

UNIVERSIDAD NACIONAL DEL LITORAL



DOCTORADO EN INGENIERÍA. MENCIÓN INTELIGENCIA
COMPUTACIONAL, SEÑALES Y SISTEMAS

MODELADO DE DEMANDA DE ENERGÍA ELÉCTRICA MEDIANTE REDES NEURONALES ARTIFICIALES A PARTIR DE INFORMACIÓN METEOROLÓGICA

Mg. Mariela Noelia Uhrig

FICH

FACULTAD DE INGENIERÍA Y CIENCIAS HÍDRICAS

CENTRO DE ESTUDIOS DE VARIABILIDAD Y CAMBIO CLIMÁTICO

sinc(i) FICH-UNL/CONICET

INSTITUTO DE INVESTIGACIÓN EN SEÑALES, SISTEMAS E
INTELIGENCIA COMPUTACIONAL

CICYTTP (CONICET-Prov. ER-UADER)

CENTRO DE INVESTIGACIÓN CIENTÍFICA Y DE TRANSFERENCIA
TECNOLÓGICA A LA PRODUCCIÓN

Doctorado en Ingeniería

Mención Inteligencia Computacional Señales y Sistemas

Título de la obra:

Modelado de demanda de energía eléctrica mediante redes neuronales artificiales a partir de información meteorológica

Autor: Mariela Noelia Uhrig

Lugar: Santa Fe, Argentina

Palabras Claves:

Redes neuronales artificiales

Predicción de demanda eléctrica

Aprendizaje profundo

Redes Neuronales Recurrentes



UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Ingeniería y Ciencias Hídricas

Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional

MODELADO DE DEMANDA DE ENERGÍA ELÉCTRICA MEDIANTE REDES NEURONALES ARTIFICIALES A PARTIR DE INFORMACIÓN METEOROLÓGICA

Mg. Mariela Noelia Uhrig

Lugar de Trabajo:

sinc(*i*)

Instituto de Señales, Sistemas e Inteligencia Computacional

Facultad de Ingeniería y Ciencias Hídricas

Universidad Nacional del Litoral

Director:

Dr. Omar Vicente Müller

CEVARCAM-UNL-CONICET

Co-director:

Dr. Leandro Daniel Vignolo

sinc(*i*)-UNL-CONICET

Jurado Evaluador:

Dr. Rodrigo Echeveste

sinc(*i*)-CONICET-UNL

Dr. Juan José Ruiz

CIMA-CONICET-UBA

Dra. Georgina Stegmayer

sinc(*i*)-CONICET-UNL



ACTA DE EVALUACIÓN DE TESIS DE DOCTORADO

En la sede de la Facultad de Ingeniería y Ciencias Hídricas de la Universidad Nacional del Litoral, a los veintiséis días del mes de agosto del año dos mil veintiséis, se reúnen en forma online sincrónica los miembros del Jurado designado para la evaluación de la Tesis de Doctorado en Ingeniería, Mención Inteligencia Computacional, Señales y Sistemas, titulada ***“Modelado de demanda de energía eléctrica mediante redes neuronales artificiales a partir de información meteorológica”***, desarrollada por la Mag. Mariela UHRIG, DNI: 29.795.109, bajo la dirección del Dr. Omar Müller y la codirección del Dr. Leandro Vignolo. Ellos son: el Dr. Rodrigo Echeveste, el Dr. Juan José Ruiz, y la Dra. Georgina Stegmayer.-----

La Presentación oral y la defensa de la Tesis se efectúan bajo la modalidad online sincrónica según lo establecido por Resolución CS N° 382/21.

Luego de escuchar la Defensa Pública y de evaluar la Tesis, el Jurado considera:

Que la temática desarrollada es relevante y el enfoque es innovador superando el estado del arte y aportando interpretabilidad a los resultados.

La presentación oral y escrita ha sido clara y completa. El volumen de trabajo es considerable con publicaciones ya realizadas y otras en curso.

La tesista ha respondido con solvencia y precisión las preguntas del Jurado demostrando un amplio manejo del tema.

Por ello, el Jurado aprueba la Tesis con calificación 10 (Diez) Sobresaliente.

Sin más, se da por finalizado el Acto Académico con la firma de los miembros del Jurado al pie de la presente. -----

Dr. Rodrigo Echeveste

Dr. Juan José Ruiz

Dra. Georgina Stegmayer



[Signature]
Dr. José Luis Macor
Director de Posgrado
FICH - UNL

Universidad Nacional del Litoral
Facultad de Ingeniería y
Ciencias Hídricas

Secretaría de Posgrado

Ciudad Universitaria
C.C. 217
Ruta Nacional N° 168 - Km. 472,4
(3000) Santa Fe
Tel: (54) (0342) 4575 229
Fax: (54) (0342) 4575 224
E-mail: posgrado@fich.unl.edu.ar

DECLARACIÓN LEGAL DEL AUTOR

Esta tesis ha sido remitida como parte de los requisitos para la obtención del grado académico de Magíster en Computación Aplicada a la Ciencia y la Ingeniería ante la Universidad Nacional del Litoral y ha sido depositada en la Biblioteca de la Facultad de Ingeniería y Ciencias Hídricas para que esté a disposición de sus lectores bajo las condiciones estipuladas por el Reglamento de la mencionada Biblioteca. Citaciones breves de esta tesis son permitidas sin la necesidad de un permiso especial, en la suposición de que la fuente sea correctamente citada. Solicitudes de permiso para la citación extendida o para la reproducción parcial o total de este manuscrito serán concebidos por el portador legal del derecho de propiedad intelectual de la obra.

Agradecimientos

Al escribir estas palabras, no puedo evitar recordar a todas las personas que han sido parte fundamental de este recorrido.

Quiero expresar mi más profundo agradecimiento al Dr. Omar Vicente Müller, mi director, por proponerme este desafiante tema de investigación, por su guía constante a lo largo de mi formación de posgrado, y especialmente por su empatía y comprensión en los momentos más difíciles. Su humanidad, que trasciende lo académico, ha sido una luz en este camino.

Al Dr. Leandro Daniel Vignolo, mi codirector, por sumarse al equipo con entusiasmo y disposición, comprendiendo las circunstancias particulares de este trayecto. Su apoyo incondicional y su acompañamiento profesional han sido esenciales para concretar este trabajo. Además de su rol como guía académico, puso a disposición su equipamiento para que los experimentos pudieran llevarse adelante sin demoras. A ambos les agradezco abrirme las puertas del CEVARCAM y del sinc(i), brindándome un espacio de aprendizaje y colaboración invaluable. Y por sobre todo, les agradezco la paciencia que tuvieron a lo largo de todo mi proceso doctoral.

Al Instituto sinc(i) de la Facultad de Ingeniería y Ciencias Hídricas de la Universidad Nacional del Litoral, por facilitar el acceso al servidor con GPU que hizo posible la ejecución de los experimentos de mayor demanda computacional de esta tesis.

A Gabriela Müller, quien con su perseverancia y generosidad me brindó una oportunidad única para avanzar en este camino, demostrando que creer en alguien puede marcar toda la diferencia. Su confianza en mí fue el impulso que necesitaba para continuar.

A Alexis, colega y amigo de la cátedra que compartimos en la Facultad de Ciencias y Tecnología, por poner a disposición su GPU personal para correr los experimentos que le enviaba. Gracias a eso, mientras mi equipo ejecutaba un experimento,

el suyo corría otro en paralelo, reduciendo sensiblemente los tiempos de la investigación. Ese gesto desinteresado, repetido cada vez que lo necesité, me ahorró un tiempo invaluable y se agradece profundamente.

A mi familia y amigos, gracias por estar siempre ahí, en cada momento.

A todos ustedes, mi más sincero y profundo agradecimiento.

Mariela Noelia Uhrig
Santa Fe, 2026

Índice general

Agradecimientos	IX
Resumen	XIX
Abstract	XXI
1. Introducción	1
1.1. Motivación	1
1.2. Antecedentes	4
1.3. Objetivos y Alcance	5
1.3.1. Objetivo General	5
1.3.2. Objetivos específicos	5
1.3.3. Alcance	6
1.4. Organización de la tesis	6
2. Formulación del problema	9
2.1. Introducción	9
2.2. Propuesta general del enfoque	10
2.3. Características de la demanda eléctrica	13
2.3.1. Demanda anual	13
2.3.2. Demanda mensual	13
2.3.3. Demanda semanal	15
2.3.4. Relación entre estado del tiempo y demanda eléctrica	17
2.4. Horizontes de predicción de la demanda	19
3. Marco teórico	21
3.1. Introducción	21
3.2. Modelos clásicos	23
3.2.1. Modelo ARIMA	23
3.2.2. Modelo SARIMAX	24
3.3. Modelos de aprendizaje profundo	25
3.3.1. Redes Neuronales Recurrentes	26

3.3.2. Modelos basados en mecanismos de atención	29
3.4. Desafíos actuales en la predicción de demanda	30
4. Modelos propuestos	33
4.1. Introducción	33
4.2. Punto de partida: modelo DA-SKIP	36
4.2.1. Arquitectura del modelo DA-SKIP	36
4.3. Modelo DA-LSTM	46
4.4. Modelo AMBER	48
4.4.1. Motivación para la incorporación de atención sobre variables meteorológicas	48
4.4.2. Arquitectura del modelo AMBER	49
5. Experimentos y resultados	53
5.1. Conjunto de datos utilizado	53
5.2. Metodología de evaluación y validación	56
5.2.1. Métricas para la evaluación del desempeño	56
5.2.2. Metodología de validación	58
5.3. Determinación de hiperparámetros	61
5.3.1. SARIMAX	61
5.3.2. LSTM	62
5.3.3. DA-SKIP	64
5.3.4. DA-LSTM	64
5.3.5. AMBER	66
5.4. Análisis comparativo y discusión de resultados	67
5.4.1. Modelos de aprendizaje profundo	68
5.4.2. Comparación global de modelos	73
5.4.3. Atención meteorológica	82
6. Conclusiones y trabajo futuro	85
6.1. Objetivos y contribuciones	85
6.1.1. Objetivo 1: Determinar las variables más influyentes en la demanda eléctrica de Entre Ríos	85
6.1.2. Objetivo 2: Proponer modelos de aprendizaje profundo para la predicción demanda energética	86
6.1.3. Objetivo 3: Analizar el efecto de distintas configuraciones temporales en el desempeño de los modelos	87
6.1.4. Objetivo 4: Evaluar y comparar el desempeño de los modelos y su capacidad explicativa	88
6.2. Trabajo futuro	89

6.3. Publicaciones y reconocimientos derivados de esta tesis	90
6.3.1. Premios y reconocimientos	90
6.3.2. Publicaciones científicas	90
6.3.3. Presentaciones en congresos científicos	90
6.4. Disponibilidad de datos y códigos	91
Bibliografía	92

Índice de figuras

2.1.	Esquema general del enfoque metodológico utilizado para la predicción de la demanda eléctrica, incluyendo las etapas de obtención de datos, preprocesamiento, construcción del dataset multivariado y desarrollo del modelo de aprendizaje profundo.	11
2.2.	Distribución espacial de la infraestructura eléctrica en la provincia de Entre Ríos. Se representan las 24 estaciones transformadoras ($\blacklozenge$) y las 65 subestaciones ($\circ$) operadas por ENERSA en su área de concesión.	12
2.3.	Evolución mensual y anual de demanda de energía provincial [MWh].	14
2.4.	Ciclo anual medio de la demanda eléctrica provincial para el período 2012–2023, mostrando la media mensual y el rango intercuartílico (IQR).	15
2.5.	Distribución diaria de la demanda eléctrica representada mediante diagramas de violín.	16
2.6.	Diagramas de dispersión de variables meteorológicas diarias frente a la demanda eléctrica diaria, incluyendo los coeficientes de correlación lineal (r).	17
3.1.	Evolución de los enfoques de modelado en la predicción de la demanda eléctrica.	22
3.2.	Estructura interna de una unidad LSTM, mostrando el flujo de información a través del estado de celda y las compuertas de olvido, entrada y salida.	27
4.1.	Evolución de las arquitecturas de aprendizaje profundo.	34
4.2.	Arquitectura general del modelo DA-SKIP	37
4.3.	Arquitectura del modelo DA-LSTM.	46
4.4.	Arquitectura del modelo propuesto, AMBER.	49
5.1.	Esquema de construcción de patrones. (a) FTW=1. (b) FTW=7. (c) FTW=14.	55
5.2.	Partición temporal del conjunto de datos utilizada en los experimentos.	59

5.3. Comparación entre la demanda observada y las predicciones generadas por el modelo SARIMAX para FTW=7.	68
5.4. Comparación entre los modelos LSTM y SARIMAX para la predicción de la demanda eléctrica en FTW=7.	70
5.5. Comparación entre los valores observados y las predicciones obtenidas por los modelos LSTM, DA-SKIP y DA-LSTM para un horizonte de predicción de siete días.	72
5.6. Comparación entre los valores observados, DA-LSTM y AMBER para un horizonte de predicción de 7 días.	74
5.7. Comparación entre los valores observados y las predicciones obtenidas por los modelos LSTM y AMBER para distintos horizontes de predicción.	76
5.8. Relación entre valores observados y predichos por el modelo AMBER.	77
5.9. Evolución del coeficiente de determinación y $RMSE$	81
5.10. Mapa de calor de los pesos promedio de atención meteorológica para los horizontes de predicción de 1, 7 y 14 días.	82

Índice de tablas

5.1. Resumen de las variables de entrada y salida utilizadas para la predicción de la demanda eléctrica	54
5.2. Configuración de búsqueda en grilla para SARIMAX	61
5.3. Mejores configuraciones obtenidas para el modelo SARIMAX	62
5.4. Configuración de búsqueda en grilla para LSTM	63
5.5. Mejores configuraciones obtenidas para el modelo LSTM	63
5.6. Configuración de búsqueda en grilla para DA-SKIP	64
5.7. Mejores configuraciones obtenidas para el modelo DA-SKIP	65
5.8. Configuración de búsqueda en grilla para DA-LSTM	66
5.9. Mejores configuraciones obtenidas para el modelo DA-LSTM	66
5.10. Configuración de búsqueda en grilla para el modelo AMBER	67
5.11. Mejores configuraciones obtenidas para el modelo AMBER: métricas en conjunto de validación	67
5.12. Comparación entre los modelos LSTM y SARIMAX	69
5.13. Comparación entre los modelos LSTM, DA-SKIP y DA-LSTM para los tres horizontes de predicción.	71
5.14. Comparación de desempeño entre DA-LSTM y AMBER para los tres horizontes de predicción.	73
5.15. Desempeño comparativo de los modelos: métricas en conjunto de prueba.	75
5.16. Métricas de error calculadas exclusivamente sobre eventos de alta demanda (percentil 90, $\approx 15\ 150$ MWh) para cada horizonte de predicción.	78

Resumen

La predicción de la demanda de energía eléctrica constituye un componente estratégico en la planificación y operación de los sistemas energéticos modernos, dado que permite optimizar la asignación de recursos, garantizar la estabilidad del sistema y reducir costos operativos. Este problema presenta una elevada complejidad intrínseca debido a la interacción simultánea de múltiples factores (meteorológicos, temporales y socioeconómicos) que introducen dinámicas no lineales y patrones de variabilidad multiescala en el consumo eléctrico. Los enfoques estadísticos clásicos, como los modelos ARIMA y sus extensiones, han demostrado limitaciones estructurales para capturar relaciones no lineales y dependencias temporales de largo alcance. En respuesta a estas restricciones, los modelos de aprendizaje profundo (en particular las redes neuronales recurrentes con arquitectura LSTM, Long Short-Term Memory) han evidenciado mejoras significativas en la capacidad predictiva, aunque persisten dificultades para modelar de manera explícita la contribución diferencial de variables exógenas. En esta tesis se aborda la predicción de la demanda eléctrica diaria en la provincia de Entre Ríos (Argentina) mediante un enfoque multivariado que integra registros históricos de consumo, variables meteorológicas provenientes del reanálisis ERA5-ECMWF y componentes temporales derivados del calendario. Se consideran horizontes de predicción de 1, 7 y 14 días, permitiendo evaluar el comportamiento de los modelos ante distintos niveles de complejidad e incertidumbre temporal.

Como primera contribución, se desarrolla y evalúa sistemáticamente un conjunto de arquitecturas LSTM para la predicción de demanda eléctrica regional. Para ello, se diseñó un esquema experimental basado en múltiples configuraciones y una búsqueda exhaustiva en grilla de hiperparámetros, evaluada mediante validación cruzada. Este proceso permitió identificar modelos LSTM robustos, capaces de superar consistentemente a SARIMAX y DA-SKIP (Dual Attention Skip) en distintos horizontes de predicción, incluso con arquitecturas más parsimoniosas y computacionalmente eficientes.

A partir de estos resultados, la tesis profundiza el estudio de mecanismos de atención aplicados a series temporales multivariadas. En primer lugar, se implementa y evalúa la arquitectura DA-SKIP, modelo basado en atención dual para la

predicción de series temporales multivariadas. Sobre esta base, se propone la arquitectura DA-LSTM (Dual Attention Long Short-Term Memory), que conserva el esquema encoder-decoder con atención dual e incorpora un componente recurrente LSTM. Luego, se propone AMBER (Attention-based Meteorological Bi-Attention Energy Regressor), que incorpora atención específica sobre variables meteorológicas para modelar su relevancia tanto en el contexto histórico como en el horizonte de predicción.

Los resultados experimentales muestran que el modelo LSTM multivariado optimizado alcanza mejoras sustanciales respecto a los métodos de referencia. Para el horizonte de 1 día obtiene un coeficiente de determinación $R^2 = 0,94$ y un RMSE de 727,2 MWh, además de una mayor estabilidad en escenarios de alta demanda. Por su parte, DA-LSTM evidencia que la incorporación de mecanismos de atención permite alcanzar un desempeño competitivo respecto de modelos estadísticos clásicos, obteniendo para el mismo horizonte un $R^2 = 0,87$ y un RMSE de 892,4 MWh. En conjunto, estos resultados evidencian que arquitecturas recurrentes correctamente optimizadas e integrando variables meteorológicas pueden ofrecer una elevada capacidad predictiva en problemas complejos de demanda eléctrica.

Sobre esta base, AMBER mejora adicionalmente la precisión y robustez predictiva mediante mecanismos de atención meteorológica. Para el horizonte de 1 día alcanza $R^2 = 0,94$ y un RMSE de 608,86 MWh, reduciendo el error del modelo LSTM base en un 16,3 % y registrando el menor sesgo sistemático ($\% \text{Bias} = -0,70$). En el horizonte de 14 días mantiene un desempeño robusto, alcanzando $R^2 = 0,69$. Además, el análisis de los pesos de atención identifica las variables térmicas y de radiación como las de mayor relevancia predictiva, aportando interpretabilidad al modelo.

En conjunto, esta investigación demuestra que la combinación de arquitecturas recurrentes multivariadas, optimización sistemática y atención meteorológica constituye una estrategia efectiva para mejorar la precisión, la robustez y la interpretabilidad en escenarios complejos de predicción de demanda eléctrica, con aplicabilidad directa en sistemas regionales de distribución.

Abstract

Electricity demand forecasting constitutes a strategic component in the planning and operation of modern energy systems, since it enables resource allocation optimization, ensures system stability, and reduces operational costs. This problem presents high intrinsic complexity due to the simultaneous interaction of multiple factors (meteorological, temporal, and socioeconomic), which introduce nonlinear dynamics and multiscale variability patterns in electricity consumption.

Classical statistical approaches, such as ARIMA models and their extensions, have shown structural limitations in capturing nonlinear relationships and long-range temporal dependencies. In response to these restrictions, deep learning models (particularly recurrent neural networks based on Long Short-Term Memory, LSTM architecture) have demonstrated significant improvements in predictive capability, although challenges remain in explicitly modeling the differential contribution of exogenous variables.

This thesis addresses daily electricity demand forecasting in the province of Entre Ríos (Argentina) through a multivariate approach integrating historical consumption records, meteorological variables obtained from the ERA5-ECMWF reanalysis, and temporal components derived from calendar information. Prediction horizons of 1, 7, and 14 days are considered, allowing evaluation of model behavior under different levels of complexity and temporal uncertainty.

As a first contribution, a set of LSTM architectures for regional electricity demand forecasting is systematically developed and evaluated. To this end, an experimental framework based on multiple configurations and an exhaustive hyperparameter grid search evaluated through cross-validation was designed. This process enabled the identification of robust LSTM models capable of consistently outperforming SARIMAX and DA-SKIP (Dual Attention Skip) across different prediction horizons, even with more parsimonious and computationally efficient architectures.

Building on these results, this thesis further investigates attention mechanisms applied to multivariate time series. First, the DA-SKIP architecture, a dual-attention-based model for multivariate time series forecasting, is implemented and evaluated. Based on this architecture, DA-LSTM (Dual Attention Long Short-Term Memory) is proposed, preserving the encoder-decoder structure with dual attention while in-

corporating a recurrent LSTM component. Subsequently, AMBER (Attention-based Meteorological Bi-Attention Energy Regressor) is proposed, incorporating specific attention over meteorological variables to model their relevance both in the historical context and for the prediction horizon.

Experimental results show that the optimized multivariate LSTM model achieves substantial improvements over baseline methods. For the 1-day horizon, it reaches a coefficient of determination of $R^2 = 0,94$ and an RMSE of 727,2 MWh, while also exhibiting greater stability under high-demand scenarios. In turn, DA-LSTM demonstrates that the incorporation of attention mechanisms enables competitive performance with respect to classical statistical models, achieving for the same horizon an $R^2 = 0,87$ and an RMSE of 892,4 MWh. Overall, these results indicate that properly optimized recurrent architectures integrating meteorological variables can provide high predictive capability in complex electricity demand forecasting problems.

Building on this foundation, AMBER further improves predictive accuracy and robustness through meteorological attention mechanisms. For the 1-day horizon, it achieves $R^2 = 0,94$ and an RMSE of 608,86 MWh, reducing the error of the baseline LSTM model by 16,3% and yielding the lowest systematic bias (%Bias = -0,70). For the 14-day horizon, it maintains robust performance, achieving $R^2 = 0,69$. Furthermore, attention weight analysis identifies thermal and radiation-related variables as the most relevant for prediction, contributing to model interpretability.

Overall, this research demonstrates that the combination of multivariate recurrent architectures, systematic optimization, and meteorological attention constitutes an effective strategy for improving accuracy, robustness, and interpretability in complex electricity demand forecasting scenarios, with direct applicability to regional electricity distribution systems.

Capítulo 1

Introducción

Este capítulo se organiza en cuatro secciones. En la Sección 1.1 se presenta la motivación del trabajo, destacando la relevancia de la predicción de la demanda eléctrica en el contexto actual y la complejidad asociada a la interacción de múltiples factores meteorológicos, energéticos y socioeconómicos. En la Sección 1.2 se introducen los principales antecedentes en el área, abarcando desde modelos estadísticos tradicionales hasta enfoques más recientes basados en aprendizaje profundo, que consideran mecanismos de atención e interpretabilidad. La Sección 1.3 expone los objetivos de la presente tesis, tanto el general como los específicos, junto con el alcance del estudio. Finalmente, en la Sección 1.4 se describe la organización general de los capítulos de la tesis.

1.1. Motivación

En todo el mundo, los países buscan optimizar el uso de la energía en sectores clave como la industria, el transporte, la agricultura y los entornos urbanos. En el caso particular de la energía eléctrica, prever su demanda constituye un elemento esencial para la planificación eficiente de los sistemas de generación, transporte y distribución. Anticipar las necesidades energéticas permite ajustar la oferta a la demanda de forma oportuna, programar adecuadamente la operación de las centrales de distribución, optimizar el uso de la infraestructura existente y planificar ampliaciones del sistema eléctrico. Asimismo, la disponibilidad de pronósticos confiables posibilita implementar medidas preventivas que minimicen problemas operativos, sobrecargas de equipos, interrupciones del servicio y pérdidas económicas asociadas a desbalances entre distribución y consumo.

Sin embargo, la predicción de la demanda eléctrica presenta un desafío significativo debido a la influencia simultánea de múltiples factores dinámicos y externos que afectan los patrones de consumo energético. Entre ellos, las condiciones meteorológicas desempeñan un papel particularmente relevante. Variables como la temperatura

del aire, la humedad relativa, la velocidad del viento, la radiación solar y los ciclos estacionales influyen directamente en el comportamiento de los usuarios y en los requerimientos energéticos asociados a calefacción, refrigeración e iluminación. A escala regional, la demanda eléctrica exhibe una marcada variabilidad temporal asociada a patrones meteorológicos estacionales, extremos de temperatura y ciclos de actividad humana, factores que condicionan fuertemente los niveles de consumo energético [61, 26, 44]. Asimismo, esta relación no es homogénea entre regiones, sino que depende del contexto socioenergético particular, especialmente del tipo de fuente utilizada para cubrir necesidades térmicas. En zonas donde la calefacción se basa predominantemente en energía eléctrica, las bajas temperaturas tienden a generar incrementos en la demanda, mientras que en regiones donde predominan combustibles alternativos, como gas natural u otros, este efecto puede verse atenuado. En este sentido, el comportamiento de la demanda resulta de la interacción entre variables meteorológicas, características del sistema energético y patrones de consumo de los usuarios, lo cual introduce un nivel adicional de complejidad en su modelado y predicción. En este contexto, resulta fundamental analizar en profundidad los enfoques existentes para la predicción de la demanda eléctrica, evaluando sus capacidades y limitaciones frente a la creciente complejidad del problema, con el objetivo de identificar las oportunidades de mejora que motivan el desarrollo de nuevas arquitecturas de modelado.

Además de los factores meteorológicos, la demanda eléctrica está influenciada por variables socioeconómicas, tecnológicas y productivas que introducen dinámicas adicionales en los patrones de consumo. Entre ellas se encuentran el crecimiento poblacional, el desarrollo urbano, la actividad económica dominante y su evolución y la adopción de tecnologías eléctricas en distintos sectores de la sociedad. Estos factores generan ciclos diarios, semanales y estacionales de consumo que varían según el tipo de usuario y la estructura económica de cada región.

En particular, los distintos sectores de consumo presentan comportamientos energéticos diferenciados. En el sector residencial y comercial, la demanda suele responder con mayor sensibilidad a las variaciones meteorológicas, debido al uso intensivo de equipos de climatización y electrodomésticos durante períodos de temperaturas extremas. En el sector industrial el consumo eléctrico está más estrechamente asociado a los niveles de actividad productiva y a los ciclos económicos que a las condiciones meteorológicas. En regiones turísticas, el consumo muchas veces presenta un comportamiento contracíclico respecto de los sectores anteriores, con mayor demanda en temporadas altas como fines de semana y feriados.

En Argentina, el sistema eléctrico se organiza bajo un esquema federal que combina la operación centralizada del Mercado Eléctrico Mayorista con la gestión provincial de la distribución. La Compañía Administradora del Mercado Mayorista

Eléctrico Sociedad Anónima (CAMMESA) coordina la operación del sistema interconectado nacional y garantiza el equilibrio continuo entre oferta y demanda de energía. A nivel provincial, las empresas distribuidoras son responsables de la operación de las redes de distribución y de la gestión comercial del servicio.

En la provincia de Entre Ríos, la empresa Energía de Entre Ríos Sociedad Anónima (ENERSA) es la responsable de la distribución eléctrica y abastece a más de 360.000 usuarios en su área de concesión, prestando servicio en un territorio de más de 56.000 km² y concentrando la mayor parte del mercado de distribución eléctrica provincial. La magnitud y complejidad de esta infraestructura requieren herramientas de planificación capaces de anticipar variaciones del consumo y posibles condiciones críticas de operación. Sin embargo, en muchos casos la planificación de la demanda eléctrica continúa apoyándose en herramientas tradicionales basadas principalmente en el análisis histórico de registros de consumo y en estimaciones operativas relativamente simples, lo que limita la capacidad de anticipar escenarios críticos y optimizar la toma de decisiones.

En términos de demanda, en la provincia de Entre Ríos la estructura de consumo industrial presenta una fuerte presencia de actividades agroindustriales y alimenticias, entre las que se destacan los frigoríficos y las plantas de producción de alimentos balanceados, que concentran una proporción significativa de la demanda energética industrial [18]. Este tipo de industrias suele mantener patrones de consumo relativamente estables asociados a los procesos productivos, aunque pueden experimentar variaciones vinculadas a la demanda de los mercados, a la estacionalidad de la producción agropecuaria o a cambios en los niveles de actividad económica. Por otro lado, la provincia tiene localidades con gran actividad turística que presentan patrones de consumo eléctrico específico. La coexistencia de estos diferentes perfiles de consumo introduce una elevada complejidad en la dinámica de la demanda eléctrica, dificultando su modelización precisa.

En este contexto, el desarrollo de modelos de predicción más precisos constituye una necesidad estratégica para mejorar la gestión de la red eléctrica y fortalecer la planificación energética a nivel provincial. La incorporación de técnicas avanzadas de inteligencia artificial ofrece nuevas oportunidades para abordar la complejidad inherente al pronóstico de la demanda eléctrica, permitiendo modelar relaciones no lineales entre múltiples variables, capturar dependencias temporales complejas y mejorar la capacidad predictiva de los sistemas utilizados en la planificación energética.

Más allá de la precisión técnica, el desarrollo contemporáneo de modelos predictivos debe alinearse con los objetivos globales de sostenibilidad. Una predicción precisa no solo mejora la eficiencia económica, sino que contribuye a reducir el impacto ambiental del sector energético al evitar la sobreproducción y el uso ineficiente de recursos [31, 3]. No obstante, existe una preocupación creciente sobre el costo

energético asociado al entrenamiento de modelos de gran escala; por ello, la tendencia actual se orienta hacia el diseño de arquitecturas que logren un equilibrio entre rendimiento predictivo y eficiencia computacional [46, 54].

1.2. Antecedentes

Históricamente, el pronóstico de la demanda eléctrica ha sido abordado mediante modelos estadísticos y econométricos basados en series temporales. Entre los enfoques más utilizados se encuentran los modelos Autoregresivos Integrados de Media Móvil (ARIMA), sus extensiones estacionales y variantes que incorporan variables exógenas, como el modelo SARIMAX, así como técnicas de suavizamiento exponencial y modelos de regresión que incluyen variables meteorológicas o indicadores económicos. Estos modelos han sido ampliamente aplicados en la planificación energética, en gran medida por su simplicidad, facilidad de implementación y bajo requerimiento computacional, además de contar con una sólida base teórica. Sin embargo, su capacidad para representar relaciones no lineales complejas y dependencias temporales de largo alcance resulta limitada cuando se aplican a sistemas energéticos modernos caracterizados por la interacción simultánea de múltiples variables explicativas [57, 68, 61].

Con el avance de las técnicas de aprendizaje automático, comenzaron a incorporarse modelos no lineales como la regresión de vectores soporte (*Support Vector Regression*, SVR), árboles de decisión, bosques aleatorios (*Random Forest*) y métodos de potenciación (*boosting*), los cuales demostraron mejoras en la capacidad predictiva frente a los enfoques tradicionales al capturar relaciones más complejas entre las variables [16, 19]. No obstante, estos modelos suelen requerir un proceso de ingeniería de características intensivo y presentan limitaciones para modelar dependencias temporales de largo plazo de manera explícita.

En este contexto, el aprendizaje profundo ha emergido como una alternativa robusta para el análisis de series temporales complejas. En particular, las redes neuronales recurrentes y sus variantes, como las arquitecturas Long Short-Term Memory (LSTM), han demostrado un desempeño destacado en problemas de pronóstico energético, al incorporar mecanismos de memoria que permiten capturar patrones temporales tanto de corto como de largo plazo [61]. La capacidad de estos modelos para integrar simultáneamente múltiples variables exógenas los convierte en herramientas prometedoras para mejorar la precisión de los sistemas de predicción de demanda eléctrica.

Sin embargo, en escenarios reales de predicción energética, caracterizados por la interacción compleja entre variables meteorológicas, energéticas y temporales, las arquitecturas recurrentes convencionales presentan ciertas limitaciones [40, 13]. En

particular, tienden a tratar de manera homogénea la información de entrada, sin distinguir explícitamente la relevancia relativa de cada variable ni la contribución diferencial de distintos instantes temporales en el proceso de predicción.

En este contexto, la investigación reciente ha comenzado a explorar arquitecturas más avanzadas que incorporan mecanismos de atención y estructuras híbridas, permitiendo modelar de manera más flexible la interacción entre variables históricas y variables exógenas [66]. Estos enfoques resultan especialmente adecuados para problemas de naturaleza multivariada, donde la importancia de cada predictor puede variar en el tiempo y según el contexto operativo.

El desarrollo de estos modelos recientes se enmarca en dos líneas de investigación emergentes. Por un lado, la inteligencia artificial explicable (XAI, del inglés eXplainable Artificial Intelligence) que busca dotar a los modelos de mayor transparencia, permitiendo interpretar el peso relativo de las variables en la predicción [51, 5]. Por otro, la inteligencia artificial verde (Green AI), que introduce la necesidad de considerar no solo la precisión de los modelos, sino también su eficiencia computacional y su impacto energético [10].

En síntesis, la tendencia actual en la predicción de la demanda eléctrica se orienta hacia el diseño de modelos más expresivos que integren mecanismos de atención, manejo explícito de múltiples fuentes de información, interpretabilidad y eficiencia, superando así las limitaciones de las arquitecturas recurrentes tradicionales.

1.3. Objetivos y Alcance

1.3.1. Objetivo General

El objetivo general de la presente tesis es desarrollar modelos de redes neuronales artificiales que permitan predecir la demanda de energía eléctrica a partir de información meteorológica y los registros de demanda histórica de consumo energético.

1.3.2. Objetivos específicos

- Identificar las variables de mayor impacto en la demanda de energía eléctrica en la provincia de Entre Ríos.
- Proponer modelos de redes neuronales recurrentes para la predicción a corto plazo de la demanda de energía eléctrica a nivel provincial.
- Analizar el efecto de distintas configuraciones temporales de entrada y de salida en el desempeño predictivo de los modelos propuestos.

- Evaluar y comparar el desempeño de los modelos mediante métricas cuantitativas de error y capacidad explicativa.

1.3.3. Alcance

El presente trabajo se enfoca en la predicción de la demanda eléctrica de corto plazo (hasta dos semanas) para la provincia de Entre Ríos, Argentina, utilizando series temporales multivariadas que integran datos de consumo energético y variables meteorológicas. El estudio contempla etapas de análisis de datos, el diseño e implementación de modelos de inteligencia artificial, su entrenamiento y validación, así como la comparación con metodologías de referencia. Si bien la propuesta metodológica podría adaptarse a otros contextos geográficos y a la previsión de generación de energías renovables, su validación experimental se limita al caso de estudio correspondiente a la provincia de Entre Ríos.

1.4. Organización de la tesis

La presente tesis se organiza en seis capítulos. El Capítulo 1 introduce el problema de investigación, presentando la motivación del estudio, los objetivos planteados y el alcance del trabajo, así como la organización general del documento. En el Capítulo 2 se formula el problema de predicción de la demanda eléctrica, describiendo las características de la serie de demanda, sus variaciones temporales y los horizontes de predicción considerados. El Capítulo 3 presenta el marco teórico, donde se revisan los principales enfoques utilizados en la literatura para la predicción de la demanda eléctrica, incluyendo modelos clásicos de series temporales y modelos basados en aprendizaje profundo, así como las limitaciones identificadas en estos enfoques. En el Capítulo 4 se presentan los modelos propuestos de la tesis, partiendo de modelos existentes y avanzando hacia un enfoque superador, describiendo su arquitectura, los componentes que lo integran y el procedimiento de entrenamiento adoptado. El Capítulo 5 expone los resultados obtenidos, incluyendo la evaluación experimental del modelo y la comparación con otros enfoques de referencia. Finalmente, el Capítulo 6 presenta una síntesis de los principales aportes y resultados obtenidos a lo largo de la tesis, organizados en función de los objetivos específicos planteados. Asimismo, se destacan las contribuciones alcanzadas, se discuten posibles líneas de trabajo futuro y se resumen las publicaciones científicas, presentaciones en congresos

y reconocimientos académicos derivados del desarrollo de esta investigación doctoral.

Capítulo 2

Formulación del problema

Este capítulo se organiza en cuatro secciones principales. En la Sección 2.1 se presenta una introducción general a la formulación del problema, destacando la importancia de comprender la dinámica de la demanda eléctrica y los desafíos asociados a su predicción. En la Sección 2.2 se describe de manera general el enfoque metodológico adoptado en esta tesis, incluyendo las fuentes de datos utilizadas, el proceso de preprocesamiento y el enfoque metodológico general. En la Sección 2.3 se analizan en detalle las características de la demanda eléctrica, desagregadas en sus dimensiones anual, mensual y semanal, así como los principales factores externos que influyen en su comportamiento. Finalmente, en la Sección 2.4 se describen los horizontes de predicción considerados en este trabajo.

2.1. Introducción

La predicción de la demanda eléctrica constituye un problema central en la planificación y operación de sistemas energéticos modernos. Como señalan Wang et al. [65], el desarrollo de modelos avanzados basados en arquitecturas profundas permite mejorar significativamente la precisión en pronósticos de corto plazo, optimizando la asignación de recursos y reduciendo riesgos operativos. En la misma línea, Abumohsen et al. [1] destacan que la anticipación precisa del consumo eléctrico impacta directamente en la estabilidad del sistema, la eficiencia económica y la confiabilidad del suministro. Asimismo, investigaciones recientes basadas en modelos Transformer y mecanismos de atención han demostrado que la incorporación explícita de variables meteorológicas locales mejora la capacidad predictiva en contextos energé-

ticos complejos [22]. En consecuencia, el pronóstico de carga se consolida como un componente estratégico dentro de los sistemas eléctricos contemporáneos.

La complejidad del problema radica en el carácter multifactorial y estocástico del consumo eléctrico. La demanda no responde exclusivamente a tendencias históricas lineales, sino que depende de la interacción entre factores meteorológicos, socio-económicos y temporales. Estudios recientes destacan que las variaciones térmicas extremas, la estacionalidad, la actividad industrial y los patrones de comportamiento humano introducen dinámicas no lineales que desafían los enfoques estadísticos tradicionales [53, 21, 28].

En consecuencia, el problema abordado en esta tesis consiste en desarrollar un enfoque predictivo capaz de integrar información histórica, variables meteorológicas y componentes temporales en un esquema multivariado que represente adecuadamente la complejidad inherente a la demanda eléctrica.

2.2. Propuesta general del enfoque

La Figura 2.1 presenta una visión general del enfoque metodológico adoptado en esta tesis para abordar el problema de predicción de la demanda eléctrica. El esquema resume las principales etapas del proceso, desde la obtención y preparación de los datos hasta el desarrollo del modelo de aprendizaje profundo utilizado para generar las predicciones.

El proceso comienza con la recopilación de múltiples fuentes de información. En particular, se utilizan registros históricos de la demanda eléctrica diaria total de Entre Ríos provistos por la empresa distribuidora Energía de Entre Ríos S.A. (ENERSA), junto con variables meteorológicas provenientes del reanálisis ERA5 del *European Centre for Medium-Range Weather Forecasts* (ECMWF) [27]. Para cada variable meteorológica se utilizó el promedio espacial de la grilla ERA5 sobre el territorio provincial. Este producto integra observaciones y un modelo numérico mediante técnicas de asimilación de datos para reconstruir de manera consistente el estado del tiempo a lo largo de los años. La integración de estas fuentes permite construir un conjunto de datos multivariado que combina información energética, meteorológica y temporal.



Figura 2.1: Esquema general del enfoque metodológico utilizado para la predicción de la demanda eléctrica, incluyendo las etapas de obtención de datos, preprocesamiento, construcción del dataset multivariado y desarrollo del modelo de aprendizaje profundo.

ENERSA concentra el 71 % del mercado de distribución eléctrica de Entre Ríos. La provincia, que se ubica en la región centro-este del país, presenta una matriz de consumo eléctrico heterogénea, asociada a la coexistencia de zonas urbanas, industriales, rurales y turísticas. La empresa distribuye y comercializa energía eléctrica a más de 360 000 usuarios dentro de su área de concesión, y opera la red provincial a través de 24 estaciones transformadoras y 65 subestaciones distribuidas sobre el territorio, cuya localización geográfica se ilustra en la Figura 2.2.

Los datos recopilados atraviesan una etapa de preprocesamiento orientada a garantizar la consistencia y calidad de la información utilizada por los modelos predictivos. Este proceso incluye la unificación de formatos de los archivos originales, la sincronización temporal de las distintas fuentes de datos y la construcción de una serie temporal multivariada homogénea. Asimismo, se realizan tareas de limpieza de datos, tratamiento de valores faltantes y normalización de las variables, con el objetivo de asegurar estabilidad numérica durante el entrenamiento de los modelos.

A partir de la base de datos consolidada se dispone de un conjunto de variables predictoras utilizadas en el proceso de modelado, las cuales se clasifican en tres grupos principales: variables energéticas, variables meteorológicas y variables temporales derivadas del calendario. Esta representación multivariada permite cap-

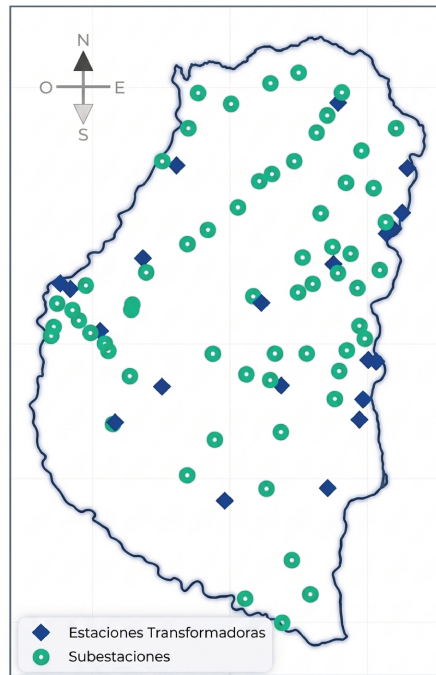


Figura 2.2: Distribución espacial de la infraestructura eléctrica en la provincia de Entre Ríos. Se representan las 24 estaciones transformadoras (◆) y las 65 subestaciones (○) operadas por ENERSA en su área de concesión.

turar simultáneamente distintos factores que influyen en el comportamiento de la demanda eléctrica.

Finalmente, sobre este conjunto de datos se aborda el problema de predicción de la demanda eléctrica mediante el análisis y la aplicación de distintos enfoques de modelado. En este sentido, se consideran alternativas que abarcan desde modelos estadísticos tradicionales hasta métodos de aprendizaje profundo, incluyendo además propuestas desarrolladas en el marco de esta tesis. Los detalles específicos de los modelos considerados, así como sus fundamentos teóricos y configuraciones, se presentan en los capítulos siguientes.

Antes de presentar en detalle el desarrollo de los modelos implementados, resulta necesario analizar las características temporales de la demanda eléctrica que se busca predecir. Por este motivo, en la siguiente sección se examinan los principales patrones del consumo eléctrico provincial a distintas escalas temporales.

2.3. Características de la demanda eléctrica

La demanda eléctrica presenta una estructura temporal multiescala. En esta sección se analiza el comportamiento de la demanda en tres niveles de análisis temporal: anual, mensual y semanal, cada uno asociado a dinámicas específicas del sistema eléctrico. Esta caracterización resulta relevante porque orienta la selección de variables predictoras y permite justificar la necesidad de modelos multivariados capaces de capturar no linealidades, dependencias temporales y cambios de régimen [53, 21, 37, 30].

2.3.1. Demanda anual

El análisis interanual de la demanda eléctrica provincial se realiza sobre el período 2012-2023 (curva negra en Figura 2.3), seleccionado en función de la disponibilidad de los datos energéticos y meteorológicos considerados, lo que permite identificar una tendencia general de crecimiento en el consumo energético, aunque acompañada de fluctuaciones de distinta magnitud a lo largo del tiempo. En particular, se observan períodos de expansión sostenida intercalados con fases de relativa estabilidad o leve desaceleración del crecimiento. Estas variaciones reflejan cambios en el nivel de consumo eléctrico del sistema provincial y pueden estar asociadas a modificaciones en la actividad económica regional, a transformaciones en los patrones de consumo energético y a factores regulatorios que afectan el comportamiento de la demanda, o incluso a cambios globales como la pandemia en 2020. Tal como se discute en [61], la evolución de la demanda eléctrica en sistemas regionales suele presentar una combinación de tendencias de largo plazo y variabilidad interanual, lo que evidencia que el consumo eléctrico no responde únicamente a su trayectoria histórica sino a la interacción de múltiples factores contextuales.

2.3.2. Demanda mensual

La evolución temporal de la demanda mensual puede observarse en la curva gris de la Figura 2.3. La serie mensual evidencia oscilaciones intra-anales pronunciadas asociadas a patrones estacionales recurrentes. En particular, el comportamiento describe un patrón característico en forma de “W”, con niveles elevados de consumo

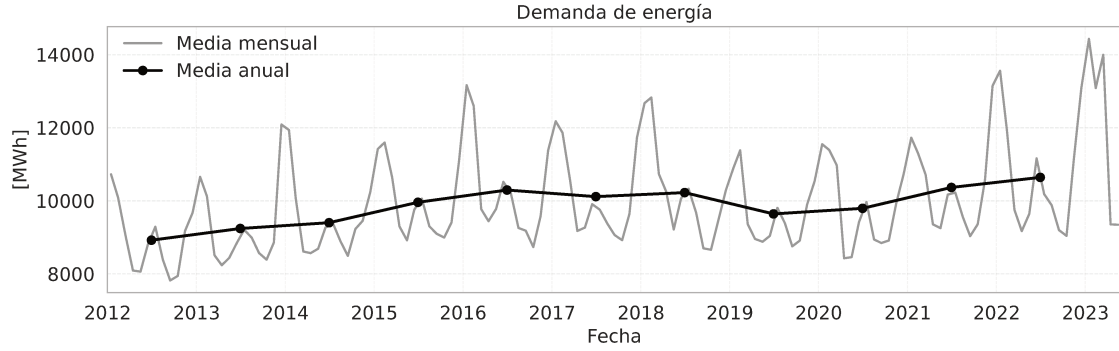


Figura 2.3: Evolución mensual y anual de demanda de energía provincial [MWh].

durante los meses de verano y de invierno, y valores relativamente menores durante las estaciones intermedias de primavera y otoño. Este comportamiento es aún más evidente en la Figura 2.4, que presenta la media mensual de la demanda junto con su rango intercuartílico (IQR), lo que permite visualizar tanto el nivel promedio de consumo en cada mes como la dispersión observada entre años.

El patrón identificado (W) resulta consistente con la influencia de condiciones térmicas sobre el consumo eléctrico, especialmente en lo que respecta al uso de sistemas de climatización, dado que en verano la demanda aumenta por el uso intensivo de sistemas de refrigeración, mientras que en invierno se incrementa debido a necesidades de calefacción [61]. No obstante, la literatura especializada indica que la relación entre temperatura y demanda no es estrictamente lineal, pudiendo presentar comportamientos asimétricos y efectos umbral en escenarios de temperaturas extremas [53, 21, 22, 63]. En este contexto, el denominado efecto umbral refiere a situaciones en las que la demanda eléctrica responde de manera relativamente estable dentro de ciertos rangos de temperatura, pero experimenta incrementos abruptos cuando la temperatura supera determinados valores críticos, por ejemplo durante olas de calor o episodios de frío intenso.

Asimismo, la amplitud del rango intercuartílico (IQR) permite analizar la variabilidad interanual del consumo eléctrico en cada mes. En particular, se observa que los meses asociados a mayor demanda presentan también una mayor dispersión, lo que sugiere que en períodos de consumo elevado el sistema eléctrico se encuentra más expuesto a fluctuaciones asociadas a factores externos, como costo del servicio o actividad productiva.

La coexistencia de estas variaciones intra-anales con la tendencia de largo plazo

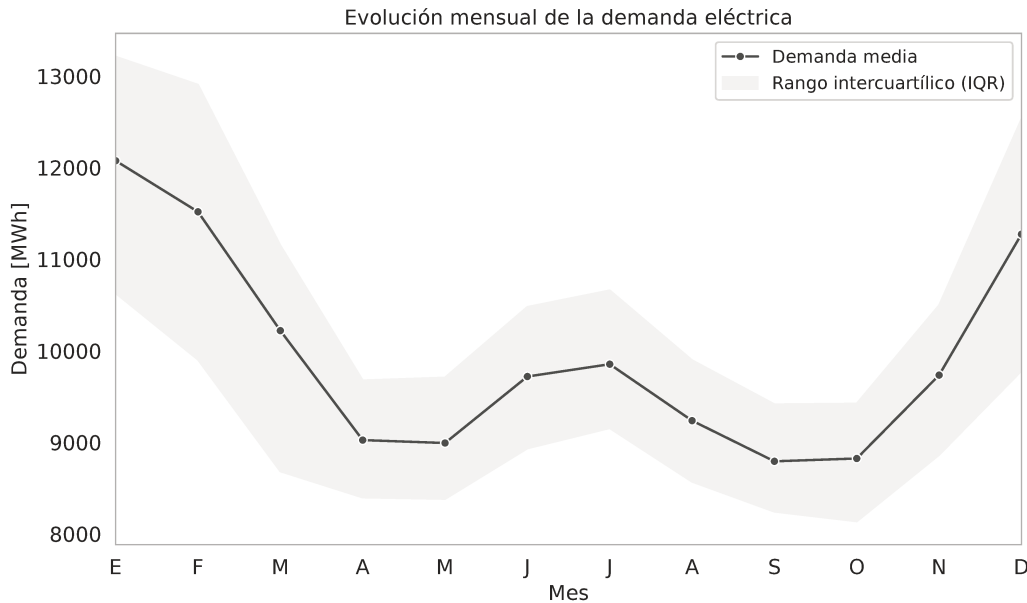


Figura 2.4: Ciclo anual medio de la demanda eléctrica provincial para el período 2012–2023, mostrando la media mensual y el rango intercuartílico (IQR).

pone de manifiesto la superposición de componentes de tendencia y estacionalidad dentro de la misma serie. En este sentido, la demanda eléctrica no constituye una serie estrictamente estacionaria ni exclusivamente determinada por su trayectoria histórica, sino que presenta simultáneamente componentes de tendencia, estacionalidad y variabilidad interanual que deben ser consideradas en el diseño de modelos predictivos. En consecuencia, el problema requiere un abordaje desde una perspectiva multivariada orientada a integrar información meteorológica y temporal, capaz de capturar relaciones complejas entre los distintos factores que influyen en la demanda eléctrica, evitando restringirse a supuestos estrictamente lineales [37, 53, 21, 60].

2.3.3. Demanda semanal

El análisis de la demanda a escala semanal permite identificar regularidades temporales asociadas a la organización operativa del sistema eléctrico provincial. A diferencia de las variaciones interanuales y estacionales analizadas en las secciones anteriores, que se vinculan principalmente con factores estructurales y condiciones térmicas, el patrón semanal refleja dinámicas recurrentes de corto plazo relacionadas con los ciclos de actividad económica y social. La Figura 2.5 muestra la distribución del consumo eléctrico por día de la semana, tanto en su forma general (de lunes a domingo) como considerando explícitamente los días feriados. Se observa que los

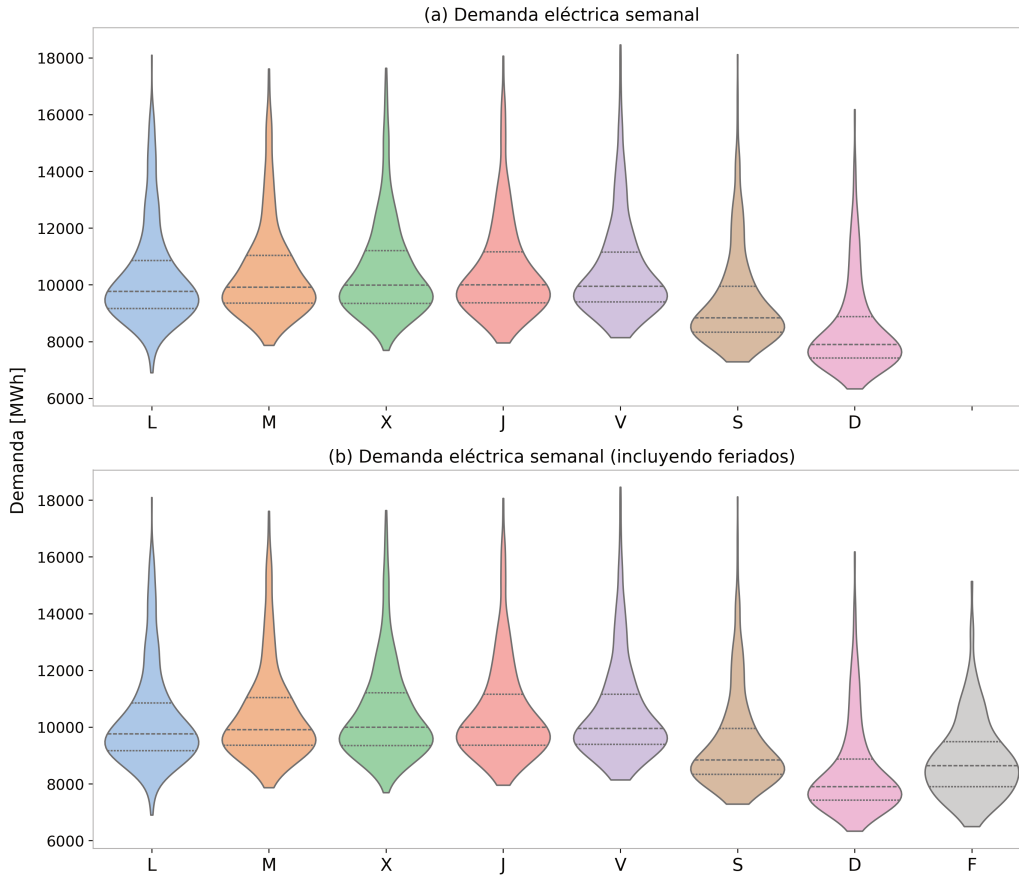


Figura 2.5: Distribución diaria de la demanda eléctrica representada mediante diagramas de violín. (a) Distribución del consumo desde lunes hasta domingo. (b) Distribución del consumo incorporando los días feriados como categoría adicional.

niveles de demanda tienden a ser más elevados durante los días hábiles, mientras que los fines de semana presentan descensos sistemáticos, particularmente los domingos. También se observa que los lunes presentan un pico hacia valores bajos, menos marcada en los restantes días hábiles. Al aislar los feriados, se observa que presentan un patrón de consumo intermedio entre el sábado y el domingo, aunque más próximo al comportamiento típico de los fines de semana que al de los días hábiles.

Este comportamiento puede interpretarse como una manifestación de los ciclos de actividad productiva y comercial que caracterizan el funcionamiento del sistema eléctrico provincial. Durante los días laborables, la demanda se ve impulsada tanto por el consumo residencial como por la actividad industrial, comercial y de servicios, mientras que durante los fines de semana y feriados gran parte de estas actividades disminuyen su intensidad o se suspenden temporalmente, reduciendo el consumo energético agregado del sistema.

Desde una perspectiva de modelado, la estabilidad de esta estructura semanal

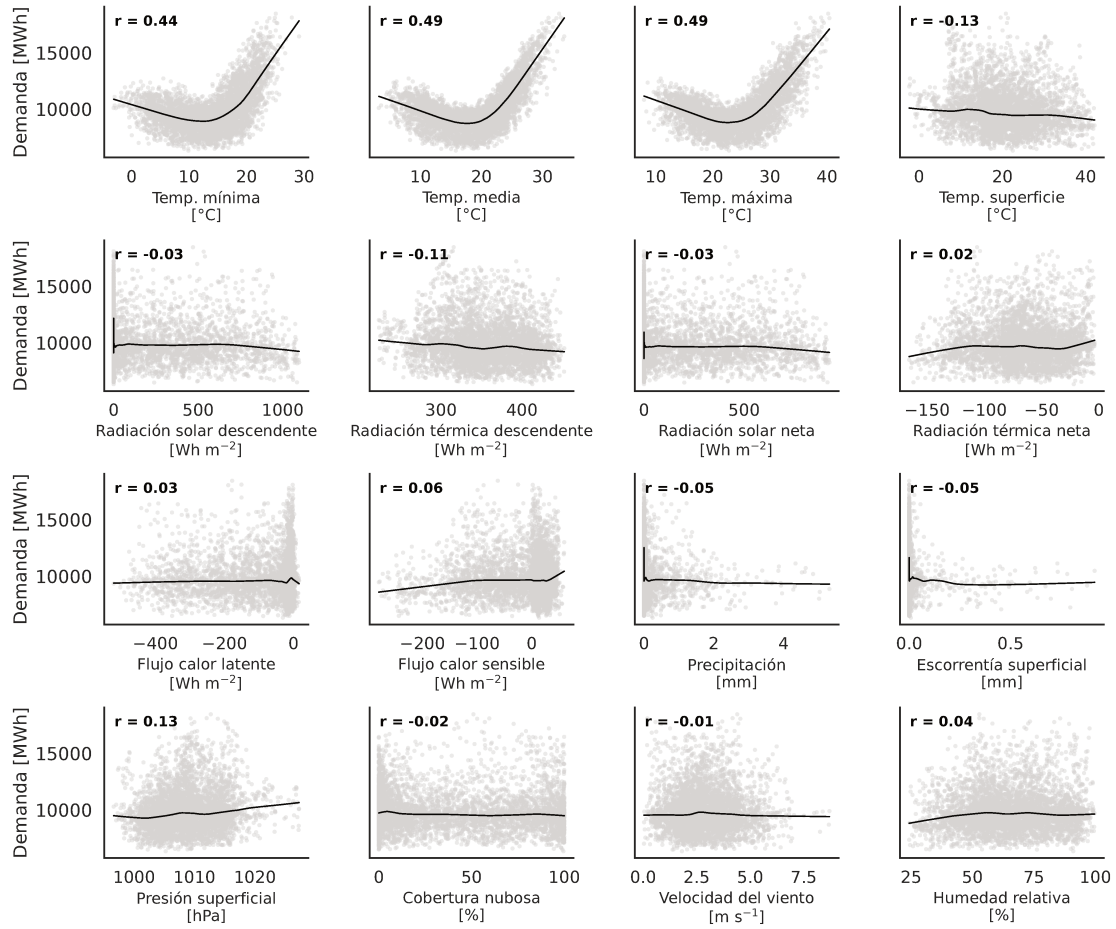


Figura 2.6: Diagramas de dispersión de variables meteorológicas diarias frente a la demanda eléctrica diaria, incluyendo los coeficientes de correlación lineal (r).

sugiere la presencia de una componente temporal determinística de corto plazo. En consecuencia, resulta conveniente incorporar variables categóricas asociadas al calendario, tales como el día de la semana y la presencia de feriados, dentro del conjunto de predictores del modelo, ya que estas permiten capturar patrones sistemáticos del consumo eléctrico que no están directamente relacionados con las variables meteorológicas.

2.3.4. Relación entre estado del tiempo y demanda eléctrica

La demanda eléctrica es el resultado de la interacción de múltiples factores de distinta naturaleza, entre los que se destacan variables meteorológicas, temporales y contextuales. En particular, las condiciones térmicas suelen desempeñar un rol relevante en sistemas eléctricos regionales debido al uso intensivo de dispositivos de climatización en escenarios de temperaturas extremas.

La Figura 2.6 presenta los diagramas de dispersión entre la demanda eléctrica diaria y distintas variables meteorológicas para ese día, junto con sus coeficientes de correlación lineal. En la figura se consideran tanto variables térmicas, como la temperatura mínima (t_{min}), media (t_{mean}) y máxima (t_{max}), así como la temperatura superficial del suelo (skt); variables radiativas como la radiación solar incidente ($ssrd$), la radiación solar neta (ssr), la radiación térmica descendente ($strd$) y la radiación térmica neta (str); variables de flujo energético superficial como el flujo de calor latente ($slhf$) y sensible ($sshf$); y otras variables meteorológicas relevantes como la precipitación total (tp), la escorrentía superficial (sro), la presión atmosférica en superficie (sp), la nubosidad total (tcc), la velocidad del viento (ws) y la humedad relativa (rh). Las siglas corresponden a las utilizadas en ERA5, el reanálisis utilizado como fuente de información.

Se observa que las temperaturas mínima, media y máxima presentan asociaciones positivas moderadas ($r \approx 0.44-0.49$), mientras que variables como radiación solar, precipitación, velocidad del viento y humedad relativa exhiben correlaciones lineales débiles. La curva negra corresponde a un suavizado LOWESS (*Locally Weighted Scatterplot Smoothing*), que permite visualizar posibles tendencias no lineales. En particular, puede observarse una tendencia con forma aproximada de “U” asimétrica, donde los niveles de demanda tienden a incrementarse tanto en rangos de temperaturas elevadas como en escenarios de temperaturas bajas, mientras que valores intermedios se asocian a consumos más moderados. Este comportamiento es consistente con el patrón estacional previamente descrito en forma de “W”, en el cual el consumo eléctrico aumenta durante los meses de verano e invierno debido al uso intensivo de sistemas de refrigeración y calefacción. Como cada gráfico considera una variable meteorológica por vez, no muestra posibles efectos combinados, por ejemplo, entre temperaturas elevadas y humedad relativa alta.

La literatura ha señalado que las interacciones entre variables meteorológicas y demanda energética suelen presentar comportamientos no lineales y efectos umbral, especialmente bajo condiciones térmicas extremas [22, 23, 53, 4]. Asimismo, diversos estudios destacan que la integración conjunta de variables temporales y meteorológicas mejora el desempeño de los modelos de pronóstico de corto plazo [1, 65, 60].

2.4. Horizontes de predicción de la demanda

La definición del horizonte de predicción constituye un aspecto central en el diseño de modelos de pronóstico de demanda eléctrica, dado que las dinámicas temporales del sistema y los factores explicativos pueden variar significativamente según la escala temporal considerada. En general, los estudios de pronóstico energético clasifican los horizontes de predicción en corto, mediano y largo plazo, cada uno asociado a diferentes procesos físicos, niveles de incertidumbre y necesidades operativas del sistema eléctrico.

En términos generales, los horizontes de corto plazo abarcan desde segundos hasta algunas horas o pocos días, y se encuentran principalmente asociados a variaciones meteorológicas de rápida evolución y a la operación inmediata del sistema eléctrico, incluyendo actividades de regulación de frecuencia y seguimiento de carga. Por su parte, los horizontes de mediano plazo comprenden escalas de días a semanas, en las cuales predominan variaciones meteorológicas de mayor escala espacial y temporal, así como efectos asociados a la organización de la actividad económica y social. Finalmente, los horizontes de largo plazo abarcan meses o años, y se vinculan principalmente con tendencias estructurales del sistema energético, cambios en la actividad económica, crecimiento poblacional y procesos meteorológicos de mayor escala [2, 71, 25].

En el contexto de esta tesis, el problema se sitúa dentro del rango de predicción de corto plazo, comúnmente denominado *Short-Term Load Forecasting* (STLF), ya que el objetivo es anticipar la demanda eléctrica diaria a partir de información histórica y variables meteorológicas. En particular, se analizan horizontes de predicción de 1, 7 y 14 días, los cuales resultan relevantes para la planificación operativa y el seguimiento de la demanda en empresas distribuidoras de energía eléctrica.

El horizonte de 1 día fue considerado inicialmente como un escenario de muy corto plazo que permitió evaluar de manera eficiente el comportamiento de las arquitecturas propuestas bajo un esquema de entrenamiento computacionalmente menos exigente. Este enfoque facilita la exploración sistemática de configuraciones de hiperparámetros y el análisis preliminar del impacto de las variables del conjunto de datos multivariado de la provincia de Entre Ríos, particularmente en lo referido

a la incorporación de múltiples predictores meteorológicos más allá de la temperatura tradicionalmente utilizada. Desde una perspectiva aplicada, el horizonte diario también posee relevancia operativa para las distribuidoras eléctricas, ya que contribuye a la planificación inmediata y al seguimiento de la demanda en escenarios de variabilidad meteorológica.

Una vez identificadas las configuraciones robustas, el estudio se amplió a horizontes de 7 y 14 días. El horizonte semanal permite capturar ciclos completos asociados a efectos calendario y patrones recurrentes del consumo eléctrico, mientras que el horizonte de 14 días introduce mayores niveles de incertidumbre y exigencia en términos de estabilidad y capacidad de generalización del modelo. Esta progresión permite evaluar no sólo la precisión predictiva, sino también la robustez del enfoque propuesto frente a distintos niveles de complejidad temporal.

La literatura indica que el desempeño de los modelos de aprendizaje automático y profundo puede variar sustancialmente en función del horizonte de predicción considerado, particularmente en contextos donde la demanda presenta componentes estacionales y relaciones no lineales [22, 65, 53, 15]. En este sentido, la evaluación de múltiples horizontes permite no sólo comparar la precisión predictiva, sino también analizar la capacidad de generalización del modelo frente a distintos escenarios temporales.

Adicionalmente, la definición del problema predictivo no solo involucra el horizonte futuro (del inglés, *future temporal window*, FTW) a predecir, sino también la longitud de la ventana temporal histórica (del inglés, *historical temporal window*, HTW) utilizada como entrada del modelo. El presente trabajo contempla distintos esquemas de combinación entre ventanas históricas y horizontes futuros, cuya formulación e implementación detallada se presenta en el Capítulo 4.

Capítulo 3

Marco teórico

Este capítulo revisa los principales enfoques utilizados en la predicción de la demanda eléctrica. En la Sección 3.1 se introduce el problema desde una perspectiva metodológica, delimitando el marco teórico en relación con el modelado de la demanda eléctrica. La Sección 3.2 presenta los modelos clásicos de series temporales incluyendo su formulación matemática y las principales extensiones utilizadas en el ámbito energético. En la Sección 3.3 se analizan los avances en modelos de aprendizaje profundo, con énfasis en arquitecturas recurrentes y enfoques híbridos aplicados a la predicción de demanda eléctrica. Finalmente, la Sección 3.4 examina las limitaciones de los enfoques existentes, identificando los desafíos metodológicos que motivan el desarrollo de nuevas arquitecturas de predicción.

3.1. Introducción

La gestión moderna de los sistemas eléctricos enfrenta retos sin precedentes debido a la transición hacia redes inteligentes (*smart grids*) y la integración masiva de fuentes de energía renovable [20, 72]. En este contexto, el STLF se ha consolidado como una herramienta fundamental para la operación y planificación de los sistemas eléctricos modernos [31, 61].

En el ámbito específico del STLF, el desarrollo metodológico ha experimentado una transición progresiva desde modelos estadísticos lineales hacia enfoques basados en aprendizaje automático y, más recientemente, en arquitecturas profundas capaces de capturar dependencias temporales no lineales y multivariadas. Revisiones recientes destacan que las arquitecturas híbridas han mostrado un desempeño competitivo

en STLF; sin embargo, su mayor profundidad estructural y número de parámetros requieren una evaluación crítica en función del tamaño del conjunto de datos, el horizonte de predicción y los recursos computacionales disponibles [62, 9].

En el contexto del modelado de series temporales, la selección de arquitecturas predictivas no depende únicamente de la precisión alcanzada, sino también de su complejidad estructural y de los requerimientos computacionales asociados. En particular, arquitecturas profundas con un elevado número de parámetros pueden presentar dificultades de entrenamiento y mayores costos computacionales, especialmente en escenarios con conjuntos de datos de tamaño moderado. En consecuencia, el diseño de modelos de predicción de demanda eléctrica requiere considerar el equilibrio entre capacidad representativa, estabilidad del entrenamiento y eficiencia computacional.

Desde una perspectiva más amplia de modelado de series temporales, las arquitecturas profundas han mostrado ventajas en escenarios de predicción multihorizonte mediante estructuras encoder-decoder y mecanismos de atención adaptativa [39]. Esta evolución metodológica constituye el punto de partida del presente marco teórico.

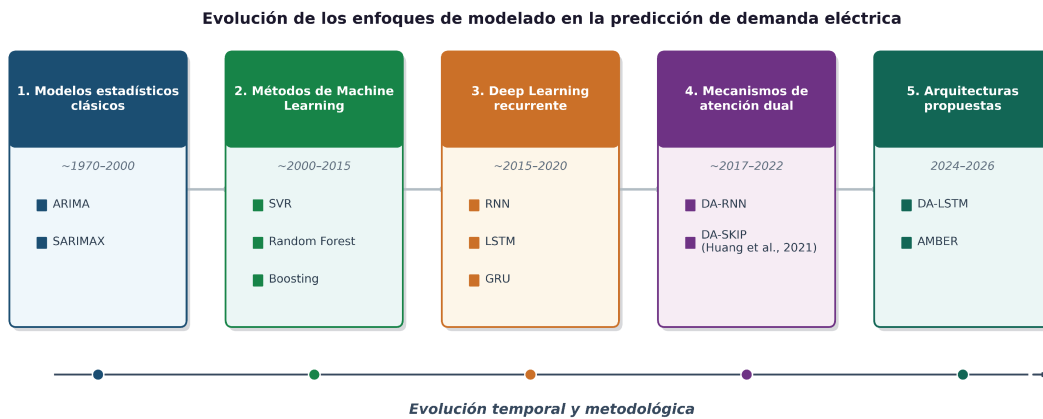


Figura 3.1: Evolución de los enfoques de modelado en la predicción de la demanda eléctrica, desde modelos estadísticos hasta arquitecturas basadas en aprendizaje profundo y mecanismos de atención, incluyendo los modelos desarrollados en esta tesis.

En este marco, la Figura 3.1 presenta de manera esquemática la evolución de los enfoques utilizados en la predicción de la demanda eléctrica, desde modelos estadísticos tradicionales como ARIMA y SARIMAX, pasando por arquitecturas de

aprendizaje profundo como LSTM, hasta modelos más recientes basados en mecanismos de atención, como DA-SKIP. A partir de estos desarrollos, en esta tesis se proponen nuevas arquitecturas que continúan esta línea de evolución, dando lugar a los modelos DA-LSTM y, posteriormente, AMBER, orientados a mejorar la capacidad de modelado en escenarios multivariados.

En este contexto, la selección de la arquitectura predictiva no debe basarse exclusivamente en la maximización de la precisión, sino también en criterios de eficiencia estructural. Esta consideración resulta particularmente relevante en sistemas eléctricos regionales, donde las restricciones de datos y recursos computacionales condicionan el diseño de los modelos. A partir de este marco, en las secciones siguientes se revisan los principales enfoques utilizados en la literatura, comenzando por los modelos clásicos de series temporales.

3.2. Modelos clásicos

Los modelos clásicos de predicción de series temporales han sido ampliamente utilizados en el ámbito energético debido a su simplicidad y capacidad interpretativa. Entre ellos, los modelos ARIMA y su extensión SARIMAX permiten capturar tendencias, estacionalidad y dependencias lineales en la serie. La incorporación de variables exógenas en el marco SARIMAX ha demostrado mejorar el desempeño cuando se consideran factores meteorológicos relevantes [37, 60, 57]. A pesar de su robustez estadística, estos métodos asumen una relación lineal que resulta insuficiente para capturar la elasticidad de la demanda, por ejemplo, ante cambios bruscos de temperatura [72].

3.2.1. Modelo ARIMA

Los modelos *Autoregressive Integrated Moving Average* (ARIMA) constituyen uno de los enfoques estadísticos más utilizados para el modelado y pronóstico de series temporales. En un modelo $\text{ARIMA}(p, d, q)$, la serie original se diferencia d veces antes de modelar sus componentes autorregresivo y de media móvil. Si B

denota el operador de retardo, la serie diferenciada se expresa como

$$w_t = (1 - B)^d y_t. \quad (3.1)$$

Sobre w_t , el modelo puede representarse mediante

$$w_t = \mu + \sum_{i=1}^p \phi_i w_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (3.2)$$

donde μ es un término constante, ϕ_i y θ_j son los coeficientes de los componentes autorregresivo (AR) y de media móvil (MA), respectivamente, y ε_t es un término de error aleatorio. Los parámetros p , d y q indican el orden autorregresivo, el orden de diferenciación y el orden de la media móvil. Cuando $d = 0$, no se aplica diferenciación y $w_t = y_t$. El término ε_t se asume como ruido blanco con media cero y varianza constante. Debido a su capacidad para modelar dependencias temporales lineales, estos modelos han sido ampliamente utilizados en el análisis y pronóstico de demanda eléctrica [11, 28, 68].

3.2.2. Modelo SARIMAX

El modelo *Seasonal AutoRegressive Integrated Moving Average with eXogenous variables* (SARIMAX) constituye una extensión del modelo ARIMA que permite incorporar variables externas capaces de influir sobre la serie temporal objetivo. Mientras que los modelos ARIMA capturan la estructura autoregresiva y de media móvil de la serie diferenciada, el modelo SARIMAX permite integrar simultáneamente dependencias internas de la serie y la influencia de factores externos relevantes.

El modelo SARIMAX introduce un término de regresión adicional que permite incorporar variables exógenas $X_{i,t}$ al modelo, dando lugar a la siguiente formulación:

$$y_t = \mu + \sum_{i=1}^k \beta_i X_{i,t} + \text{componentes ARIMA} + \varepsilon_t, \quad (3.3)$$

donde $X_{i,t}$ representa la i -ésima variable exógena en el instante t , β_i corresponde al coeficiente asociado a dicha variable, y k indica el número total de predictores externos considerados. Esta formulación permite modelar simultáneamente la diná-

mica temporal interna de la serie y la influencia directa de variables externas. En el contexto de la predicción de demanda eléctrica, estas variables exógenas suelen corresponder a factores meteorológicos, temporales o socioeconómicos que afectan el comportamiento del consumo energético. En particular, los componentes autoregresivos y de media móvil permiten modelar dependencias temporales lineales de la demanda eléctrica, mientras que el término de regresión exógeno captura la influencia directa de variables meteorológicas. Bajo esta estructura, el modelo SARIMAX puede representar variaciones de la demanda inducidas por factores meteorológicos, siempre que las relaciones subyacentes puedan aproximarse mediante dependencias lineales y que la serie resultante cumpla las condiciones de estacionariedad tras el proceso de diferenciación.

No obstante, estos enfoques presentan limitaciones estructurales. En primer lugar, asumen relaciones predominantemente lineales entre variables. En segundo lugar, su capacidad para modelar múltiples predictores simultáneamente resulta restringida. Finalmente, diversos estudios reportan una degradación significativa del desempeño a medida que aumenta el horizonte de predicción [58, 15]. Estas limitaciones impulsaron la incorporación de técnicas de aprendizaje automático capaces de modelar relaciones no lineales sin requerir, en general, transformaciones explícitas orientadas a garantizar la estacionariedad de la serie, como ocurre en los modelos ARIMA tradicionales. No obstante, la simplicidad estructural de estos modelos constituye una ventaja en términos de interpretabilidad y bajo requerimiento computacional. Este equilibrio entre precisión y complejidad resulta especialmente relevante en sistemas eléctricos regionales donde los recursos de cómputo pueden ser limitados.

3.3. Modelos de aprendizaje profundo

El desarrollo de arquitecturas de aprendizaje profundo ha transformado la predicción de la demanda eléctrica al permitir la modelización de dependencias temporales complejas y relaciones no lineales multivariadas. Si bien estas arquitecturas han demostrado mejoras sustanciales en precisión, su implementación puede implicar mayores requerimientos computacionales y tiempos de entrenamiento prolongados,

especialmente en escenarios multivariados de alta dimensionalidad [39, 7].

Dentro de estas arquitecturas, las redes neuronales recurrentes y sus variantes han adquirido especial relevancia debido a su capacidad para modelar secuencias temporales. En particular, las redes Long Short-Term Memory (LSTM) introducen mecanismos de compuertas que permiten capturar dependencias de corto y largo plazo de manera más estable [58, 53]. Diversos estudios han demostrado su eficacia en la integración conjunta de datos históricos de demanda eléctrica y variables meteorológicas [22, 1].

3.3.1. Redes Neuronales Recurrentes

Las Redes Neuronales Artificiales (ANN, *del inglés Artificial Neural Networks*) constituyen modelos computacionales inspirados en el funcionamiento del cerebro humano, capaces de aprender relaciones complejas entre variables de entrada y salida mediante procesos de entrenamiento iterativo [48]. Si bien las redes neuronales de tipo *feedforward* pueden aplicarse a problemas de predicción de series temporales, su estructura no permite capturar dependencias secuenciales a lo largo del tiempo.

Las Redes Neuronales Recurrentes (RNN, *del inglés Recurrent Neural Networks*) fueron desarrolladas para abordar esta limitación mediante la incorporación de conexiones recurrentes que permiten mantener información de pasos temporales previos. Sin embargo, las RNN tradicionales presentan dificultades para aprender dependencias de largo plazo debido a problemas de desvanecimiento y explosión del gradiente durante el proceso de entrenamiento.

Las Memorias de Corto y Largo Plazo (LSTM, *del inglés Long-Short Term Memories*) fueron introducidas para superar estas limitaciones mediante la incorporación de una estructura de memoria interna basada en compuertas (*gates*) que regulan el flujo de información a través del tiempo [37, 43, 29]. Estas compuertas permiten al modelo decidir qué información debe conservarse, actualizarse o descartarse durante el procesamiento de la secuencia.

La Figura 3.2 muestra la estructura interna de una unidad LSTM, la cual incorpora un conjunto de compuertas que regulan el flujo de información a lo largo del tiempo. Estas compuertas permiten decidir qué información del pasado debe ser olvidada, qué nueva información debe ser incorporada y qué parte del estado interno

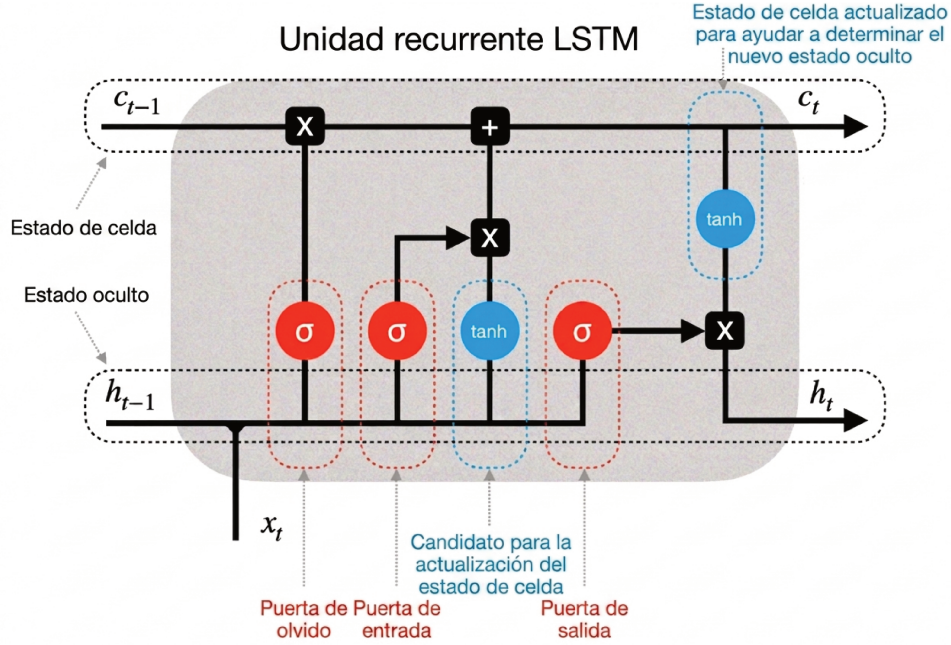


Figura 3.2: Estructura interna de una unidad LSTM, mostrando el flujo de información a través del estado de celda y las compuertas de olvido, entrada y salida.

se utiliza para generar la salida.

En particular, la compuerta de olvido determina qué parte del estado de la celda anterior se descarta, mientras que la compuerta de entrada controla la incorporación de nueva información proveniente de la entrada actual y del estado oculto previo. Esta información es combinada con un estado candidato generado mediante una función no lineal. Finalmente, la compuerta de salida define qué parte del estado de la celda se utiliza para calcular la salida del modelo en cada instante de tiempo.

Formalmente, la arquitectura de una celda LSTM incluye tres compuertas principales: la compuerta de olvido (*forget gate*), la compuerta de entrada (*input gate*) y la compuerta de salida (*output gate*). La función sigmoide $\sigma(\cdot)$ acota la salida de cada compuerta al intervalo $[0, 1]$, permitiendo interpretarla como una proporción de información que se conserva o descarta.

La compuerta de olvido determina qué parte del estado de memoria previo debe conservarse y se define como:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f), \quad (3.4)$$

donde W_f representa la matriz de pesos aplicada a la entrada x_t , U_f corresponde a la matriz de pesos recurrentes asociada al estado oculto previo h_{t-1} y b_f es el término

de sesgo de la compuerta.

La compuerta de entrada controla la cantidad de información nueva que se incorpora al estado de memoria y se define como:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i), \quad (3.5)$$

donde W_i y U_i corresponden a las matrices de pesos asociadas a la entrada actual y al estado oculto previo, respectivamente, y b_i representa el sesgo de la compuerta de entrada.

El estado candidato de la celda en el instante t se calcula como:

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C), \quad (3.6)$$

donde la función de activación tangente hiperbólica asegura que los valores resultantes se encuentren en el intervalo $[-1, 1]$.

El estado de memoria de la celda se actualiza combinando la información retenida del estado previo con la nueva información candidata:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (3.7)$$

donde $f_t \odot C_{t-1}$ representa la información retenida del estado de memoria previo e $i_t \odot \tilde{C}_t$ corresponde a la nueva información incorporada. El operador $\odot$ denota el producto elemento a elemento.

Finalmente, la compuerta de salida determina qué parte del estado de memoria será utilizada para generar la salida de la celda:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o), \quad (3.8)$$

donde el estado oculto h_t se calcula como:

$$h_t = o_t \odot \tanh(C_t). \quad (3.9)$$

y representa la salida de la celda LSTM en el instante t .

En conjunto, estas ecuaciones describen los mecanismos internos de memoria de

la arquitectura LSTM. La compuerta de olvido f_t regula la información proveniente del pasado, mientras que la compuerta de entrada i_t y el estado candidato $\tilde{C}_t$ controlan la incorporación de nueva información. El estado de memoria C_t integra ambos componentes, permitiendo preservar dependencias temporales de largo plazo. Finalmente, la compuerta de salida o_t determina qué parte de la memoria interna se expone para generar el estado oculto h_t .

Gracias a esta estructura, las redes LSTM pueden capturar relaciones temporales no lineales y dependencias de largo plazo presentes en series multivariadas, como en el caso de la demanda eléctrica. Esto convierte a este modelo en una herramienta particularmente adecuada para problemas de pronóstico energético influenciados por variables meteorológicas.

Las Gated Recurrent Units (GRU) constituyen una variante simplificada de las LSTM, reduciendo la complejidad computacional sin comprometer significativamente el desempeño predictivo [1, 12]. Asimismo, se han propuesto modelos híbridos que combinan arquitecturas recurrentes con técnicas de ensemble learning o algoritmos evolutivos, mostrando mejoras en escenarios de alta variabilidad [55]. Revisiones recientes destacan que las arquitecturas híbridas basadas en redes neuronales convolucionales (*Convolutional Neural Networks*, CNN) combinadas con LSTM han mostrado un desempeño competitivo en STLFF, aunque su complejidad estructural debe evaluarse en función del tamaño del conjunto de datos y los recursos disponibles [62, 35].

3.3.2. Modelos basados en mecanismos de atención

En los últimos años, las arquitecturas basadas en mecanismos de atención han ganado relevancia, permitiendo ponderar dinámicamente la importancia de distintas variables y pasos temporales dentro de la secuencia. Estos enfoques buscan superar limitaciones de las arquitecturas recurrentes tradicionales al capturar interacciones complejas entre múltiples fuentes de información [23].

Dentro de las arquitecturas recurrentes avanzadas, estos modelos han adquirido especial relevancia en la predicción de series temporales multivariadas. En particular, el modelo Dual-Stage Attention-Based Recurrent Neural Network (DA-RNN) introdujo un esquema de doble atención que permite ponderar dinámicamente tanto las

variables de entrada como los estados temporales relevantes dentro de la secuencia, mejorando la capacidad de capturar relaciones no lineales complejas. Sobre esta base, Huang et al. (2021) propusieron el modelo DA-SKIP, que incorpora un mecanismo recurrente adicional orientado a capturar patrones periódicos en series temporales, combinando atención dual con conexiones de salto (skip connections). Este enfoque, que se aborda con mayor detalle en la Sección 4.2, mostró mejoras en escenarios con componentes estacionales marcadas y periodicidad recurrente [32]. No obstante, la incorporación de bloques adicionales y mecanismos recurrentes múltiples puede incrementar la complejidad arquitectónica, lo cual plantea la necesidad de evaluar cuidadosamente el balance entre mejora predictiva y costo computacional.

En términos generales, los mecanismos de atención permiten asignar pesos dinámicos tanto a las variables de entrada como a los estados temporales relevantes dentro de la secuencia. En el contexto energético, modelos con atención en dos etapas han demostrado mejoras en la extracción adaptativa de características y en la modelización de dependencias de largo plazo [47, 49, 64]. Desde una perspectiva de diseño, la atención dual ofrece una estrategia estructural eficiente para incrementar la capacidad representativa del modelo sin recurrir necesariamente a incrementos significativos en profundidad o número de parámetros.

Estos desarrollos evidencian una evolución desde modelos recurrentes tradicionales hacia arquitecturas más estructuradas capaces de integrar información multivariada y patrones temporales de distinta escala mediante mecanismos de ponderación adaptativa. En particular, las estructuras de atención de doble etapa permiten al modelo identificar, en una primera instancia, qué variables exógenas resultan más relevantes en cada instante temporal. Posteriormente, permiten determinar qué pasos históricos deben priorizarse para la predicción, mitigando la degradación del gradiente en horizontes extendidos y mejorando la estabilidad del entrenamiento sin necesidad de incrementar excesivamente la profundidad arquitectónica [49].

3.4. Desafíos actuales en la predicción de demanda

A pesar de los avances en modelos de predicción de la demanda eléctrica, persisten desafíos relevantes que limitan su desempeño y aplicabilidad en contextos

operativos reales. Uno de los principales problemas es la degradación de la precisión a medida que aumenta el horizonte de predicción, como consecuencia de la acumulación de errores y de la incertidumbre asociada a variables exógenas, en particular las de origen meteorológico [58], cuya influencia se vuelve más crítica en horizontes medios y largos [17, 73, 67]. Asimismo, la predicción de eventos extremos o picos abruptos de demanda continúa siendo una tarea compleja, debido al carácter no lineal, no estacionario y altamente dependiente del contexto de estos fenómenos, lo que sigue representando un desafío incluso para modelos avanzados basados en aprendizaje profundo [14, 45].

Por otra parte, si bien los modelos de aprendizaje profundo han demostrado una elevada capacidad predictiva, presentan limitaciones en términos de interpretabilidad, lo que dificulta su adopción en entornos donde la trazabilidad de las decisiones es un requisito fundamental. En este contexto, la inteligencia artificial explicable (*Explainable Artificial Intelligence*, XAI) ha surgido como una línea de investigación orientada a proporcionar interpretaciones sobre el comportamiento de los modelos. Sin embargo, muchos de los enfoques existentes se basan en técnicas *post-hoc*, como SHAP o LIME, que operan de manera externa al modelo y presentan dificultades para capturar de forma integrada la influencia de variables exógenas en el proceso de predicción [51, 41].

En respuesta a estas limitaciones, se han propuesto arquitecturas que incorporan mecanismos de atención, las cuales permiten ponderar dinámicamente la relevancia de distintas variables y de diferentes instantes temporales. Estos enfoques ofrecen una vía prometedora para mejorar simultáneamente la precisión y la interpretabilidad de los modelos. No obstante, muchas de estas arquitecturas presentan una elevada complejidad estructural y requieren grandes volúmenes de datos y recursos computacionales, lo que puede limitar su aplicación en contextos regionales. La integración de mecanismos de atención en arquitecturas *encoder-decoder* ha demostrado ser efectiva tanto en el dominio del procesamiento del lenguaje natural como en la predicción de series temporales [6, 42].

En este marco, la literatura reciente ha explorado arquitecturas que incorporan mecanismos de atención en estructuras *encoder-decoder*, como el modelo DA-SKIP [32], el cual introduce un esquema de doble atención orientado a capturar tanto

la relevancia de las variables de entrada como las dependencias temporales en la serie. Este tipo de enfoques ha demostrado ser efectivo para el modelado de series multivariadas, aunque su complejidad y configuración no siempre resultan óptimas para escenarios con recursos limitados o conjuntos de datos de tamaño moderado. Tomando DA-SKIP como punto de partida, en el siguiente capítulo se proponen dos modelos orientados a mejorar la eficiencia, la interpretabilidad y la capacidad predictiva del modelo.

Capítulo 4

Modelos propuestos

Este capítulo presenta el desarrollo de los modelos propuestos para la predicción de la demanda eléctrica. En la Sección 4.1 se introduce el enfoque de modelado adoptado, describiendo el diseño de las arquitecturas propuestas a partir de un modelo de referencia de la literatura. En la Sección 4.2 se presenta el modelo DA-SKIP como punto de partida arquitectónico, detallando sus fundamentos y arquitectura. Posteriormente, las secciones siguientes describen dos alternativas propuestas para mejorar el desempeño del modelo base: DA-LSTM (Sección 4.3) y AMBER (Sección 4.4).

4.1. Introducción

A partir de los desafíos identificados en el capítulo anterior, este capítulo presenta el modelo base DA-SKIP y describe posteriormente dos nuevas arquitecturas de aprendizaje profundo propuestas para la predicción de la demanda eléctrica: *DA-LSTM* (Dual Attention Long-Short Term Memory) y *AMBER* (Attention-based Meteorological Bi-Attention Energy Regressor). En particular, el término *Bi-Attention* utilizado en el acrónimo AMBER hace referencia a la incorporación de dos mecanismos de atención diferenciados dentro de la arquitectura propuesta. Por un lado, el modelo emplea un esquema de atención asociado al bloque encoder-decoder inspirado en la arquitectura DA-RNN. Por otro lado, incorpora un mecanismo específico de atención meteorológica destinado a modelar la relevancia relativa de las variables exógenas históricas y futuras durante el proceso de predicción. En este contexto, el prefijo “Bi” no refiere a procesamiento bidireccional temporal, sino a la coexistencia

de dos componentes de atención dentro del modelo. El diseño de ambas propuestas se enmarca en la creciente adopción de mecanismos de atención en el modelado de secuencias, que han demostrado mejorar sustancialmente la capacidad de los modelos para ponderar de forma dinámica la información relevante [64, 6]. En línea con esta tendencia, tanto DA-LSTM como AMBER toman como punto de partida la arquitectura DA-SKIP y exploran, desde perspectivas diferentes, estrategias orientadas a mejorar el desempeño predictivo, la eficiencia del modelo y la incorporación de variables exógenas dentro del esquema de predicción.

La Figura 4.1 ilustra la evolución de las arquitecturas consideradas en esta tesis. El recorrido parte de las redes recurrentes clásicas (RNN, LSTM, GRU), incorpora luego los mecanismos de atención dual presentes en DA-RNN y DA-SKIP, y culmina con los modelos DA-LSTM y AMBER, desarrollados en el marco de esta tesis como alternativas superadoras al modelo base.

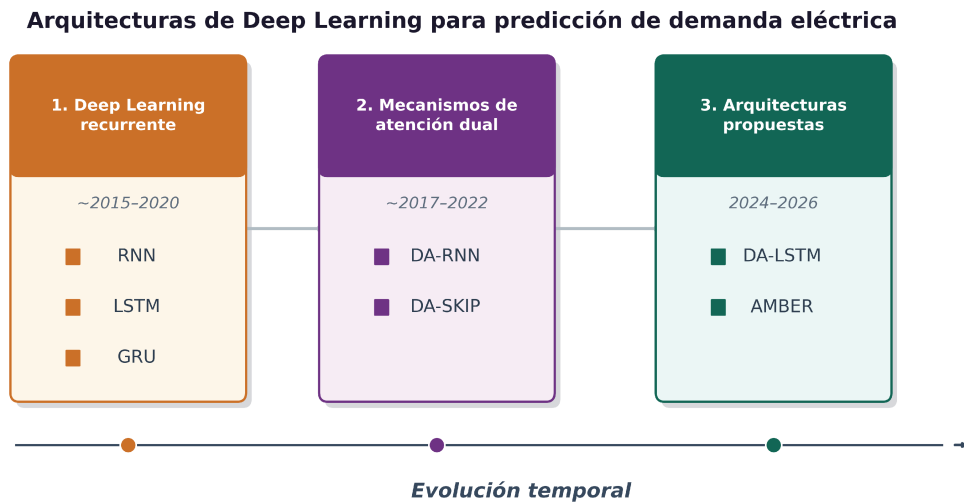


Figura 4.1: Evolución de las arquitecturas de aprendizaje profundo consideradas en este capítulo, desde las redes recurrentes hasta los modelos propuestos DA-LSTM y AMBER.

El diseño de los modelos propuestos se fundamenta en dos antecedentes clave de la literatura. Por un lado, el modelo DA-SKIP [32] introduce mecanismos de atención dual y conexiones de salto para preservar información temporal relevante en horizontes de largo alcance. Por otro lado, el modelo TimeXer [66, 70] propone una arquitectura basada en atención que integra explícitamente variables exógenas, permitiendo modelar de manera conjunta la dinámica de la serie objetivo y la influencia de factores externos. Ambos enfoques inspiran las arquitecturas propuestas

en este trabajo.

El desarrollo de los modelos propuestos en esta tesis parte inicialmente de una exploración sistemática de arquitecturas LSTM aplicadas a la predicción de demanda eléctrica, realizada sobre el conjunto de datos multivariado considerado en este trabajo. Esta primera etapa, de carácter experimental, permitió identificar configuraciones robustas capaces de superar consistentemente a modelos clásicos y a arquitecturas recurrentes más complejas reportadas en la literatura. A partir de estos resultados, se incorporó el modelo del estado del arte DA-SKIP [32] como referencia arquitectónica para el diseño de las propuestas de esta tesis, dada su capacidad para integrar mecanismos de doble atención y modelado periódico de series temporales.

Sobre esta base, se proponen posteriormente dos alternativas arquitectónicas independientes. La primera corresponde al modelo DA-LSTM, en el cual los componentes GRU-SKIP y autorregresivo del DA-SKIP son reemplazados por un bloque LSTM, con el objetivo de capturar de manera unificada las dinámicas no lineales de la serie temporal bajo el esquema de doble atención heredado de DA-RNN [50]. La segunda alternativa es el modelo AMBER, que incorpora un mecanismo de atención específico para variables meteorológicas, considerando tanto la información histórica como la correspondiente al horizonte de predicción, con el fin de mejorar la integración de variables exógenas en el proceso de predicción.

Un aspecto central en el diseño de AMBER es la interpretabilidad. Más allá de la precisión predictiva, los entornos operativos de planificación energética requieren comprender cómo influyen las variables exógenas en la demanda [51, 52]. Los mecanismos de atención empleados en la arquitectura propuesta ofrecen una forma de explicabilidad intrínseca, al permitir visualizar y analizar la relevancia relativa de cada variable en cada instante del proceso de predicción.

En síntesis, los modelos DA-LSTM y AMBER han sido diseñados como propuestas alternativas para la predicción de la demanda eléctrica diaria bajo un enfoque multivariado que integra variables temporales, energéticas y meteorológicas. Mientras DA-LSTM simplifica la arquitectura de DA-SKIP unificando sus componentes bajo un bloque LSTM con doble atención, AMBER extiende dicho esquema incorporando un mecanismo de atención específico sobre variables meteorológicas, brindando tanto un pronóstico de demanda energética como la relevancia relativa

de las variables de entrada. Las salidas de los componentes de cada arquitectura se integran para producir la predicción final, tal como se describe en detalle en las secciones siguientes.

4.2. Punto de partida: modelo DA-SKIP

El modelo DA-SKIP se basa en un enfoque que divide la tarea de predicción de series temporales en tres componentes fundamentales: una componente no lineal, una componente lineal y una componente periódica [32]. En particular, la componente no lineal es modelada mediante una arquitectura encoder-decoder basada en DA-RNN [50], para capturar dependencias complejas entre variables. La componente lineal es abordada mediante un modelo autorregresivo, mientras que la componente periódica es modelada a través del bloque GRU-SKIP, diseñado para capturar patrones recurrentes en la serie temporal.

La elección de este modelo como base se fundamenta en su capacidad para integrar de manera explícita distintos patrones temporales presentes en series reales, combinando en una única arquitectura la modelación de dinámicas no lineales, tendencias lineales y comportamientos periódicos. Este enfoque resulta particularmente atractivo en el contexto de la predicción de la demanda eléctrica, donde coexisten efectos estacionales, tendencias y relaciones complejas con variables exógenas, tal como se mostró en el Capítulo 2.

Dado que la implementación original del modelo DA-SKIP no se encuentra disponible públicamente, fue necesario reconstruir, codificar e implementar su arquitectura a partir de la descripción presentada en la literatura para poder evaluar el modelo sobre el conjunto de datos considerado en esta tesis. Este proceso permitió no solo reproducir su funcionamiento, sino también analizar en profundidad el rol de cada uno de sus componentes y estudiar su comportamiento en el contexto específico del problema abordado.

4.2.1. Arquitectura del modelo DA-SKIP

El modelo DA-SKIP es una extensión del esquema DA-RNN [50], incorporando módulos adicionales para capturar la periodicidad y la componente lineal de la serie

temporal [32]. Su arquitectura puede interpretarse como la combinación de tres bloques complementarios:

1. un bloque encoder–decoder con atención, encargado de modelar relaciones no lineales y dependencias temporales de largo alcance;
2. un bloque GRU-SKIP, orientado a capturar patrones periódicos mediante conexiones de salto;
3. un bloque autorregresivo lineal, destinado a modelar relaciones lineales simples en la serie objetivo.

La estructura general del modelo DA-SKIP se presenta en la Figura 4.2, donde se observan los distintos bloques que componen la arquitectura y su interacción en el proceso de predicción.

Sea $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{T-1}\}$ la serie multivariante de entrada, donde cada $\mathbf{x}_t \in \mathbb{R}^n$ representa el vector de variables exógenas en el instante t , y sea $\mathbf{y} = \{y_1, y_2, \dots, y_{T-1}\}$ la serie objetivo. El objetivo del modelo es predecir el valor y_T a partir de la información contenida en ambas secuencias. La salida final del modelo se obtiene mediante la suma de la predicción generada por el bloque encoder–decoder y la producida por el bloque GRU-SKIP junto con el componente autorregresivo.

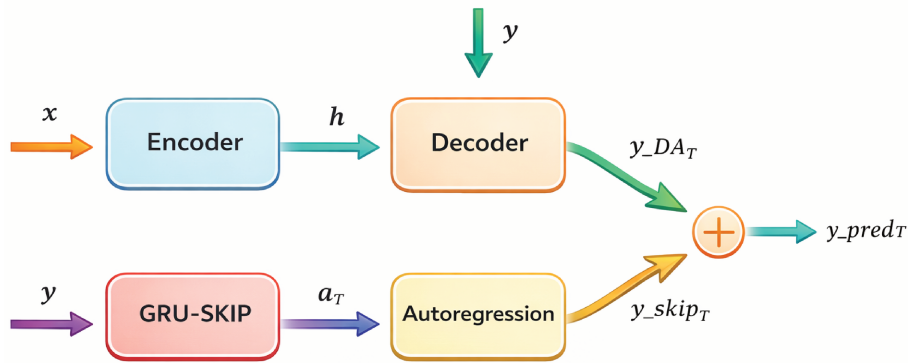


Figura 4.2: Arquitectura general del modelo DA-SKIP (adaptado de [32]).

Encoder con atención de entrada

El encoder del modelo DA-SKIP hereda la formulación básica propuesta en DA-RNN [50]. Su objetivo es transformar la secuencia multivariante de entrada en una representación latente que preserve la información más relevante para la predicción.

Específicamente, el encoder procesa secuencialmente $\mathbf{X}$ y genera una secuencia

de estados ocultos $\mathbf{H} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T\}$, donde cada $\mathbf{h}_t$ resume la información relevante acumulada hasta el instante t . En forma general, la dinámica del encoder puede expresarse como:

$$\mathbf{h}_t = f_1(\mathbf{h}_{t-1}, \tilde{\mathbf{x}}_t), \quad (4.1)$$

donde $\mathbf{h}_{t-1}$ es el estado oculto previo, $\tilde{\mathbf{x}}_t$ es la entrada ponderada por atención en el instante t , y $f_1(\cdot)$ representa la transformación no lineal implementada por la unidad LSTM del encoder. La celda LSTM del encoder se define mediante las siguientes ecuaciones:

$$\mathbf{s}_t = \mathbf{f}_t \odot \mathbf{s}_{t-1} + \mathbf{u}_t \odot \tanh(\mathbf{W}_s[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_s), \quad (4.2)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_f), \quad (4.3)$$

$$\mathbf{u}_t = \sigma(\mathbf{W}_u[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_u), \quad (4.4)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}; \mathbf{x}_t] + \mathbf{b}_o), \quad (4.5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{s}_t), \quad (4.6)$$

donde:

- $\mathbf{s}_t$ es el estado de celda en el instante t , que almacena la memoria interna de la LSTM;
- $\mathbf{f}_t$ es la compuerta de olvido, que controla qué proporción de la memoria previa $\mathbf{s}_{t-1}$ se conserva;
- $\mathbf{u}_t$ es la compuerta de actualización o entrada, que regula cuánta información nueva se incorpora al estado de celda;
- $\mathbf{o}_t$ es la compuerta de salida, que determina qué parte del estado de celda se utiliza para construir el estado oculto;
- $\mathbf{W}_s$, $\mathbf{W}_f$, $\mathbf{W}_u$ y $\mathbf{W}_o$ son matrices de pesos entrenables;
- $\mathbf{b}_s$, $\mathbf{b}_f$, $\mathbf{b}_u$ y $\mathbf{b}_o$ son vectores de sesgo;
- $[\mathbf{h}_{t-1}; \mathbf{x}_t]$ denota la concatenación del estado oculto previo con la entrada actual;

- $\sigma(\cdot)$ representa la función sigmoide;
- $\tanh(\cdot)$ es la función tangente hiperbólica;
- $\odot$ denota el producto elemento a elemento.

A diferencia de un encoder recurrente convencional, DA-SKIP incorpora un mecanismo de atención de entrada, cuyo propósito es asignar una importancia relativa distinta a cada variable de entrada en cada instante temporal. De este modo, el modelo puede seleccionar dinámicamente cuáles variables resultan más relevantes para actualizar el estado oculto.

Sea $X^k = \{x_1^k, x_2^k, \dots, x_T^k\}$ la serie temporal correspondiente a la variable k , con $k = 1, \dots, n$. Para cada instante t y para cada variable k , se calcula primero una puntuación de atención:

$$e_t^k = \mathbf{V}_e^\top \tanh(\mathbf{W}_e[\mathbf{h}_{t-1}; \mathbf{s}_{t-1}] + \mathbf{U}_e \mathbf{x}^k), \quad (4.7)$$

donde:

- e_t^k es la puntuación no normalizada asociada a la variable k en el instante t ;
- $\mathbf{h}_{t-1}$ y $\mathbf{s}_{t-1}$ son, respectivamente, el estado oculto y el estado de celda previos del encoder;
- $\mathbf{x}^k$ representa la trayectoria temporal asociada a la variable k ;
- $\mathbf{W}_e$, $\mathbf{U}_e$ y $\mathbf{V}_e$ son parámetros entrenables del mecanismo de atención. Estos permiten combinar la información temporal previa del encoder con las variables de entrada, generando un puntaje de relevancia que determina la importancia relativa de cada variable en cada instante temporal.

Estas puntuaciones se normalizan luego mediante una función *softmax*:

$$\alpha_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^n \exp(e_t^i)}, \quad (4.8)$$

donde α_t^k representa el peso de atención asignado a la variable k en el instante t . Por construcción, se cumple que $\sum_{k=1}^n \alpha_t^k = 1$, por lo que los coeficientes α_t^k pueden interpretarse como una distribución de relevancia entre las variables de entrada.

A partir de estos pesos, la entrada ponderada del encoder se define como:

$$\tilde{\mathbf{x}}_t = (\alpha_t^1 x_t^1, \alpha_t^2 x_t^2, \dots, \alpha_t^n x_t^n)^\top. \quad (4.9)$$

En esta expresión, cada componente de $\mathbf{x}_t$ es reescalado por su correspondiente peso de atención, de manera que el encoder recibe una versión adaptativa de la entrada original, enfatizando las variables más influyentes en cada paso temporal. Como resultado, el encoder genera la secuencia de estados ocultos $\mathbf{H}$, que resume la evolución temporal de la serie y constituye la entrada principal del decoder.

Decoder con atención temporal

El decoder recibe como entrada la secuencia de estados ocultos del encoder, $\mathbf{H}$, junto con la serie objetivo histórica $\mathbf{y} = \{y_1, y_2, \dots, y_{T-1}\}$. Su función es construir una representación contextual capaz de seleccionar los instantes más relevantes de la secuencia codificada para producir la predicción final y_T .

Al igual que en el encoder, el decoder se implementa mediante una LSTM. Sus estados ocultos y de celda se denotan como $\mathbf{h}'_t$ y $\mathbf{s}'_t$, respectivamente. Sin embargo, en este caso el mecanismo de atención no actúa sobre las variables de entrada, sino sobre la dimensión temporal, es decir, sobre los estados ocultos generados por el encoder.

Para cada estado oculto $\mathbf{h}_i$ del encoder, se calcula una puntuación de atención temporal dada por:

$$l_t^i = \mathbf{V}_d^\top \tanh(\mathbf{W}_d[\mathbf{h}'_{t-1}; \mathbf{s}'_{t-1}] + \mathbf{U}_d \mathbf{h}_i), \quad (4.10)$$

donde:

- l_t^i es la puntuación no normalizada asociada al estado oculto $\mathbf{h}_i$ del encoder en el paso t del decoder;
- $\mathbf{h}'_{t-1}$ y $\mathbf{s}'_{t-1}$ son el estado oculto y el estado de celda previos del decoder;
- $\mathbf{h}_i$ es el i -ésimo estado oculto producido por el encoder;
- $\mathbf{W}_d$, $\mathbf{U}_d$ y $\mathbf{V}_d$ son parámetros entrenables del mecanismo de atención temporal. Estos permiten combinar la información previa del decoder con los estados

ocultos generados por el encoder, produciendo un puntaje de relevancia que determina qué instantes temporales históricos resultan más importantes para la predicción.

Luego, estas puntuaciones se normalizan mediante *softmax*:

$$\beta_t^i = \frac{\exp(l_t^i)}{\sum_{j=1}^T \exp(l_t^j)}, \quad (4.11)$$

donde β_t^i representa el peso de atención temporal asignado al estado oculto $\mathbf{h}_i$. Al igual que en el encoder, estos coeficientes satisfacen $\sum_{i=1}^T \beta_t^i = 1$.

Con estos pesos se construye el vector de contexto temporal:

$$\mathbf{c}_t = \sum_{i=1}^T \beta_t^i \mathbf{h}_i. \quad (4.12)$$

El vector $\mathbf{c}_t$ resume, en el paso t del decoder, la información más relevante de toda la secuencia codificada por el encoder. En otras palabras, permite que el decoder se enfoque selectivamente en aquellos instantes pasados que resultan más informativos para la predicción.

A continuación, el valor objetivo previo y el vector de contexto se combinan linealmente para formar la entrada corregida del decoder:

$$\tilde{y}_{t-1} = \mathbf{w}^\top [y_{t-1}; \mathbf{c}_{t-1}] + b, \quad (4.13)$$

donde:

- y_{t-1} es el valor previo de la serie objetivo;
- $\mathbf{c}_{t-1}$ es el vector de contexto calculado en el paso anterior;
- $\mathbf{w}$ es un vector de pesos entrenable;
- b es un sesgo escalar.

Esta combinación permite integrar en un único valor tanto la historia reciente de la variable objetivo como la información contextual proveniente del encoder.

Las ecuaciones internas de la LSTM del decoder son:

$$\mathbf{s}'_t = \mathbf{f}'_t \odot \mathbf{s}'_{t-1} + \mathbf{u}'_t \odot \tanh(\mathbf{W}'_s [\mathbf{h}'_{t-1}; \tilde{y}_{t-1}] + \mathbf{b}'_s), \quad (4.14)$$

$$\mathbf{f}'_t = \sigma(\mathbf{W}'_f[\mathbf{h}'_{t-1}; \tilde{y}_{t-1}] + \mathbf{b}'_f), \quad (4.15)$$

$$\mathbf{u}'_t = \sigma(\mathbf{W}'_u[\mathbf{h}'_{t-1}; \tilde{y}_{t-1}] + \mathbf{b}'_u), \quad (4.16)$$

$$\mathbf{o}'_t = \sigma(\mathbf{W}'_o[\mathbf{h}'_{t-1}; \tilde{y}_{t-1}] + \mathbf{b}'_o), \quad (4.17)$$

$$\mathbf{h}'_t = \mathbf{o}'_t \odot \tanh(\mathbf{s}'_t). \quad (4.18)$$

En estas ecuaciones, $\mathbf{f}'_t$, $\mathbf{u}'_t$ y $\mathbf{o}'_t$ representan, respectivamente, las compuertas de olvido, actualización y salida del decoder; $\mathbf{s}'_t$ es el estado de celda; y $\mathbf{h}'_t$ es el estado oculto correspondiente al instante t .

Finalmente, la salida del bloque encoder–decoder, que representa la componente no lineal del modelo DA-SKIP, se calcula como:

$$y_{DA_T} = \mathbf{V}_y^\top (\mathbf{W}_y[\mathbf{h}'_T; \mathbf{c}_T] + \mathbf{b}_w) + b_v. \quad (4.19)$$

Aquí, y_{DA_T} representa la predicción generada por el bloque DA-RNN; $\mathbf{h}'_T$ es el último estado oculto del decoder; $\mathbf{c}_T$ es el último vector de contexto temporal; $\mathbf{W}_y$ y $\mathbf{V}_y$ son parámetros entrenables; y $\mathbf{b}_w$, b_v son términos de sesgo. Esta salida constituye la contribución no lineal del modelo completo, la cual posteriormente se combina con las componentes periódica y lineal de DA-SKIP.

Bloque GRU-SKIP

El bloque GRU-SKIP fue incorporado en DA-SKIP con el objetivo de capturar explícitamente patrones periódicos presentes en la serie temporal objetivo. A diferencia de un modelo recurrente convencional, este componente introduce conexiones de salto (*skip connections*) que permiten modelar dependencias entre observaciones separadas por un período fijo.

Para ello, se define una secuencia periódica o *jumping sequence*, construida a partir de valores históricos de la serie objetivo y_t , separados por un intervalo constante k :

$$\mathbf{p} = \{y_{T-km}, y_{T-k(m-1)}, \dots, y_{T-k}\}, \quad km \leq T. \quad (4.20)$$

En esta expresión:

- $\mathbf{p}$ representa la secuencia periódica utilizada como entrada al bloque GRU-

SKIP;

- y_t es el valor de la serie objetivo en el instante t ;
- T es el último instante temporal disponible;
- k es el período de salto, que define la distancia entre observaciones (por ejemplo, $k = 7$ para capturar patrones semanales);
- m es la cantidad de saltos considerados, tal que $km \leq T$.

Sobre esta secuencia se aplica una red GRU modificada, en la cual el estado oculto en el instante t depende del estado oculto ubicado k pasos antes, es decir, $\mathbf{a}_{t-k}$, en lugar de $\mathbf{a}_{t-1}$ como en una GRU estándar.

La dinámica interna de la GRU-SKIP se define mediante las siguientes ecuaciones. En primer lugar, se calcula el estado candidato:

$$\hat{\mathbf{a}}_t = \tanh(\mathbf{W}_a[\mathbf{r}_t \odot \mathbf{a}_{t-k}; y_t] + \mathbf{b}_a), \quad (4.21)$$

donde:

- $\hat{\mathbf{a}}_t$ es el estado oculto candidato en el instante t ;
- $\mathbf{a}_{t-k}$ es el estado oculto previo con salto de k pasos;
- $\mathbf{r}_t$ es la compuerta de reinicio;
- y_t es la entrada en el instante t (valor de la serie objetivo);
- $[\cdot; \cdot]$ denota concatenación de vectores;
- $\mathbf{W}_a$ es una matriz de pesos entrenable;
- $\mathbf{b}_a$ es un vector de sesgo;
- $\odot$ representa el producto elemento a elemento;
- $\tanh(\cdot)$ es la función tangente hiperbólica.

La compuerta de actualización se calcula como:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z[\mathbf{a}_{t-k}; y_t] + \mathbf{b}_z), \quad (4.22)$$

y la compuerta de reinicio como:

$$\mathbf{r}_t = \sigma(\mathbf{W}_r[\mathbf{a}_{t-k}; y_t] + \mathbf{b}_r), \quad (4.23)$$

donde:

- $\mathbf{z}_t$ es la compuerta de actualización, que controla cuánto del nuevo estado candidato se incorpora al estado final;
- $\mathbf{r}_t$ es la compuerta de reinicio, que regula la influencia del estado oculto previo;
- $\mathbf{W}_z, \mathbf{W}_r$ son matrices de pesos entrenables;
- $\mathbf{b}_z, \mathbf{b}_r$ son vectores de sesgo;
- $\sigma(\cdot)$ es la función sigmoide.

El estado oculto final se obtiene como una combinación convexa entre el estado candidato y el estado previo con salto:

$$\mathbf{a}_t = \mathbf{z}_t \odot \hat{\mathbf{a}}_t + (1 - \mathbf{z}_t) \odot \mathbf{a}_{t-k}. \quad (4.24)$$

En esta ecuación:

- $\mathbf{a}_t$ es el estado oculto final en el instante t ;
- el término $\mathbf{z}_t \odot \hat{\mathbf{a}}_t$ representa la información nueva incorporada;
- el término $(1 - \mathbf{z}_t) \odot \mathbf{a}_{t-k}$ preserva la información periódica previa;
- la compuerta $\mathbf{z}_t$ actúa como un mecanismo de balance entre ambas contribuciones.

La salida periódica del bloque GRU-SKIP se obtiene proyectando linealmente el último estado oculto:

$$y_{skip_T} = \mathbf{W}_j \mathbf{a}_T + b_j, \quad (4.25)$$

donde:

- y_{skip_T} es la contribución del componente periódico a la predicción final;
- $\mathbf{a}_T$ es el último estado oculto de la GRU-SKIP;
- $\mathbf{W}_j$ es una matriz de pesos entrenable;
- b_j es un término de sesgo escalar.

Componente autorregresivo lineal

Además del modelado no lineal y periódico, DA-SKIP incorpora un bloque autorregresivo lineal cuyo objetivo es capturar dependencias lineales simples de la serie

temporal, que podrían no ser modeladas eficientemente por los componentes no lineales. Cuando este componente se activa, la salida del bloque periódico se modifica como:

$$y_{skip_T} = \mathbf{W}_j \mathbf{a}_T + b_j + \mathbf{W}_i \mathbf{y} + b_i, \quad (4.26)$$

donde:

- $\mathbf{y}$ representa un vector de valores históricos recientes de la serie objetivo;
- $\mathbf{W}_i$ es un vector de pesos asociado al modelo autorregresivo;
- b_i es un término de sesgo;
- el término $\mathbf{W}_i \mathbf{y}$ captura la dependencia lineal directa entre valores pasados y el valor futuro.

Salida final del modelo

La predicción final del modelo DA-SKIP se obtiene mediante la suma de las contribuciones de sus distintos componentes:

$$y_{pred_T} = y_{DA_T} + y_{skip_T}. \quad (4.27)$$

En esta expresión:

- y_{pred_T} es la predicción final en el instante T ;
- y_{DA_T} corresponde a la salida del bloque encoder-decoder (componente no lineal);
- y_{skip_T} integra las contribuciones periódicas y lineales.

De este modo, el modelo combina explícitamente tres tipos de patrones: no lineales, periódicos y lineales. Es decir, DA-SKIP combina en una misma arquitectura tres mecanismos complementarios: atención sobre variables de entrada, atención temporal sobre los estados codificados, y modelado explícito de patrones periódicos y lineales. Esta combinación motivó su elección como arquitectura base para el desarrollo de la propuesta. Sin embargo, como se analizará en la siguiente subsección, la interacción entre estos componentes también introduce limitaciones cuando el modelo se aplica al problema específico de predicción de demanda eléctrica abordado en esta tesis.

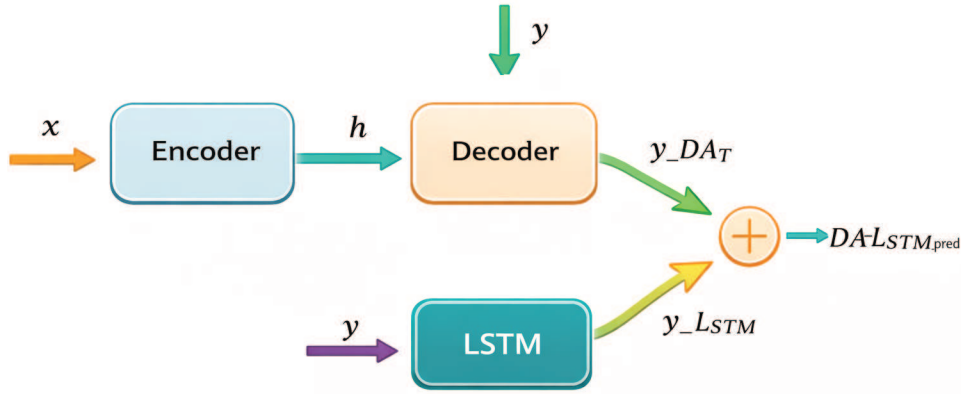


Figura 4.3: Arquitectura del modelo DA-LSTM.

4.3. Modelo DA-LSTM

La evaluación del modelo DA-SKIP sobre el conjunto de datos considerado en esta tesis evidenció un desempeño inferior al alcanzado por una arquitectura LSTM de menor complejidad. Esta evidencia motivó la propuesta de DA-LSTM, una variante más compacta que retiene el esquema encoder–decoder con doble atención e incorpora un bloque LSTM en reemplazo de los componentes GRU-SKIP y autorregresivo.

La arquitectura resultante se presenta en la Figura 4.3 y se compone de tres bloques principales: un encoder con atención sobre las variables de entrada, un decoder con atención temporal, y un componente recurrente basado en LSTM para la predicción final. Los dos primeros bloques mantienen la formulación presentada en la sección anterior.

Componente encoder

El encoder recibe como entrada un tensor $\mathbf{X} \in \mathbb{R}^{B \times T \times N}$, donde B es el tamaño del lote, T la cantidad de pasos temporales y N el número de variables. El tamaño del lote se selecciona durante el entrenamiento; en la inferencia puede utilizarse $B = 1$ sin modificar los parámetros del modelo. A partir de esta entrada, el encoder genera una secuencia de estados ocultos $\mathbf{H} = \{\mathbf{h}_1, \dots, \mathbf{h}_T\}$ de dimensión M , utilizando una LSTM con atención sobre las variables de entrada, tal como se describió en la Sección 4.2.

Componente decoder

Como se describió en la Sección 4.2, el decoder recibe como entrada la secuencia de estados ocultos del encoder junto con la serie objetivo histórica, y genera una representación latente mediante una LSTM con atención temporal. Como resultado, se obtiene una secuencia de estados ocultos $\mathbf{H}' = \{\mathbf{h}'_1, \dots, \mathbf{h}'_T\}$ de dimensión P , que sintetiza la información relevante para la predicción.

Componente LSTM y predicción final

La principal modificación introducida en DA-LSTM es la incorporación de un componente recurrente LSTM, cuya función es recibir la representación generada por el decoder y realizar la predicción final del modelo. El componente LSTM está compuesto por una o más capas, definidas por los parámetros *hidden_dim* y *num_layers*, permitiendo ajustar la capacidad del modelo en función de la complejidad del problema. La formulación matemática utilizada en este componente corresponde a la presentada en la Sección 3.3, por lo que no se repiten aquí sus ecuaciones.

Finalmente, el último estado oculto, $\mathbf{h}_T^c$, se proyecta mediante una transformación lineal para obtener las F predicciones del horizonte:

$$\hat{\mathbf{y}}_{T+1:T+F} = \mathbf{W}_{out} \mathbf{h}_T^c + \mathbf{b}_{out}. \quad (4.28)$$

Aquí, $\mathbf{W}_{out}$ y $\mathbf{b}_{out}$ son los parámetros de la capa de salida. En los experimentos se consideró $F = 1, 7$ y 14 , y las predicciones de cada horizonte se generaron conjuntamente. De este modo, la arquitectura DA-LSTM simplifica la estructura original de DA-SKIP manteniendo capacidad de modelado temporal, al integrar en un único componente recurrente la representación de dinámicas no lineales y dependencias temporales, eliminando la necesidad de módulos adicionales específicos para periodicidad o linealidad.

4.4. Modelo AMBER

4.4.1. Motivación para la incorporación de atención sobre variables meteorológicas

En el problema de predicción de la demanda eléctrica, las variables meteorológicas desempeñan un rol central, no solo en la explicación del comportamiento histórico de la serie, sino también en la determinación de su evolución futura. En particular, la disponibilidad de información meteorológica en el horizonte de predicción constituye una fuente adicional de información que, como se analizará en detalle en la Sección 5.4.1 del capítulo siguiente, no está siendo plenamente explotada por la arquitectura DA-LSTM.

Además de contribuir a la precisión predictiva, la incorporación de mecanismos de atención sobre variables exógenas permite obtener información interpretable acerca de la relevancia relativa de cada variable meteorológica para la predicción. Esto representa un valor agregado significativo, ya que posibilita analizar cuáles son las variables meteorológicas que ejercen mayor influencia sobre la demanda eléctrica en distintos horizontes temporales, aportando una capa adicional de interpretabilidad que no está presente en los modelos previamente evaluados.

En este contexto, el análisis de enfoques recientes en la literatura, como el modelo TimeXer [66], pone de manifiesto la relevancia de incorporar mecanismos capaces de modelar explícitamente la interacción entre variables endógenas y exógenas. En dichos enfoques, los mecanismos de atención permiten asignar pesos dinámicos a las variables externas, favoreciendo una integración más efectiva de su contribución en el proceso de predicción.

Aquí se propone extender la arquitectura DA-LSTM mediante la incorporación de un mecanismo de atención específico sobre variables meteorológicas. A diferencia de las arquitecturas basadas en Transformer, el mecanismo propuesto se diseña de manera acorde a la estructura del modelo desarrollado, priorizando una formulación más simple y computacionalmente eficiente. Esta modificación da lugar al modelo AMBER, que integra la información meteorológica histórica y la correspondiente al horizonte futuro. En los experimentos, estos valores futuros provienen de ERA5.

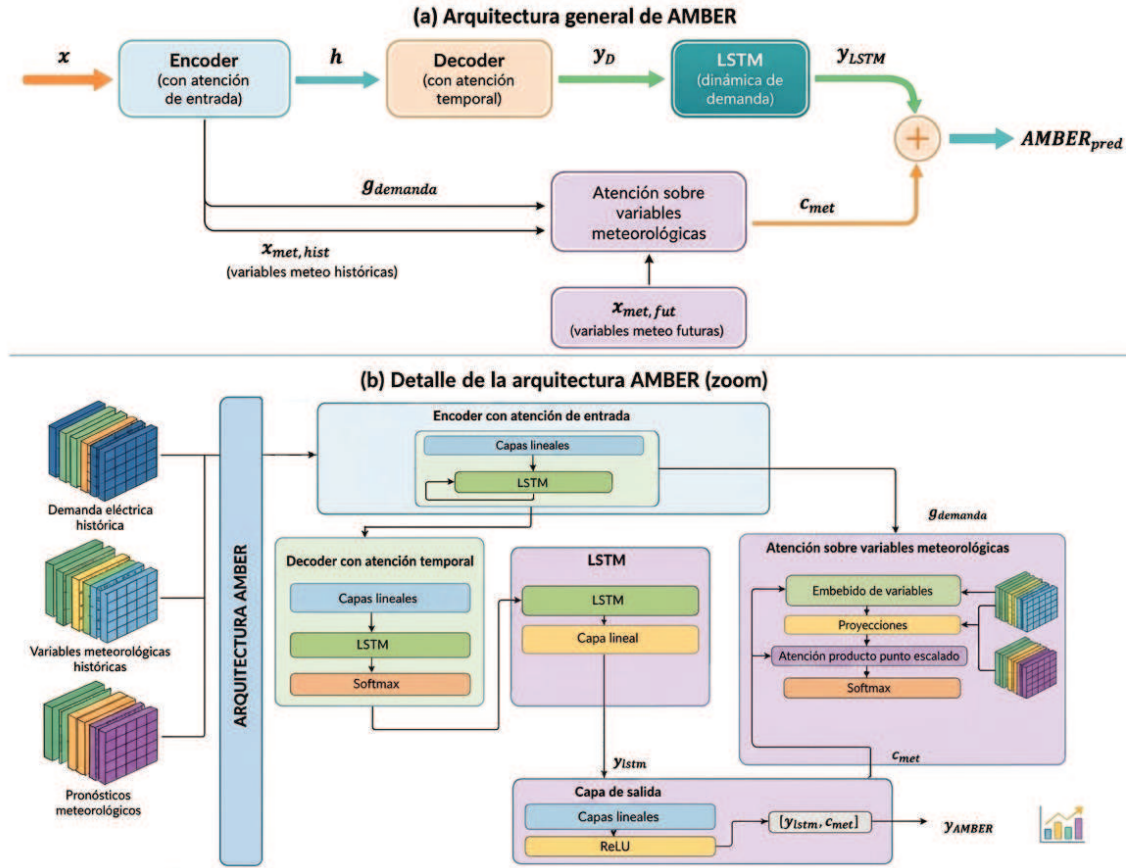


Figura 4.4: Arquitectura del modelo propuesto, AMBER. (a) Estructura general del modelo y el flujo completo de procesamiento desde las variables de entrada hasta la predicción final de la demanda eléctrica. (b) Desglose modular de la arquitectura, detallando la organización interna de los componentes de atención (temporal, de entrada y meteorológica) y los bloques basados en LSTM.

4.4.2. Arquitectura del modelo AMBER

La arquitectura del modelo AMBER se presenta en la Figura 4.4. Este modelo extiende la estructura de DA-LSTM mediante la incorporación de un mecanismo de atención específico sobre variables meteorológicas, permitiendo integrar información histórica de la demanda eléctrica con variables meteorológicas correspondientes al horizonte de predicción.

El modelo se compone de cuatro bloques principales: un encoder con atención sobre variables de entrada, un decoder con atención temporal, un bloque recurrente LSTM para el modelado de la dinámica de la demanda y un componente de atención sobre variables meteorológicas.

Encoder - Decoder

El encoder y el decoder mantienen la formulación previamente descrita. A partir de la salida del decoder, se aplica el bloque LSTM, cuya función es modelar la evolución temporal de la demanda en el horizonte de predicción. La salida de este bloque se denota como:

$$\mathbf{y}_{lstm} \in \mathbb{R}^{B \times F}, \quad (4.29)$$

donde B es el tamaño del batch y F el horizonte de predicción.

Componente de atención sobre variables meteorológicas

El componente de atención meteorológica tiene como objetivo modelar la relevancia relativa de las variables meteorológicas, considerando tanto la información histórica como los valores de ERA5 correspondientes al horizonte futuro.

Sea $\mathbf{X}_{met}^{hist} \in \mathbb{R}^{B \times T \times K}$ la matriz de variables meteorológicas históricas y $\mathbf{X}_{met}^{fut} \in \mathbb{R}^{B \times F \times K}$ la correspondiente al horizonte de predicción, donde B representa el tamaño del lote (*batch size*), K es el número de variables meteorológicas, T la ventana histórica y F el horizonte futuro.

Ambas matrices se concatenan en la dimensión temporal:

$$\mathbf{X}_{met} = [\mathbf{X}_{met}^{hist}, \mathbf{X}_{met}^{fut}] \in \mathbb{R}^{B \times (T+F) \times K}. \quad (4.30)$$

Luego, la representación es reorganizada:

$$\mathbf{X}_{met}^T \in \mathbb{R}^{B \times K \times (T+F)}, \quad (4.31)$$

de modo que cada variable meteorológica queda representada como una secuencia temporal completa. En particular, $\mathbf{x}_k \in \mathbb{R}^{T+F}$ representa la secuencia asociada a la variable meteorológica k . Sobre esta representación, cada variable es proyectada a un espacio latente mediante una transformación lineal seguida de una función no lineal:

$$\mathbf{e}_k = \phi(\mathbf{W}_{emb}\mathbf{x}_k + \mathbf{b}_{emb}), \quad (4.32)$$

donde $\mathbf{e}_k \in \mathbb{R}^{d_{model}}$ es la representación embebida de la variable k .

Mecanismo de atención

Se define un vector de consulta:

$$\mathbf{q} = \mathbf{h}_T, \quad (4.33)$$

donde $\mathbf{h}_T$ es el último estado oculto del encoder, y a partir de este vector se construyen:

$$\mathbf{Q} = \mathbf{W}_Q \mathbf{q}, \quad \mathbf{K} = \mathbf{W}_K \mathbf{E}, \quad \mathbf{V}_{val} = \mathbf{W}_V \mathbf{E}, \quad (4.34)$$

donde:

- $\mathbf{E}$ contiene las representaciones embebidas de todas las variables;
- $\mathbf{Q}$ (query) representa el estado actual de la demanda;
- $\mathbf{K}$ (keys) describe cada variable meteorológica;
- $\mathbf{V}_{val}$ (values) contiene la información a combinar.

Los pesos de atención se calculan como:

$$\alpha = \text{softmax} \left(\frac{\mathbf{QK}^\top}{\sqrt{d_{model}}} \right), \quad (4.35)$$

y el contexto meteorológico como:

$$\mathbf{c}_{met} = \alpha \mathbf{V}_{val}. \quad (4.36)$$

Predicción final

La salida del modelo se obtiene integrando la dinámica de la demanda y la información meteorológica. En particular, ambas representaciones se combinan mediante concatenación:

$$\mathbf{z} = [\mathbf{y}_{lstm}; \mathbf{c}_{met}]. \quad (4.37)$$

Finalmente, la predicción se obtiene mediante un bloque de fusión implementado como una red neuronal *feedforward*:

$$\hat{\mathbf{y}}_{AMBER} = f_{fusion}(\mathbf{z}), \quad (4.38)$$

donde $f_{fusion}(\cdot)$ corresponde a una red compuesta por capas lineales y una función de activación no lineal intermedia.

Este mecanismo permite integrar de manera explícita la información meteorológica y la dinámica temporal de la demanda mediante una aproximación funcional cuyos parámetros se optimizan de forma *end-to-end* (es decir, de manera conjunta y directa a partir de los datos, sin etapas de optimización independientes) durante el entrenamiento. La incorporación conjunta de información histórica y meteorológica futura enriquece la representación del modelo, favoreciendo una mejor adaptación frente a variaciones meteorológicas y cambios en la dinámica temporal de la demanda. Como se evidenciará en el capítulo siguiente, esta estrategia no solo mejora el desempeño predictivo respecto de DA-LSTM en todos los horizontes considerados, sino que además otorga al modelo una mayor robustez frente al incremento del horizonte de predicción, aspecto en el que los modelos basados únicamente en arquitecturas recurrentes muestran una degradación más pronunciada. Asimismo, el mecanismo de atención meteorológica permite identificar de manera transparente qué variables exógenas resultan más relevantes en cada escenario de predicción, aportando una dimensión interpretable que complementa la capacidad predictiva del modelo.

Capítulo 5

Experimentos y resultados

Este capítulo presenta los experimentos realizados y los resultados obtenidos mediante los modelos propuestos en esta tesis para la predicción de la demanda eléctrica, comparando su desempeño con el de modelos de referencia. En la Sección 5.1 se describe el conjunto de datos utilizado y los principales aspectos de implementación. La Sección 5.2 detalla la metodología de evaluación y validación, incluyendo las métricas empleadas y la estrategia de validación adoptada. En la Sección 5.3 se presentan los experimentos realizados, diferenciando entre modelos clásicos y modelos de aprendizaje profundo. Finalmente, en la Sección 5.4 se comparan las diferentes variantes propuestas en esta tesis.

5.1. Conjunto de datos utilizado

El conjunto de datos utilizado en este trabajo corresponde al período comprendido entre el 1 de enero de 2012 y el 31 de mayo de 2023 (4169 días), compuesto por registros de la demanda eléctrica diaria total de Entre Ríos y variables meteorológicas representativas de la provincia. Las variables de energía fueron provistas por ENERSA y las variables meteorológicas fueron obtenidas del reanálisis atmosférico ERA5 y promediadas espacialmente sobre el territorio provincial (más detalles en la Sección 2.2).

El conjunto de datos final está conformado por 21 variables de entrada y la variable objetivo del modelo, correspondiente al consumo eléctrico diario. La Tabla 5.1 presenta el detalle de ellas, organizadas en tres categorías principales: variables temporales, variables energéticas y variables meteorológicas. Los nombres abreviados

Tabla 5.1: Resumen de las variables de entrada y salida utilizadas para la predicción de la demanda eléctrica

Categoría	Tipo	Variable	Descripción
Entrada	Temporal	Año	Año de observación (2012–2023)
		Mes	Mes del año (1–12)
		Día	Día del mes (1–31)
		Día codificado	Día de la semana codificado (1–8)*
	Energética	Usuarios	Cantidad de usuarios del servicio
		Temperatura mínima (tmin)	Temperatura mínima del aire a 2 m
		Temperatura media (tmean)	Temperatura media del aire a 2 m
		Temperatura máxima (tmax)	Temperatura máxima del aire a 2 m
	Meteorológica	Temperatura de superficie (skt)	Temperatura media en superficie
		Radiación solar entrante (ssrd)	Rad. de onda corta entrante en sup.
		Radiación térmica entrante (strd)	Rad. de onda larga entrante en sup.
		Radiación solar neta (ssr)	Balance neto de radiación de onda corta
		Radiación térmica neta (str)	Balance neto de radiación de onda larga
		Flujo de calor latente (slhf)	Flujo de calor latente en superficie
		Flujo de calor sensible (sshf)	Flujo de calor sensible en superficie
		Precipitación total (tp)	Precipitación acumulada diaria
		Escorrentía superficial (sro)	Escurrimiento superficial de agua
		Presión superficial (sp)	Presión atmosférica en superficie
		Cobertura nubosa (tcc)	Fracción de cobertura de nubes
		Velocidad del viento (ws)	Velocidad del viento a 10 m
		Humedad relativa (rh)	Humedad relativa
Salida	Objetivo	Demanda energética	Consumo eléctrico diario provincial

* 1=lunes, ..., 7=domingo, 8=feriado

de las variables meteorológicas corresponden a la nomenclatura utilizada en ERA5, basada en siglas o abreviaturas en inglés (ej., *ssrd* = *surface solar radiation downwards*).

La selección de variables meteorológicas se realizó con el objetivo de representar de manera más completa la influencia de las condiciones meteorológicas sobre la demanda eléctrica. Si bien numerosos trabajos se centran en la temperatura y las variables temporales, estudios previos muestran que el consumo puede estar condicionado por múltiples variables meteorológicas y por relaciones no lineales entre ellas [63, 4, 30]. Por ello se incluyeron variables térmicas, pero también las variables principales de los balances de radiación, de agua y de energía, junto a variables atmosféricas como viento y humedad que inciden en la sensación térmica.

Por otra parte, la construcción de los datos de entrada se realizó mediante ventanas deslizantes, utilizando una HTW de 7 días para predecir FTW de 1, 7 y 14 días, de acuerdo con las definiciones introducidas en la Sección 2.4. En todos los modelos, las ventanas combinaron la información histórica con los valores meteorológicos correspondientes al horizonte futuro. En esta evaluación retrospectiva, esos valores del

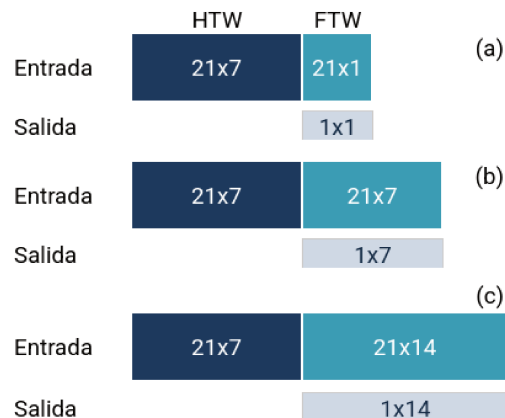


Figura 5.1: Esquema de construcción de patrones. (a) FTW=1. (b) FTW=7. (c) FTW=14.

horizonte futuro provienen de ERA5, por lo que los resultados no incluyen el error de un pronóstico meteorológico operativo. La Figura 5.1 ilustra la generación de los patrones para los tres horizontes.

Aspectos de Implementación

La implementación de los modelos se realizó en lenguaje Python, utilizando librerías estándar para análisis de series temporales y aprendizaje automático, entre ellas *NumPy*, *Pandas*, *Scikit-learn*, *Statsmodels* y *PyTorch* para los modelos de aprendizaje profundo.

Los experimentos correspondientes a los modelos clásicos y a las arquitecturas base de aprendizaje profundo, incluyendo SARIMAX, LSTM, DA-SKIP y AMBER fueron ejecutados en una notebook DELL Latitude 3520, equipada con un procesador Intel Core i7-1165G7, 32 GB de memoria RAM y un disco de estado sólido de 736 GB, bajo sistema operativo Fedora.

Debido al mayor costo computacional asociado al entrenamiento y exploración de hiperparámetros de la arquitectura más compleja, los experimentos correspondientes al modelo DA-LSTM fueron ejecutados utilizando GPUs NVIDIA GeForce RTX 3090 disponibles en servidores externos, incluyendo infraestructura del instituto sinc(i) y otros recursos de cómputo de colegas. Esta estrategia permitió disponer de una mayor capacidad de procesamiento para los experimentos de aprendizaje profundo y reducir los tiempos de entrenamiento.

5.2. Metodología de evaluación y validación

5.2.1. Métricas para la evaluación del desempeño

Para evaluar el desempeño de los modelos se emplearon métricas ampliamente utilizadas en problemas de predicción de series temporales, las cuales permiten cuantificar diversos aspectos de la capacidad del modelo para simular la demanda eléctrica [33, 69].

Raíz del error cuadrático medio (RMSE)

La raíz del error cuadrático medio (*Root Mean Squared Error*) mide la magnitud promedio del error de predicción, penalizando en mayor medida los errores de mayor magnitud. Se define como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (5.1)$$

donde y_i representa el valor real y $\hat{y}_i$ el valor predicho. Esta métrica se expresa en las mismas unidades que la variable objetivo (MWh), lo que facilita su interpretación en términos operativos.

Coefficiente de determinación (R^2)

El coeficiente de determinación evalúa la proporción de la varianza de la variable objetivo que es explicada por el modelo. Se calcula como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5.2)$$

donde $\bar{y}$ es el valor medio de la serie observada. Valores de R^2 cercanos a 1 indican un alto desempeño, mientras que valores cercanos a 0 reflejan un bajo desempeño, pudiendo incluso tomar valores negativos. A diferencia del coeficiente de correlación lineal, R^2 evalúa el error de las predicciones respecto de la variabilidad observada.

Índice de concordancia de Willmott (WI)

El índice de concordancia de Willmott (*Willmott's Index*) mide el grado de acuerdo entre los valores observados y predichos, considerando tanto los errores de predicción como la variabilidad de la serie. Se define como:

$$WI = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2}, \quad (5.3)$$

El valor de WI varía entre 0 y 1, donde valores cercanos a 1 indican una alta concordancia entre las predicciones y los valores reales.

Sesgo porcentual (%Bias)

El porcentaje de sesgo evalúa la tendencia sistemática del modelo a sobreestimar o subestimar la variable objetivo. Se calcula como:

$$\%Bias = 100 \cdot \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{\sum_{i=1}^n y_i}, \quad (5.4)$$

Valores positivos indican una sobreestimación promedio, mientras que valores negativos reflejan una subestimación del modelo.

Error absoluto medio (MAE)

El error absoluto medio (*Mean Absolute Error*) mide la magnitud promedio de los errores de predicción, sin considerar su dirección. Se define como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (5.5)$$

donde y_i representa el valor real y $\hat{y}_i$ el valor predicho. A diferencia del RMSE, el MAE pondera todos los errores por igual, lo que lo hace menos sensible a valores extremos. Se expresa en las mismas unidades que la variable objetivo (MWh).

Error porcentual absoluto medio (MAPE)

El error porcentual absoluto medio (*Mean Absolute Percentage Error*) expresa el error de predicción en términos relativos, independientemente de la escala de la

variable objetivo. Se calcula como:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (5.6)$$

Valores más bajos indican un mejor desempeño. Al expresarse en porcentaje, facilita la comparación entre distintos horizontes de predicción o conjuntos de datos con diferente escala. Para FTW=7 y FTW=14, las métricas se calcularon considerando todos los pasos de cada horizonte y no únicamente el último.

5.2.2. Metodología de validación

Particionamiento

Para la evaluación de los modelos, se utilizó una estrategia de partición temporal del conjunto de datos. En la Figura 5.2, el panel superior ilustra la dinámica de la demanda eléctrica a lo largo de todo el período, evidenciando una alta variabilidad, caracterizada por mayores niveles de consumo durante los meses de verano e invierno. El panel inferior presenta un zoom al año 2019, donde se observan las dinámicas semanales de la demanda, con un mayor consumo durante los días hábiles y una marcada disminución durante los fines de semana.

Considerando estas dinámicas temporales de la demanda eléctrica, el conjunto de datos se dividió inicialmente de forma cronológica, conservando continuidad temporal: un 80 % para entrenamiento y un 20 % para prueba (ver panel superior de la Figura 5.2). Esta partición fija fue utilizada en todos los experimentos desarrollados en esta tesis.

El conjunto de entrenamiento abarca desde enero de 2012 hasta abril de 2022, mientras que el conjunto de prueba comprende desde mayo de 2022 hasta mayo de 2023. La partición cronológica preserva la estructura temporal de los datos en ambos subconjuntos, incluyendo la marcada estacionalidad de la demanda eléctrica. Mantener esta coherencia temporal resulta fundamental para capturar adecuadamente las tendencias estacionales y otros patrones dependientes del tiempo que influyen en el desempeño predictivo.

Antes del entrenamiento de los modelos, todas las variables de entrada fueron escaladas al rango $[0, 1]$ mediante normalización Min–Max, utilizando exclusivamente

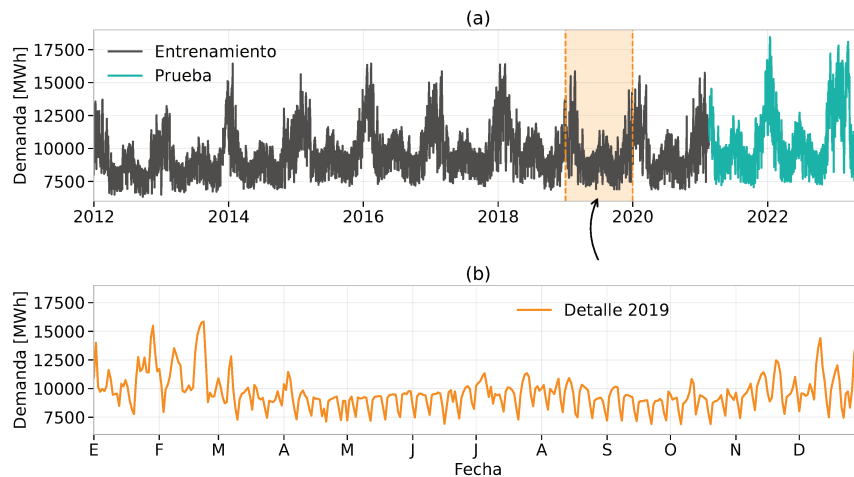


Figura 5.2: Partición temporal del conjunto de datos utilizada en los experimentos. (a) Serie temporal completa de demanda eléctrica, mostrando la división cronológica entre los subconjuntos de entrenamiento y prueba, junto con el período resaltado correspondiente al año 2019. (b) Detalle de la dinámica de la demanda eléctrica durante el año 2019.

las estadísticas del conjunto de entrenamiento. Posteriormente, la misma transformación fue aplicada al conjunto de prueba. Finalmente, las predicciones obtenidas fueron reescaladas a las unidades originales para su interpretación.

Este proceso resulta fundamental en el entrenamiento de redes neuronales artificiales, dado que los algoritmos de optimización basados en gradiente son sensibles a la escala de las variables de entrada. En ausencia de normalización, las variables con magnitudes elevadas pueden dominar el proceso de aprendizaje, generando gradientes inestables y afectando negativamente el desempeño del modelo. La normalización Min-Max permite equilibrar la contribución relativa de cada variable, favoreciendo una convergencia más estable durante el entrenamiento. Este enfoque es ampliamente utilizado en estudios de predicción de demanda energética basados en redes neuronales [24, 38].

Validación cruzada

Los métodos de remuestreo constituyen una herramienta fundamental en el aprendizaje, ya que permiten estimar el desempeño de un modelo a partir de los datos disponibles sin necesidad de disponer de un conjunto de prueba independiente de gran tamaño [34, 8]. En particular, la validación cruzada se utiliza para aproximar el error de generalización, es decir, el error esperado al aplicar el modelo sobre

nuevas observaciones no utilizadas durante el entrenamiento.

En este contexto, resulta importante distinguir entre el error de entrenamiento y el error de prueba. Mientras que el primero mide el ajuste del modelo sobre los datos utilizados para su entrenamiento, el segundo refleja su capacidad de generalización. En la práctica, el error de entrenamiento tiende a subestimar el error de prueba, por lo que es necesario recurrir a estrategias que permitan obtener estimaciones más realistas del desempeño del modelo [34].

Una de las técnicas más utilizadas con este propósito es la validación cruzada de tipo k -fold. En su formulación clásica, el conjunto de datos se divide en k subconjuntos de tamaño aproximadamente uniforme. En cada iteración, uno de estos subconjuntos se utiliza como conjunto de validación, mientras que los $k - 1$ restantes se emplean para entrenar el modelo. Este procedimiento se repite k veces, de modo que cada subconjunto actúa una única vez como conjunto de validación. La estimación final del error se obtiene promediando los errores individuales calculados en cada partición [34].

En este trabajo se empleó una estrategia de *5-fold cross-validation*, aplicada exclusivamente sobre el conjunto de entrenamiento definido previamente (80 % inicial de la serie temporal). De este modo, el conjunto de prueba permaneció completamente aislado durante todo el proceso de ajuste y selección de hiperparámetros.

Tal como se describió en la sección anterior, las particiones utilizadas en la validación cruzada respetaron la estructura cronológica de la serie temporal, evitando que información futura intervenga en el ajuste de los modelos. Este enfoque redujo la dependencia de una única partición durante la selección de hiperparámetros y mantuvo aislado el conjunto de prueba. No obstante, no se evaluó la variabilidad asociada a distintas inicializaciones de los pesos.

Finalmente, el entrenamiento de los modelos se llevó a cabo minimizando el error cuadrático medio (*Mean Squared Error*, *MSE*), incorporando además un mecanismo de parada temprana (*early stopping*) con el objetivo de evitar el sobreajuste y mejorar la capacidad de generalización [24].

Tabla 5.2: Configuración de búsqueda en grilla para SARIMAX

Parámetro	Valores considerados
p (autoregresivo)	$\{7\}$
d (diferenciación)	$\{0, 1, 2\}$
q (media móvil)	$\{1, 2, 3, 7, 14\}$

5.3. Determinación de hiperparámetros

La determinación de hiperparámetros constituyó una etapa central del proceso experimental desarrollado en esta tesis. Debido a la complejidad y sensibilidad de los modelos de aprendizaje profundo frente a diferentes configuraciones arquitectónicas y de entrenamiento, se implementó un proceso progresivo de exploración y refinamiento de hiperparámetros.

En una primera etapa se realizaron exploraciones preliminares orientadas a comprender el comportamiento de las arquitecturas propuestas bajo diferentes configuraciones temporales y de entrenamiento. A partir de los resultados obtenidos, se definieron posteriormente espacios de búsqueda más acotados y específicos para cada modelo, permitiendo llevar a cabo evaluaciones sistemáticas mediante búsqueda en grilla y validación cruzada. En todos los experimentos realizados se mantuvo una ventana temporal histórica fija (HTW) de 7 días y se consideraron horizontes futuros de predicción (FTW) de 1, 7 y 14 días, con el objetivo de garantizar condiciones homogéneas de comparación entre modelos.

5.3.1. SARIMAX

Para este modelo se fijó $p = 7$ para representar el ciclo semanal de la demanda y se evaluaron todas las combinaciones de d y q indicadas en la Tabla 5.2, incorporando variables exógenas meteorológicas.

Como resultado del proceso de exploración de hiperparámetros, la configuración que presentó el mejor desempeño para el horizonte FTW=1 correspondió a ($p = 7, d = 0, q = 1$), alcanzando un R^2 de 0.40 y un RMSE de 1878.25 MWh en el conjunto de prueba. La Tabla 5.3 resume las mejores configuraciones obtenidas para cada horizonte futuro evaluado, junto con sus métricas de desempeño correspondientes.

Tabla 5.3: Mejores configuraciones obtenidas para el modelo SARIMAX

FTW	p	d	q	RMSE [MWh]	R^2
1	7	0	1	1878.2	0.40
7	7	0	7	1866.1	0.41
14	7	0	14	1803.1	0.45

El análisis del espacio de hiperparámetros muestra que, dentro de las configuraciones con $d = 0$, los resultados fueron relativamente estables para valores cercanos de q . Por otro lado, las configuraciones con $d \geq 1$ produjeron resultados inferiores; por lo tanto, $d = 0$ fue la mejor opción dentro de las configuraciones evaluadas.

Estos resultados corresponden al espacio de hiperparámetros evaluado. La comparación con los modelos no lineales se presenta en las secciones siguientes. La leve disminución del RMSE entre FTW=1 y FTW=14 no implica una mejora con el horizonte, ya que cada FTW fue optimizado de manera independiente y las métricas reúnen todos sus pasos.

5.3.2. LSTM

El modelo LSTM constituyó la primera arquitectura de aprendizaje profundo evaluada de manera sistemática en esta investigación y sirvió como punto de partida para la definición del esquema experimental adoptado posteriormente. En una etapa preliminar se realizaron exploraciones orientadas a comprender el comportamiento del modelo bajo distintas configuraciones temporales y de entrenamiento. A partir de los resultados obtenidos se definió el esquema experimental común presentado en la Sección 5.3, utilizado posteriormente para la evaluación y comparación sistemática entre modelos.

Esta etapa demandó un extenso proceso experimental que incluyó la evaluación sistemática de múltiples configuraciones arquitectónicas y de entrenamiento: distintas profundidades de red, dimensiones de la capa oculta, tasas de aprendizaje, tamaños de batch y criterios de parada temprana. A lo largo de este proceso se acumularon evidencias sobre el comportamiento del modelo frente a diferentes condiciones, permitiendo identificar patrones de estabilidad y degradación que orientaron las decisiones de diseño adoptadas en etapas posteriores. Entre los hallazgos más relevantes se destaca que configuraciones con tamaños de batch elevados tendían a

Tabla 5.4: Configuración de búsqueda en grilla para LSTM

Hiperparámetro	Valores considerados
Número de capas	{1, 2, 3}
Dimensión oculta (HD)	{5, 10, 15, 20, 25, 30, 35, 40, 45, 50}
Tasa de aprendizaje (lr)	{0,0001, 0,001}
Tamaño de batch	{4, 8, 10, 16, 32}
Paciencia (Early Stopping)	{10, 50, 100, 150}

Tabla 5.5: Mejores configuraciones obtenidas para el modelo LSTM

FTW	Batch	Capas	HD	lr	E.S.	RMSE [MWh]	R^2
1	8	1	15	0.001	10	727.2	0.94
7	8	3	10	0.001	50	848.4	0.91
14	8	1	10	0.001	100	1696.8	0.66

introducir inestabilidad durante el entrenamiento y a deteriorar la capacidad de generalización, motivo por el cual las etapas subsiguientes priorizaron configuraciones más compactas. Asimismo, el optimizador Adam mostró un comportamiento consistentemente más robusto que SGD, por lo que fue adoptado como estándar en todos los experimentos posteriores.

A partir de esta etapa preliminar, se implementó un esquema sistemático y automatizado de evaluación experimental. Todos los modelos LSTM fueron entrenados utilizando un máximo de 300 épocas y aplicando el criterio de *early stopping* para reducir el riesgo de sobreajuste durante el entrenamiento. Asimismo, se evaluaron los algoritmos de optimización SGD y Adam [36]. A partir de los resultados obtenidos, el optimizador Adam mostró un comportamiento más robusto y estable durante el entrenamiento, alcanzando una convergencia más consistente y un mejor desempeño predictivo, por lo que fue adoptado en los experimentos posteriores.

Luego, se llevó a cabo una búsqueda exhaustiva en grilla (*grid search*) evaluando todas las combinaciones de número de capas, cantidad de neuronas ocultas, tasa de aprendizaje, tamaño de batch y paciencia del mecanismo de *early stopping*. Los valores considerados se resumen en la Tabla 5.4.

La Tabla 5.5 resume las mejores configuraciones obtenidas para el modelo LSTM. Los hiperparámetros de mejor desempeño mantuvieron algunos valores comunes entre los distintos horizontes de predicción. En particular, la tasa de aprendizaje óptima fue de 0.001 y el tamaño de batch fue de 8 en los tres casos evaluados. Las

Tabla 5.6: Configuración de búsqueda en grilla para DA-SKIP

Hiperparámetro	Valores considerados
Tamaño de salto periódico (<i>skip</i>)	{7, 14}
Dimensión encoder (M)	{15, 20, 32}
Dimensión decoder (P)	{15, 20, 32}
Tasa de aprendizaje (lr)	{0,0005, 0,001, 0,08}
Tamaño de batch	{8, 16}
Épocas	{300}
Paciencia (early stopping)	{50}

diferencias se observaron principalmente en la arquitectura interna y en la paciencia del mecanismo de *early stopping*: para el horizonte de 1 día, la mejor configuración correspondió a 1 capa con 15 neuronas y una paciencia de 10; para el horizonte de 7 días, a 3 capas con 10 neuronas y una paciencia de 50; y para el horizonte de 14 días, a 1 capa con 10 neuronas y una paciencia de 100.

5.3.3. DA-SKIP

El modelo DA-SKIP fue implementado siguiendo la configuración propuesta en la literatura [32]. Dado que los autores no detallan exhaustivamente el espacio de búsqueda de hiperparámetros, se definió una búsqueda en grilla adaptada al esquema experimental de este trabajo. En particular, para el hiperparámetro *skip* se evaluaron valores de 7 y 14, definidos en función de la longitud de la ventana temporal histórica, con el objetivo de permitir al modelo capturar dependencias periódicas asociadas a uno y dos ciclos semanales. Estas configuraciones temporales fueron adoptadas también en los modelos desarrollados posteriormente, con el fin de mantener criterios homogéneos de evaluación. La Tabla 5.6 presenta las configuraciones de hiperparámetros exploradas para DA-SKIP.

La Tabla 5.7 resume las mejores configuraciones obtenidas para cada horizonte. En todos los casos se utilizó el optimizador Adam, un máximo de 300 épocas y una paciencia de 50 para el mecanismo de *early stopping*.

5.3.4. DA-LSTM

Como etapa previa a la evaluación formal del modelo DA-LSTM, se llevó a cabo una exploración experimental inicial tomando como punto de partida las configu-

Tabla 5.7: Mejores configuraciones obtenidas para el modelo DA-SKIP

FTW	M	P	lr	Batch	RMSE [MWh]	R^2
1	15	15	0.001	8	942.3	0.83
7	32	32	0.001	8	1832.5	0.34
14	32	15	0.001	8	1883.1	0.31

raciones de hiperparámetros que habían mostrado mejor desempeño en el modelo LSTM utilizado como línea base. Adicionalmente, se consideraron configuraciones reportadas para el modelo DA-RNN [50], cuyas decisiones de diseño fueron utilizadas como referencia para definir los espacios iniciales de búsqueda.

Durante esta etapa exploratoria se evaluaron distintos algoritmos de optimización, incluyendo Adam, AdamW y Adagrad con estrategias de *learning rate decay*, así como variantes con regularización mediante *weight decay*. Los resultados obtenidos mostraron un comportamiento más estable utilizando el optimizador Adam sin mecanismos adicionales de decaimiento ni regularización explícita. Asimismo, se observó que tasas de aprendizaje reducidas tendían a generar estancamiento durante el entrenamiento, con escasa mejora en las métricas de desempeño. En contraste, configuraciones con tasas de aprendizaje moderadas permitieron una convergencia más efectiva.

Por otra parte, configuraciones con dimensiones elevadas en los módulos encoder y decoder evidenciaron comportamientos más inestables y menor capacidad de generalización. En consecuencia, arquitecturas intermedias, particularmente configuraciones con dimensiones $(M, P) = (20, 20)$, mostraron un desempeño más robusto y estable para el modelo DA-LSTM. A partir de estas exploraciones preliminares se definieron los espacios de búsqueda utilizados en la evaluación formal de este modelo, presentados en la Tabla 5.8.

La evaluación de las configuraciones definidas en la Tabla 5.8 permitió identificar las combinaciones de hiperparámetros de mejor desempeño para cada horizonte de predicción, resumidas en la Tabla 5.9. Se observa una configuración relativamente estable en las dimensiones del encoder y decoder, mientras que las principales variaciones se asocian al número de capas y a la dimensión del componente LSTM. En particular, horizontes más extensos tendieron a requerir una mayor capacidad de representación temporal.

Tabla 5.8: Configuración de búsqueda en grilla para DA-LSTM

Hiperparámetro	Valores considerados
Número de capas	{1, 2, 3}
Dimensión oculta (HD)	{5, 10, 15}
Dimensión encoder (M)	{20, 32}
Dimensión decoder (P)	{20, 32}
Tasa de aprendizaje (lr)	{0,0005, 0,001, 0,005}
Tamaño de batch	{8, 14, 16, 32, 64}
Épocas	{300, 500}
Paciencia (early stopping)	{80, 150}

Tabla 5.9: Mejores configuraciones obtenidas para el modelo DA-LSTM

FTW	Capas	HD	M	P	lr	RMSE [MWh]	R^2
1	1	10	32	20	0.0010	892.4	0.87
7	2	10	32	20	0.0005	1296.2	0.41
14	1	15	32	20	0.0010	1356.0	0.35

5.3.5. AMBER

En el caso del modelo AMBER, las exploraciones iniciales se realizaron tomando como base las configuraciones previamente validadas para el modelo DA-LSTM. De este modo, se mantuvo el optimizador Adam y se conservaron las configuraciones de entrenamiento que habían mostrado mayor estabilidad en las etapas previas. Esta decisión permitió focalizar el análisis experimental en el impacto del mecanismo de atención meteorológica incorporado en AMBER, evitando introducir variabilidad adicional asociada a cambios simultáneos en la estrategia de optimización.

En particular, las exploraciones se concentraron en el análisis de distintas configuraciones asociadas al bloque de atención meteorológica, incluyendo la dimensión interna de representación y el tamaño de la capa de fusión. Asimismo, se evaluaron variantes con diferentes dimensiones ocultas en el componente LSTM y distintas profundidades de la arquitectura. Los resultados obtenidos durante esta etapa exploratoria permitieron identificar configuraciones más estables y parsimoniosas, las cuales fueron posteriormente utilizadas para definir los espacios de búsqueda considerados en la evaluación formal del modelo AMBER, presentados en la Tabla 5.10.

La evaluación de las configuraciones definidas en la Tabla 5.10 permitió identificar las de mejor desempeño, resumidas en la Tabla 5.11. Los módulos encoder y decoder

Tabla 5.10: Configuración de búsqueda en grilla para el modelo AMBER

Hiperparámetro	Valores considerados
Número de capas	{1, 2, 3}
Dimensión oculta (HD)	{5, 10, 15}
Dimensión encoder (M)	{20, 32}
Dimensión decoder (P)	{20, 32}
Tasa de aprendizaje (lr)	{0,0005, 0,001}
Tamaño de batch	{16}
Dimensión atención meteorológica	{16, 24, 32}
Dimensión capa de fusión	{16, 24, 32}
Épocas	{500}
Paciencia (early stopping)	{150}

Tabla 5.11: Mejores configuraciones obtenidas para el modelo AMBER: métricas en conjunto de validación

FTW	Capas	HD	M	P	Atención	Fusión	lr	RMSE [MWh]	R^2
1	1	10	20	20	24	32	0.0005	529.7	0.90
7	1	10	20	20	24	24	0.0010	767.2	0.79
14	2	15	20	20	32	32	0.0010	843.58	0.74

mantuvieron configuraciones similares, mientras que para FTW=14 se seleccionaron mayores dimensiones en el componente LSTM, la atención meteorológica y la capa de fusión. Este resultado describe las configuraciones elegidas, pero no permite aislar el efecto de cada hiperparámetro.

En términos comparativos, DA-LSTM representa una mejora respecto de modelos previos como DA-SKIP, particularmente para horizontes cortos e intermedios. Sobre esta arquitectura base, la incorporación del mecanismo de atención meteorológica en AMBER permitió mejorar de forma consistente la capacidad explicativa y el desempeño predictivo, especialmente a medida que aumenta el horizonte de predicción.

5.4. Análisis comparativo y discusión de resultados

SARIMAX

La Figura 5.3 muestra la comparación entre la demanda observada y las predicciones del modelo SARIMAX optimizado para FTW=7. Se observa que el modelo tiende a suavizar la serie y presenta dificultades para reproducir los picos estacio-

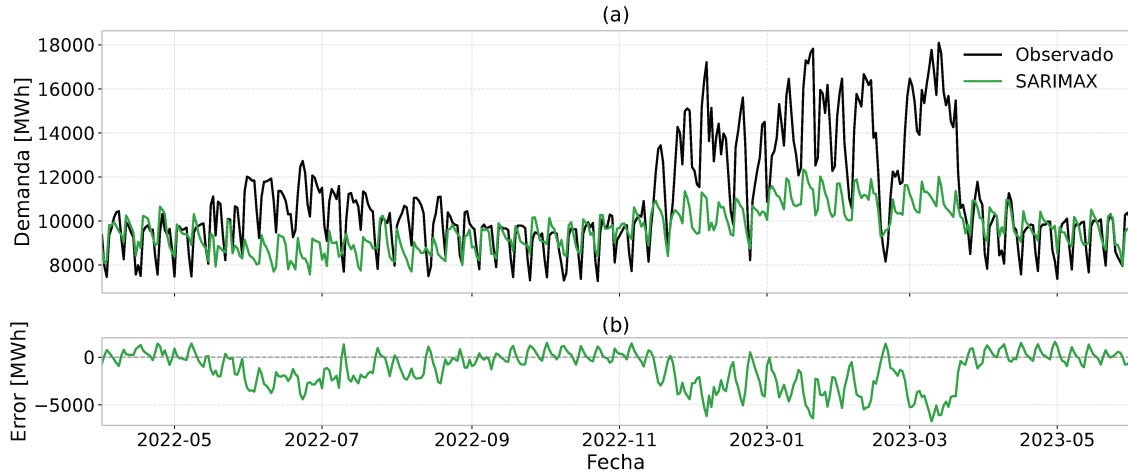


Figura 5.3: Comparación entre la demanda observada y las predicciones generadas por el modelo SARIMAX. (a) Muestra la serie temporal observada y las predicciones del modelo, mientras que (b) presenta el error de predicción correspondiente.

nales y la dinámica semanal del consumo. Si bien representa el aumento al inicio de la semana, anticipa la caída de la demanda antes de los fines de semana, lo que se refleja en el error del panel inferior.

Los resultados obtenidos evidencian una capacidad limitada para representar la dinámica de la demanda eléctrica, alcanzando un coeficiente de determinación de $R^2 = 0,41$ y un error $RMSE = 1866,1$ MWh. Asimismo, el sesgo negativo es de $-11,6\%$, lo que indica una subestimación sistemática de la demanda. Este comportamiento refleja la dificultad del modelo para capturar dinámicas no lineales asociadas a variaciones meteorológicas, efectos estacionales y períodos de alta demanda, lo que motivó la evaluación de modelos basados en aprendizaje profundo.

5.4.1. Modelos de aprendizaje profundo

LSTM

La red LSTM constituyó la primera arquitectura de aprendizaje profundo desarrollada y evaluada de manera sistemática en esta investigación, representando una etapa central en la construcción metodológica posterior de esta tesis. A diferencia de utilizar configuraciones reportadas directamente en la literatura, el modelo fue diseñado y optimizado específicamente para el problema de predicción de demanda eléctrica abordado, mediante un proceso progresivo de exploración experimental, búsqueda en grilla y validación cruzada.

Tabla 5.12: Comparación entre los modelos LSTM y SARIMAX

FTW	Modelo	R^2	RMSE [MWh]
1	SARIMAX	0.40	1878.2
	LSTM	0.94	727.2
7	SARIMAX	0.41	1866.1
	LSTM	0.91	848.4
14	SARIMAX	0.45	1803.1
	LSTM	0.66	1696.8

Como resultado de este proceso se identificaron configuraciones robustas capaces de capturar relaciones temporales complejas entre variables energéticas y meteorológicas. Esta etapa permitió definir criterios de diseño experimental, acotar espacios de búsqueda y establecer una línea base profunda optimizada específicamente para el dominio estudiado, dando lugar además a resultados posteriormente consolidados en publicaciones científicas derivadas de esta línea de trabajo.

La Tabla 5.12 presenta la comparación entre los modelos SARIMAX y LSTM para distintos horizontes de predicción, considerando una ventana histórica fija de 7 días. En cada bloque de filas se comparan ambos enfoques bajo condiciones experimentales equivalentes, permitiendo analizar el efecto de incorporar mecanismos de memoria temporal profunda frente a un enfoque estadístico tradicional.

Como se observa en la Tabla 5.12, el modelo LSTM supera consistentemente al modelo SARIMAX para todos los horizontes considerados. En FTW=1 se alcanza el mejor desempeño global, con un R^2 de 0.94 y un RMSE de 727.2 MWh, mejorando sustancialmente la capacidad explicativa y reduciendo el error respecto del modelo estadístico. Esta diferencia se mantiene para FTW=7 y FTW=14, aunque con una degradación gradual esperable a medida que aumenta el horizonte de predicción.

La Figura 5.4 muestra una comparación visual entre ambos enfoques para el horizonte de siete días. Puede observarse que el modelo LSTM reproduce con mayor fidelidad la evolución temporal de la demanda eléctrica, incluyendo períodos de alta variabilidad y eventos de demanda elevada.

En particular, el modelo LSTM mantiene una representación más estable del comportamiento temporal de la serie, reduciendo desviaciones tanto en períodos regulares como durante eventos de mayor demanda. En contraste, SARIMAX evidencia una tendencia más marcada a subestimar los cambios abruptos y presenta

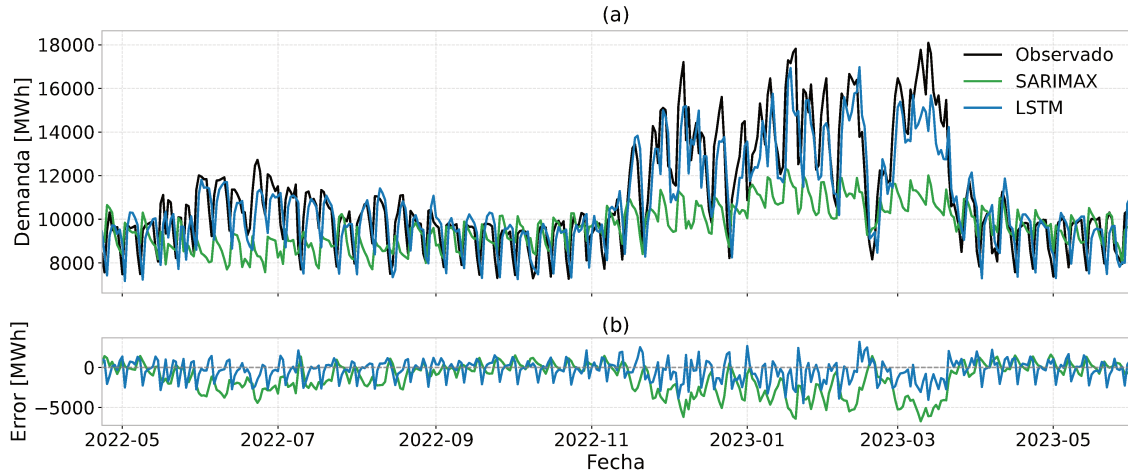


Figura 5.4: Comparación entre los modelos LSTM y SARIMAX para la predicción de la demanda eléctrica en $FTW=7$. (a) Muestra los valores observados y predichos por ambos modelos. (b) Presenta los errores de predicción calculados como la diferencia entre valores predichos y observados.

errores de mayor amplitud.

Estos resultados muestran que la incorporación de mecanismos de memoria temporal permite modelar de forma más efectiva las relaciones no lineales presentes en la demanda eléctrica, superando las limitaciones observadas en enfoques estadísticos tradicionales. Más allá del desempeño alcanzado, esta etapa constituyó un aporte metodológico en sí mismo, al establecer una línea base profunda optimizada específicamente para el problema abordado y generar evidencia experimental sobre el efecto de distintas configuraciones temporales y de entrenamiento.

El conocimiento obtenido durante esta fase permitió posteriormente avanzar hacia arquitecturas más especializadas orientadas a incorporar mecanismos de atención y estrategias de modelado más complejas, presentadas en las secciones siguientes.

DA-SKIP y DA-LSTM

Luego de establecer una línea base mediante el modelo LSTM, se evaluaron arquitecturas basadas en mecanismos de atención con el objetivo de analizar si una estructura más especializada permitía mejorar el desempeño predictivo. En particular, se comparó el modelo DA-SKIP [32] con DA-LSTM, cuya formulación y fundamentos fueron presentados previamente en la Sección 4.3.

La Tabla 5.13 presenta el desempeño de los modelos LSTM, DA-SKIP y DA-LSTM para distintos horizontes de predicción utilizando una ventana histórica fija

Tabla 5.13: Comparación entre los modelos LSTM, DA-SKIP y DA-LSTM para los tres horizontes de predicción.

FTW	Modelo	R^2	RMSE [MWh]
1	LSTM	0.94	727.2
	DA-SKIP	0.83	942.3
	DA-LSTM	0.87	892.4
7	LSTM	0.91	848.4
	DA-SKIP	0.34	1832.5
	DA-LSTM	0.50	1738.1
14	LSTM	0.66	1696.8
	DA-SKIP	0.31	1883.1
	DA-LSTM	0.44	1841.1

de 7 días. Se mantiene el modelo LSTM como referencia con el objetivo de contextualizar el desempeño alcanzado por las arquitecturas basadas en atención.

Los resultados muestran que DA-LSTM logra mejorar el desempeño respecto del modelo DA-SKIP para horizontes cortos e intermedios. En FTW=1 se observa un incremento de aproximadamente cuatro puntos porcentuales en R^2 junto con una reducción del RMSE cercana a 50 MWh. Esta tendencia también se mantiene en FTW=7, donde DA-LSTM incrementa el poder explicativo y reduce el error respecto de la arquitectura original.

Sin embargo, para horizontes más extensos (FTW=14), la modificación propuesta no logró sostener el mismo nivel de desempeño y presentó una degradación significativa respecto tanto de LSTM como de DA-SKIP.

La Figura 5.5 complementa el análisis cuantitativo mostrando una comparación visual entre los modelos LSTM, DA-SKIP y DA-LSTM. Se observa que DA-LSTM logra aproximarse en mayor medida a los valores observados respecto de DA-SKIP, reduciendo parte de las desviaciones presentes en la arquitectura original. Este comportamiento resulta especialmente visible en períodos de mayor variabilidad de la demanda, donde DA-LSTM presenta una mejor adaptación a la dinámica temporal de la serie.

El análisis de los errores permite observar que las mejoras introducidas por DA-LSTM se traducen en una reducción de la amplitud y dispersión del error respecto de DA-SKIP para todos los horizontes evaluados. No obstante, el comportamiento obtenido continúa siendo inferior al alcanzado por LSTM, que mantiene una mayor estabilidad predictiva y una representación más precisa de la evolución temporal de

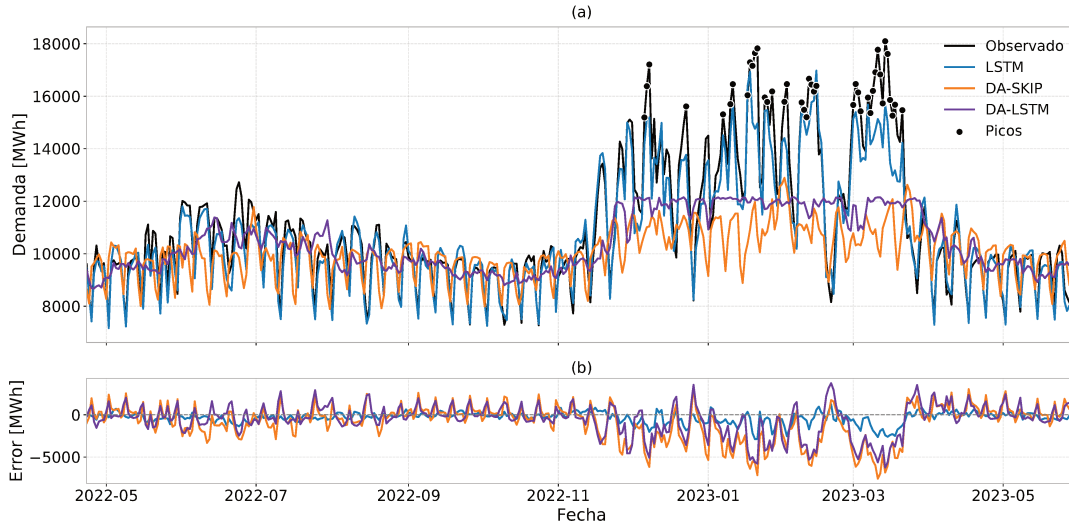


Figura 5.5: Comparación entre los valores observados y las predicciones obtenidas por los modelos LSTM, DA-SKIP y DA-LSTM para un horizonte de predicción de siete días ($FTW=7$). (a) Muestra las series observadas y las predicciones generadas por cada modelo. (b) Presenta el error de predicción respecto de los valores reales.

la demanda eléctrica. En conjunto, estos resultados muestran que las modificaciones incorporadas en DA-LSTM permitieron mejorar consistentemente el desempeño del modelo de referencia, constituyendo una evidencia favorable del rediseño arquitectónico propuesto. Al mismo tiempo, las limitaciones observadas motivaron el desarrollo posterior de una nueva arquitectura orientada a integrar de forma más efectiva la información meteorológica disponible, presentada en la siguiente subsección.

AMBER

Como se observa en la Tabla 5.14, el modelo AMBER presenta un desempeño consistentemente superior al de DA-LSTM en todos los horizontes de predicción considerados, con una ventaja que se acentúa a medida que aumenta la complejidad temporal del problema.

En el horizonte $FTW=1$, ambos modelos alcanzan niveles de precisión elevados, aunque AMBER logra una reducción sustancial del error respecto de DA-LSTM, lo que sugiere que el mecanismo de atención meteorológica contribuye incluso en escenarios de baja incertidumbre. A medida que el horizonte se extiende, las diferencias se vuelven más pronunciadas: para $FTW=7$, la brecha en R^2 se amplía considerablemente, evidenciando una mayor capacidad de AMBER para mantener coherencia predictiva frente al incremento de la incertidumbre asociada a las varia-

Tabla 5.14: Comparación de desempeño entre DA-LSTM y AMBER para los tres horizontes de predicción.

FTW	Modelo	R^2	RMSE [MWh]
1	DA-LSTM	0.87	892.4
	AMBER	0.94	608.9
7	DA-LSTM	0.50	1738.1
	AMBER	0.78	1157.1
14	DA-LSTM	0.44	1841.1
	AMBER	0.69	1360.6

bles meteorológicas. El caso más ilustrativo es FTW=14, donde DA-LSTM presenta un colapso en su capacidad explicativa, mientras que AMBER conserva un desempeño aceptable, con valores de R^2 y RMSE que indican una degradación gradual y controlada. Este comportamiento sugiere que la incorporación de atención sobre variables meteorológicas no solo mejora la precisión, sino que además dota al modelo de mayor robustez estructural frente a horizontes extendidos.

La Figura 5.6a permite visualizar el comportamiento temporal de ambos modelos para un horizonte de 7 días. AMBER sigue con mayor precisión la dinámica de la demanda, mientras que DA-LSTM presenta una respuesta más suavizada, en particular para demandas superiores a 12.000 MWh. Este amesetamiento puede estar asociado con el uso de MSE, que puede favorecer predicciones más próximas a valores medios, aunque esta causa no fue evaluada por separado.

La Figura 5.6b muestra la evolución de los errores de predicción. Se aprecia que AMBER mantiene errores de menor magnitud y menor dispersión a lo largo del conjunto de prueba, mientras que DA-LSTM exhibe desviaciones más pronunciadas, especialmente en períodos de elevada demanda. Estos resultados son consistentes con las métricas presentadas en la Tabla 5.14, donde AMBER alcanza un mejor desempeño tanto en términos de R^2 como de RMSE para este horizonte de predicción.

5.4.2. Comparación global de modelos

En esta sección, se presenta una comparación global de los modelos evaluados incluyendo todas las métricas. Como se observa en la Tabla 5.15, el modelo AMBER presenta un desempeño competitivo frente al resto de los modelos, destacándose especialmente en términos de reducción del error y mejor calibración de las predic-

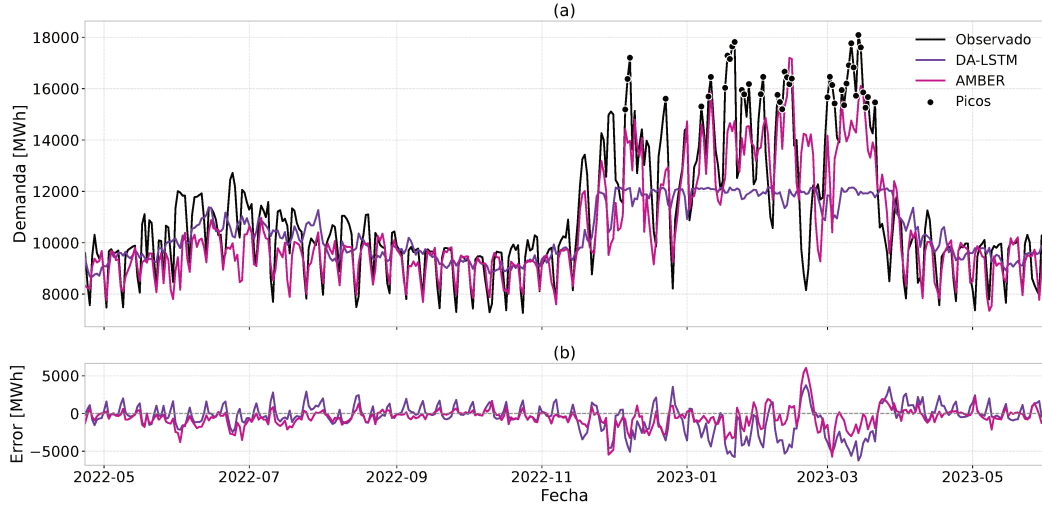


Figura 5.6: Comparación entre los valores observados, DA-LSTM y AMBER para un horizonte de predicción de 7 días. (a) Muestra las series temporales predichas y observadas. (b) Presenta los errores de predicción de ambos modelos.

ciones. En $FTW=1$, AMBER alcanza un $R^2 = 0,94$, igualando el desempeño del modelo LSTM, pero logrando una reducción del error absoluto de aproximadamente un 16.3 %. Asimismo, presenta el menor sesgo entre todos los modelos evaluados, indicando una menor tendencia a la subestimación sistemática.

A medida que el horizonte de predicción se incrementa, se observa una degradación general del desempeño en todos los modelos [56]. Sin embargo, AMBER mantiene una mayor robustez relativa, superando a DA-LSTM y DA-SKIP en todos los horizontes, y alcanzando el mejor desempeño en $FTW=14$ días.

Si bien el modelo AMBER obtuvo el mejor desempeño global para los horizontes $FTW=1$ y $FTW=14$, los resultados muestran que para el horizonte intermedio $FTW=7$ el modelo LSTM alcanzó métricas superiores, obteniendo un coeficiente de determinación $R^2 = 0,91$ y un $RMSE = 848,4$, frente a $R^2 = 0,78$ y $RMSE = 1157,0$ obtenidos por AMBER. Este resultado evidencia que una arquitectura más compleja no necesariamente garantiza mejoras sistemáticas en todos los escenarios temporales. En particular, para horizontes intermedios de predicción, la dinámica temporal de la demanda eléctrica podría ser suficientemente capturada por una arquitectura recurrente multivariada convencional, sin requerir mecanismos adicionales de atención meteorológica.

Asimismo, la incorporación de múltiples componentes de atención puede incrementar la complejidad del proceso de optimización y favorecer cierto grado de sobre-

Tabla 5.15: Desempeño comparativo de los modelos: métricas en conjunto de prueba.

FTW	Modelo	R^2	RMSE [MWh]	WI	%Bias
1	SARIMAX	0.40	1878.2	0.55	-11.6
	LSTM	0.94	727.2	0.98	-1.0
	DA-SKIP	0.83	942.3	0.93	-3.0
	DA-LSTM	0.87	892.4	0.96	-2.0
	AMBER	0.94	608.9	0.98	-0.7
7	SARIMAX	0.41	1866.1	0.53	-12.1
	LSTM	0.91	848.4	0.97	-3.0
	DA-SKIP	0.34	1832.5	0.55	-8.4
	DA-LSTM	0.50	1738.1	0.76	-4.1
	AMBER	0.78	1157.1	0.94	-4.5
14	SARIMAX	0.45	1803.1	0.57	-12.0
	LSTM	0.66	1696.8	0.90	-10.7
	DA-SKIP	0.31	1883.1	0.54	-9.2
	DA-LSTM	0.44	1841.1	0.73	-4.1
	AMBER	0.69	1360.6	0.90	-6.6

ajuste o inestabilidad en horizontes donde las dependencias temporales aún conservan una estructura relativamente predecible mediante modelos recurrentes clásicos. No obstante, los resultados obtenidos para FTW=14 muestran que AMBER presenta una mayor robustez frente al incremento del horizonte de predicción, superando al modelo LSTM tanto en RMSE como en R^2 . Esto sugiere que los mecanismos de atención meteorológica incorporados en AMBER resultan particularmente útiles en escenarios de mayor incertidumbre temporal y degradación progresiva de la información histórica. Como no se repitieron los entrenamientos con distintas semillas ni se realizaron pruebas estadísticas, estas diferencias no deben interpretarse como estadísticamente significativas.

La Figura 5.7 presenta la comparación entre los valores observados y las predicciones de LSTM y AMBER para los tres horizontes. Para FTW=7 y FTW=14, las series se construyeron alineando las predicciones por fecha y promediando las estimaciones disponibles para un mismo día. Este promedio se utilizó sólo para la visualización; las métricas se calcularon sobre la totalidad de las predicciones de cada FTW.

El desempeño de ambos modelos se evalúa a través de cuatro métricas: el MAE y el RMSE cuantifican el error absoluto promedio y penalizan en mayor medida los errores de gran magnitud respectivamente, el Bias refleja la dirección sistemática del error —donde valores negativos indican subestimación— y el MAPE expresa el



Figura 5.7: Comparación entre los valores observados y las predicciones obtenidas por los modelos LSTM y AMBER para distintos horizontes de predicción. En (a), (c) y (e) se muestran las series temporales observadas y predichas para horizontes de 1, 7 y 14 días, respectivamente, mientras que en (b), (d) y (f) se presentan los errores de predicción asociados a cada horizonte.

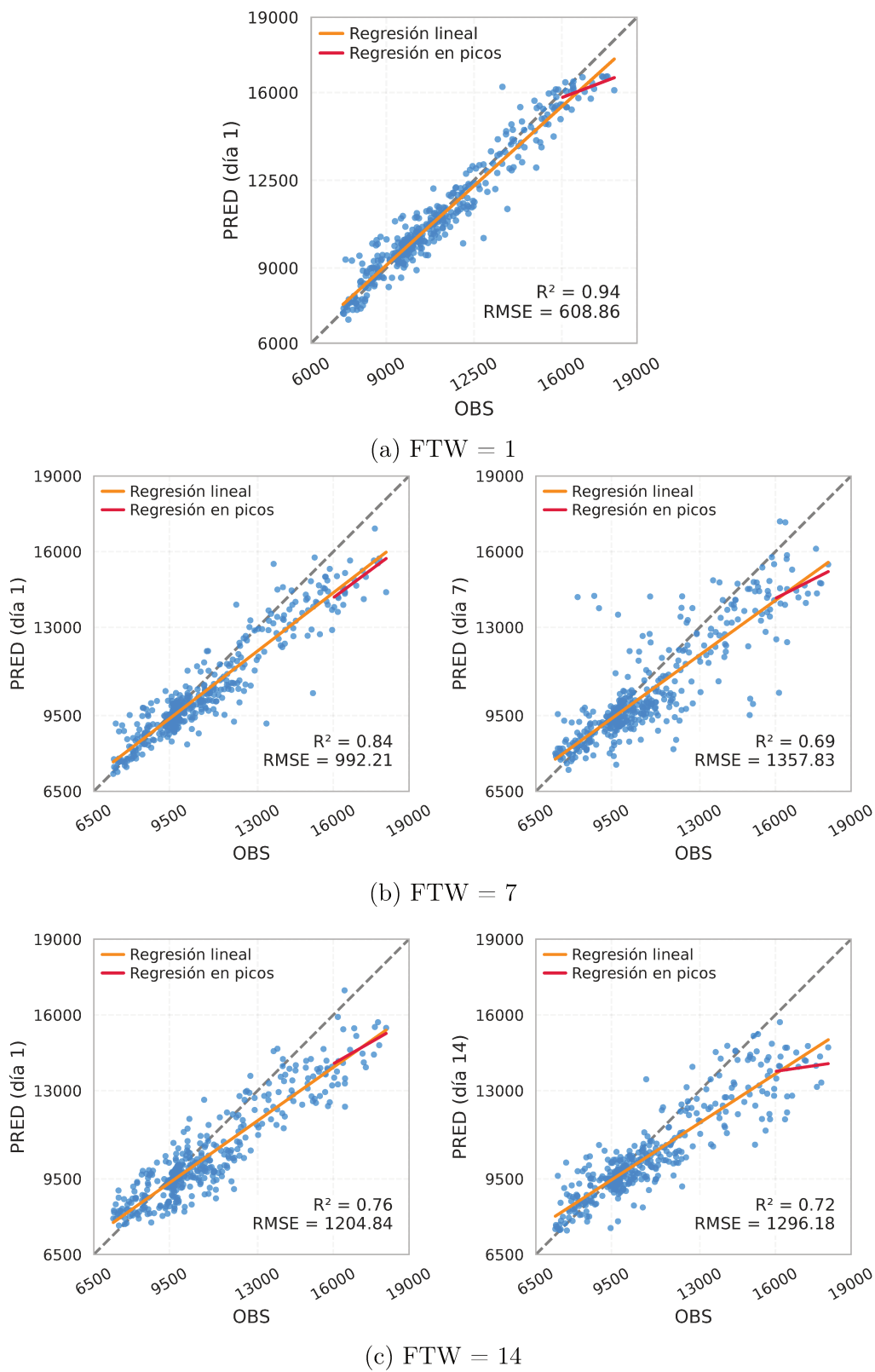


Figura 5.8: Relación entre valores observados y predichos por el modelo AMBER.

Tabla 5.16: Métricas de error calculadas exclusivamente sobre eventos de alta demanda (percentil 90, $\approx 15\,150$ MWh) para cada horizonte de predicción. MAE, RMSE y Bias en MWh; MAPE en %. La columna *Mejora AMBER* indica la reducción relativa del error respecto de LSTM; valores negativos indican que LSTM supera a AMBER.

FTW	LSTM				AMBER				Mejora AMBER	
	MAE	RMSE	Bias	MAPE	MAE	RMSE	Bias	MAPE	RMSE	MAE
1	1084.94	1416.75	-819.78	6.65	624.63	773.94	-535.27	3.78	+45.37 %	+42.43 %
7	1442.14	1668.46	-1396.86	8.78	1923.51	2243.16	-1836.27	11.71	-34.44 %	-33.38 %
14	2414.30	2933.38	-2319.84	14.92	2669.33	2897.59	-2669.33	16.28	+1.22 %	-10.56 %

error relativo en porcentaje, independientemente de la escala de la demanda.

Para $FTW = 1$, ambos modelos reproducen adecuadamente la dinámica general de la demanda. Sin embargo, AMBER muestra una mejor alineación con los valores observados en los períodos de pico, con errores de menor magnitud y distribución más simétrica respecto del cero, lo que indica una menor tendencia a la subestimación sistemática que caracteriza al modelo LSTM en este horizonte.

Para $FTW = 7$, las diferencias entre ambos modelos se hacen más notorias y favorables al modelo LSTM. Este alcanza un coeficiente de determinación R^2 de 0,91 y un RMSE de 848,4 MWh, frente a un R^2 de 0,78 y un RMSE de 1.157,0 MWh obtenidos por AMBER, evidenciando un mejor desempeño global del modelo recurrente convencional para este horizonte intermedio. En términos de sesgo, ambos modelos presentan una tendencia a la subestimación, aunque esta resulta más pronunciada en AMBER ($\%Bias = -4,5\%$) que en LSTM ($\%Bias = -3,0\%$), en concordancia con el panel de errores de la Figura 5.7(d), donde se observan errores predominantemente negativos en ambos casos y una mayor dispersión en las predicciones de AMBER. Estos resultados sugieren que, para horizontes intermedios de predicción, la información temporal disponible podría ser suficientemente capturada por LSTM, sin que la incorporación de mecanismos adicionales de atención aporte mejoras sistemáticas en el desempeño predictivo.

Para $FTW = 14$, la subestimación de LSTM se intensifica y se vuelve el rasgo dominante de su comportamiento: la curva predicha corre de manera persistente por debajo de la observada, con errores negativos de gran magnitud en los períodos de mayor demanda. AMBER, si bien también exhibe degradación respecto de horizontes más cortos, mantiene una mayor proximidad a los valores reales y una distribución de errores más equilibrada, evidenciando una mayor robustez frente al incremento

de la incertidumbre acumulada en horizontes extendidos.

Con el objetivo de cuantificar el comportamiento diferencial de ambos modelos, específicamente sobre los eventos de mayor demanda, la Tabla 5.16 reporta las métricas de error calculadas exclusivamente sobre los días que superaron el percentil 90 de la serie observada, definidos como eventos de pico. Para cada horizonte se identificaron entre 40 y 41 eventos de este tipo, con un umbral de aproximadamente 15150 MWh.

Para $FTW=1$, los resultados confirman y cuantifican lo observado visualmente: AMBER reduce el MAE en un 42 % y el $RMSE$ en un 45 % respecto de LSTM, obteniendo valores de $MAE = 624,6$ y $RMSE = 773,9$, frente a $MAE = 1084,9$ y $RMSE = 1416,7$ del modelo recurrente. El sesgo negativo, presente en ambos modelos, es considerablemente menor en AMBER ($Bias = -535,3$ MWh vs. $Bias = -819,8$ MWh), lo que corrobora la menor tendencia a la subestimación sistemática ya señalada en el análisis de la Figura 5.7. El $MAPE$ de AMBER en picos alcanza 3,78 %, prácticamente la mitad del 6,65 % de LSTM, indicando que la ventaja no es marginal sino sustancial en este horizonte.

Para $FTW=7$, la situación se invierte de manera marcada. AMBER presenta $MAE = 1923,5$ y $RMSE = 2243,2$, valores aproximadamente entre un 33 % y 34 % superiores a los de LSTM ($MAE = 1442,1$ y $RMSE = 1668,5$). El sesgo negativo de AMBER se intensifica notablemente hasta $Bias = -1836,3$ MWh, más del doble que en $FTW=1$, lo que sugiere que el modelo pierde capacidad para anticipar la magnitud de los picos cuando el horizonte se extiende a una semana. Este deterioro selectivo —más pronunciado en los extremos que en el desempeño global— es consistente con la mayor dispersión de errores visible en el panel (d) de la Figura 5.7, y podría estar vinculado a la dinámica del mecanismo de atención en escenarios de incertidumbre intermedia.

Para $FTW=14$, los resultados son mixtos. En términos de $RMSE$ los modelos convergen ($RMSE = 2897,6$ para AMBER vs. $RMSE = 2933,4$ para LSTM, diferencia aproximada del 1,2 %), lo que indica que ambos cometen errores de magnitud similar en los picos más extremos. Sin embargo, el MAE de AMBER ($MAE = 2669,3$) supera en un 10,6 % al de LSTM ($MAE = 2414,3$), señalando que AMBER incurre en errores moderados más frecuentes sobre los días de alta demanda. El sesgo

de AMBER en este horizonte ($Bias = -2669,3$ MWh) iguala numéricamente a su propio MAE, lo que implica que prácticamente la totalidad del error en picos tiene dirección negativa: el modelo subestima sistemáticamente sin instancias compensatorias de sobreestimación. Este comportamiento es coherente con el desempeño global similar que ambos modelos exhiben para $FTW=14$, visible en la Figura 5.9, y sugiere que las diferencias entre AMBER y LSTM en este horizonte se manifiestan principalmente en la dirección del error más que en su magnitud.

Las Figuras 5.7 y 5.9 permiten vincular estos resultados con el comportamiento visual de los modelos: la principal limitación de LSTM frente al incremento del horizonte de predicción no es la pérdida de resolución temporal sino la subestimación progresiva y sistemática de la demanda. A medida que el horizonte se extiende, AMBER mantiene errores más acotados en $FTW=14$ y una ventaja clara sobre los eventos de pico en $FTW=1$, aunque en $FTW=7$ muestra una mayor dificultad específica sobre los extremos, lo que resulta coherente con la mejora en las métricas presentadas en la Tabla 5.14 y con el análisis diferencial sobre eventos de pico de la Tabla 5.16.

Un aspecto particularmente relevante es el análisis de la regresión en picos de demanda. El análisis de la relación entre los valores observados y predichos por el modelo AMBER, presentado en las Figuras 5.8a, 5.8b y 5.8c, permite evaluar en detalle la calidad de ajuste y el comportamiento del modelo frente a distintos horizontes de predicción. Para el horizonte de corto plazo ($FTW=1$), se observa una alta concentración de los puntos en torno a la recta de identidad, con una pendiente de regresión cercana a la unidad, lo que indica una elevada capacidad del modelo para reproducir la dinámica real de la demanda eléctrica. A medida que el horizonte de predicción se incrementa, se evidencia una mayor dispersión de los puntos, particularmente en los valores más altos de demanda. Este comportamiento es consistente con el aumento de la incertidumbre inherente a la predicción multihorizonte en series temporales. No obstante, el modelo mantiene una estructura lineal clara, lo que indica que conserva la coherencia global de la dinámica del sistema incluso en horizontes extendidos.

En todos los horizontes considerados, se observa que la pendiente asociada a los picos es menor que la de la regresión global, lo que confirma la existencia de

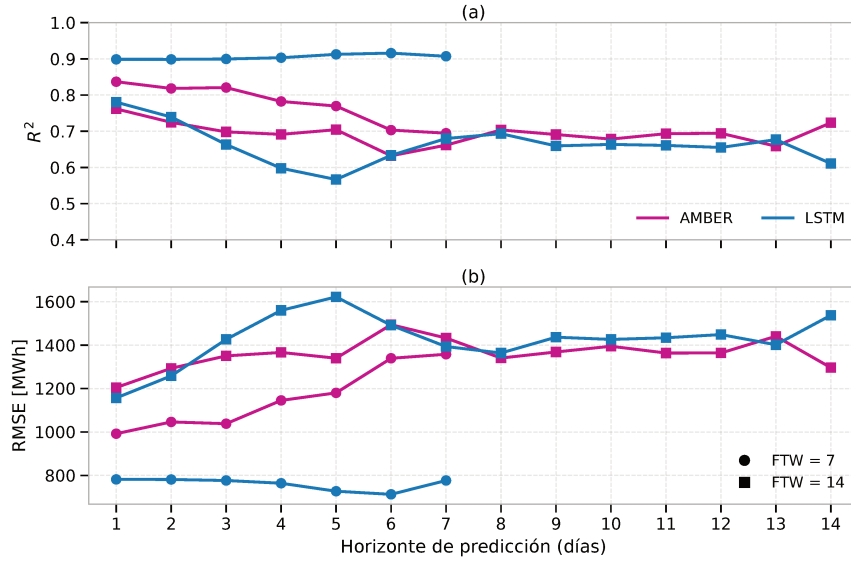


Figura 5.9: (a) Evolución del coeficiente de determinación (R^2). (b) $RMSE$ para los modelos AMBER y LSTM en función del día de predicción para $FTW = 7$ y 14 días.

un sesgo leve de subestimación en condiciones extremas. Sin embargo, esta subestimación resulta significativamente más controlada que en modelos tradicionales, lo cual es consistente con el comportamiento reportado en modelos basados en redes neuronales recurrentes en la literatura.

La Figura 5.9 presenta la evolución del RMSE y de R^2 para cada paso de los modelos entrenados con $FTW=7$ y $FTW=14$. Cada FTW corresponde a un modelo optimizado de manera independiente. Para $FTW=7$, LSTM alcanza valores de R^2 superiores a los de AMBER a lo largo de todo el horizonte y también presenta menores valores de RMSE.

Para $FTW = 14$ días, ambos modelos presentan un comportamiento más estable a lo largo del horizonte. AMBER mantiene valores de R^2 en torno a 0,70 y alcanza, en varios pasos de predicción, valores de RMSE inferiores a los de LSTM, lo que evidencia una degradación más controlada frente al aumento del horizonte temporal. LSTM, por su parte, presenta valores de RMSE que superan los 1400 MWh en la mayor parte del horizonte, con un incremento notable hacia el final del mismo. La comparación entre ambos casos de FTW revela que la degradación del desempeño de AMBER resulta más acotada en el horizonte de 14 días que en el de 7 días, lo que sugiere una mayor robustez relativa del modelo ante escenarios de mayor incertidumbre temporal. Este comportamiento es esperable en problemas de predicción

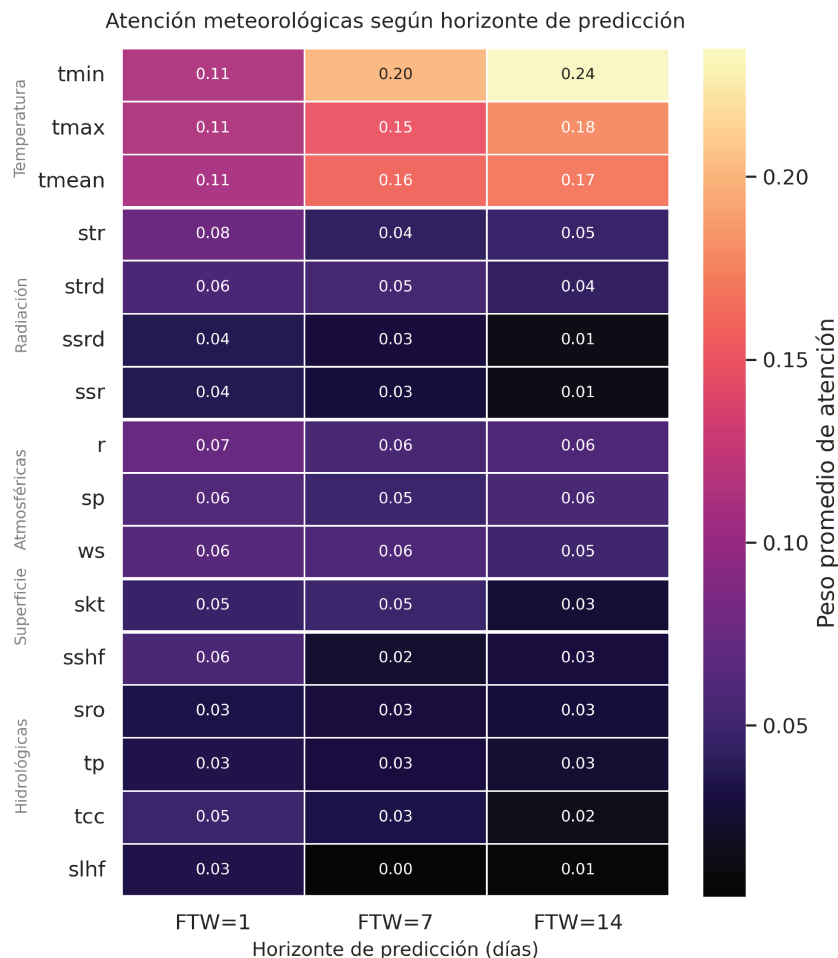


Figura 5.10: Mapa de calor de los pesos promedio de atención meteorológica para los horizontes de predicción de 1, 7 y 14 días.

multihorizonte, donde la acumulación de incertidumbre —en particular la asociada a las variables meteorológicas— afecta gradualmente la precisión de las estimaciones.

En conjunto, LSTM presenta el mejor desempeño para FTW=7, mientras que AMBER alcanza mejores métricas globales para FTW=14.

5.4.3. Atención meteorológica

El análisis del mecanismo de atención del modelo AMBER permite interpretar el proceso de predicción en términos de la relevancia asignada a las variables meteorológicas. La Figura 5.10 presenta los pesos promedio asignados por AMBER para los distintos horizontes. Al tratarse de promedios sobre todo el conjunto de prueba, no muestran su variación temporal ni estacional.

Las variables térmicas ((tmin), (tmax) y (tmean)) concentran los mayores pesos

de atención en los horizontes considerados, en concordancia con el análisis presentado en el Capítulo 2. Asimismo, se evidencia un incremento en la importancia relativa de estas variables a medida que aumenta el horizonte de predicción, destacándose el caso de (t_{min}). Este comportamiento refuerza la hipótesis de que, en escenarios de mayor incertidumbre, el modelo incrementa su dependencia de variables meteorológicas más estables, que capturan el comportamiento base del sistema. En cambio, las variables radiativas, hidrológicas y atmosféricas presentan pesos menores. Sin embargo, determinar si alguna de ellas puede eliminarse sin afectar el desempeño requiere experimentos específicos.

En conjunto, estos resultados muestran que el mecanismo de atención no solo contribuye a mejorar el desempeño predictivo del modelo, sino que además aporta interpretabilidad, permitiendo identificar de manera explícita las variables meteorológicas más relevantes en cada horizonte de predicción, en concordancia con el comportamiento físico esperado del sistema analizado.

En síntesis, los resultados obtenidos evidencian que los modelos basados en aprendizaje profundo superan ampliamente a los enfoques clásicos en la predicción de la demanda eléctrica, particularmente en contextos caracterizados por relaciones no lineales y alta variabilidad. En este marco, la arquitectura propuesta AMBER no solo mejora la precisión respecto de modelos como DA-SKIP y DA-LSTM, sino que además demuestra una mayor robustez frente al incremento del horizonte de predicción, manteniendo un desempeño competitivo incluso en los escenarios más exigentes. Asimismo, la incorporación de mecanismos de atención meteorológica permite no solo optimizar el desempeño predictivo, sino también aportar interpretabilidad al modelo, identificando de manera coherente las variables más relevantes para la dinámica del sistema. Estos resultados consolidan el aporte de la presente tesis y establecen una base sólida para las conclusiones y líneas de trabajo futuro que se desarrollan en el capítulo siguiente.

Capítulo 6

Conclusiones y trabajo futuro

Este capítulo presenta una síntesis integral de los principales aportes y hallazgos derivados de la presente investigación doctoral. En la Sección 6.1 se analizan los resultados obtenidos en relación con cada uno de los objetivos específicos planteados en esta tesis, destacando las principales contribuciones metodológicas y experimentales alcanzadas. Posteriormente, la Sección 6.2 expone posibles líneas futuras de investigación y potenciales aplicaciones operativas derivadas del trabajo desarrollado. Finalmente, la Sección 6.3 resume las publicaciones científicas, presentaciones en congresos y reconocimientos académicos asociados al desarrollo de esta tesis doctoral.

6.1. Objetivos y contribuciones

6.1.1. **Objetivo 1: Determinar las variables más influyentes en la demanda eléctrica de Entre Ríos**

Uno de los principales objetivos de esta tesis consistió en identificar las variables con mayor influencia sobre la demanda de energía eléctrica en la provincia de Entre Ríos. Para ello, se integraron variables meteorológicas provenientes del reanálisis ERA5 junto con variables temporales y registros históricos de demanda energética.

El análisis exploratorio realizado permitió observar relaciones significativas entre determinadas variables meteorológicas y el comportamiento de la demanda eléctrica. En particular, las variables asociadas a temperatura mostraron una fuerte correlación con el consumo energético, especialmente durante períodos estivales e invernales,

donde la utilización de sistemas de climatización produce incrementos importantes en la demanda provincial.

Posteriormente, el modelo AMBER permitió profundizar este análisis mediante mecanismos de atención sobre las variables meteorológicas. Las temperaturas máxima, mínima y media recibieron los mayores pesos, mientras que las variables relacionadas con radiación y flujos de calor presentaron valores intermedios, y las asociadas a precipitación y escorrentía, valores menores. Estos resultados aportan información sobre la ponderación asignada por el modelo a las variables exógenas, cuya contribución debe confirmarse mediante experimentos específicos.

6.1.2. Objetivo 2: Proponer modelos de aprendizaje profundo para la predicción demanda energética

El segundo objetivo de esta tesis estuvo orientado al desarrollo de modelos de redes neuronales recurrentes para la predicción a corto plazo de la demanda eléctrica provincial.

En una primera etapa se implementó un modelo LSTM multivariado utilizando variables meteorológicas, energéticas y temporales como entradas. El desarrollo de este modelo incluyó un proceso sistemático de optimización de hiperparámetros mediante búsqueda en grilla y validación cruzada, considerando múltiples configuraciones de entrenamiento, arquitectura y horizonte de predicción. Este proceso permitió construir una línea base robusta y metodológicamente controlada para la comparación posterior con arquitecturas más complejas. Los resultados obtenidos demostraron que las redes recurrentes permiten modelar adecuadamente relaciones no lineales presentes en la serie temporal de demanda eléctrica, superando claramente el desempeño alcanzado por modelos clásicos como SARIMAX.

Posteriormente, se analizaron arquitecturas más complejas basadas en mecanismos de atención, incluyendo DA-SKIP y una propuesta desarrollada en esta tesis, denominada DA-LSTM, concebida como una modificación de dicha arquitectura. Estas arquitecturas permitieron explorar estrategias de atención sobre variables de entrada y dependencias temporales. No obstante, los resultados no alcanzaron el desempeño logrado con LSTM.

Finalmente, se propuso el modelo AMBER, principal aporte original de esta tesis.

La arquitectura desarrollada incorpora mecanismos de atención sobre variables meteorológicas históricas y futuras, integrando información exógena correspondiente tanto a la ventana temporal de entrada como al horizonte de predicción. Asimismo, el modelo combina componentes encoder-decoder basados en atención con una red LSTM compacta destinada al refinamiento de la predicción final. Los resultados obtenidos demostraron que la incorporación explícita de atención meteorológica permitió mejorar la robustez predictiva del modelo, particularmente en horizontes de predicción más extensos.

6.1.3. Objetivo 3: Analizar el efecto de distintas configuraciones temporales en el desempeño de los modelos

Con el fin de estudiar el comportamiento de los modelos ante diferentes configuraciones temporales, se evaluaron múltiples horizontes de predicción futura manteniendo fija una ventana histórica de entrada de siete días.

Los experimentos consideraron horizontes de predicción de 1, 7 y 14 días mediante modelos independientes, optimizados para cada FTW. En general, el desempeño disminuyó en los horizontes más extensos; sin embargo, en el caso de AMBER, esta disminución fue moderada y el modelo mantuvo una capacidad predictiva adecuada incluso para FTW=14.

Para horizontes cortos de predicción, particularmente FTW=1, tanto LSTM como AMBER alcanzaron elevados valores de capacidad explicativa y bajos errores de predicción. En horizontes intermedios y largos, los modelos comenzaron a presentar una mayor dificultad para reproducir correctamente los picos de demanda, observándose una tendencia al suavizado de valores extremos.

No obstante, para el horizonte de predicción más extenso evaluado (FTW=14), el modelo AMBER mostró una mayor estabilidad en horizontes de predicción más largos, obteniendo mejores resultados que otros enfoques basados en atención evaluados en esta tesis. En particular, para FTW=14, el modelo AMBER logró superar el desempeño del modelo LSTM en términos de RMSE y coeficiente R^2 , evidenciando una mejor capacidad para modelar dependencias complejas entre variables meteorológicas y demanda energética.

Estos resultados confirman la importancia de analizar distintas configuraciones

temporales de entrada y salida al momento de diseñar modelos de predicción de demanda eléctrica.

6.1.4. Objetivo 4: Evaluar y comparar el desempeño de los modelos y su capacidad explicativa

El último objetivo específico consistió en evaluar comparativamente los distintos modelos desarrollados mediante métricas cuantitativas de desempeño.

Para ello se utilizaron métricas ampliamente empleadas en problemas de predicción de series temporales, incluyendo RMSE, coeficiente de determinación (R^2), índice de Willmott y porcentaje de sesgo (%Bias). Estas métricas permitieron analizar tanto la magnitud de los errores como la capacidad explicativa y estabilidad de los modelos.

Los resultados obtenidos mostraron que los enfoques basados en aprendizaje profundo superaron ampliamente el desempeño del modelo clásico evaluado. SARIMAX presentó limitaciones importantes para modelar relaciones no lineales y capturar adecuadamente la complejidad de la demanda eléctrica provincial.

Entre los modelos de aprendizaje profundo evaluados, el modelo LSTM logró un desempeño competitivo en horizontes cortos e intermedios. Sin embargo, AMBER obtuvo resultados más robustos en horizontes de predicción largos, alcanzando menores errores de predicción y una mejor estabilidad frente a variaciones de la demanda.

Asimismo, el análisis de las curvas de error permitió observar comportamientos diferenciados según el horizonte de predicción. Para $FTW = 1$ día, AMBER y LSTM presentaron niveles de error similares, aunque AMBER logró reducir levemente el RMSE. En $FTW = 7$ días, el modelo LSTM obtuvo errores más contenidos y alcanzó el mejor desempeño global. Sin embargo, para $FTW = 14$ días, AMBER mostró una degradación más controlada del error y una mejor capacidad para mantener la estructura general de la demanda frente al incremento del horizonte predictivo. Estos resultados sugieren que la incorporación de mecanismos de atención específicos sobre variables meteorológicas resulta especialmente beneficiosa en escenarios de mayor incertidumbre temporal.

En conjunto, los resultados obtenidos permitieron cumplir satisfactoriamente

los objetivos planteados en esta tesis y demostrar el potencial de las arquitecturas basadas en atención y redes neuronales recurrentes para la predicción a corto plazo de demanda eléctrica a nivel provincial.

6.2. Trabajo futuro

Como línea futura se propone reemplazar los valores futuros de ERA5 utilizados en esta evaluación por pronósticos de modelos meteorológicos y extender la arquitectura hacia horizontes más amplios.

Asimismo, futuras investigaciones podrían incorporar nuevas variables exógenas asociadas a factores socioeconómicos, indicadores de actividad industrial, tarifas energéticas o eventos extremos, con el objetivo de mejorar la representación de escenarios complejos de demanda.

Otra línea relevante consiste en la incorporación de mecanismos avanzados de interpretabilidad y explicabilidad basados en técnicas de XAI (*Explainable Artificial Intelligence*), permitiendo una comprensión más profunda de las decisiones generadas por los modelos de aprendizaje profundo. En este sentido, también se propone analizar la variación temporal y estacional de los pesos de atención y evaluar el modelo utilizando conjuntos reducidos de variables.

Otra línea futura consiste en evaluar la distribución espacial de las variables meteorológicas, ya que en esta tesis se utilizó un promedio provincial para cada variable. Asimismo, podría analizarse la transferencia del modelo hacia escalas más locales utilizando datos de demanda de estaciones transformadoras.

Finalmente, se considera de interés explorar arquitecturas híbridas basadas en Transformers y modelos de atención multiescala, así como evaluar estrategias orientadas a *Green AI* que permitan reducir aún más el costo computacional y energético asociado al entrenamiento de modelos predictivos.

6.3. Publicaciones y reconocimientos derivados de esta tesis

6.3.1. Premios y reconocimientos

El trabajo presentado inicialmente en ASAD 2024 fue seleccionado entre los mejores artículos del simposio, recibiendo una invitación para su publicación extendida en el *Electronic Journal of SADIO (EJS)*, revista indexada perteneciente al Núcleo Básico de Revistas Científicas Argentinas.

Asimismo, el trabajo presentado en la *CPAM 2024 – Conferência Pan-Americana de Meteorologia e Simpósio em Clima, Água, Energia e Alimentos*, realizada en São Paulo, Brasil, fue seleccionado entre los mejores pósters del evento y postulado a la instancia final de premiación.

6.3.2. Publicaciones científicas

Los resultados parciales y finales de esta tesis dieron lugar a las siguientes publicaciones científicas:

- Uhrig, M. N., Vignolo, L. D., & Müller, O. V. (2026). *Multivariate forecasting of energy demand using recurrent neural networks*. Advanced Engineering Informatics, 71, 104419. <https://doi.org/10.1016/j.aei.2026.104419>
- Uhrig, M. N., Vignolo, L. D., & Müller, O. V. (2025). *Modelado y predicción de la demanda eléctrica: comparación de enfoques estadísticos y de aprendizaje automático*. SADIO Electronic Journal of Informatics and Operations Research (EJS), 24(1). <https://doi.org/10.24215/15146774e067>

Asimismo, como continuidad de esta línea de investigación, se encuentra en preparación un manuscrito destinado a publicar el modelo propuesto *AMBER*.

6.3.3. Presentaciones en congresos científicos

Los avances y resultados obtenidos durante el desarrollo de esta tesis fueron presentados en distintos eventos científicos nacionales e internacionales:

- First B2IST Project Workshop, Universidad Federal de Santa María, Brasil, 2023. Presentación del trabajo *Modelado de la demanda de energía eléctrica en base a pronósticos meteorológicos utilizando redes neuronales artificiales*.
- ASaid 2024 – Simposio Argentino de Inteligencia Artificial y Ciencia de Datos, en el marco de las 53 JAIIO, Bahía Blanca, Argentina, 2024. Presentación del trabajo *Forecast Model of Electricity Demand Based on Meteorological and Historical Demand Data Using Artificial Neural Networks*.
- CPAM 2024 – Conferência Pan-Americana de Meteorologia e Simpósio em Clima, Água, Energia e Alimentos, São Paulo, Brasil, 2024. Presentación del trabajo *Un enfoque de redes neuronales recurrentes para la previsión de la demanda de energía a partir de datos meteorológicos y energéticos*.
- KHIPU 2025 – Latin American Meeting in Artificial Intelligence, 2025. Presentación del póster *Energy, Weather and AI: Short-Term Electricity Demand Forecasting Model*, donde se expusieron resultados vinculados al uso de redes neuronales recurrentes, validación cruzada y variables meteorológicas para la predicción de demanda eléctrica en Entre Ríos.

6.4. Disponibilidad de datos y códigos

Los códigos fuente desarrollados durante esta tesis se encuentran disponibles públicamente en el repositorio [GitHub](#)¹.

Asimismo, el conjunto de datos utilizado en esta investigación se encuentra registrado en el Banco de Datos de Investigación del CONICET bajo el título *Datos de Clima y Energía de la provincia de Entre Ríos* [59]. El dataset se encuentra disponible en el [repositorio](#)² institucional del CONICET.

¹<https://github.com/mnuhrig/Energy-Weather-and-AI-Short-Term-Electricity-Demand-Forecasting-Model>

²<http://hdl.handle.net/11336/253952>

Bibliografía

- [1] Mobarak Abumohsen, Amani Yousef Owda, and Majdi Owda. Electrical load forecasting using lstm, gru, and rnn algorithms. *Energies*, 16(5), 2023. ISSN 1996-1073. doi: 10.3390/en16052283. URL <https://www.mdpi.com/1996-1073/16/5/2283>.
- [2] T. Acker. Iea wind task 24 integration of wind and hydropower systems; volume 1: Issues, impacts, and economics of wind and hydropower integration. Technical Report NREL/TP-5000-50181, National Renewable Energy Lab. (NREL), Golden, CO, United States, 2011.
- [3] Tanveer Ahmad, Dongdong Zhang, Chao Huang, Hongcai Zhang, Ning Dai, Yonghua Song, and Huanxin Chen. Artificial intelligence in sustainable energy industry: Status quo, challenges and opportunities. *Journal of Cleaner Production*, 289:125834, 2021. doi: 10.1016/j.jclepro.2021.125834.
- [4] Fabio Apadula, Andrea Bassini, Alberto Elli, and Silvio Scapin. Relationships between meteorological variables and monthly electricity demand. *Applied Energy*, 98:346–356, 2012. doi: 10.1016/j.apenergy.2012.03.053.
- [5] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bernetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020. doi: 10.1016/j.inffus.2019.12.012.
- [6] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International*

- Conference on Learning Representations (ICLR 2015)*, 2015. URL <https://arxiv.org/abs/1409.0473>.
- [7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013. doi: 10.1109/TPAMI.2013.50.
- [8] Christoph Bergmeir and José M. Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012. doi: 10.1016/j.ins.2011.12.028.
- [9] Filippo Maria Bianchi, Enrico Maiorino, Michael C. Kampffmeyer, Antonello Rizzi, and Robert Jenssen. *Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis*. SpringerBriefs in Computer Science. Springer, Cham, 2017. doi: 10.1007/978-3-319-70338-1.
- [10] Verónica Bolón-Canedo, Laura Morán-Fernández, Brais Cancela, and Amparo Alonso-Betanzos. A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing*, 599:128096, 2024. ISSN 0925-2312. doi: 10.1016/j.neucom.2024.128096.
- [11] Peter J. Brockwell and Richard A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 3 edition, 2016. doi: 10.1007/978-1-4757-2526-1.
- [12] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, 2014. doi: 10.3115/v1/D14-1179.
- [13] Razvan-Gabriel Cirstea, Chenjuan Guo, Bin Yang, Tung Kieu, Xuanyi Dong, and Shirui Pan. Triformer: Triangular, variable-specific attentions for long sequence multivariate time series forecasting. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22)*, pages 1994–2001, 2022. doi: 10.24963/ijcai.2022/277.

- [14] Shuang Dai et al. Machine learning in peak demand forecasting. *Renewable and Sustainable Energy Reviews*, 2025. doi: 10.1016/j.rser.2025.116500.
- [15] Chirag Deb, Fan Zhang, Junjing Yang, Siew Eang Lee, and Kwok Wei Shah. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74:902–924, 2017. doi: 10.1016/j.rser.2017.02.085.
- [16] Federico Divina, Aude Gilson, Francisco Gómez-Vela, Miguel García Torres, and José F. Torres. Stacking ensemble learning for short-term electricity consumption forecasting. *Energies*, 11(4):949, 2018. doi: 10.3390/en11040949.
- [17] Qi Dong et al. Short-term electricity-load forecasting by deep learning: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, 2024. survey.
- [18] Energía de Entre Ríos S.A. (ENERSA). Memoria y estados contables correspondientes al ejercicio finalizado el 31 de diciembre de 2023, 2023. URL <https://www.enersa.com.ar>. Informe institucional.
- [19] Cheng Fan, Fu Xiao, and Yang Zhao. A short-term building cooling load prediction method using deep learning algorithms. *Applied Energy*, 195:222–233, 2017. doi: 10.1016/j.apenergy.2017.03.064.
- [20] Xi Fang, Satyajayant Misra, Guoliang Xue, and Dejun Yang. Smart grid — the new and improved power grid: A survey. *IEEE Communications Surveys & Tutorials*, 14(4):944–980, 2012. doi: 10.1109/SURV.2011.101911.00087.
- [21] N. Ghani, A. El-Shafie, and Z. M. Yaseen. Hybrid machine learning models for energy demand forecasting: A state-of-the-art review. *Energy Reports*, 8: 569–589, 2022. doi: 10.1016/j.egyr.2022.01.042.
- [22] Suján Ghimire, Thong Nguyen-Huy, Mohanad S. AL-Musaylh, Ravinesh C. Deo, David Casillas-Pérez, and Sancho Salcedo-Sanz. Integrated multi-head self-attention transformer model for electricity demand prediction incorporating local climate variables. *Energy and AI*, 14:100302, 2023. doi: 10.1016/j.egyai.2023.100302.

- [23] Sujan Ghimire, Mohanad S. Al-Musaylh, Thong Nguyen-Huy, Ravinesh C. Deo, Rajendra Acharya, Zaher Mundher Yaseen, David Casillas-Pérez, and Sancho Salcedo-Sanz. Explainable deeply-fused nets electricity demand prediction model: Factoring climate predictors for accuracy and deeper insights with probabilistic confidence interval and point-based forecasts. *Applied Energy*, 378:124763, 2025. doi: 10.1016/j.apenergy.2024.124763.
- [24] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL <http://www.deeplearningbook.org>.
- [25] Stephen Haben, Siddharth Arora, Georgios Giasemidis, Marcus Voss, and Danica Vukadinović Greetham. Review of low voltage load forecasting: Methods, applications, and recommendations. *Applied Energy*, 304:117798, 2021. doi: 10.1016/j.apenergy.2021.117798.
- [26] Luis Hernández, Carlos Baladrón, Javier M. Aguiar, Belén Carro, Antonio Sánchez-Esguevillas, and Jaime Lloret. Short-term load forecasting for microgrids based on artificial neural networks. *Energies*, 6(3):1385–1408, 2013. doi: 10.3390/en6031385.
- [27] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803.
- [28] H.S. Hippert, C.E. Pedreira, and R.C. Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16(1):44–55, 2001. doi: 10.1109/59.910780.
- [29] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.
- [30] Tao Hong and Shu Fan. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016. doi: 10.1016/j.ijforecast.2015.11.011.

- [31] Qingbo Hua, Zengliang Fan, Wei Mu, Jiqiang Cui, Rongxin Xing, Huabo Liu, and Junwei Gao. A short-term power load forecasting method using cnn-gru with an attention mechanism. *Energies*, 18(1):106, 2025. doi: 10.3390/en18010106.
- [32] Bingqing Huang, Haonan Zheng, Xinbo Guo, Yi Yang, and Ximing Liu. A novel model based on da-rnn network and skip gated recurrent neural network for periodic time series forecasting. *Sustainability*, 13(21), 2021. doi: 10.3390/su132112120.
- [33] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006. doi: 10.1016/j.ijforecast.2006.03.001.
- [34] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An Introduction to Statistical Learning: with Applications in Python*. Springer, 2023.
- [35] Tae-Young Kim and Sung-Bae Cho. Predicting residential energy consumption using cnn-lstm neural networks. *Energy*, 182:72–81, 2019. doi: 10.1016/j.energy.2019.05.230.
- [36] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations (ICLR 2015)*, 2015. URL <https://arxiv.org/abs/1412.6980>.
- [37] Vaia I. Kontopoulou, Athanasios D. Panagopoulos, Ioannis Kakkos, and George K. Matsopoulos. A review of arima vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, 15, 2023. ISSN 1999-5903. URL <https://www.mdpi.com/1999-5903/15/8/255>.
- [38] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521 (7553):436–444, 2015. doi: 10.1038/nature14539.
- [39] Bryan Lim and Stefan Zohren. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A: Mathematical,*

- Physical and Engineering Sciences*, 379(2194):20200209, 02 2021. ISSN 1364-503X. doi: 10.1098/rsta.2020.0209.
- [40] Bryan Lim, Sercan Ö. Arik, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021. ISSN 0169-2070. doi: 10.1016/j.ijforecast.2021.03.012.
- [41] Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 4765–4774, 2017. URL <https://arxiv.org/abs/1705.07874>.
- [42] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1412–1421, 2015. doi: 10.18653/v1/D15-1166.
- [43] Ivan Malashin, Vadim Tynchenko, Andrei Gantimurov, Vladimir Nelyub, and Aleksei Borodulin. Applications of long short-term memory (lstm) networks in polymeric sciences: A review. *Polymers*, 16(18), 2024. ISSN 2073-4360. doi: 10.3390/polym16182607. URL <https://www.mdpi.com/2073-4360/16/18/2607>.
- [44] Johanna L. Mathieu, Stephan Koch, and Duncan S. Callaway. State estimation and control of electric loads to manage real-time energy imbalance. *IEEE Transactions on Power Systems*, 28(1):430–440, 2013. doi: 10.1109/TPWRS.2012.2204074.
- [45] K. Misiurek et al. Review of methods and models for forecasting electricity demand. *Energies*, 18(15):4032, 2025. doi: 10.3390/en18154032.
- [46] Lukas Moosbrugger, Valentin Seiler, Philipp Wohlgenannt, Sebastian Hegenbart, Sashko Ristov, Elias Eder, and Peter Kepplinger. Load forecasting for households and energy communities: Are deep learning models worth the effort? *arXiv preprint arXiv:2501.05000*, 2025. URL <https://arxiv.org/abs/2501.05000>. Code available at: <https://github.com/rce-flh/loadforecasting/>.

- [47] Alper Ozcan, Cagatay Catal, and Ahmet Kasif. Energy load forecasting using a dual-stage attention-based recurrent neural network. *Sensors*, 21(21), 2021. doi: 10.3390/s21217115.
- [48] Keshab K. Parhi and Nanda K. Unnikrishnan. Brain-inspired computing: Models and architectures. *IEEE Open Journal of Circuits and Systems*, 1:185–204, 2020. doi: 10.1109/OJCAS.2020.3032092.
- [49] Jieyang Peng, Andreas Kimmig, Jiahai Wang, Xiufeng Liu, Zhibin Niu, and Jivka Ovtcharova. Dual-stage attention-based long-short-term memory neural networks for energy demand prediction. *Energy and Buildings*, 249:111211, 2021. doi: 10.1016/j.enbuild.2021.111211.
- [50] Yao Qin, Dongjin Song, Haifeng Chen, Wei Cheng, Guofei Jiang, and Garrison Cottrell. A dual-stage attention-based recurrent neural network for time series prediction. *arXiv*, 2017. doi: 10.48550/arXiv.1704.02971.
- [51] Daniel Ramos, Pedro Faria, and Zita Vale. Linking short-term electricity demand forecasting and explainable ai: A review for building energy applications. *Applied Energy*, 412:127658, 2026. ISSN 0306-2619. doi: 10.1016/j.apenergy.2026.127658.
- [52] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019. doi: 10.1038/s42256-019-0048-x.
- [53] Jason Runge and Radu Zmeureanu. A review of deep learning techniques for forecasting energy use in buildings. *Energies*, 14(3):608, 2021. doi: 10.3390/en14030608. URL <https://www.mdpi.com/1996-1073/14/3/608>.
- [54] Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. Green ai. *Communications of the ACM*, 63(12):54–63, 2020. doi: 10.1145/3381831.
- [55] Leo Semmelmann, Sarah Henni, and Christof Weinhardt. Load forecasting for energy communities: a novel lstm-xgboost hybrid model based on smart meter data. *Energy Informatics*, 5(1), 2022. ISSN 2520-8942. doi: 10.1186/s42162-022-00212-9.

- [56] Souhaib Ben Taieb, Gianluca Bontempi, Amir F. Atiya, and Antti Sorjamaa. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Engineering Applications of Artificial Intelligence*, 25(8):1580–1595, 2012. doi: 10.1016/j.engappai.2012.02.009.
- [57] James W. Taylor. Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8):799–805, 2003. doi: 10.1057/palgrave.jors.2601589.
- [58] Joaquín F. Torres, Francisco Martínez-Álvarez, and Alicia Troncoso. A deep lstm network for the spanish electricity consumption forecasting. *Neural Computing and Applications*, 34:10533–10545, 2022. doi: 10.1007/s00521-021-06773-2.
- [59] Mariela N. Uhrig, Leandro D. Vignolo, and Omar V. Müller. Datos de clima y energía de la provincia de entre ríos, 2025. URL <http://hdl.handle.net/11336/253952>. Banco de Datos de Investigación del CONICET.
- [60] Mariela N. Uhrig, Leandro D. Vignolo, and Omar V. Müller. Modelling and forecasting electricity demand: comparing statistical and machine learning approaches. *SADIO Electronic Journal of Informatics and Operations Research*, 24(1), 2025. doi: 10.24215/15146774e067.
- [61] Mariela N. Uhrig, Leandro D. Vignolo, and Omar V. Müller. Multivariate forecasting of energy demand using recurrent neural networks. *Advanced Engineering Informatics*, 71:104419, 2026. doi: 10.1016/j.aei.2026.104419.
- [62] Kaleem Ullah, Muhammad Ahsan, Syed Muhammad Hasanat, Muhammad Harris, Hamza Yousaf, Syed Faraz Raza, Ritesh Tandon, Samain Abid, and Zahid Ullah. Short-term load forecasting: A comprehensive review and simulation study with cnn-lstm hybrids approach. *IEEE Access*, 12:111858–111881, 2024. doi: 10.1109/ACCESS.2024.3440631.
- [63] Enric Valor, Vicente Meneu, and Vicente Caselles. Daily air temperature and electricity load in spain. *Journal of Applied Meteorology*, 40(8):1413–1421, 2001. doi: 10.1175/1520-0450(2001)040<1413:DATAEL>2.0.CO;2.

- [64] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 5998–6008, 2017. URL <https://arxiv.org/abs/1706.03762>.
- [65] Chong Wang, Xiang Li, Yan Shi, Wenshuai Jiang, Qiong Song, and Xueling Li. Load forecasting method based on cnn and extended lstm. *Energy Reports*, 12: 2452–2461, 2024. ISSN 2352-4847. doi: 10.1016/j.egy.2024.07.030.
- [66] Yuxuan Wang, Chenglong Li, Jinyuan Chen, Zezhi Shao, Cheng Zhang, Defu Cao, Yuntong Wang, Gaofeng Luan, Yanhua Liu, James Y. Wang, et al. Timexer: Empowering transformers for time series forecasting with exogenous variables. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, pages 1–24, 2024. URL <https://arxiv.org/abs/2402.19072>.
- [67] Zhixian Wang et al. Diffload: Uncertainty quantification in electrical load forecasting with the diffusion model. *IEEE Transactions on Power Systems*, 2023. doi: 10.1109/TPWRS.2024.3449032. preprint.
- [68] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014. doi: 10.1016/j.ijforecast.2014.08.008.
- [69] Cort J. Willmott. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, 63(11):1309–1313, 1982. doi: 10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2.
- [70] Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 22419–22430, 2021. URL <https://arxiv.org/abs/2106.13008>.
- [71] B. Xue and J. Geng. Dynamic transverse correction method of middle and long term energy forecasting based on statistic of forecasting errors. In *2012 10th International Power & Energy Conference (IPEC)*, pages 253–256. IEEE, 2012.

- [72] Diming Zhang, Yuchen Xu, and Yuanjiang Li. Electric load forecasting based on kernel extreme learning machine optimized by improved sparrow search algorithm. *Scientific Reports*, 2025. doi: 10.1038/s41598-025-07755-3.
- [73] Monika Zimmermann and Florian Ziel. Spatial weather, socio-economic and political risks in probabilistic load forecasting. *arXiv*, 2024. doi: <https://doi.org/10.48550/arXiv.2408.00507>. preprint.