). Estas métricas fueron calculadas de manera diaria para realizar un análisis acerca de la distribución y variabilidad del rendimiento de los modelos. En seguida se mencionan las conclusiones que se obtuvieron del análisis realizado:

- En el caso del pronóstico horario del IUV para el siguiente día y para la siguiente semana, la diferencia de rendimiento entre los modelos que utilizan la descomposición EWT y los modelos que emplean el método VMD no es significativa. Sin embargo, los modelos basados en la descomposición EWT obtiene un mayor número de rendimientos superiores al 0.6 en la métrica *R²* en comparación al método VMD, lo que representa un mejor resultado en el pronóstico para la siguiente semana.
- El modelo Arclight obtiene los mejores rendimientos y resultados estables en el pronóstico para el día siguiente y la siguiente semana. Además, debido al bajo número de parámetros y tiempo de ejecución que requiere este modelo, presenta una gran ventaja en términos de su aplicabilidad en dispositivos electrónicos con bajos recursos computacionales. Es por ello que es preferible implementar el modelo Arclight utilizando la descomposición EWT. El modelo MLP puede utilizarse como alternativa, ya que obtiene rendimientos buenos y estables, además, el tiempo de ejecución de este modelo es el más rápido. Una desventaja del modelo MLP es su alto número de parámetros.
- Los modelos analizados obtienen un buen rendimiento cuando se presenta una condición de cielo despejado de manera consecutiva, en otros casos, los modelos tienden a subestimar el IUV puesto que no son capaces de pronosticar las disminuciones en la intensidad del IUV debido a la presencia de nubes.

- Comparando los rendimientos de los modelos que utilizan la descomposición EWT y VMD con modelos que no emplean métodos de descomposición, se obtiene que los rendimientos para el pronóstico del siguiente día son semejantes en las tres metodologías. Sin embargo, los modelos que utilizan alguna descomposición obtienen un mejor rendimiento en el pronóstico de la siguiente semana en comparación a los modelos que no la realizan. Las tablas A.1 y A.2 muestra la mediana de las métricas R^2 diarias obtenidas por los modelos en cada metodología para el pronóstico del IUV del siguiente día y la siguiente semana respectivamente.

Trabajo a futuro

- Entrenar los modelos propuestos para cada temporada del año. Esto para que los modelos puedan destacar patrones especializados de cada temporada y obtener mejores rendimientos.
- Desarrollar una versión de los modelos donde la salida esperada sean las descomposiciones del IUV obtenida por el método EWT. Esto con el fin de ayudar al modelo a tener una mayor variabilidad en los valores numéricos de sus parámetros.
- Abordar la aumentación de datos siguiendo un modelo basado en autoencoders variacionales (Goodfellow y cols., 2014).
- Realizar comparaciones con otros modelos de pronóstico que utilizan datos satelitales como el Tropospheric Emission Monitoring Internet Service (TEMIS) (The European Space Agency, 2023).

Referencias

- Acosta, L. R. (1997). Diseño de la red uv del valle de méxico.
- Ahlberg, J. H., Nilson, E. N., y Walsh, J. L. (2016). *The theory of splines and their applications: Mathematics in science and engineering: A series of monographs and textbooks, vol. 38* (Vol. 38). Elsevier.
- Allen, R. G., y cols. (Eds.). (2005). *The ASCE standardized reference evapotranspiration equation*. American Society of Civil Engineers. doi: 10.1061/9780784408056
- Alzahrani, A., Shamsi, P., Dagli, C., y Ferdowsi, M. (2017). Solar irradiance forecasting using deep neural networks. *Procedia Computer Science*, 114, 304–313. doi: 10.1016/j.procs.2017.09.045
- Anh, N.-L. H., Ravinesh, C. D., Mumtaz, A., Shahab, A., y Nawin, R. (2021, sep). Novel short-term solar radiation hybrid model: Long short-term memory network integrated with robust local mean decomposition. *Applied Energy*, 298, 117193. doi: 10.1016/j.apenergy.2021.117193
- Atmosphere US Standard. (1976). *Us standard atmosphere*. National Oceanic and Atmospheric Administration.
- Avilov, O., Rimbert, S., Popov, A., y Bougrain, L. (2020, jul). Deep learning techniques to improve intraoperative awareness detection from electroencephalographic signals. En *2020 42nd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*. IEEE. doi: 10.1109/embc44109.2020.9176228
- Ba, J. L., Kiros, J. R., y Hinton, G. E. (2016). *Layer normalization*. arXiv. doi: 10.48550/ARXIV.1607.06450
- Bengio, Y., Simard, P., y Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157-166. doi: 10.1109/72.279181
- Black, J. N., Bonython, C. W., y Prescott, J. A. (1954, apr). Solar radiation and the duration of sunshine. *Q.J Royal Met. Soc.*, 80(344), 231–235. doi: 10.1002/qj.49708034411
- Boland, J. (2008). Time series modelling of solar radiation. En V. Badescu (Ed.), *Modeling solar radiation at the earth's surface: Recent advances* (pp. 283–312). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-77455-6_11
- Box, G. E., Jenkins, G. M., Reinsel, G. C., y Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Cannizzaro, D., Aliberti, A., Bottaccioli, L., Macii, E., Acquaviva, A., y Patti, E. (2021, nov). Solar radiation forecasting based on convolutional neural network and en-

- semble learning. *Expert Systems with Applications*, 181, 115167. doi: 10.1016/j.eswa.2021.115167
- Carvalho, V. R., Moraes, M. F., Braga, A. P., y Mendes, E. M. (2020, sep). Evaluating five different adaptive decomposition methods for EEG signal seizure detection and classification. *Biomedical Signal Processing and Control*, 62, 102073. doi: 10.1016/j.bspc.2020.102073
- Chicco, D., Warrens, M. J., y Jurman, G. (2021, jul). The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. doi: 10.7717/peerj-cs.623
- Christoffersen, P., y Jacobs, K. (2004, may). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2), 291–318. doi: 10.1016/j.jfineco.2003.02.001
- CIE. (2014). *Rationalizing nomenclature for UV doses and effects on humans: Joint publication of CIE and WMO (world meteorological organization)*.
- Collobert, R., y Weston, J. (2008). A unified architecture for natural language processing. En *Proceedings of the 25th international conference on machine learning*. ACM Press. doi: 10.1145/1390156.1390177
- CONAPO-INEGI-SEDATU. (2018). *Delimitación de las zonas metropolitanas de México 2015*.
- Cornaro, C., Pierro, M., y Bucci, F. (2015, jan). Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast. *Solar Energy*, 111, 297–312. doi: 10.1016/j.solener.2014.10.036
- Cunningham, P., Cord, M., y Delany, S. J. (2008). Supervised learning. En M. Cord y P. Cunningham (Eds.), *Machine learning techniques for multimedia: Case studies on organization and retrieval* (pp. 21–49). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-75171-7_2
- Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics. doi: 10.1137/1.9781611970104
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., y Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. En *2009 ieee conference on computer vision and pattern recognition* (p. 248-255). doi: 10.1109/CVPR.2009.5206848
- D’Orazio, J., Jarrett, S., Amaro-Ortiz, A., y Scott, T. (2013). Uv radiation and the skin. *International Journal of Molecular Sciences*, 14(6), 12222–12248. doi: 10.3390/ijms140612222
- Dragomiretskiy, K., y Zosso, D. (2014). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531-544. doi: 10.1109/TSP.2013.2288675

- Gamboa, J. C. B. (2017). *Deep learning for time-series analysis*.
- Gehring, J., Auli, M., Grangier, D., Yarats, D., y Dauphin, Y. N. (2017). *Convolutional sequence to sequence learning*. arXiv. doi: 10.48550/ARXIV.1705.03122
- Gilaberte, Y., Aguilera, J., Carrascosa, J., Figueroa, F., de Gabriel, J. R., y Nagore, E. (2011, octubre). La vitamina d: evidencias y controversias. *Actas Dermosifiliográficas*, 102(8), 572-588.
- Gilles, J. (2013). Empirical wavelet transform. *IEEE Transactions on Signal Processing*, 61(16), 3999-4010. doi: 10.1109/TSP.2013.2265222
- Gilles, J. (2023). *Empirical wavelet transforms*. Descargado de <https://www.mathworks.com/matlabcentral/fileexchange/42141-empirical-wavelet-transforms>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). *Generative adversarial networks*.
- Gueymard, C. (2009). Simple model of the atmospheric radiative transfer of sunshine (smarts). *NREL, smarts*, 2(5).
- Han, Z., Zhao, J., Leung, H., Ma, K. F., y Wang, W. (2021). A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6), 7833-7848. doi: 10.1109/JSEN.2019.2923982
- Hastie, T., Tibshirani, R., Friedman, J. H., y Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2). Springer.
- He, K., Zhang, X., Ren, S., y Sun, J. (2015). Deep residual learning for image recognition. arXiv. doi: 10.48550/ARXIV.1512.03385
- Heckman, C. J., Liang, K., y Riley, M. (2019, jun). Awareness, understanding, use, and impact of the uv index: A systematic review of over two decades of international research. *Preventive Medicine*, 123, 71–83. doi: 10.1016/j.ypmed.2019.03.004
- Hochreiter, S., y Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. doi: 10.1162/neco.1997.9.8.1735
- Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., ... Smirnov, A. (1998, oct). Aeronet- a federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, 66(1), 1–16. doi: 10.1016/s0034-4257(98)00031-5
- Honeyman, J. (2002). Efectos de las radiaciones ultravioleta en la piel. *Rev. Peru. Dermatol*, 12(2).
- Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., y Zeng, Z. (2021, apr). Solar radiation prediction using different machine learning algorithms and implications for extreme climate events. *Front. Earth Sci.*, 9. doi: 10.3389/feart.2021.596860
- Husein, M., y Chung, I.-Y. (2019). Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach.

- Energies*, 12(10). doi: 10.3390/en12101856
- INEGI. (2021). *Panorama sociodemográfico de nuevo león. censo de población y vivienda 2020*.
- Inman, R. H., Pedro, H. T., y Coimbra, C. F. (2013, dec). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6), 535–576. doi: 10.1016/j.pecs.2013.06.002
- Ipiña, A., López-Padilla, G., Retama, A., Piacentini, R. D., y Madronich, S. (2021, aug). Ultraviolet radiation environment of a tropical megacity in transition: Mexico city 2000–2019. *Environmental Science & Technology*, 55(16), 10946–10956. doi: 10.1021/acs.est.0c08515
- Iqbal, M. (1983). *An introduction to solar radiation*.
- Jiménez, J. A. M., Moya, B. C., y Jiménez, M. T. M. (2015). Factores nutricionales en la prevención de la osteoporosis. *Nutrición Hospitalaria*, 32(1), 49–55.
- Jurado-Santa Cruz, F., Medina-Bojórquez, A., Gutiérrez-Vidrio, R. M., y Ruiz-Rosillo, J. M. (2011). Prevalencia del cáncer de piel en tres ciudades de México. *Revista Médica del Instituto Mexicano del Seguro Social*, 49(3), 253–258.
- Kapoor, P., y Bedi, S. S. (2013, dec). Weather forecasting using sliding window algorithm. *ISRN Signal Processing Signal Processing*, 2013, 1–5. doi: 10.1155/2013/156540
- Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., ... Zhang, W. (2019, dec). A comparative study on transformer vs RNN in speech applications. En *2019 IEEE automatic speech recognition and understanding workshop (ASRU)*. IEEE. doi: 10.1109/asru46091.2019.9003750
- Krzyścin, J. W., y Puchalski, S. (1998, jul). Aerosol impact on the surface uv radiation from the ground-based measurements taken at belsk. *J. Geophys. Res.*, 103(D13), 16175–16181. doi: 10.1029/98jd00899
- Lara del Valle, D. R. (2010). *Hábitos y conocimientos de fotoprotección de madres de familia y profesores de alumnos de primer año de escuelas primarias oficiales de la delegación benito juárez del distrito federal* (Tesis Doctoral no publicada). Universidad Nacional Autónoma de México.
- Li, H., Lian, Y., Wang, X., Ma, W., y Zhao, L. (2011, mar). Solar constant values for estimating solar radiation. *Energy*, 36(3), 1785–1789. doi: 10.1016/j.energy.2010.12.050
- Masip, T. C., Morell, N. A., y Serra, J. D. (2008). Nuevas recomendaciones diarias de ingesta de calcio y vitamina d: prevención del raquitismo nutricional. *Acta Pediatr Esp*, 66(5), 233–236.
- McCullough, E. C. (1977). *Optical radiations in medicine: A survey of uses, measurement and sources* (Inf. Téc.). doi: 10.37206/2

- Miljanovic, M. (2012, 02). Comparative analysis of recurrent and finite impulse response neural networks in time series prediction. *Indian Journal of Computer Science and Engineering*, 3.
- Modenese, A., Bisegna, F., Borra, M., Grandi, C., Gugliermetti, F., Militello, A., y Gobba, F. (2016, oct). Outdoor work and solar radiation exposure: Evaluation method for epidemiological studies. *Med Pr*, 67(5), 577–587. doi: 10.13075/mp.5893.00461
- NASA EOS/Aura Validation Data Center (AVDC) - Correlative data, Field of View Predictions, Data Subsets, GEOMS, DCIO. (2023). Descargado de https://avdc.gsfc.nasa.gov/pub/most_popular/overpass/OMI/
- National Renewable Energy Laboratory. (2006). *Smarts: Simple model of the atmospheric radiative transfer of sunshine*. Descargado de <https://www.nrel.gov/grid/solar-resource/smarts.html>
- Neale, R., Lucas, R., Byrne, S., Hollestein, L., Rhodes, L., Yazar, S., ... Olsen, C. (2023). The effects of exposure to solar radiation on human health. *Photochemical & Photobiological Sciences*, 1–37.
- Pang, Z., Niu, F., y O'Neill, Z. (2020, aug). Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons. *Renewable Energy*, 156, 279–289. doi: 10.1016/j.renene.2020.04.042
- Paulescu, M., Paulescu, E., Gravila, P., y Badescu, V. (2013). *Weather modeling and forecasting of pv systems operation*. Springer London. doi: 10.1007/978-1-4471-4649-0
- Phyo, P.-P., Byun, Y.-C., y Park, N. (2022, jan). Short-term energy forecasting using machine-learning-based ensemble voting regression. *Symmetry*, 14(1), 160. doi: 10.3390/sym14010160
- Pounds, J. A., Fogden, M. P. L., y Campbell, J. H. (1999, apr). Biological response to climate change on a tropical mountain. *Nature*, 398(6728), 611–615. doi: 10.1038/19297
- Prechelt, L. (1998, jun). Automatic early stopping using cross validation: quantifying the criteria. *Neural Networks*, 11(4), 761–767. doi: 10.1016/s0893-6080(98)00010-0
- Retama, A., Ramos-Cerón, M., Rivera-Hernández, O., Allen, G., y Velasco, E. (2022). Aerosol optical properties and brown carbon in mexico city. *Environ. Sci.: Atmos.*, 2(3), 315–334. doi: 10.1039/d2ea00006g
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. doi: 10.1037/h0042519
- Rosso, S., Zanetti, R., Martinez, C., Tormo, M., Schraub, S., Sancho-Garnier, H., ... Wechsler, J. (1996, jun). The multicentre south european study 'helios'. II: Different

- sun exposure patterns in the aetiology of basal cell and squamous cell carcinomas of the skin. *Br J Cancer*, 73(11), 1447–1454. doi: 10.1038/bjc.1996.275
- Secretaría del Medio Ambiente de la Ciudad de México. (2023). Descargado de <http://www.aire.cdmx.gob.mx/default.php?opc='aqBjnmM='>
- Sharda, S., Singh, M., y Sharma, K. (2021). Rsam: Robust self-attention based multi-horizon model for solar irradiance forecasting. *IEEE Transactions on Sustainable Energy*, 12(2), 1394-1405. doi: 10.1109/TSTE.2020.3046098
- Sherstinsky, A. (2020, mar). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404, 132306. doi: 10.1016/j.physd.2019.132306
- Shettle, E. P., y Fenn, R. W. (1979). *Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties* (n.º 676). Optical Physics Division, Air Force Geophysics Laboratory.
- Sistema Integral de Monitoreo Ambiental. (2015). *Normas empleadas en las mediciones del sima*. Descargado de http://aire.nl.gob.mx/nor_metodos.html
- Solano, E. S., Dehghanian, P., y Affonso, C. M. (2022, sep). Solar radiation forecasting using machine learning and ensemble feature selection. , 15(19), 7049. doi: 10.3390/en15197049
- Solarmeter. (2023). *Solarmeter®model 5.0 standard uva+b meter*. Descargado de <https://www.solarmeter.com/wp-content/uploads/Solarmeter-Model-5.0-UVA-UVB-01252023.pdf>
- Stern, R. S. (1986, may). Risk reduction for nonmelanoma skin cancer with childhood sunscreen use. *Arch Dermatol*, 122(5), 537. doi: 10.1001/archderm.1986.01660170067022
- Still, C. J., Foster, P. N., y Schneider, S. H. (1999, apr). Simulating the effects of climate change on tropical montane cloud forests. *Nature*, 398(6728), 608–610. doi: 10.1038/19293
- Sutskever, I., Vinyals, O., y Le, Q. V. (2014). Sequence to sequence learning with neural networks. En Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, y K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27). Curran Associates, Inc.
- The European Space Agency. (2023). *Tropospheric emission monitoring internet service*. Descargado de <https://www.temis.nl/>
- TrentUniv, Televisa, y Instituto Tecnológico y de Estudios Superiores de Monterrey. (1993). *Collaboration agreement for environmental information and monitoring system*.
- Tsanekidis, A., Passalis, N., Tefas, A., Kannainen, J., Gabbouj, M., y Iosifidis, A. (2017,

- jul). Forecasting stock prices from the limit order book using convolutional neural networks. En *2017 IEEE 19th conference on business informatics (CBI)*. IEEE. doi: 10.1109/cbi.2017.23
- van den Oord, A., Dieleman, S., y Schrauwen, B. (2013). Deep content-based music recommendation. En C. Burges, L. Bottou, M. Welling, Z. Ghahramani, y K. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 26). Curran Associates, Inc.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017a). *Attention is all you need*. arXiv. doi: 10.48550/ARXIV.1706.03762
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017b). Attention is all you need. *Advances in neural information processing systems*, 30.
- Viatte, C., Abeed, R., Yamanouchi, S., Porter, W. C., Safieddine, S., Damme, M. V., ... Clerbaux, C. (2022, oct). NH_3 spatiotemporal variability over paris, mexico city, and toronto, and its link to $\text{PM}_{2.5}$ during pollution events. *Atmos. Chem. Phys.*, 22(19), 12907–12922. doi: 10.5194/acp-22-12907-2022
- Wang, X., Wang, Y., Yuan, P., Wang, L., y Cheng, D. (2021). An adaptive daily runoff forecast model using vmd-lstm-psp hybrid approach. *Hydrological Sciences Journal*, 66(9), 1488–1502. doi: 10.1080/02626667.2021.1937631
- Wang, Y., Guo, P., Ma, N., y Liu, G. (2022, dec). Robust wavelet transform neural-network-based short-term load forecasting for power distribution networks. *Sustainability*, 15(1), 296. doi: 10.3390/su15010296
- Webb, A. R., y Holick, M. F. (1988, jul). The role of sunlight in the cutaneous production of vitamin d3. *Annu. Rev. Nutr.*, 8(1), 375–399. doi: 10.1146/annurev.nu.08.070188.002111
- WHO. (2016). *Radiation: Ultraviolet (uv) radiation*. Descargado de [https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-\(uv\)](https://www.who.int/news-room/questions-and-answers/item/radiation-ultraviolet-(uv))
- WHO, WMO, UNEP, y ICNIRP. (2002). *Global solar uv index : a practical guide*. World Health Organization.
- Willmott, C. J., y Matsuura, K. (2006, jan). On the use of dimensioned measures of error to evaluate the performance of spatial interpolators. *International Journal of Geographical Information Science*, 20(1), 89–102. doi: 10.1080/13658810500286976
- Ying, X. (2019, feb). An overview of overfitting and its solutions. *J. Phys.: Conf. Ser.*, 1168, 022022. doi: 10.1088/1742-6596/1168/2/022022
- Yu, Y., Cao, J., y Zhu, J. (2019). An LSTM short-term solar irradiance forecasting under complicated weather conditions. *IEEE Access*, 7, 145651–145666. doi: 10.1109/

- access.2019.2946057
- Zang, H., Liu, L., Sun, L., Cheng, L., Wei, Z., y Sun, G. (2020, nov). Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renewable Energy*, 160, 26–41. doi: 10.1016/j.renene.2020.05.150
- Zhang, J., He, T., Sra, S., y Jadbabaie, A. (2020). *Why gradient clipping accelerates training: A theoretical justification for adaptivity.*
- Zhou, Y., Li, Y., Wang, D., y Liu, Y. (2023, jan). A multi-step ahead global solar radiation prediction method using an attention-based transformer model with an interpretable mechanism. *International Journal of Hydrogen Energy*. doi: 10.1016/j.ijhydene.2023.01.068
- Zosso, D. (2023). *Variational mode decomposition.* Descargado de <https://www.mathworks.com/matlabcentral/fileexchange/44765-variational-mode-decomposition>

Anexo

Estación	Modelo	EWT	VMD	Sin descomposición
Noreste	MLP	0.885	0.881	0.877
	Attention	0.884	0.885	0.882
	Arclight	0.877	0.885	0.888
	Residual Attention	0.878	0.877	0.880
	Residual MLP	0.878	0.879	0.877
	Transformer	0.874	0.880	0.873
	LSTM	0.878	0.886	0.883
	CNN LSTM	0.873	0.875	0.875
	LSTM Attention	0.878	0.876	0.877
	CNN LSTM Attention	0.863	0.873	0.872
Noroeste	MLP	0.868	0.867	0.875
	Attention	0.869	0.864	0.871
	Arclight	0.876	0.863	0.859
	Residual Attention	0.863	0.866	0.869
	Residual MLP	0.890	0.882	0.877
	Transformer	0.861	0.859	0.835
	LSTM	0.856	0.867	0.861
	CNN LSTM	0.863	0.866	0.863
	LSTM Attention	0.855	0.863	0.862
	CNN LSTM Attention	0.864	0.859	0.862
Sureste	MLP	0.883	0.884	0.887
	Attention	0.882	0.882	0.884
	Arclight	0.883	0.883	0.877
	Residual Attention	0.879	0.877	0.874
	Residual MLP	0.879	0.881	0.875
	Transformer	0.867	0.875	0.864
	LSTM	0.878	0.883	0.879
	CNN LSTM	0.879	0.870	0.877
	LSTM Attention	0.875	0.877	0.880
	CNN LSTM Attention	0.873	0.88	0.875
Suroeste	MLP	0.876	0.871	0.887
	Attention	0.889	0.878	0.880
	Arclight	0.880	0.865	0.871
	Residual Attention	0.866	0.875	0.871
	Residual MLP	0.882	0.881	0.883
	Transformer	0.862	0.858	0.874
	LSTM	0.871	0.874	0.878
	CNN LSTM	0.880	0.873	0.879
	LSTM Attention	0.867	0.863	0.874
	CNN LSTM Attention	0.866	0.875	0.875

Tabla A.1: Mediana de las métricas R^2 diarias del pronóstico del IUV para el siguiente día de los modelos que utilizan la descomposición EWT y VMD, y los modelos que no realizan descomposiciones al GHI . En negritas se muestra el mejor rendimiento obtenido en cada metodología y estación de monitoreo.

Estación	Modelo	EWT	VMD	Sin descomposición
Noreste	MLP	0.683	0.740	0.725
	Attention	0.123	0.245	0.704
	Arclight	0.760	0.238	0.757
	Residual Attention	0.760	0.576	0.732
	Residual MLP	0.422	0.569	0.753
	Transformer	0.681	0.503	0.738
	LSTM	0.204	0.086	0.724
	CNN LSTM	0.607	0.492	0.741
	LSTM Attention	0.188	0.208	0.734
	CNN LSTM Attention	0.304	-0.757	0.725
Noroeste	MLP	0.709	0.712	0.651
	Attention	0.459	0.717	0.603
	Arclight	0.738	0.771	0.668
	Residual Attention	0.440	0.776	0.672
	Residual MLP	0.771	0.647	0.698
	Transformer	0.545	0.737	0.648
	LSTM	0.526	0.584	0.651
	CNN LSTM	0.695	0.605	0.681
	LSTM Attention	0.683	0.543	0.667
	CNN LSTM Attention	0.745	0.757	0.671
Sureste	MLP	0.738	0.712	0.662
	Attention	0.629	0.623	0.635
	Arclight	0.751	0.652	0.685
	Residual Attention	0.727	0.620	0.689
	Residual MLP	0.614	0.738	0.704
	Transformer	0.705	0.736	0.689
	LSTM	0.688	0.730	0.691
	CNN LSTM	0.738	0.632	0.659
	LSTM Attention	0.646	0.581	0.682
	CNN LSTM Attention	0.723	0.677	0.670
Suroeste	MLP	0.763	0.746	0.680
	Attention	-0.215	0.582	0.664
	Arclight	0.718	0.696	0.725
	Residual Attention	0.612	0.497	0.712
	Residual MLP	0.201	0.153	0.735
	Transformer	0.546	0.469	0.712
	LSTM	0.511	0.406	0.688
	CNN LSTM	0.677	0.452	0.699
	LSTM Attention	0.727	0.508	0.697
	CNN LSTM Attention	0.766	0.534	0.680

Tabla A.2: Mediana de las métricas R^2 diarias del pronóstico del IUV para la siguiente semana de los modelos que utilizan la descomposición EWT y VMD, y los modelos que no realizan descomposiciones al *GHI*. En negritas se muestra el mejor rendimiento obtenido en cada metodología y estación de monitoreo.