



DEPARTAMENTO DE
INGENIERÍA ELÉCTRICA
UNIVERSIDAD DE CONCEPCIÓN

ESTIMACIÓN DEL CONTENIDO DE COBRE Y RAZÓN HIERRO SÍLICE EN LA ESCORIA DE UN CONVERTIDOR FLASH USANDO DATOS DE PROCESO Y DEEP LEARNING

POR

Kevin Darren Navarrete Zurita

Memoria de Título presentada a la Facultad de Ingeniería de la Universidad de Concepción
para optar al título profesional de Ingeniero Civil en Telecomunicaciones.

Profesor Guía

Dr. Francisco Pérez Venegas

Concepción,

26 de septiembre de 2025

© 2025 Kevin Navarrete Zurita

© 2025 Kevin Navarrete Zurita

Ninguna parte de esta tesis puede reproducirse o transmitirse bajo ninguna forma o por ningún medio o procedimiento, sin permiso por escrito del autor.

A mis padres, la más grande suerte de mi vida es tenerlos

Resumen

El presente trabajo aborda el desarrollo de una herramienta de monitoreo que permita el seguimiento de la ley de cobre (Cu) y la razón hierro-sílice (Fe/SiO₂) en la escoria de un horno de conversión flash. Esta problemática surge de la dependencia actual de métodos de análisis químicos, invasivos y con desfase temporal, los cuales impiden un monitoreo en tiempo real, no invasivo y en línea.

Para solucionar esta problemática se propone un sistema no invasivo basado en el uso de datos de proceso y modelos de *Deep Learning* programados en Python. La metodología consistió en la construcción de dos conjuntos de datos: uno endógeno, basado en el pasado de las variables objetivo y otro exógeno, compuesto por variables de proceso operacional, simulando así un escenario industrial real donde no siempre se dispone de mediciones químicas pasadas. Esto permitió evaluar la capacidad de los modelos en ambos contextos.

Se implementaron y compararon tres arquitecturas predictivas: un modelo de regresión lineal (que sirvió como línea base), una red neuronal recurrente LSTM (*Long Short-Term Memory*) y un *Temporal Fusion Transformer* (TFT). El desempeño se evaluó utilizando como métrica principal *Mean Absolute Scaled Error* (MASE), y las métricas como MAE, sesgo de pronóstico (FB), coeficiente de correlación de Pearson, R² e información mutua normalizada (NMI).

Los resultados obtenidos muestran que la regresión lineal constituye una base sólida para capturar tendencias generales, pero resulta insuficiente frente a la alta variabilidad de las series. El modelo LSTM alcanzó el mejor desempeño global, demostrando robustez en la estimación tanto del cobre como de la razón hierro-sílice en el conjunto endógeno. Por su parte, el TFT presentó un mejor comportamiento en el conjunto exógeno, al explotar relaciones no lineales en las variables de proceso, aunque en el conjunto endógeno exhibió

dificultades de generalización.

En conclusión, este trabajo analiza la factibilidad de construir un sistema de monitoreo no invasivo para la industria del cobre, con proyección a su integración en sistemas de control en línea. Este enfoque contribuye a la optimización de los procesos pirometalúrgicos, ofreciendo beneficios en seguridad, reducción de costos y monitoreo operacional.

Abstract

This work addresses the development of a monitoring tool that enables the tracking of copper grade (Cu) and the iron–silica ratio (Fe/SiO₂) in the slag of a flash converting furnace. This problem arises from the current reliance on chemical analysis methods, which are invasive and delayed, preventing real-time, non-invasive, and online monitoring.

To solve this issue, a non-invasive system is proposed based on the use of process data and *Deep Learning* models programmed in Python. The methodology consisted of constructing two datasets: an endogenous one, based on the past values of the target variables, and an exogenous one, composed of operational process variables, thus simulating a real industrial scenario where past chemical measurements are not always available. This approach allowed evaluating the capability of the models in both contexts.

Three predictive architectures were implemented and compared: a linear regression model (serving as the baseline), a recurrent neural network LSTM (*Long Short-Term Memory*), and a *Temporal Fusion Transformer* (TFT). Performance was assessed using *Mean Absolute Scaled Error* (MASE) as the main metric, along with MAE, forecast bias (FB), Pearson’s correlation coefficient, R^2 , and normalized mutual information (NMI).

The results show that linear regression provides a solid basis for capturing general trends, but is insufficient when facing the high variability of the series. The LSTM model achieved the best overall performance, demonstrating robustness in estimating both copper grade and the iron–silica ratio in the endogenous dataset. Meanwhile, the TFT exhibited better behavior in the exogenous dataset by exploiting non-linear relationships among process variables, although it showed generalization difficulties in the endogenous case.

In conclusion, this work analyzes the feasibility of building a non-invasive monitoring system for the copper industry, with the potential to be integrated into online control systems. This approach contributes to the optimization of pyrometallurgical processes, offering benefits in safety, cost reduction, and operational monitoring.

Agradecimientos

Falta espacio y faltan palabras para dar agradecimientos, soy una persona afortunada. Quiero agradecer a mis padres, que siempre me han apoyado, que me han educado y convertido en la persona que soy hoy. Llevo en mis recuerdos cada una de las enseñanzas que me han dado. Creo que la mejor forma que tengo para agradecerles todo es dando lo mejor de mí y seguir avanzando (algo que realmente me gusta; no soy de quedarme sentado esperando que pase la vida).

Aún recuerdo los viajes al río-playa con un plátano y un Super 8 jajajaj, cuando viajamos en el auto y teníamos que bajar a empujarlo, cuando me enseñaron a nadar a la vieja escuela. . . Me enseñaron a apreciar lo sencillo de la vida y a no impresionarme por los lujos que la gente quiere ostentar.

Cuando mamá me espantaba a mis "pretendientes", cuando papá me acompañaba a hacer deporte (siempre molestando que lo hiciera con más ganas jajaja). Nunca hubo seriedad en nuestra vida; cada momento, por más duro y tenso que fuera, siempre tuvo una risa. Me enseñaron que no hay que ahogarse en un vaso de agua; no importa que sea difícil, hay que hacerlo igual, y si nos podemos reír en el trayecto, mucho mejor.

Los amo mucho, padres, y amo mucho a mi hermanito. Espero verlo crecer con ansias, ya que ya puedo ver que heredó el humor de la familia jajaja. Agustín, si algún día lees esto, quiero que sepas que te molestaré toda la vida y siempre te recordaré que te cambiaba los pañales y que te gustaba la Elif jajajja. Siempre puedes contar conmigo, hermano. Te amo mucho.

Agradezco también a mis amigos. Siempre he intentado, y creo que sigo siendo la misma persona que hace 20 años. En el trayecto muchos fallaron, pero también muchos quedaron. Los quiero mucho y, sin ustedes, todo esto hubiera sido más difícil (no nombraré a ninguno para que no anden cobrando sentimientos jajajaj).

Gracias, vida, por dejarme vivirte. Este extracto es lo único informal de esta memoria, pero a la vez es lo que se encuentra en lo más profundo de mí. Los quiero mucho a todos.

Índice General

Resumen	I
Abstract	III
Agradecimientos	v
Índice de Figuras	xI
Índice de Tablas	xv
1. Introducción	1
1.1. Trabajos previos	3
1.1.1. Discusión	5
1.2. Definición del problema	6
1.3. Objetivos	7
1.3.1. Objetivo general	7
1.3.2. Objetivos específicos	7
1.4. Metodología	8
1.4.1. Metodología general de trabajo	8
1.4.2. Definición del conjunto de datos	8
1.4.3. Desarrollo de modelos	9
1.4.4. Comparativa de resultados	10
1.4.5. Conclusiones	10
1.4.6. Incorporación industrial de la herramienta	10
1.5. Alcances y limitaciones	11
1.5.1. Alcances	11
1.5.2. Limitaciones	12

2. Marco Teórico	13
2.1. Proceso químico	13
2.2. Ley cobre	14
2.3. Razon hierro silice	14
2.4. Machine learning	15
2.5. Deep learning	15
2.6. Interpolación	15
2.7. Time delay Embedding	16
2.8. Validación cruzada	16
2.9. Grid Search	17
2.10. Métricas	17
2.10.1. MAE	17
2.10.2. MASE	18
2.10.3. Sesgo de pronóstico (FB)	19
2.10.4. Coeficiente de correlación lineal de Pearson	19
2.10.5. Coeficiente de determinación (R^2)	20
2.10.6. Mutual Information (MI)	21
2.11. Correlación cruzada	22
2.12. Regresión lineal múltiple	23
2.13. Long Short-Term Memory (LSTM)	23
2.14. Temporal Fusion Transformer (TFT)	24
3. Desarrollo	27
3.1. Introducción	27
3.2. Definición del conjunto de datos	28
3.2.1. Análisis estadístico del conjunto de datos	29
3.2.2. Preprocesamiento	33
3.2.3. Creación del conjunto de datos endógenos	37
3.2.4. Creación del conjunto de datos exógenos	39
3.3. Desarrollo de modelos	40
3.4. Regresión lineal	41
3.4.1. Curvas de perdida.	42
3.5. Long Short-Term Memory	45
3.5.1. Curvas de pérdida	46

3.6. Temporal Fusion Transformer (TFT)	47
3.6.1. Curvas de pérdida	48
3.7. Modelos endógenos	49
3.7.1. Ley de cobre	49
3.7.2. Razón hierro-sílice	51
3.8. Modelos suavizados	53
3.8.1. Ley de cobre	55
3.8.2. Razón hierro-sílice	56
4. Resultados	59
4.1. Resultados Regresión lineal para pronóstico de ley de cobre	59
4.1.1. Correlación cruzada	60
4.1.2. Diagrama de dispersión	61
4.1.3. Serie temporal	62
4.2. Resultados Regresión lineal para pronóstico de la razón Fe/SiO ₂	63
4.2.1. Correlación cruzada	64
4.2.2. Diagrama de dispersión	65
4.2.3. Serie temporal	66
4.3. Resultados LSTM para pronóstico de ley de cobre	67
4.3.1. Correlación cruzada	68
4.3.2. Diagrama de dispersión	69
4.3.3. Serie temporal	70
4.4. Resultados LSTM para pronóstico de la razón Fe/SiO ₂	71
4.4.1. Correlación cruzada	72
4.4.2. Diagrama de dispersión	73
4.4.3. Serie temporal	74
4.5. Resultados TFT para pronóstico de ley de cobre	75
4.5.1. Correlación cruzada	76
4.5.2. Diagrama de dispersión	77
4.5.3. Serie temporal	78
4.6. Resultados TFT para pronóstico de la razón Fe/SiO ₂	79
4.6.1. Correlación cruzada	80
4.6.2. Diagrama de dispersión	81
4.6.3. Serie temporal	82

4.7. Discusión de resultados	83
4.8. Resultados conjunto endógeno	84
4.8.1. Ley de cobre	84
4.8.2. Razón hierro-sílice	86
4.8.3. Discusión de resultados modelo endógeno	88
4.9. Resultados modelo suavizado	89
4.9.1. Ley de cobre	89
4.9.2. Razón hierro-sílice	91
4.9.3. Discusión de resultados modelo suavizado	93
5. Conclusiones	94
5.1. Consideraciones en aplicación a la industria	96
5.1.1. Costes computacionales	96
5.1.2. Tasas de error	97
5.1.3. Retrasos en las predicciones	97
5.2. Trabajos futuros	97
Bibliografía	101
Appendices	102
A. ANEXO: Figuras	103
B. ANEXO: Código	107

Índice de Figuras

1.1. Diagrama general de trabajo.	8
1.2. Esquema de integración de la herramienta en la industria.	10
2.1. Time delay Embedding	16
3.1. Series temporales de datos crudos del proceso.	30
3.2. Serie temporal consumo de concentrado	31
3.3. Diagrama de caja consumo de concentrado	31
3.4. Serie temporal de la Temperatura de la escoria.	32
3.5. Histograma de la Temperatura de la escoria.	33
3.6. Diagrama de flujo del preprocesamiento	34
3.7. Series temporales de datos del proceso con frecuencia de 30 minutos.	35
3.8. Serie temporal de consumo de concentrado preprocesado.	36
3.9. Serie temporal de la Temperatura de la escoria.	37
3.10. Histograma de la Temperatura de la escoria.	37
3.11. Curvas de pérdida para el modelo de regresión lineal para el pronóstico de ley de cobre.	43
3.12. Curvas de pérdida para el modelo de regresión lineal para el pronóstico de razón hierro-sílice	44
3.13. Curvas de pérdida para el modelo LSTM aplicado a la ley de cobre.	46
3.14. Curvas de pérdida para el modelo LSTM aplicado a la razón hierro-sílice.	47
3.15. Curvas de pérdida del modelo TFT aplicado a la ley de cobre.	48
3.16. Curvas de pérdida del modelo TFT aplicado a la razón hierro-sílice.	49
3.17. Curvas de pérdida del modelo endógeno de regresión lineal para el pronóstico de la ley de cobre.	50

3.18. Curvas de pérdida del modelo endógeno LSTM para el pronóstico de la ley de cobre.	51
3.19. Curvas de pérdida del modelo endógeno TFT para el pronóstico de la ley de cobre.	51
3.20. Curvas de pérdida del modelo de regresión lineal para el pronóstico de la razón hierro-sílice (conjunto endógeno).	52
3.21. Curvas de pérdida del modelo híbrido LSTM para el pronóstico de la razón hierro-sílice (conjunto endógeno).	52
3.22. Curvas de pérdida del modelo TFT para el pronóstico de la razón hierro-sílice (conjunto endógeno).	53
3.23. Ley de cobre suavizado mediante EMA.	54
3.24. Razón hierro-sílice suavizado mediante EMA.	54
3.25. Curvas de pérdida del modelo suavizado de regresión lineal para el pronóstico de la ley de cobre.	55
3.26. Curvas de pérdida del modelo suavizado LSTM para el pronóstico de la ley de cobre.	56
3.27. Curvas de pérdida del modelo suavizado TFT para el pronóstico de la ley de cobre.	56
3.28. Curvas de pérdida del modelo suavizado de regresión lineal para el pronóstico de la razón hierro-sílice.	57
3.29. Curvas de pérdida del modelo suavizado LSTM para el pronóstico de la razón hierro-sílice.	57
3.30. Curvas de pérdida del modelo suavizado TFT para el pronóstico de la razón hierro-sílice.	58
4.1. Correlación cruzada para la ley de cobre utilizando regresión lineal. . .	60
4.2. Correlación cruzada interpolada para la ley de cobre utilizando regresión lineal	61
4.3. Diagrama de dispersión entre valores objetivo y predichos de la ley de cobre utilizando el modelo de regresión lineal.	62
4.4. Comparación entre valores objetivo y predicciones de la ley de cobre en el conjunto de prueba utilizando el modelo de regresión lineal.	63
4.5. Correlación cruzada para la razón Fe/SiO ₂ utilizando regresión lineal. .	64

4.6. Correlación cruzada interpolada para la razón Fe/SiO ₂ utilizando regresión lineal	65
4.7. Diagrama de dispersión entre valores objetivo y predichos de la razón Fe/SiO ₂ utilizando el modelo de regresión lineal.	66
4.8. Comparación entre valores objetivo y predicciones de la razón Fe/SiO ₂ en el conjunto de prueba utilizando el modelo de regresión lineal. . . .	67
4.9. Correlación cruzada para la ley de cobre utilizando el modelo LSTM. . .	68
4.10. Correlación cruzada interpolada para la ley de cobre utilizando el modelo LSTM.	69
4.11. Diagrama de dispersión entre valores objetivo y predichos de la ley de cobre para el modelo LSTM.	70
4.12. Comparación entre valores objetivo y predicciones de la ley de cobre en el conjunto de prueba utilizando el modelo LSTM.	71
4.13. Correlación cruzada para la razón Fe/SiO ₂ utilizando el modelo LSTM.	72
4.14. Correlación cruzada interpolada para la razón Fe/SiO ₂ utilizando el modelo LSTM.	73
4.15. Diagrama de dispersión entre valores objetivo y predichos de la razón Fe/SiO ₂	74
4.16. Comparación entre valores objetivo y predicciones de la razón Fe/SiO ₂ en el conjunto de prueba utilizando el modelo LSTM.	75
4.17. Correlación cruzada para la ley de cobre utilizando el modelo TFT. . .	76
4.18. Correlación cruzada interpolada para la ley de cobre utilizando el modelo TFT.	77
4.19. Diagrama de dispersión entre valores objetivo y predichos de la ley de cobre utilizando el modelo TFT.	78
4.20. Comparación entre valores objetivo y predicciones de la ley de cobre en el conjunto de prueba utilizando el modelo TFT.	79
4.21. Correlación cruzada para la razón Fe/SiO ₂ utilizando el modelo TFT. .	80
4.22. Correlación cruzada interpolada para la razón Fe/SiO ₂ utilizando el modelo TFT.	81
4.23. Diagrama de dispersión entre valores objetivo y predichos de la razón Fe/SiO ₂ utilizando el modelo TFT.	82
4.24. Comparación entre valores objetivo y predicciones de la razón Fe/SiO ₂ en el conjunto de prueba utilizando el modelo TFT.	83

4.25. Correlación cruzada entre valores objetivo y predicciones para la ley de cobre (Modelo LSTM).	85
4.26. Diagrama de dispersión de valores objetivo vs. predicciones de la ley de cobre (Modelo LSTM).	85
4.27. Serie temporal entre valores objetivo y predicciones de la ley de cobre (Modelo LSTM).	86
4.28. Correlación cruzada entre valores objetivo y predicciones para la razón hierro-sílice.	87
4.29. Diagrama de dispersión de valores objetivo vs. predichos de la razón hierro-sílice.	87
4.30. Comparación temporal entre valores objetivo y predicciones de la razón hierro-sílice.	88
4.31. Correlación cruzada para la ley de cobre modelo TFT suavizado.	90
4.32. Diagrama de dispersión para la ley de cobre modelo TFT suavizado	90
4.33. Serie temporal para la ley de cobre modelo TFT suavizado	91
4.34. Correlación cruzada para la razón Fe/SiO_2 modelo TFT suavizado.	92
4.35. Diagrama de dispersión para la razón Fe/SiO_2 modelo TFT suavizado	92
4.36. Serie temporal para la razón Fe/SiO_2 modelo TFT suavizado	93
A.1. Series temporales con frecuencia de 1 hora.	103
A.2. Series temporales con frecuencia de 1 hora.	104
A.3. Histogramas con frecuencia de 1 hora.	104
A.4. Histogramas con frecuencia de 1 hora.	105
A.5. Diagramas de caja con frecuencia de 1 hora.	105
A.6. Diagramas de caja con frecuencia de 1 hora.	106

Índice de Tablas

3.1. Variables y cantidad de datos	32
3.2. Variables y cantidad de datos	36
3.3. Conjunto de datos Endógeno.	38
3.4. Descripción de columnas en conjunto de datos endógeno.	38
3.5. Descripción de columnas en conjunto de datos exógenos.	40
4.1. Métricas de desempeño del modelo de regresión lineal en el conjunto de prueba para la ley de cobre.	60
4.2. Métricas de desempeño del modelo de regresión lineal en el conjunto de prueba para la razón Fe/SiO ₂	64
4.3. Métricas de desempeño del modelo LSTM en el conjunto de prueba para la ley de cobre.	68
4.4. Métricas de desempeño del modelo híbrido LSTM en el conjunto de prueba para la razón Fe/SiO ₂	72
4.5. Métricas de desempeño del modelo TFT en el conjunto de prueba para la ley de cobre.	76
4.6. Métricas de desempeño del modelo TFT en el conjunto de prueba para la razón Fe/SiO ₂	80
4.7. Métricas de desempeño en el conjunto endógeno para la ley de cobre.	84
4.8. Métricas de desempeño en el conjunto endógeno para la razón hierro-sílice.	86
4.9. Métricas de desempeño (suavizado) para la ley de cobre.	89
4.10. Métricas de desempeño (suavizado) para la razón hierro-sílice.	91

Siglas

DL Deep Learning

FB Forecast Bias

LSTM Long Short Term Memory

MAE Mean Absolute Error

MASE Mean Absolute Scaled Error

MI Mutual Information

ML Machine learning

MSE Mean Squared Error

TCN Temporal Convolutional Network

TFT Temporal Fusion Transformer

Capítulo 1

Introducción

Durante el siglo XIX, la minería del cobre se consolidó como una de las principales actividades económicas de Chile, posicionando al país como un líder mundial en su producción. A lo largo del tiempo, el cobre ha sido clave para el desarrollo económico del país, especialmente tras la Guerra del Pacífico y la expansión de la minería. [1]

En la actualidad, Chile continúa siendo un actor estratégico en el mercado global del cobre, concentrando cerca del 21 % de las reservas mundiales en estado metálico. Según el Anuario de la Minería publicado por SERNAGEOMIN, la minería del cobre representó 10,9 % del PIB chileno en 2022. Por lo tanto, afirmar que la extracción del cobre es el pilar de nuestra economía no resulta una exageración [2].

Sin embargo, en las fundiciones de cobre aún no existen métodos estandarizados que permitan evaluar, de forma integral y en línea, la productividad, calidad y eficiencia global de los procesos de fundición. Los análisis en los hornos generalmente se realizan en dos etapas: una antes del inicio del proceso y otra periódicamente, entregando información sobre la ley de cobre y la razón hierro-sílice de la escoria, con ciertos desfases temporales. Este procedimiento requiere de un laboratorio especializado para efectuar el análisis químico de las muestras, lo que lo convierte en un método invasivo y sin capacidad de proporcionar datos en tiempo real sobre el estado del proceso.

Un estudio realizado en 2019 por investigadores de la Universidad de Concepción, pertenecientes a los departamentos de Ingeniería Civil en Telecomunicaciones y de Ingeniería Metalúrgica [3], describe la implementación de un sensor radiométrico en

el proceso de fundición de cobre. Este sensor ofrece múltiples ventajas comparativas respecto a los métodos tradicionales, ya que mide la emisión espectral generada durante el proceso, detectando cambios en la temperatura y en la composición de los materiales. De este modo, se establece un sistema de monitoreo no invasivo, mejorando la supervisión y el control del proceso.

En este contexto, las herramientas basadas en Machine Learning (ML) representan una oportunidad relevante para superar las limitaciones actuales en la predicción y el análisis de patrones complejos en los datos obtenidos. Estas herramientas permiten, por ejemplo, detectar variaciones en la composición química de la escoria o monitorear en línea parámetros críticos en procesos pirometalúrgicos. Su capacidad para manejar datos ruidosos y capturar relaciones no lineales las convierte en aliadas ideales para analizar información de los procesos [4]. Particularmente, las técnicas avanzadas de Deep Learning (DL) permiten mejorar la precisión en la estimación de parámetros críticos, como la ley de cobre y la razón hierro-sílice en dichos procesos.

A diferencia de los trabajos anteriores, en la actual memoria utilizaremos una combinación de datos de proceso con algoritmos de aprendizaje profundo (DL), lo cual busca mejorar el monitoreo de convertidores flash, proporcionando datos más precisos.

Este trabajo busca contribuir al desarrollo de esta tecnología mediante un análisis comparativo de distintos modelos de regresión clásicos y modelos predictivos de aprendizaje profundo programados en Python. Se evaluarán aspectos como la precisión y exactitud de las predicciones, la eficiencia computacional y la capacidad de generalización en diferentes conjuntos de datos, con el objetivo de proponer una herramienta en línea que mejore la eficiencia y sostenibilidad de los procesos industriales en la minería del cobre.

Cabe destacar que, en el desarrollo de los modelos predictivos, se adoptó un enfoque metodológico de un conjunto de datos exógenos que prescinde del uso de valores pasados de las variables objetivo (ley de cobre y razón hierro-sílice). Esta decisión se fundamenta en dos aspectos centrales. En primer lugar, responde a una restricción operacional real en una planta industrial, el análisis químico de la escoria conlleva un retraso inherente, por lo que, para realizar una predicción en un instante determinado, solo se dispone de los datos de proceso capturados por sensores en tiempo real. En segundo lugar, este enfoque permite evaluar la capacidad real de los modelos para aprender y explotar las

relaciones causales entre las variables exógenas de entrada y las variables objetivo, sin depender de la autocorrelación temporal de la serie objetivo.

En este sentido, el presente trabajo no solo evalúa el desempeño predictivo de distintas arquitecturas, sino que también busca validar la factibilidad de implementar un sistema de control en línea en un escenario realista, donde la predicción debe generarse exclusivamente a partir de las variables de proceso disponibles en tiempo real.

Adicionalmente, se considera el desarrollo de los modelos mediante el uso de un conjunto de datos endógeno, basado únicamente en los valores pasados de la variable objetivo, con el fin de analizar la capacidad de predicción mediante su propia autorrelación temporal.

1.1. Trabajos previos

La implementación de diversos modelos de regresión mediante Machine Learning para predecir la ley de cobre en el proceso pirometalúrgico de fusión de cobre. La obtención de datos mediante la utilización de sensores ópticos y la obtención de parámetros de temperatura y emisividad total mediante la ley de Planck, que son utilizados para el entrenamiento del modelo de regresión.

- B. A. Rossel Morales, *Machine Learning, Sensores Ópticos en Hornos de Fusión de Baño*, Memoria de Título, Universidad de Concepción, Concepción, Chile, 2022.[5]

Basándose en datos obtenidos mediante un sensor radiométrico y la ley de Planck se plantea un modelo para calcular simultáneamente la emisividad y la temperatura de un proceso pirometalúrgico, el modelo basado en n longitudes espectrales prueba ser más robusto y preciso en procesos con presencia de ruido significativo en comparación con el modelo de dos longitudes de onda.

- J. L. Torres Sanhueza, *Cálculo de Temperatura y Emisividad en Datos de Espectrometría de Procesos Pirometalúrgicos por Método Inverso*, Memoria de Título, Universidad de Concepción, Concepción, Chile, 2022.[6]

Medición espectral en la fundición de cobre mediante el método de flash smelting con el uso de sensores radiométricos. Este estudio propone una alternativa no invasiva y en tiempo real para medir la temperatura mediante el método de dos longitudes de

onda y el uso de sensores radiométricos para mejorar la precisión en el monitoreo de la temperatura y composición de los materiales durante el proceso de fundición.

- L. Arias, S. Torres, C. Toro, E. Balladares, R. Parra, C. Loeza, C. Villagrán, y P. Coelho, “Flash Smelting Copper Concentrates Spectral Emission Measurements,” *Sensors*, vol. 18, no. 7, p. 2009, Jun. 2018, doi: 10.3390/s18072009.[3]

Mediante la utilización de un espectrómetro que opera en el espectro visible y cercano al infrarrojo (VIS-NIR) y la utilización de la ley de Planck, se hace el cálculo de la temperatura radiométrica y la emisividad espectral utilizando el método de multicanal de longitud de onda (MWM), este modelo se compara con el método de dos longitudes de onda (TWM).

- J. Torres-Sanhueza, F. Rivas, F. G. Pérez, R. Parra, S. N. Torres, and S. E. Godoy, “Radiometric methods for high temperature flame estimation: Robustness analysis and application to an industrial flash smelting furnace,” *IEEE Trans. Instrum. Meas.*, vol. 72, no. 2023, pp. 1–20, Aug. 2023, doi: 10.1109/TIM.2023.3302381.[7]

Libro enfocado en el aprendizaje de diversas técnicas modernas en el ámbito del aprendizaje automático de máquinas para la predicción de series temporales. Entre los métodos se destacan modelos básicos como ARIMA y suavizado espacial, y modelos más avanzados como Random Forest, Gradient Boosting, y redes neuronales profundas como LSTM, Transformers y Temporal Fusion Transformers (TFT).

- M. Joseph, *Modern Time Series Forecasting with Python: Explore Industry-Ready Time Series Forecasting Using Modern Machine Learning and Deep Learning*, 1st ed. Birmingham: Packt Publishing, 2022.[4]

Se presentan los principios de la radiación térmica, fenómeno fundamental en la transferencia de calor. Se describe la ley de Planck cómo la radiación emitida por un cuerpo que depende de su temperatura y de la longitud de onda. Este concepto es clave para entender la eficiencia energética en procesos industriales y de fundición, además de permitir diseñar sistemas de monitoreo para controlar distintos KPI's de un proceso industrial.

- Yunus A. Cengel, *Transferencia de Calor y Masa: Un Enfoque Práctico*, 3ra ed., México: McGraw-Hill Interamericana, 2007.[8]

Utilizando espectrometría en tiempo real, se monitorea y analiza la influencia de la temperatura y la concentración de oxígeno en las características de la combustión. Se simula las condiciones de un horno de fusión flash a escala utilizando un horno tipo droptube.

- R. N. Bandak Ramírez, “Caracterización espectral de la combustión *flash* a nivel de laboratorio de especies mineralógicas individuales y un concentrado real,” Memoria de Título, Ingeniería Civil Metalúrgica, Univ. de Concepción, Concepción, Chile, 2018.[9]

1.1.1. Discusión

En los últimos años, se han realizado avances significativos en la incorporación de monitoreo no invasivo en procesos metalúrgicos, mediante el uso de sensores radiométricos. Diversos estudios han aplicado modelos de regresión utilizando técnicas de Machine Learning (ML) en plataformas como Matlab, enfocándose en la estimación de propiedades fisicoquímicas de materiales a partir de datos espectrales.

Sin embargo, el presente trabajo identifica limitaciones en el estado del arte. Primero, existe una fuerte dependencia de modelos de regresión tradicionales, los cuales, si bien son interpretables, poseen una capacidad limitada para capturar relaciones no lineales altamente complejas y dinámicas propias de un horno de conversión flash. Segundo, la operación en ecosistemas cerrados como Matlab limita la escalabilidad, integración y reproducibilidad de las soluciones, aspectos cruciales para su implementación industrial. Finalmente, muchos de estos estudios se realizan en condiciones de laboratorio o con datos preprocesados, lo que no prepara a los modelos para el ruido sensorial, los datos faltantes y las perturbaciones inherentes a una planta real.

Por lo tanto, esta memoria aborda estas limitaciones mediante la implementación y evaluación comparativa de modelos predictivos de Deep Learning, utilizando datos de proceso, con el fin de desarrollar una herramienta robusta y escalable para el monitoreo en tiempo real de variables objetivo en hornos de conversión flash.

1.2. Definición del problema

Como se discutió en la revisión bibliográfica, los sistemas actuales de monitoreo de propiedades fisicoquímicas en la escoria, tales como el contenido de cobre y la razón hierro-sílice, dependen en gran medida de técnicas invasivas y/o de análisis de laboratorio. Estos procedimientos requieren la extracción física de muestras y su análisis posterior en laboratorio, lo que implica altos costos operacionales, baja frecuencia de muestreo y ausencia de información en tiempo real sobre el estado del proceso.

Si bien se han desarrollado modelos de regresión utilizando *Machine Learning* implementados en *Matlab* [5], estas propuestas presentan limitaciones relevantes:

- No incorporan datos de proceso.
- Se restringen al uso de arquitecturas clásicas, sin explorar modelos más complejos y robustos como los de *Deep Learning*.
- Se limitan en la predicción de solo una variable objetivo.
- Suelen implementarse en entornos privativos, lo que reduce la flexibilidad, escalabilidad y replicabilidad industrial.

Ante esta carencia, el presente trabajo propone la exploración y comparación de distintos modelos predictivos basados en *Deep Learning* para evaluar su efectividad en la estimación de variables fisicoquímicas críticas del proceso pirometalúrgico, específicamente la **ley de cobre** y la **razón hierro-sílice**.

Para abordar este problema, se plantea la construcción y evaluación de dos enfoques de conjuntos de datos:

1. Un enfoque **endógeno**, que utiliza únicamente el historial de las variables objetivo, transformado mediante *Time Delay Embedding*, para evaluar la predictibilidad basada en pura autocorrelación temporal.
2. Un enfoque **exógeno**, que incorpora series temporales multivariadas de variables de operación del proceso, para explotar relaciones causales y proveer estimaciones basadas en el estado actual del horno.

Este planteamiento busca determinar no sólo qué arquitectura de *Deep Learning* ofrece

mejor desempeño para este problema, sino también una estrategia de estructuración de datos adecuada para un sistema de monitoreo no invasivo, escalable y aplicable en entornos industriales reales.

1.3. Objetivos

1.3.1. Objetivo general

Desarrollar una herramienta de monitoreo no invasiva que permita estimar en tiempo real el contenido de cobre y la razón hierro-sílice en la escoria de un horno de conversión flash, utilizando datos de proceso y modelos de Deep Learning implementados en Python para procesar dicha información y generar predicciones de las variables objetivo del proceso (ley de cobre y razón hierro-sílice).

1.3.2. Objetivos específicos

En base al objetivo general nuestros objetivos específicos serán los siguientes:

1. Definir un conjunto de datos de proceso mediante la definición, extracción y preprocesamiento de variables relevantes, incluyendo normalización, alineación temporal y segmentación del conjunto de datos para entrenamiento, validación y prueba.
2. Establecer una línea base de desempeño utilizando modelos clásicos de regresión, que sirvan de referencia para la evaluación de modelos avanzados.
3. Desarrollar modelos predictivos basados en *Deep Learning* para la estimación de la ley de cobre y la razón hierro-sílice, seleccionando e implementando arquitecturas adecuadas para series de tiempo.
4. Evaluar el desempeño de los modelos desarrollados utilizando métricas cuantitativas, analizando su precisión y capacidad de generalización sobre datos no vistos previamente.
5. Comparar los modelos implementados en cuanto a precisión, tiempo de entrenamiento y robustez, mediante la elaboración de una tabla comparativa.

6. Interpretar los resultados obtenidos mediante un análisis de las predicciones obtenidas y los valores reales del proceso.

1.4. Metodología

1.4.1. Metodología general de trabajo

El trabajo se desarrolló siguiendo una metodología general compuestas por cuatro etapas principales que van desde la definición de los conjuntos de datos hasta las conclusiones de los resultados.

En la Figura 1.1 se muestra el esquema general que resume el flujo de trabajo.

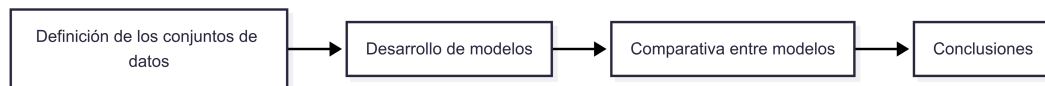


Fig. 1.1: Diagrama general de trabajo.

1.4.2. Definición del conjunto de datos

Objetivo específico 1: Preparar el conjunto de datos que servirá para el entrenamiento y evaluación de los modelos.

1. **Definición de variables relevantes.** Investigar las propiedades fisicoquímicas relacionadas con el proceso pirometalúrgico del cobre de conversión flash y exploración de los datos de proceso, para seleccionar cuales son las variables aptas para entrenar los modelos predictivos.
2. **Análisis estadístico.** Se calcularon estadísticas como cantidad de datos, media y varianza. Además, se realizaron histogramas y diagramas de caja para observar la distribución y detectar valores atípicos..
3. **Preprocesamiento de los datos.** Alineación temporal de los datos debido a las diferencias de sensorización de los datos de proceso y los datos fisicoquímicos, además de la normalización y eliminación de valores atípicos.

4. **Definición de conjuntos** Se construyen dos conjuntos de datos para entrenamiento:

- Conjunto **Endógeno**: Datos de ley de cobre y razón Fe/SiO₂ transformados mediante la técnica de *time embedding* para incorporar información histórica y patrones temporales en el entrenamiento.
- Conjunto **Exógeno**: variables externas que pueden influir en el comportamiento del proceso.

1.4.3. Desarrollo de modelos

Segmentación de datos. El conjunto de datos se dividió en entrenamiento (70%), validación (15%) y prueba (15%), asegurando que los subconjuntos fueran representativos.

Objetivo específico 2: Establecer una línea base con modelos clásicos.

1. **Modelos utilizados.** Regresión lineal.
2. **Entrenamiento.** Se aplicaron *pipelines* con estandarización y validación cruzada para evitar sobreajuste.
3. **Evaluación.** Se usaron métricas como MAE, MASE, sesgo de pronóstico (FB), Pearson, información mutua (MI) y R^2 .
4. **Documentación.** Se registraron hiperparámetros, configuraciones y resultados.

Objetivo específico 3: Implementar modelos de *Deep Learning*.

1. **Modelos utilizados.** LSTM y TFT.
2. **Entrenamiento.** Configuración de hiperparámetros como épocas, tasa de aprendizaje y optimizador, entre otras.
3. **Evaluación.** Se aplicaron las mismas métricas que en los modelos clásicos para poder compararlos.
4. **Documentación.** Se registraron hiperparámetros, configuraciones y resultados.

Objetivo específico 4: Evaluar la capacidad de los modelos para generalizar sobre datos no vistos.

Mediante la definición de un conjunto de datos de prueba correspondiente al 15 % de la totalidad del conjunto de datos.

1.4.4. Comparativa de resultados

Objetivo específico 5: Comparar todos los modelos según precisión, tiempo de entrenamiento y robustez, presentando los resultados en una tabla comparativa.

Mediante la compración de métrica como MAE, MASE, sesog de pronóstico (FB), Pearson, información mutua (MI) y R^2 .

1.4.5. Conclusiones

Objetivo específico 6: Analizar los resultados y compararlos con el comportamiento real del proceso.

Se hace un análisis profundo de los resultados obtenidos con el modelo *Temporal Fusion Trnasformer*(TFT), visualizando además su correlación cruzada y dando un analisis fisico de los valores obtenidos.

1.4.6. Incorporación industrial de la herramienta

El sistema propuesto busca integrarse de manera directa al entorno industrial de un convertidor flash, transformando los datos obtenidos por sensores en pronósticos confiables de las variables objetivo del proceso. La incorporación en la industria sigue el siguiente flujo de trabajo. 1.2



Fig. 1.2: Esquema de integración de la herramienta en la industria.

1.5. Alcances y limitaciones

1.5.1. Alcances

Este trabajo aborda el desarrollo de modelos predictivos para el monitoreo no invasivo de la estimación de contenido de cobre y la razón hierro-sílice en la escoria del proceso de conversión flash.

Se incluye el análisis del proceso pirometalúrgico, el preprocesamiento de los datos de proceso, el estudio del estado del arte, técnicas de aprendizaje profundo aplicadas a series de tiempo y la implementación y evaluación comparativa de diferentes arquitecturas.

Los principales resultados esperados son:

- Un conjunto de datos de proceso preprocesados.
- Modelo predictivo clásico de regresión entrenado para estimar las variables objetivo.
- Modelos predictivos basados en técnicas de deep learning entrenados para estimar las variables objetivo.
- Tabla comparativa de desempeño entre los modelos implementados.
- Visualizaciones de resultados y análisis de las predicciones frente a datos reales.

Este trabajo se realiza bajo el supuesto que existe una correlación significativa entre las variables de proceso y las variables químicas de interés.

Esta solución se sitúa en la industria minera, específicamente en plantas de fundición de cobre que operan con convertidores flash. Los modelos predictivos propuestos buscan integrarse como una herramienta de monitoreo como complemento de los sistemas de control, permitiendo un monitoreo continuo, remoto y digital del estado químico de la escoria.

1.5.2. Limitaciones

Este trabajo no aborda la predicción de otros parámetros de operación del proceso como presión o flujo de gases, ni estima el rendimiento global de la conversión metalúrgica. Tampoco se consideran eventos operacionales extremos ni fluctuaciones rápidas del sistema.

Las restricciones del trabajo incluyen la dependencia de datos históricos ya registrados, debido a eso tenemos una cantidad limitada de datos de proceso con los que trabajar ya que estos deben estar alineados con datos fisicoquímicos de nuestras variables objetivos. Asimismo, la capacidad computacional disponible impone límites al tamaño y complejidad de los modelos que pueden ser entrenados.

Los resultados de este estudio son válidos para las condiciones particulares del convertidor y los datos de proceso. La implementación y generalización de los modelos a otros entornos o plantas requiere ajustes adicionales, incluyendo reentrenamiento, recalibración y requerimientos computacionales adicionales.

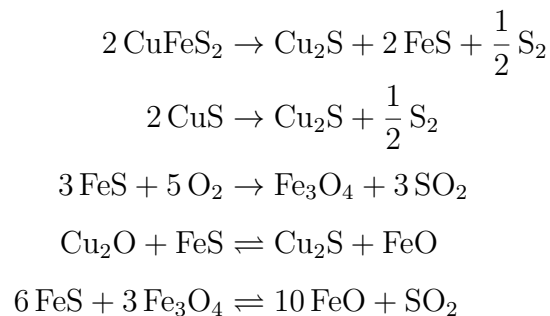
Capítulo 2

Marco Teórico

2.1. Proceso químico

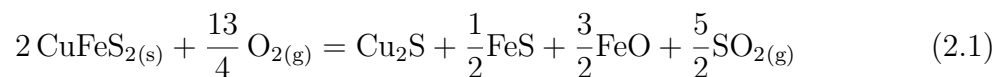
Durante el proceso de fusión de concentrados de cobre, cuyo principal mineral es la calcopirita (CuFeS_2), se alcanza una temperatura cercana a los 1250°C . Bajo estas condiciones, el azufre y el hierro presentes en el mineral se oxidan mediante el suministro de aire enriquecido, generando una fase fundida compuesta principalmente por sulfuros de cobre y hierro (Cu_2S - FeS), denominada *mata*. Esta fase líquida contiene aproximadamente un 75 % de cobre y coexiste con una escoria fundida de naturaleza oxidada, así como con gases ricos en SO_2 .

Las principales reacciones químicas que ocurren durante esta etapa del proceso son las siguientes:



Las reacciones globales que describen la producción de mata y escoria a partir de la

calcopirita se resumen como sigue:



La ecuación (2.1) representa la obtención de la mata rica en cobre, mientras que la ecuación (2.2) refleja la formación de escoria silicatada, producto de la reacción entre el óxido de hierro y la sílice.[5]

2.2. Ley cobre

La *ley de cobre* se refiere al porcentaje de cobre que encierra una determinada muestra. Cuando se habla de una ley del 1 % esto quiere decir que en cada 100 kilogramos de roca mineralizada hay 1 kilogramo de cobre puro.[10]

2.3. Razon hierro silice

La *razón hierro-sílice*, comúnmente expresada como Fe/SiO_2 , corresponde a la relación en peso entre el contenido total de hierro ($\text{Fe}^{2+} + \text{Fe}^{3+}$) y el contenido de dióxido de silicio (SiO_2) presente en una escoria pirometalúrgica. Este parámetro es crucial para determinar la viscosidad y estructura del fundido, ya que influye directamente en la despolimerización del sistema silicato. Un mayor valor de Fe/SiO_2 tiende a reducir la viscosidad de la escoria, facilitando su separación y mejorando la eficiencia del proceso de fundición.[11]

La razón se puede expresar como:

$$\frac{\text{Fe}}{\text{SiO}_2} = \frac{w_{\text{Fe}(\text{total})}}{w_{\text{SiO}_2}}$$

donde $w_{\text{Fe}(\text{total})}$ y w_{SiO_2} son los contenidos en peso de hierro total y sílice, respectivamente.

2.4. Machine learning

El *aprendizaje automático* es un campo de estudio que otorga a las computadoras la capacidad de aprender sin estar explícitamente programadas, tal como lo define Arthur Samuel (1959). A diferencia de la programación tradicional, donde un programador debe proporcionar todas las reglas explícitas, en aprendizaje automático se parte de los datos y los resultados deseados, permitiendo que el algoritmo infiera relaciones entre las entradas y salidas, generando modelos predictivos o de clasificación a partir de ejemplos.[4]

2.5. Deep learning

Es un subcampo del aprendizaje automático que se centra en construir representaciones jerárquicas y distribuidas de los datos. Según Yoshua Bengio (2021), este enfoque explota la estructura desconocida de los datos de entrada para descubrir relaciones útiles en múltiples niveles, donde las características de nivel superior se definen en función de las de nivel inferior. Peter Norvig lo describe como aprendizaje con múltiples niveles de abstracción, y Yann LeCun lo conceptualiza como la construcción de modelos mediante el ensamblaje de módulos parametrizados optimizados con métodos basados en gradiente, destacando la composicionalidad y la capacidad de optimización como elementos clave.[4]

2.6. Interpolación

Es el proceso de estimar valores faltantes dentro de una secuencia de datos usando los valores existentes. En el contexto de series de tiempo, consiste en generar estimaciones para puntos temporales donde no se tiene información, asegurando que el resultado preserve la regularidad temporal de la serie. [4]

- **Interpolación lineal:** consiste en conectar con una línea dos puntos observados y completar los valores faltantes para que se encuentren en dicha línea.

2.7. Time delay Embedding

Incrustación de retardo temporal (*Time Delay Embedding*, TDE) es un método para transformar una serie temporal en una matriz compuesta por segmentos de datos dependientes en el tiempo. Este procedimiento convierte una secuencia larga de observaciones en un conjunto de segmentos más cortos, de manera que el análisis se centre en la **dependencia temporal** de los datos, en lugar de predecir un valor puntual en un instante específico. En otras palabras, cada fila de la matriz generada mediante TDE resume la estructura temporal de un segmento de la serie original [?].

En la figura 2.1 se muestra el esquema general.[12]

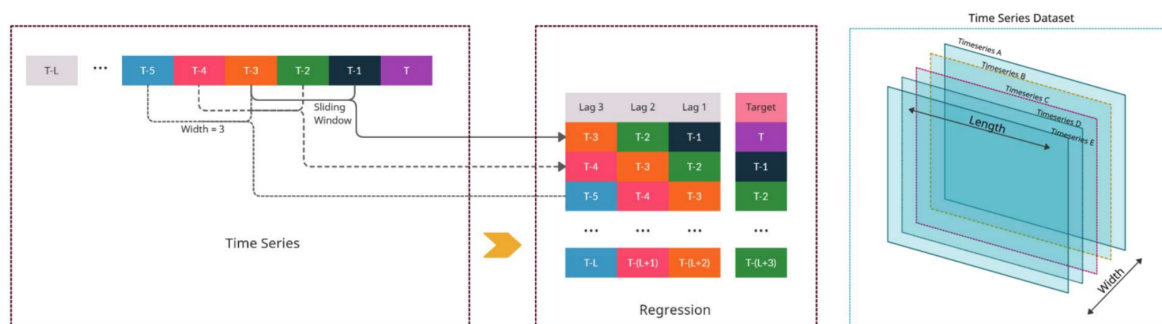


Fig. 2.1: Time delay Embedding

2.8. Validación cruzada

Es una técnica estadística utilizada para evaluar el rendimiento de un modelo predictivo y estimar su capacidad de generalización a datos no vistos. El procedimiento consiste en dividir el conjunto de datos en múltiples particiones denominadas *folds*. En cada iteración, se entrena el modelo con un subconjunto de los datos y se evalúa con el subconjunto restante, repitiendo este proceso hasta que todas las particiones hayan sido utilizadas como conjunto de prueba. Finalmente, se calcula una métrica promedio de desempeño a partir de todas las iteraciones, reduciendo así la varianza asociada a una única partición entrenamiento/prueba y proporcionando una estimación más robusta del rendimiento real del modelo [13].

- **TimeSeriesSplit:** es un método de validación cruzada diseñado específicamente

para datos de series temporales con observaciones ordenadas cronológicamente. A diferencia de métodos tradicionales como K-Fold, *TimeSeriesSplit* garantiza que el conjunto de entrenamiento siempre contiene únicamente datos anteriores al conjunto de prueba, evitando así el entrenamiento con información futura. En cada una de las k particiones, los primeros k bloques se usan como entrenamiento y el siguiente bloque como prueba, de modo que los conjuntos de entrenamiento crecen de manera acumulativa en cada iteración. Además, cada conjunto de prueba cubre un periodo temporal del mismo tamaño, lo que permite una estimación consistente de la capacidad de generalización del modelo en entornos de datos secuenciales. [14]

2.9. Grid Search

Es un procedimiento de búsqueda exhaustiva sobre un conjunto predefinido de valores de hiperparámetros para un estimador. El proceso implica definir una “rejilla” (*grid*) que combina todas las combinaciones posibles de los valores deseados para los hiperparámetros, y evaluar cada combinación mediante validación cruzada para encontrar la configuración óptima [15].

GridSearchCV, implementado en el módulo `sklearn.model_selection`, automatiza este proceso. Se proporciona un estimador (modelo), una rejilla de hiperparámetros (`param_grid`), y una estrategia de validación cruzada (`cv`). Por cada combinación de parámetros, se entrena y evalúa el modelo, y al final se selecciona la configuración que arroja el mejor rendimiento promedio. El estimador resultante puede reenfocarse (`refit`) sobre todo el conjunto de datos con los mejores parámetros encontrados [15].

2.10. Métricas

2.10.1. MAE

El *Mean Absolute Error* (MAE) es una métrica simple que calcula el error absoluto promedio entre las predicciones realizadas por un modelo y los valores observados correspondientes.[4]

$$\text{MAE} = \frac{1}{N \cdot L} \sum_i^N \sum_j^L |f_{i,j} - y_{i,j}| \quad (2.3)$$

donde:

- N es el número de series temporales evaluadas,
- L es la longitud del período de prueba (número de predicciones por serie),
- $f_{i,j}$ es el valor pronosticado en el instante j para la serie i ,
- $y_{i,j}$ es el valor observado en el mismo instante.

2.10.2. MASE

El *Mean Absolute Scaled Error* (MASE) es una métrica ligeramente más complicada que el MSE o MAE, pero ofrece una mejor medida para superar la naturaleza dependiente de la escala de estas métricas. Si se tienen múltiples series temporales con diferentes valores promedio, tanto MAE como MSE tienden a mostrar errores más altos para las series con valores grandes en comparación con las de valores pequeños. MASE resuelve este problema escalando los errores utilizando el MAE dentro de la muestra obtenido con el *método de pronóstico ingenuo*, que consiste en usar el valor anterior como predicción (es decir, y_{j-1} para predecir y_j).[4]

$$\text{MASE} = \frac{\frac{1}{L} \sum_{i=1}^L |f_i - y_i|}{\frac{1}{L-1} \sum_{j=2}^L |y_j - y_{j-1}|} \quad (2.4)$$

donde:

- L es la longitud del período de prueba (número de predicciones),
- f_i es el valor pronosticado en el instante i ,
- y_i es el valor observado en el instante i ,
- el denominador representa el MAE del pronóstico ingenuo dentro de la muestra.

2.10.3. Sesgo de pronóstico (FB)

El *Forecast Bias* (FB) es una métrica que considera aspectos ligeramente diferentes a los de otras métricas vistas previamente. Mientras que otras métricas evalúan la *corrección* del pronóstico, independientemente de la dirección del error, el sesgo de pronóstico nos permite entender el sesgo general en el modelo. Esta métrica nos ayuda a identificar si el modelo pronostica sistemáticamente por encima o por debajo de los valores observados.

Calculamos el sesgo de pronóstico como la diferencia entre la suma de los valores pronosticados y la suma de los valores observados, expresada como porcentaje sobre la suma de todos los valores observados:[4]

$$FB = \frac{\sum_i^N \sum_j^L f_{i,j} - \sum_i^N \sum_j^L y_{i,j}}{\sum_i^N \sum_j^L y_{i,j}} \quad (2.5)$$

donde:

- N es el número de series temporales evaluadas,
- L es la longitud del período de prueba (número de predicciones por serie),
- $f_{i,j}$ es el valor pronosticado en el instante j para la serie i ,
- $y_{i,j}$ es el valor observado en el mismo instante.

2.10.4. Coeficiente de correlación lineal de Pearson

Es un índice estadístico diseñado para variables cuantitativas (con escala mínima de intervalo) que mide el grado de covariación lineal entre dos variables. Evalúa en qué medida los cambios en una variable están asociados linealmente con cambios en la otra. Sus valores oscilan entre -1 y $+1$: un valor de $+1$ indica correlación lineal positiva perfecta, un valor de -1 indica correlación lineal negativa perfecta y un valor de 0 indica ausencia de correlación lineal. Es importante señalar que una correlación alta no implica necesariamente causalidad entre las variables. [16]

El coeficiente de correlación de Pearson se define como:

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}$$

donde $\text{cov}(X, Y)$ representa la covarianza entre las variables X y Y , y $\text{var}(\cdot)$ es la varianza de cada una.

Cuando se trabaja con datos muestrales, se utiliza la siguiente expresión equivalente:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

donde:

- x_i, y_i son los valores observados de cada variable,
- \bar{x}, \bar{y} son sus medias muestrales,
- n es el número de observaciones.

Este estadístico es adimensional y no depende de la escala de medida de las variables, lo que permite comparar correlaciones entre diferentes conjuntos de datos.

2.10.5. Coeficiente de determinación (R^2)

Es un estadístico que mide la proporción de la variabilidad total de la variable dependiente que es explicada por el modelo de regresión. Sus valores se encuentran en el rango $0 \leq R^2 \leq 1$, donde $R^2 = 1$ indica que el modelo explica el 100 % de la variabilidad observada, mientras que $R^2 = 0$ significa que no explica ninguna variabilidad. Un valor intermedio refleja un grado parcial de explicación, y $1 - R^2$ corresponde a la proporción de variabilidad no explicada.

Se define como:

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

donde:

- $\sigma_r^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$ es la **varianza residual** de la predicción,

- $\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$ es la **varianza total** de la variable dependiente,
- y_i es el valor observado,
- \hat{y}_i es el valor estimado por el modelo,
- \bar{y} es la media muestral de la variable dependiente,
- N es el número de observaciones.

Este estadístico es independiente de la escala de medida de las variables y se utiliza ampliamente para evaluar la calidad de ajuste de un modelo predictivo.

2.10.6. Mutual Information (MI)

La **Información Mutua** (*Mutual Information*, MI) es una medida de dependencia entre dos variables aleatorias que proviene de la teoría de la información. A diferencia del coeficiente de correlación de Pearson, que solo captura relaciones lineales, la MI es capaz de detectar tanto dependencias lineales como no lineales entre las variables [17].

Formalmente, la MI se define como la diferencia entre las entropías marginales y la entropía conjunta, o equivalentemente:

$$I(X; Y) = \int_{\mathbb{R}^2} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right) d(x, y) \quad (2.6)$$

donde $p(x, y)$ es la densidad conjunta de X y Y , y $p(x)$, $p(y)$ son las densidades marginales correspondientes. En el caso discreto, la integral se reemplaza por una suma.

Una propiedad fundamental es:

$$I(X; Y) \geq 0 \quad (2.7)$$

y $I(X; Y) = 0$ si y solo si X y Y son independientes.

La MI tiene un rango en $[0, \infty)$, pero puede normalizarse para obtener un coeficiente de correlación en el intervalo $[0, 1]$ mediante la expresión:

$$\rho_{I(X; Y)} = \sqrt{1 - \exp(-2 I(X; Y))} \quad (2.8)$$

Si (X, Y) sigue una distribución normal bivariada, entonces $\rho_{I(X;Y)}$ es equivalente al coeficiente de correlación de Pearson.

Ventajas:

- Detecta relaciones no lineales y no paramétricas.
- Es invariante bajo transformaciones monótonas de las variables.
- Admite versiones condicionales $I(X; Y|Z)$, análogas a la correlación parcial.

En análisis exploratorio de datos, la MI normalizada (NMI) se interpreta de forma similar a un coeficiente de correlación tradicional, pero con mayor sensibilidad frente a patrones complejos en los datos.[17]

2.11. Correlación cruzada

Es una técnica estadística que cuantifica la similitud entre dos series temporales en función de un desplazamiento (*lag*) aplicado a una de ellas. A través de este análisis, es posible detectar patrones comunes, determinar relaciones temporales y evaluar si una serie antecede o sigue a la otra en el tiempo. Su valor varía entre -1 y $+1$, donde $+1$ indica una correlación perfecta en la misma dirección, 0 ausencia de correlación y -1 una correlación perfecta en dirección opuesta.

En el análisis de series temporales, esta herramienta permite identificar el desfase óptimo en el que ambas secuencias se asemejan más, lo que es de utilidad en ámbitos como el procesamiento de señales, la ingeniería, las finanzas y el modelado predictivo. Además, puede representarse gráficamente mediante un *correlograma*, que facilita la visualización de la correlación en función de los distintos *lags* [18].

- **Máxima correlación:** corresponde al valor más alto del coeficiente de correlación cruzada obtenido al desplazar una serie temporal respecto de otra a lo largo de distintos desfases (*lags*). Este valor refleja el grado máximo de similitud alcanzable entre ambas señales y permite determinar el desfase óptimo que genera la alineación temporal más precisa. Un valor cercano a $+1$ indica una fuerte relación directa, mientras que uno cercano a -1 señala una fuerte relación inversa en el desfase correspondiente [18].

2.12. Regresión lineal múltiple

La *regresión lineal múltiple* es un método estadístico utilizado para modelar la relación entre una variable dependiente y un conjunto de variables independientes. A diferencia de la regresión lineal simple, que considera un solo predictor, esta técnica permite analizar de manera simultánea el efecto de múltiples factores sobre la variable objetivo.

El modelo se expresa como una combinación lineal de los predictores:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (2.9)$$

donde:

- Y es la variable dependiente (objetivo del modelo),
- X_1, X_2, \dots, X_p son las variables independientes (predictores),
- β_0 es el intercepto,
- $\beta_1, \beta_2, \dots, \beta_p$ son los coeficientes de regresión que cuantifican el impacto de cada predictor sobre Y ,
- ε es el término de error aleatorio.

Los parámetros β se estiman mediante el método de mínimos cuadrados, que minimiza la suma de los errores al cuadrado entre los valores observados y los predichos.[19]

2.13. Long Short-Term Memory (LSTM)

La *Long Short-Term Memory* (LSTM) es una arquitectura avanzada de redes neuronales recurrentes (RNN), diseñada para abordar las limitaciones de las RNN tradicionales, especialmente el problema del desvanecimiento del gradiente. Introducida por Hochreiter y Schmidhuber, las LSTM son capaces de capturar dependencias a largo plazo en secuencias de datos, lo que las hace particularmente efectivas para tareas como traducción automática, reconocimiento del habla y pronóstico de series temporales [20].

Cada unidad LSTM incluye una célula de memoria y tres puertas (gates) que regulan el flujo de información:

- **Puerta de entrada (input gate)**, que controla cuánto de la nueva información se almacena en la célula.
- **Puerta de olvido (forget gate)**, que decide qué información previa debe descartarse.
- **Puerta de salida (output gate)**, que regula qué información de la célula se envía como salida en el paso actual.

Esta estructura de gates permite a las LSTM retener y descartar información de manera selectiva, superando así el problema del gradiente que se desvanece en secuencias largas [20].

La inferencia dentro de una célula LSTM puede formalizarse mediante el siguiente conjunto de ecuaciones, donde \odot representa el producto elemento a elemento:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \quad (2.10)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \quad (2.11)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \quad (2.12)$$

$$\tilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (2.13)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (2.14)$$

$$h_t = o_t \odot \tanh(c_t). \quad (2.15)$$

Estas ecuaciones reflejan cómo las puertas deciden qué información guardar, eliminar o transmitir, permitiendo que el modelo mantenga dependencias temporales a lo largo de muchas etapas de tiempo .

2.14. Temporal Fusion Transformer (TFT)

El *Temporal Fusion Transformer* (TFT) es un modelo de aprendizaje profundo diseñado para el pronóstico de series temporales multivariantes y multihorizonte, que combina

redes recurrentes y mecanismos de atención para capturar dependencias locales y de largo alcance de manera interpretable. Su arquitectura integra distintos tipos de variables: covariables estáticas, entradas observadas en el pasado y entradas conocidas en el futuro, lo que le permite adaptarse a escenarios complejos de predicción [21].

El objetivo del modelo es estimar pronósticos probabilísticos mediante cuantiles:

$$\hat{y}_i(q, t, \tau) = f_q(\tau, y_{i, t-k:t}, \mathbf{z}_{i, t-k:t}, \mathbf{x}_{i, t-k:t+\tau}, \mathbf{s}_i), \quad (2.16)$$

donde:

- $\hat{y}_i(q, t, \tau)$ es el pronóstico del cuantil q en el horizonte τ para la serie i .
- $y_{i, t-k:t}$ son los valores históricos de la variable objetivo.
- $\mathbf{z}_{i, t-k:t}$ representan las entradas exógenas desconocidas (futuros inciertos).
- $\mathbf{x}_{i, t-k:t+\tau}$ corresponden a las entradas conocidas.
- \mathbf{s}_i son las covariables estáticas asociadas a la serie i .
- f_q es la función aprendida por la red neuronal para el cuantil q .

La arquitectura de TFT se compone de los siguientes módulos principales:

1. **Selección de variables:** redes neuronales especializadas ponderan la relevancia de cada variable en cada instante, adaptando dinámicamente el conjunto de entradas.
2. **Codificación recurrente:** una LSTM procesa las entradas pasadas para capturar dependencias temporales locales.
3. **Atención interpretable:** un mecanismo de auto-atención enmascarada permite identificar qué instantes históricos son más relevantes para los pronósticos futuros.
4. **Gated Residual Networks (GRN):** mecanismos de compuertas que permiten omitir información redundante y mejorar la generalización.
5. **Pronóstico por cuantiles:** las salidas del decodificador se proyectan sobre diferentes cuantiles ($q = 0,1, 0,5, 0,9, \dots$), proporcionando estimaciones con intervalos de confianza.

Gracias a la combinación de LSTM, auto-atención y mecanismos de selección, el TFT logra un balance entre precisión de pronóstico e interpretabilidad, permitiendo comprender qué variables y momentos en el tiempo influyen más en las predicciones.

Capítulo 3

Desarrollo

3.1. Introducción

Este capítulo detalla el desarrollo llevado a cabo para la estimación del contenido de cobre y la razón hierro/sílice en la escoria de un Convertidor Flash, utilizando datos de proceso y técnicas de *Deep Learning*.

El trabajo se realizó sobre un conjunto de datos de proceso proporcionado por terceros. El desarrollo comenzó con la etapa de selección y preprocesamiento de estos datos, en la que se aplicaron técnicas de limpieza y normalización de las series temporales. Posteriormente, se entrenaron y evaluaron diversos modelos de regresión y de aprendizaje profundo.

Una parte fundamental del desarrollo fue la selección de métricas cuantitativas para evaluar el desempeño predictivo de cada modelo, considerando tanto la precisión como la capacidad de generalización. A partir de los resultados obtenidos, se elaboró una tabla comparativa con el objetivo de contrastar las arquitecturas utilizadas.

Resulta relevante destacar la importancia de los enfoques endógeno y exógeno en la construcción de los modelos. Mientras que los conjuntos endógenos permiten analizar la capacidad predictiva de los modelos a partir de la autocorrelación de la propia serie de la variable objetivo, los conjuntos exógenos ofrecen un escenario más cercano a la operación real de la planta, al prescindir del uso de valores pasados de la ley de cobre y de la razón hierro-sílice e incorporar únicamente variables de proceso medidas en

línea. Este enfoque no solo responde a la limitación práctica derivada del retraso en los análisis de laboratorio, sino que también posibilita evaluar con mayor realismo la capacidad de los modelos de *Deep Learning* para aprender relaciones causales entre las condiciones operacionales y las variables fisicoquímicas objetivo.

3.2. Definición del conjunto de datos

Se utilizó una base de datos histórica del proceso pirometalúrgico del Convertidor Flash, correspondiente a los años 2024 y 2025. A partir de esta fuente, se seleccionó un subconjunto continuo de registros comprendido entre el **3 de agosto de 2024 a las 20:00 horas** y el **17 de octubre de 2024 a las 14:00 horas**, un total de 75 días, garantizando una consistencia temporal para el análisis de series de tiempo.

El conjunto de datos incluyó las variables objetivo del modelo: *ley de cobre* y *razón hierro/sílice*. Además, se consideraron seis variables de entrada asociadas al comportamiento del proceso:

- **Temperatura de la escoria** (*Slag Temperature*): corresponde a la temperatura del material fundido presente en el horno flash. Es un indicador clave del balance térmico del proceso, ya que influye directamente en el comportamiento de la escoria dentro del horno. Valores demasiado bajos generan acreciones en las paredes refractarias y dificultan la separación entre las fases escoria y cobre blister; en cambio, temperaturas muy altas dañan los refractarios por su agresividad, diluyendo la magnetita e infiltrándose en los ladrillos.
- **Flujo de alimentación de concentrado** (*Cons Feed Rate*): representa la tasa de alimentación de concentrado hacia el horno flash. Su impacto radica en la generación de calor por las reacciones exotérmicas asociadas a los sulfuros presentes y en la relación Cu:S, que determina el carácter fuertemente exotérmico de la fusión. La capacidad de procesamiento no depende directamente de este flujo, sino del dimensionamiento del reactor.
- **Alimentación de polvo** (*Dust Feed Rate*): corresponde a la reincorporación de material particulado recirculado proveniente del sistema de limpieza de gases o tren de gases. Este material posee un alto contenido de cobre, junto con hierro,

sílice e impurezas del concentrado, por lo que su reintegración afecta el balance de masa y la composición química de la carga.

- **Alimentación de fundente** (*Flux Feed Rate*): consiste en la adición de fundente (principalmente sílice), cuya función es facilitar el manejo de la fase escoria al disminuir su punto de fusión y mantenerla líquida a temperaturas de operación. A través del control químico del fundente se favorece la separación entre la escoria y el cobre blister en el reposador.
- **Coeficiente de oxígeno** (*Oxygen Coefficient*): expresa la relación entre la cantidad de oxígeno inyectado y el oxígeno teórico requerido para la oxidación completa. En la práctica industrial, se refiere a la cantidad de oxígeno utilizado por tonelada de concentrado procesado.
- **Enriquecimiento de oxígeno** (*Oxygen Enrichment*): corresponde al porcentaje de oxígeno presente en el flujo de aire soplado al horno flash. Un mayor enriquecimiento aumenta la velocidad de reacción y la temperatura alcanzada, modificando el comportamiento térmico y químico de la operación.

Estas variables fueron seleccionadas por su relevancia en el proceso metalúrgico y su impacto en la composición química de la escoria.

Previo a su uso en los modelos de aprendizaje profundo, los datos fueron sometidos a una fase de preprocesamiento. Se eliminaron datos atípicos y se aplicó normalización para cada variable, asegurando una escala uniforme durante el entrenamiento de los modelos.

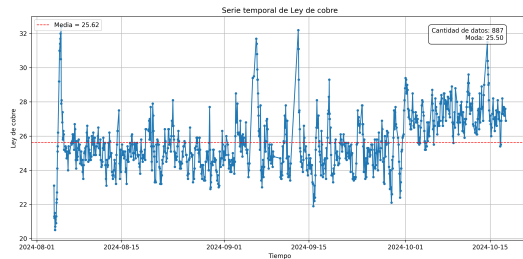
3.2.1. Análisis estadístico del conjunto de datos

Se realiza un análisis estadístico inicial con el objetivo de identificar patrones de periodicidad, variabilidad y distribución en los datos. Para ello, se emplearon diferentes visualizaciones y métricas que permiten identificar el comportamiento de las variables:

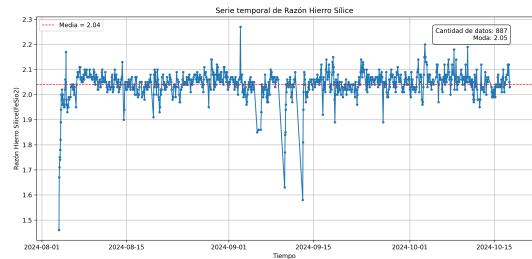
- **Serie temporal:** permite observar el comportamiento de las variables a lo largo del tiempo, identificando tendencias, estacionalidades y datos atípicos.
- **Histograma:** muestra la distribución de frecuencia de los valores, facilitando la detección de sesgos y rangos predominantes.

- **Diagrama de caja:** proporciona información sobre la dispersión, mediana y valores atípicos.

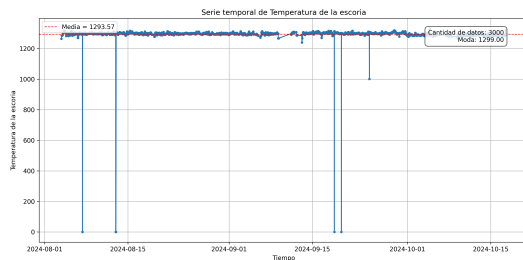
A continuación podemos ver las distintas series temporales de cada uno de nuestros datos de entrada antes de la realización de un preprocesamiento 3.1.



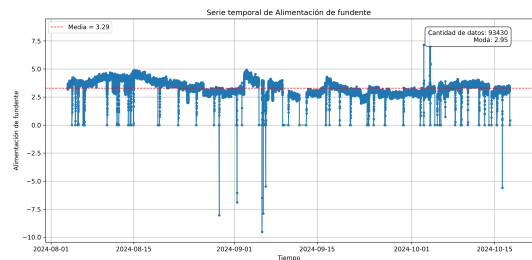
(a) Serie temporal de Ley de cobre



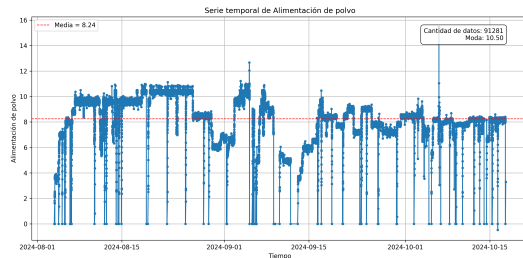
(b) Serie temporal de Razón Hierro Sílice



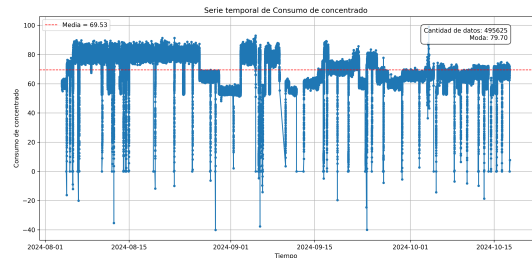
(c) Serie temporal de Temperatura de la escoria



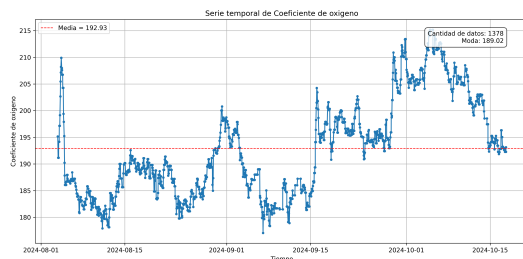
(d) Serie temporal de Alimentación de fundente



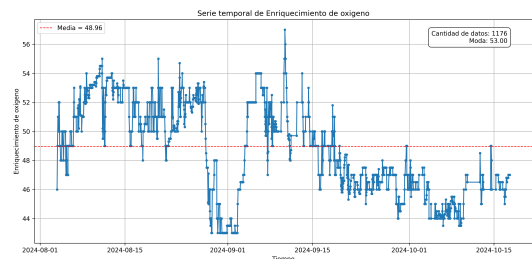
(e) Serie temporal de Alimentación de polvo



(f) Serie temporal de Consumo de concentrado



(g) Serie temporal de Coeficiente de oxígeno



(h) Serie temporal de Enriquecimiento de oxígeno

Fig. 3.1: Series temporales de datos crudos del proceso.

Podemos notar una alta variabilidad en las series (picos y caídas abruptas), además de la presencia de ruido y valores atípicos. Por ejemplo, analizando más en detalle los datos obtenidos de *Consumo de concentrado* 3.2 podemos encontrar valores negativos los cuales en términos físicos no tiene una lógica.

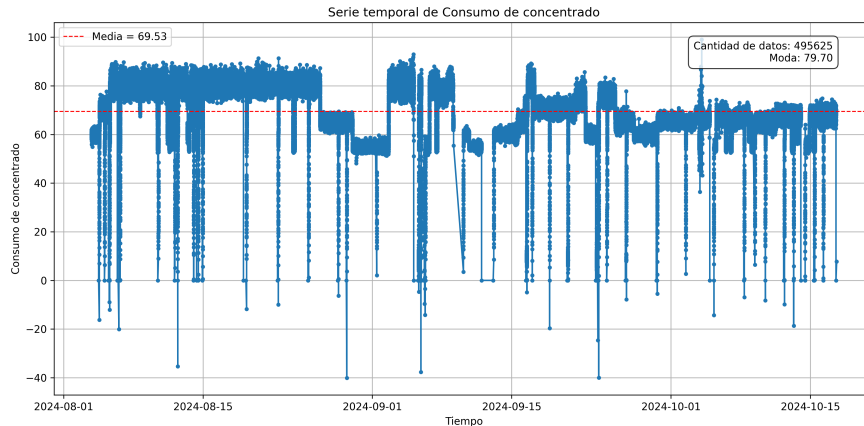


Fig. 3.2: Serie temporal consumo de concentrado

Al analizar su diagrama de caja 3.3 se confirma la presencia de valores atípicos los cuales deben ser normalizados primeramente para el correcto entrenamiento de los modelos de regresión.

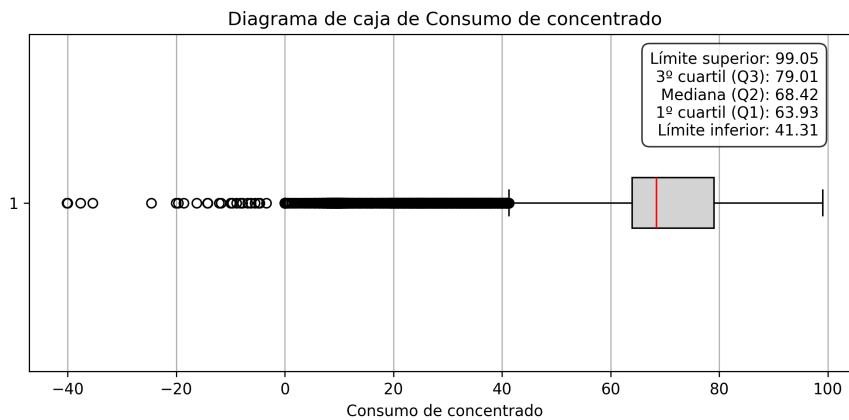


Fig. 3.3: Diagrama de caja consumo de concentrado

Además de la alta variabilidad tenemos el problema de la dimensionalidad de los datos, debido a que por ejemplo en el *Consumo de concentrado* tenemos 495.625 datos los cuales superan por mucho a los 887 datos de *Ley de cobre y Razón Hierro/Sílice*

obtenidos del proceso. Esta diferencia en cantidad de datos la podemos apreciar en la siguiente tabla comparativa 3.1.

Variable	Cantidad de datos
Ley de cobre	887
Razón hierro/sílice	887
Temperatura de la escoria	3.000
Consumo de concentrado	495.625
Alimentación de polvo	91.281
Alimentación de fundente	93.430
Coefficiente de oxígeno	1.378
Enriquecimiento de oxígeno	1.176

Tabla 3.1: Variables y cantidad de datos

Al analizar detalladamente la serie temporal de la figura 3.4 y el histograma 3.5 correspondiente a la *Temperatura de la escoria* observamos que presenta caídas abruptas hasta valores cercanos a cero, lo cual físicamente no es correcto debido a la naturaleza continua del proceso, esto nos indica una posible falla en la obtención de los datos. Debido a estos valores atípicos concluimos que el valor de varianza y distribución de los datos que se observa en el histograma 3.5 esta sesgado.

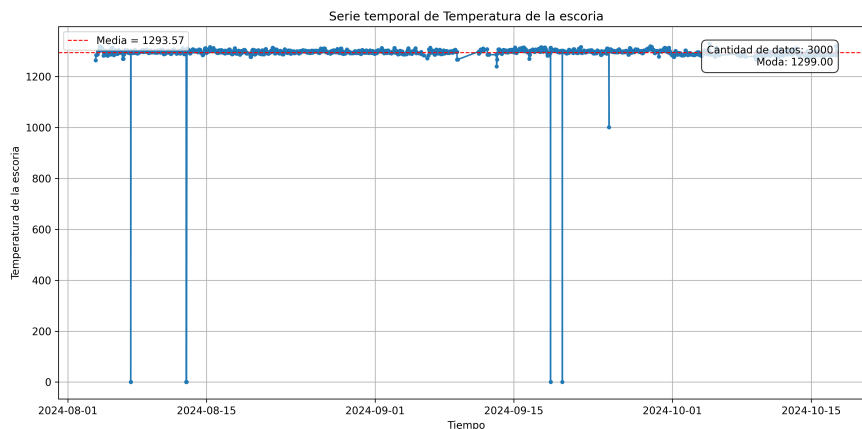


Fig. 3.4: Serie temporal de la Temperatura de la escoria.

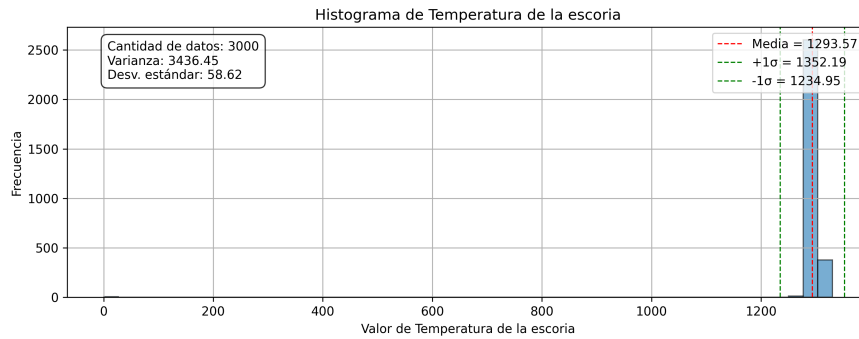


Fig. 3.5: Histograma de la Temperatura de la escoria.

Debido a las diferencias entre las cantidades de datos, la presencia de alta variabilidad y valores atípicos se emplea un preprocesamiento de los datos.

3.2.2. Preprocesamiento

El preprocesamiento tuvo como propósito unificar la periodicidad y frecuencia de los datos, así como eliminar valores atípicos y normalizar las variables, con el fin de garantizar la consistencia entre los distintos conjuntos y facilitar el entrenamiento de los modelos.

En el caso de datos temporales provenientes de diferentes fuentes, como registros de proceso y datos fisicoquímicos, encontramos variaciones en los intervalos de muestreo y en la alineación temporal. Si estas discrepancias no se corrigen, pueden introducir sesgos que afecten el rendimiento del modelo y reduzcan su capacidad predictiva.

Las principales transformaciones aplicadas fueron:

- **Definición de periodo y frecuencia:** se estableció un intervalo de muestreo uniforme para todos los conjuntos de datos, abarcando desde el 3 de agosto hasta el 17 de octubre. Se generaron dos bases de datos: una con frecuencia de 1 hora y otra con frecuencia de 30 minutos. Esta regularidad es esencial para que los modelos de series de tiempo puedan detectar patrones temporales de forma consistente.
- **Resampling:** se reagruparon los datos mediante el cálculo de promedios en los intervalos definidos (1 hora y 30 minutos). Esta técnica ayuda a reducir el ruido de alta frecuencia y a resaltar tendencias generales sin perder información

significativa.

- **Reindexado:** se aplicó un índice temporal uniforme para alinear correctamente los registros entre las distintas variables, evitando desfases y asegurando que cada observación temporal sea coherente con el resto del conjunto.
- **Interpolación lineal:** se completaron los datos faltantes mediante interpolación lineal, cuando fue necesario. Este método fue elegido por ser adecuado en variables continuas, ya que mantiene la coherencia en la evolución temporal. En este trabajo, se asume que las variables tienen un comportamiento continuo, debido a su naturaleza física y química.

En la figura 3.6 se muestra el diagrama de flujo del preprocesamiento de los datos.

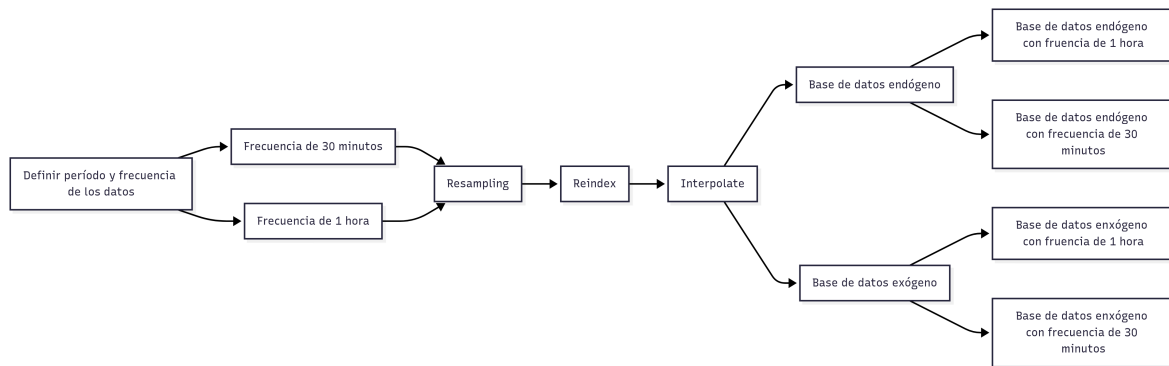
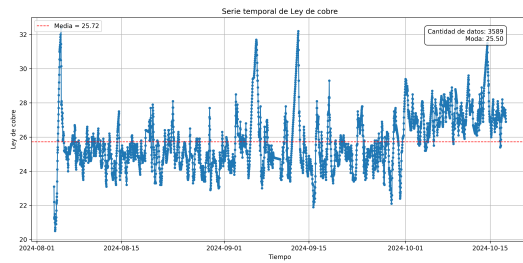
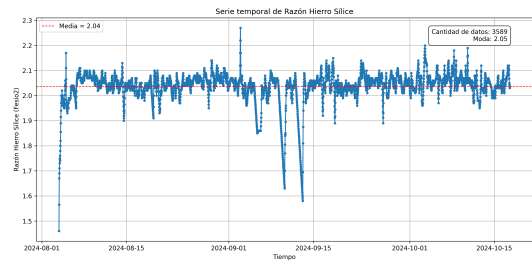


Fig. 3.6: Diagrama de flujo del preprocesamiento

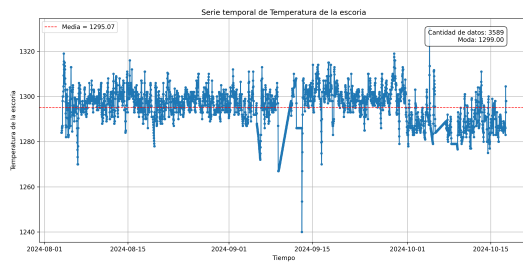
A continuación, se presentan las nuevas series temporales con una frecuencia de muestreo de 30 minutos. Esta frecuencia se adopta como base para el desarrollo de la memoria, ya que permite disponer de una mayor cantidad de datos en comparación con la frecuencia de 1 hora, lo que resulta de interés para el análisis. No obstante, los gráficos correspondientes a la frecuencia de 1 hora se incluyen en el anexo A.1 de la memoria de título para su eventual revisión y análisis complementario.



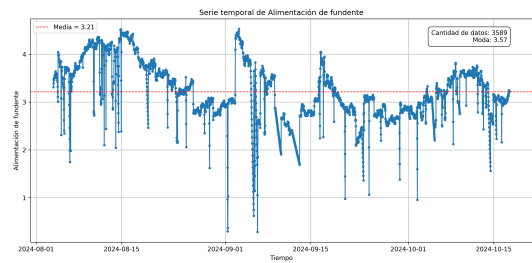
(a) Serie temporal de Ley de cobre con frecuencia de 30 minutos



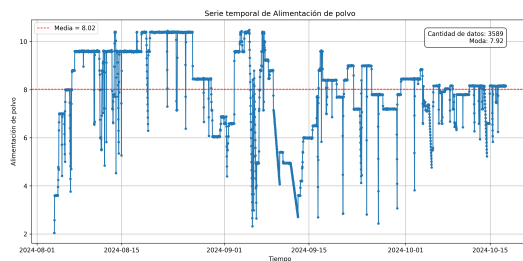
(b) Serie temporal de Razón Hierro Sílice con frecuencia de 30 minutos



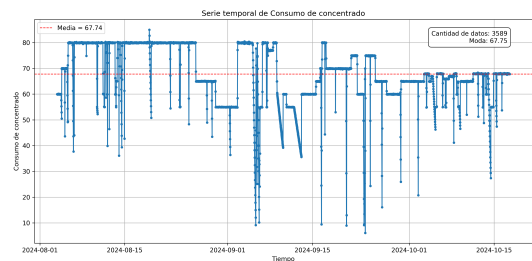
(c) Serie temporal de Temperatura de la escoria con frecuencia de 30 minutos



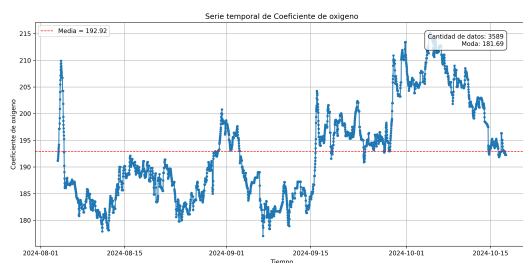
(d) Serie temporal de Alimentación de fundente con frecuencia de 30 minutos



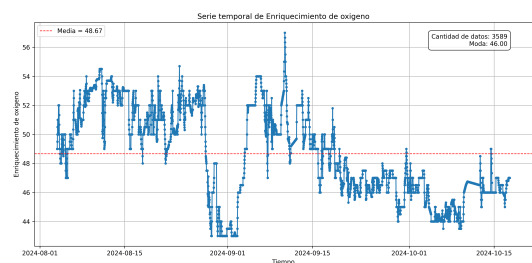
(e) Serie temporal de Alimentación de polvo con frecuencia de 30 minutos



(f) Serie temporal de Consumo de concentrado con frecuencia de 30 minutos



(g) Serie temporal de Coeficiente de oxígeno con frecuencia de 30 minutos



(h) Serie temporal de Enriquecimiento de oxígeno con frecuencia de 30 minutos

Fig. 3.7: Series temporales de datos del proceso con frecuencia de 30 minutos.

Una vez preprocesados los datos podemos analizar que en la serie temporal de consumo de concentrado 3.8 se ha atenuado el ruido y la presencia de valores atípicos, además de la eliminación de valores nulos y negativos.

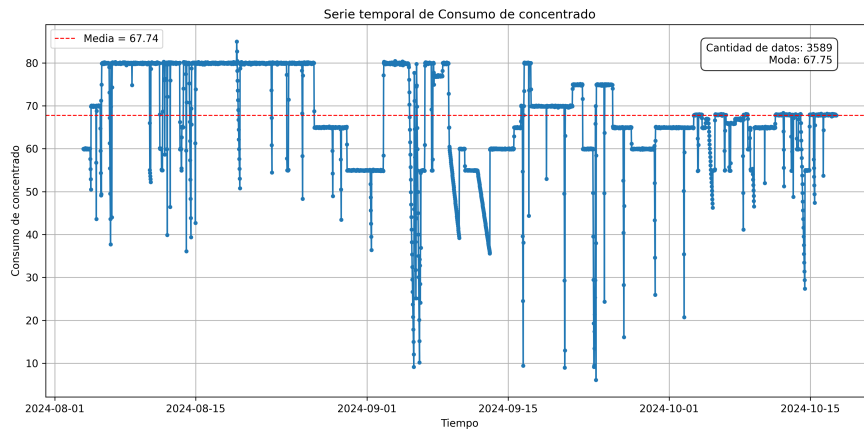


Fig. 3.8: Serie temporal de consumo de concentrado preprocesado.

Los datos han sido normalizados en sus dimensiones generando un nuevo conjunto de datos con la misma cantidad entre ellos, lo cual queda resumido en la siguiente tabla 3.2.

Variable	Cantidad de datos
Ley de cobre	3.589
Razón hierro/sílice	3.589
Temperatura de la escoria	3.589
Consumo de concentrado	3.589
Alimentación de polvo	3.589
Alimentación de fundente	3.589
Coefficiente de oxígeno	3.589
Enriquecimiento de oxígeno	3.589

Tabla 3.2: Variables y cantidad de datos

Además al analizar la serie temporal correspondiente a la *Temperatura de la escoria* 3.9 observamos una eliminación de los valores atípicos cercanos a cero, mayor información y una distribución en un rango más amplio en el histograma 3.10.

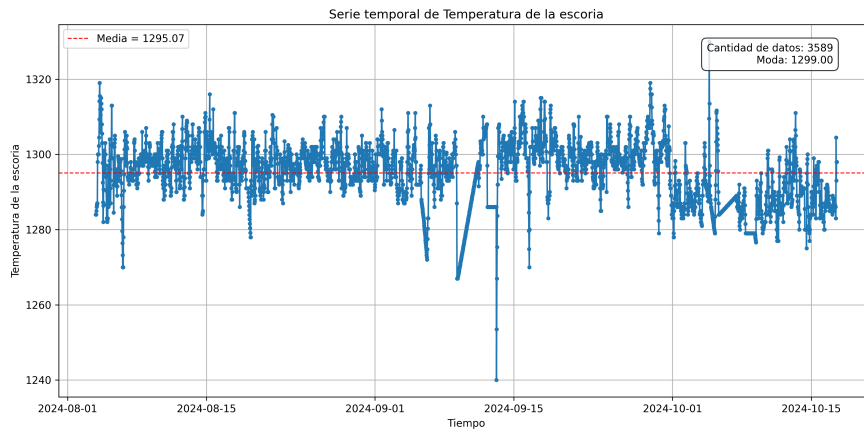


Fig. 3.9: Serie temporal de la Temperatura de la escoria.

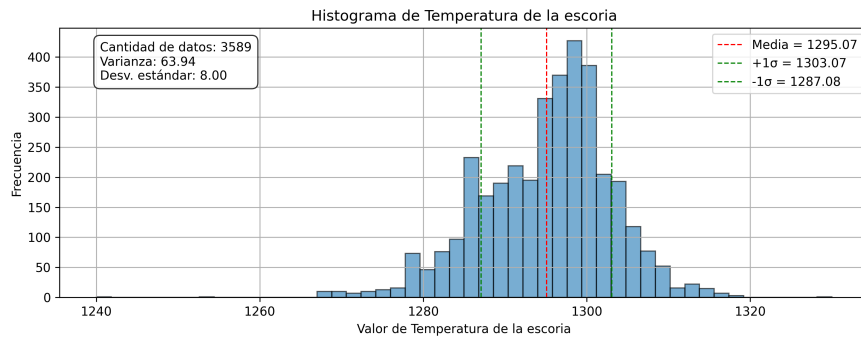


Fig. 3.10: Histograma de la Temperatura de la escoria.

3.2.3. Creación del conjunto de datos endógenos

Para el desarrollo del conjunto de datos endógenos se utilizaron exclusivamente las variables objetivo del proceso: **ley de cobre** y **razón hierro/sílice (Fe/SiO_2)**. A cada una se le aplicó por separado la técnica de *incrustación de retardo temporal* (*Time Delay Embedding*) con el objetivo de transformar la serie temporal original en un conjunto de datos autoregresivo.

Este procedimiento consiste en generar vectores de entrada a partir del historial de la variable, mediante una ventana deslizante que considera un número fijo de observaciones pasadas. En este caso, se empleó una ventana de **cinco instancias previas** al instante actual.

El resultado es una matriz donde cada fila representa un momento del tiempo, las

columnas T-5 a T-1 corresponden a los valores históricos, y la columna T contiene el valor actual, que se utiliza como salida objetivo durante el entrenamiento del modelo 3.3.. La estructura se resume en la Tabla 3.4.

Time	T-5	T-4	T-3	T-2	T-1	T
2024-08-03 20:00:00	23.1	21.3	21.2	21.5	20.5	20.7
2024-08-03 21:00:00	21.3	21.2	21.5	20.5	20.7	21.1
2024-08-03 22:00:00	21.2	21.5	20.5	20.7	21.1	21.3
2024-08-03 23:00:00	21.5	20.5	20.7	21.1	21.3	20.9
2024-08-04 00:00:00	20.5	20.7	21.1	21.3	20.9	21.3
2024-08-04 01:00:00	20.7	21.1	21.3	20.9	21.3	22.1
2024-08-04 02:00:00	21.1	21.3	20.9	21.3	22.1	22.3
2024-08-04 03:00:00	21.3	20.9	21.3	22.1	22.3	22.7
2024-08-04 04:00:00	20.9	21.3	22.1	22.3	22.7	24.1
2024-08-04 05:00:00	21.3	22.1	22.3	22.7	24.1	25.1

Tabla 3.3: Conjunto de datos Endógeno.

Tabla 3.4: Descripción de columnas en conjunto de datos endógeno.

Columna	Descripción
Time	Marca temporal correspondiente al valor objetivo a predecir.
T-5 a T-1	Valores históricos de la variable objetivo en los cinco pasos anteriores.
T	Valor actual de la variable objetivo.

Este enfoque permitió generar, para cada variable objetivo, un conjunto de datos independiente que fue empleado en el entrenamiento de modelos clásicos de regresión y *Deep Learning*.

3.2.4. Creación del conjunto de datos exógenos

A diferencia del enfoque endógeno, donde se plantea la hipótesis de que las variables objetivo presentan dependencia temporal con sus propios valores pasados, en el caso del conjunto de datos exógenos se considera que las variables objetivo (**ley de cobre y razón hierro/sílice**) pueden estar influenciadas por otras variables de proceso medidas en el convertidor flash.

Para construir este conjunto, se seleccionaron variables de proceso relevantes y se alinearon temporalmente con las variables objetivo. Esto asegura que, para cada instante de tiempo, todas las mediciones correspondan al mismo momento del proceso, evitando desfases que puedan introducir errores en el modelado.

En particular, se consideraron las siguientes variables exógenas:

- **ConsFeed**: Consumo de concentrado (t/h).
- **DustFeed**: Alimentación de polvo (t/h).
- **FluxFeed**: Alimentación de fundente (t/h).
- **OxygenCoeff**: Coeficiente de oxígeno suministrado.
- **OxygenEnrich**: Porcentaje de enriquecimiento de oxígeno en el aire de soplado.
- **SlagTemperature**: Temperatura de la escoria (°C).

Cada registro del conjunto exógeno incluye, además de estas variables, el valor de la variable objetivo correspondiente y su marca temporal asociada. La estructura general se presenta en la Tabla 3.5.

Tabla 3.5: Descripción de columnas en conjunto de datos exógenos.

Columna	Descripción
Time	Marca temporal correspondiente a las mediciones de proceso y al valor objetivo.
ConsFeed	Tasa de alimentación de concentrado en toneladas por hora.
DustFeed	Tasa de alimentación de polvo en toneladas por hora.
FluxFeed	Tasa de alimentación de fundente en toneladas por hora.
OxygenCoeff	Coficiente que representa la cantidad de oxígeno suministrada al convertidor.
OxygenEnrich	Porcentaje de enriquecimiento de oxígeno en el aire de soplado.
SlagTemperature	Temperatura de la escoria en grados Celsius.
Target	Variable objetivo (<i>Ley de cobre o Razón Fe/SiO₂</i>).

Este conjunto de datos permite evaluar la hipótesis de que las condiciones de operación del convertidor influyen de manera significativa en las variables objetivo.

3.3. Desarrollo de modelos

La metodología seguida consistió en entrenar y evaluar los modelos tanto con datos endógenos como exógenos, además de contrastar su desempeño bajo distintas métricas. De esta manera, se buscó identificar un modelo con una estrategia de entrenamiento que favorezca la generalización de las predicciones para ambas variables objetivo.

Cabe destacar que, al entrenar los modelos con el conjunto de datos exógeno, no se utilizó el historial de las variables objetivo para realizar las predicciones. Este aspecto resulta de interés a nivel industrial, ya que la posibilidad de obtener estimaciones sin recurrir al pasado de las variables objetivo permite reducir o incluso sustituir las tomas de muestras fisicoquímicas de la escoria del convertidor flash. Esto se traduce en beneficios en términos financieros, en la mejora de la calidad del proceso y en la disminución del uso de técnicas invasivas.

Dentro de las metodologías adoptadas para el entrenamiento de cada uno de los modelos se tomaron ciertas generalizaciones que se explican a continuación:

- Ingeniería de variables: Este proceso consistió en la generación y transformación de características a partir de las variables de entrada, con el objetivo de enriquecer la representación de la información y mejorar la capacidad predictiva de los modelos. En particular, se incorporaron retardos (*lags*) de las variables exógenas para capturar dependencias temporales, así como medias móviles (*rolling windows*) que permiten reflejar tendencias y suavizar fluctuaciones locales. Esta metodología asegura que el modelo trabaje con un conjunto de variables más representativo, reduciendo el riesgo de sobreajuste.
- Para la búsqueda de hiperparámetros se utilizó *GridSearchCV* en conjunto con *TimeSeriesSplit*, técnica que respeta el orden temporal en la validación cruzada al entrenar únicamente con datos pasados y evaluar con datos futuros. Esto asegura una estimación más realista del desempeño de los modelos en contexto de series temporales.
- Se utilizó como métrica principal el *Mean Absolute Scaled Error* (MASE), ya que permite comparar modelos de forma consistente al ser independiente de la escala de los datos y ofrece una interpretación clara frente a un modelo de referencia ingenuo.
- Una vez seleccionados los hiperparámetros, los modelos se entrenaron utilizando los conjuntos de entrenamiento y validación, y posteriormente se evaluaron en el conjunto de prueba.

3.4. Regresión lineal

Se empleó un modelo de *regresión lineal múltiple* con el fin de capturar las relaciones entre las variables exógenas del proceso y la ley de cobre (*Cu*). El modelo se implementó mediante un *pipeline* compuesto de la siguiente manera:

1. **Ingeniería de variables:** se generaron variables rezagadas (*lags*) y medias móviles (*rolling means*) de las variables exógenas. En particular, se consideraron rezagos de $[1, 2, 3, 6]$ pasos y ventanas móviles de $[3, 6, 12]$ para las predicciones

de ley de cobre y rezagos de [1, 2, 3, 6, 12, 24, 48, 72, 96, 120, 144, 168] pasos y ventanas móviles de [3, 6, 12, 24, 48] para las predicciones de razón Hierro-Sílice . Con ello, se busca capturar efectos retardados y promedios de corto plazo en el comportamiento de las entradas.

2. **Transformación polinomial (*PolynomialFeatures*)**: a partir de las variables resultantes de la ingeniería de variables, se generaron combinaciones polinomiales de grado 1 y grado 2, lo que permite modelar relaciones no estrictamente lineales en las variables originales. A pesar de estas expansiones, el modelo sigue siendo lineal en sus predicciones.
3. **Selección de variables (*SelectKBest*)**: se aplicó un criterio basado en información mutua para seleccionar las variables más relevantes, reduciendo dimensionalidad y evitando sobreajuste.
4. **Escalado robusto (*RobustScaler*)**: tanto las variables de entrada como la variable objetivo se escalaron empleando percentiles (10-90%), con el fin de mitigar la influencia de valores atípicos.[22]
5. **Regresión lineal regularizada (*Ridge*)**: se ajustó un modelo de regresión lineal con regularización, lo que controla la magnitud de los coeficientes y mejora la estabilidad numérica. Este modelo incluye un intercepto para absorber sesgos globales.

La validación se realizó mediante un esquema de validación cruzada (*cross-validation*) temporal y un *gap* fijo, empleando como métrica principal MASE (*Mean Absolute Scaled Error*).

3.4.1. Curvas de pérdida.

En la Figura 3.11 se muestran las curvas de pérdida (*loss curves*) obtenidas durante la validación cruzada para el pronóstico de ley de cobre. La línea azul corresponde al error en el conjunto de entrenamiento, mientras que la línea naranja corresponde al error en validación.

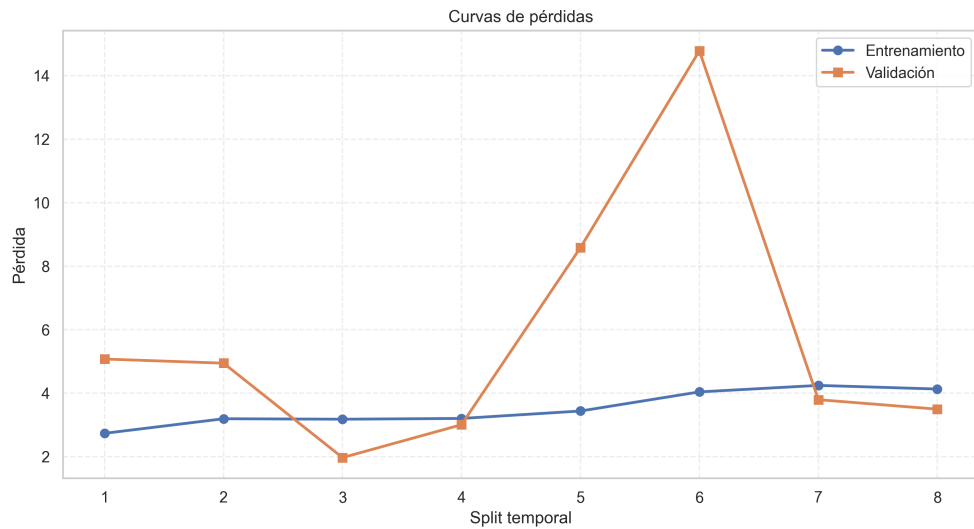


Fig. 3.11: Curvas de pérdida para el modelo de regresión lineal para el pronóstico de ley de cobre.

Se observa que el error de entrenamiento permanece relativamente estable y acotado a lo largo de los *splits*, mientras que el error de validación muestra mayor variabilidad, con picos significativos en algunos intervalos (notablemente en el sexto *split*). Este comportamiento refleja la alta variabilidad de la serie temporal y la dificultad del modelo para generalizar en ciertos segmentos.

En la Figura 3.12 se presentan las curvas de pérdida obtenidas durante la validación cruzada del modelo de regresión lineal múltiple aplicado a la razón hierro-sílice (Fe/SiO_2).

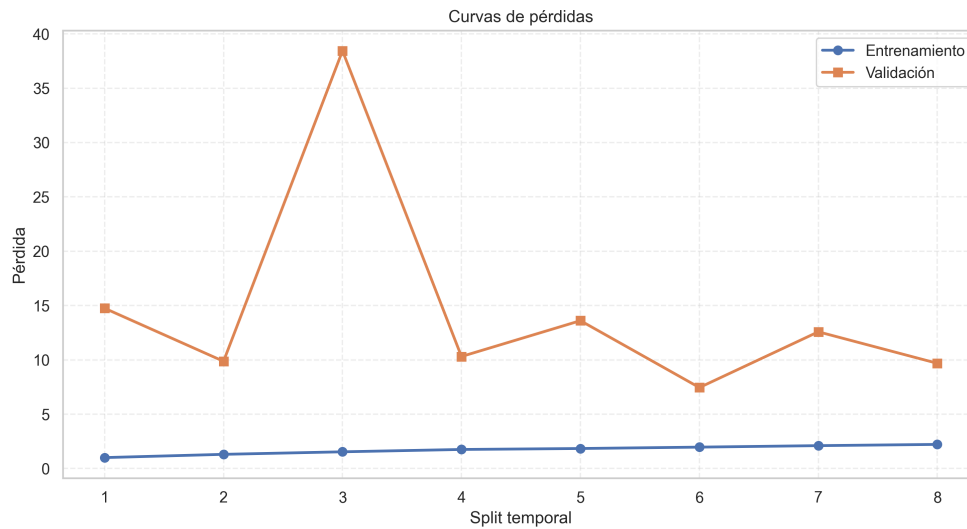


Fig. 3.12: Curvas de pérdida para el modelo de regresión lineal para el pronóstico de razón hierro-sílice

Se observa que la pérdida en el conjunto de entrenamiento (línea azul) permanece baja y relativamente estable en todos los *splits*, lo que indica que el modelo logra un buen ajuste sobre los datos de entrenamiento. Sin embargo, la pérdida en validación (línea naranja) muestra una variabilidad considerable, con picos significativos (notablemente en el tercer *split*, donde el error se eleva abruptamente por sobre 35 unidades de MASE).

Este comportamiento revela que, aunque el modelo captura adecuadamente los patrones en el conjunto de entrenamiento, su capacidad de generalización es limitada en ciertos intervalos temporales. Dichas discrepancias pueden estar asociadas a:

- **Alta variabilidad del proceso:** la razón hierro-sílice presenta fluctuaciones abruptas que dificultan el ajuste de un modelo lineal.
- **Perturbaciones externas:** algunos segmentos del subconjunto de validación muestran dinámicas diferentes a las vistas en el entrenamiento, lo que provoca un deterioro en el desempeño.
- **Limitaciones del enfoque lineal:** pese a la inclusión de lags y medias móviles, la complejidad no lineal del proceso metalúrgico puede requerir modelos más avanzados.

En síntesis, las curvas reflejan que el modelo logra estabilidad en entrenamiento, pero

presenta episodios de sobreajuste y pobre capacidad de generalización en validación. Estos resultados sugieren la necesidad de explorar enfoques más avanzados, para mejorar la robustez del pronóstico de la razón hierro-sílice.

3.5. Long Short-Term Memory

Con el propósito de superar las limitaciones observadas, se implementó un esquema híbrido que combina una *regresión lineal múltiple* como línea base y una red neuronal recurrente del tipo *Long Short-Term Memory* (LSTM) aplicada sobre los residuos. De esta forma, la componente lineal capta la tendencia general explicada por las variables exógenas, mientras que la LSTM modela patrones no lineales y dependencias temporales de mayor complejidad.

1. **Ingeniería de variables:** se construyeron rezagos ($lags = 5$) para cada variable exógena y cuatro variables temporales (seno y coseno de la hora del día y del día de la semana).
2. **Baseline lineal:** se ajustó una regresión lineal robusta sobre las variables de entrada. Su salida constituye la primera estimación de la variable objetivo.
3. **LSTM:** la diferencia entre los valores objetivo y el *baseline* fue modelada mediante una red LSTM bidireccional con normalización y capa de salida densa. La predicción final corresponde a la suma del *baseline* y el residuo estimado por la LSTM.
4. **Búsqueda de hiperparámetros:** se utilizó un *RandomizedSearch* optimizando la métrica MASE en un esquema de validación cruzada temporal. Los parámetros explorados incluyeron *lookback*, número de capas, tamaño oculto, tasa de *dropout*, bidireccionalidad, tamaño de lote, tasa de aprendizaje y regularización.
5. **Entrenamiento:** se aplicó descenso por gradiente estocástico con optimizador AdamW, función de pérdida Huber, *early stopping* y reducción adaptativa de la tasa de aprendizaje.

3.5.1. Curvas de pérdida

En la Figura 3.13 se presentan las curvas de pérdida para el modelo LSTM aplicado a la ley de cobre. Se observa un descenso progresivo de la pérdida tanto en entrenamiento como en validación, con convergencia hacia valores estables en torno a la centésima época. Las curvas se mantienen próximas, lo que refleja un entrenamiento estable sin sobreajuste.

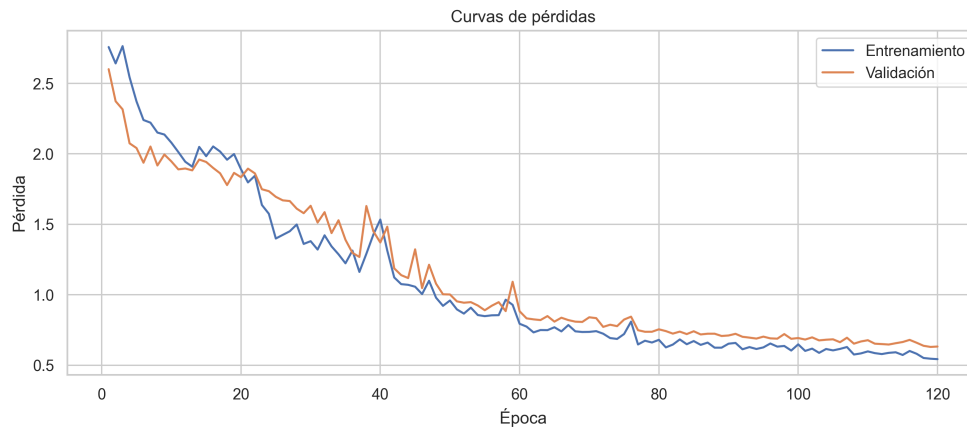


Fig. 3.13: Curvas de pérdida para el modelo LSTM aplicado a la ley de cobre.

En la Figura 3.14 se presentan las curvas de pérdida correspondientes al modelo LSTM para la razón hierro-sílice. A diferencia del caso del cobre, se observa mayor variabilidad en la pérdida de validación, con oscilaciones en los valores de pérdida. Este comportamiento evidencia la dificultad adicional para generalizar en esta variable, probablemente asociada a la mayor volatilidad y complejidad del proceso.

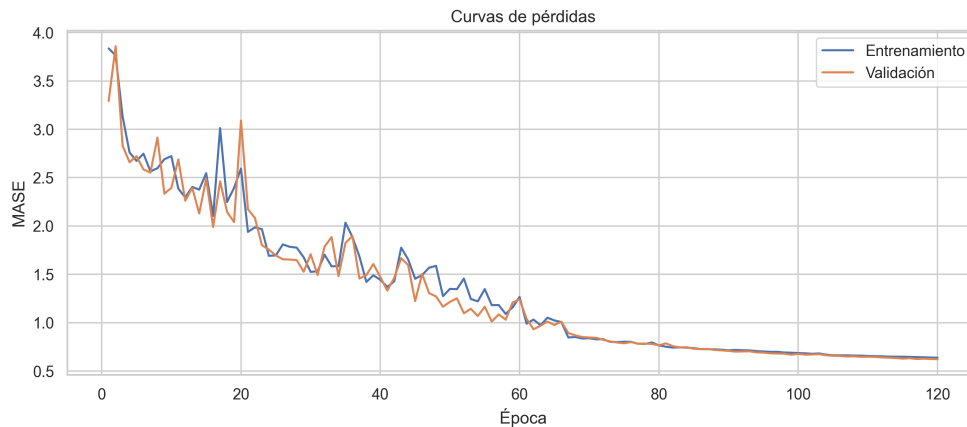


Fig. 3.14: Curvas de pérdida para el modelo LSTM aplicado a la razón hierro-sílice.

En síntesis, los resultados muestran que la incorporación de una LSTM aporta capacidad de modelar dinámicas más complejas, aunque su efectividad se ve limitada por la elevada variabilidad de las series y la posible necesidad de arquitecturas más robustas, como modelos de atención (TFT) o estrategias de calibración adicionales.

3.6. Temporal Fusion Transformer (TFT)

El *Temporal Fusion Transformer* (TFT) se aplicó con el propósito de explorar una arquitectura que combine mecanismos de atención multi-cabeza, compuertas (*gating*) y normalización, capaz de capturar dependencias de largo plazo y relaciones no lineales entre las variables exógenas y las variables objetivo (Cu y Fe/SiO_2). A diferencia de los modelos lineales o de redes recurrentes simples, el TFT integra múltiples componentes que permiten modelar de forma simultánea tendencias, estacionalidades y efectos complejos de las entradas.

1. **Ingeniería de variables:** se aplicó selección por información mutua, reteniendo las 40 variables más relevantes en el caso del cobre y 24 en el caso de la razón hierro-sílice.
2. **Configuración de entrada:** se definió una longitud de codificador (*encoder length*) de 72 pasos, con variables exógenas rezagadas y atributos temporales.
3. **Entrenamiento:** el modelo fue optimizado mediante *RandomSearch*, empleando

MASE como métrica principal y validación cruzada temporal. Se utilizó descenso por gradiente con optimizador AdamW, reducción adaptativa de la tasa de aprendizaje y *early stopping*.

3.6.1. Curvas de pérdida

En la Figura 3.15 se presentan las curvas de pérdida del modelo para la ley de cobre. Se observa una reducción sostenida de la pérdida en entrenamiento y validación, aunque con un nivel más alto y fluctuante en validación, lo que sugiere dificultades de generalización en algunos intervalos temporales.

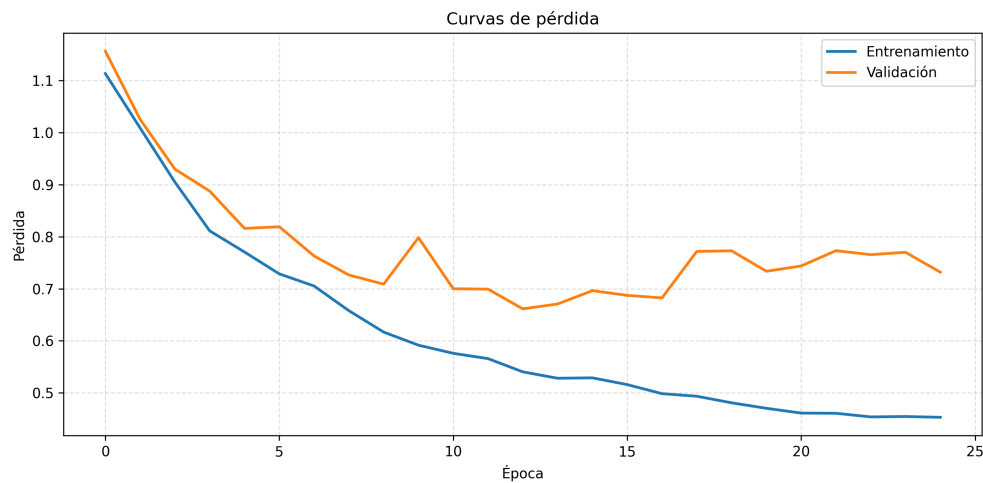


Fig. 3.15: Curvas de pérdida del modelo TFT aplicado a la ley de cobre.

En la Figura 3.16 se presentan las curvas de pérdida para la razón hierro-sílice. En este caso, se observa que la pérdida de entrenamiento desciende de manera continua, mientras que la validación se mantiene más estable pero con una brecha respecto al entrenamiento, lo que indica un ajuste moderado con riesgo de sobreajuste.

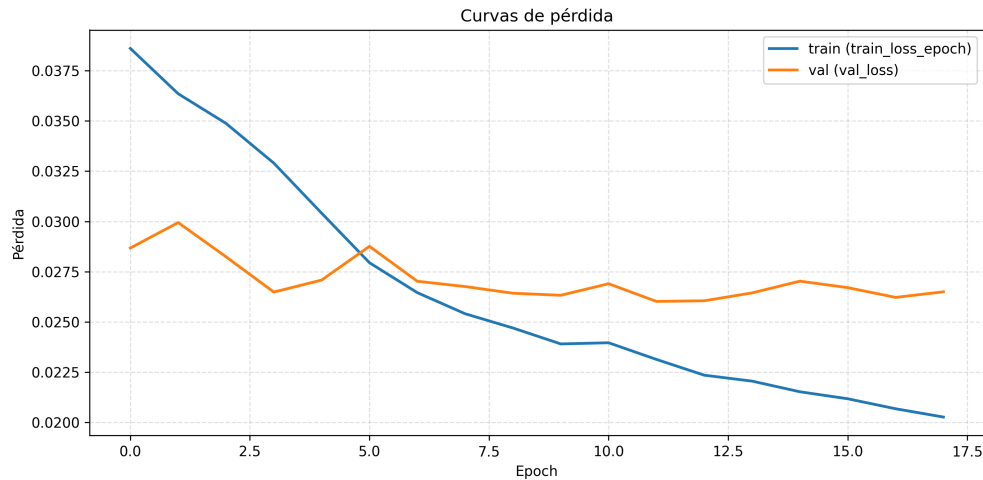


Fig. 3.16: Curvas de pérdida del modelo TFT aplicado a la razón hierro-sílice.

Los resultados indican que el TFT ofrece una mejora en la capacidad de capturar dependencias no lineales y patrones complejos en comparación con la regresión lineal y la LSTM .

3.7. Modelos endógenos

En este apartado se presentan los resultados de los modelos entrenados con el conjunto de datos endógeno, es decir, incluyendo el pasado de las variables objetivo (Cu y Fe/SiO_2). Este enfoque aprovecha la dependencia temporal directa de las series, lo cual generalmente incrementa la capacidad de predicción frente al conjunto exógeno.

3.7.1. Ley de cobre

Los modelos aplicados bajo este esquema (regresión lineal, LSTM y Temporal Fusion Transformer) compartieron un mismo *pipeline* que incluyó: ingeniería de variables mediante rezagos y medias móviles, expansión polinomial y selección de variables por información mutua.

En términos generales, los tres enfoques lograron un buen desempeño, destacando la capacidad de los modelos para capturar dependencias temporales al utilizar el pasado de la propia variable. El modelo de regresión lineal alcanzó un error de ($MASE = 0,545$,

$R^2 = 0,982$) 3.17 , confirmando que las relaciones lineales explican buena parte de la serie temporal. El modelo LSTM entregó el mejor rendimiento global, con $MASE = 0,459$, $MAE = 0,0797$ y correlación de Pearson de 0.992 3.18 , lo que demuestra la capacidad de este modelo para complementar la componente lineal con patrones no lineales. Por su parte, el TFT, pese a su mayor flexibilidad, mostró un desempeño más limitado en este caso ($MASE = 1,20$, $R^2 = 0,925$) 3.19, lo que indica dificultades de generalización bajo la alta variabilidad de la serie.

Las curvas de pérdida ilustradas en las Figuras 3.17, 3.18 y 3.19 reflejan un comportamiento coherente con estos resultados: estabilidad relativa en la regresión lineal, convergencia progresiva y estable en la LSTM, y mayor brecha entre entrenamiento y validación en el TFT.

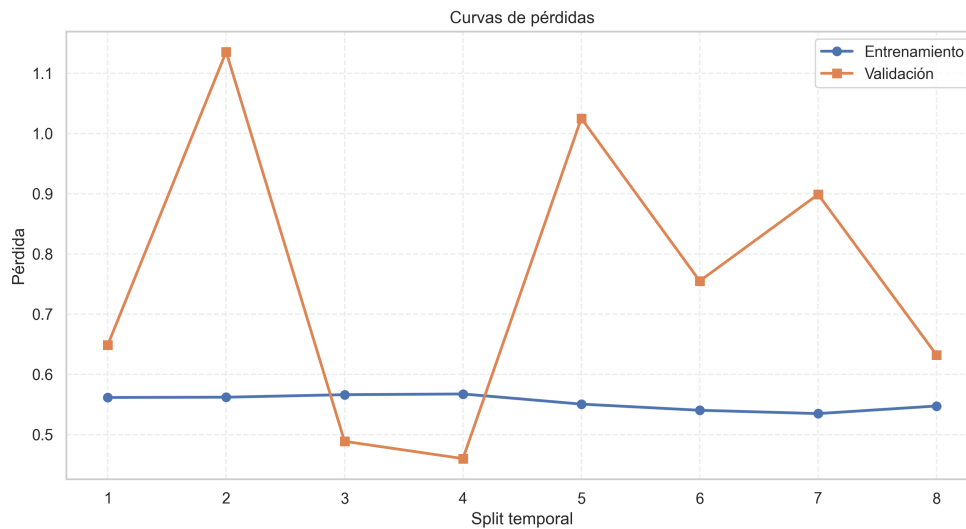


Fig. 3.17: Curvas de pérdida del modelo endógeno de regresión lineal para el pronóstico de la ley de cobre.

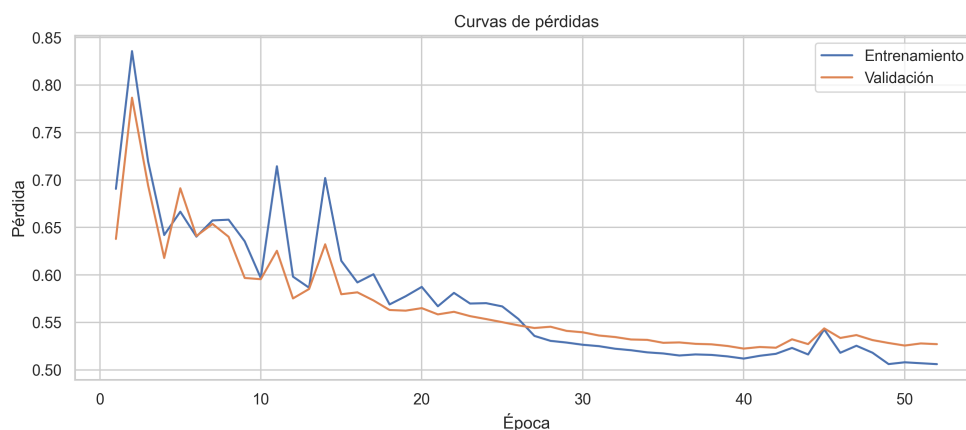


Fig. 3.18: Curvas de pérdida del modelo endógeno LSTM para el pronóstico de la ley de cobre.

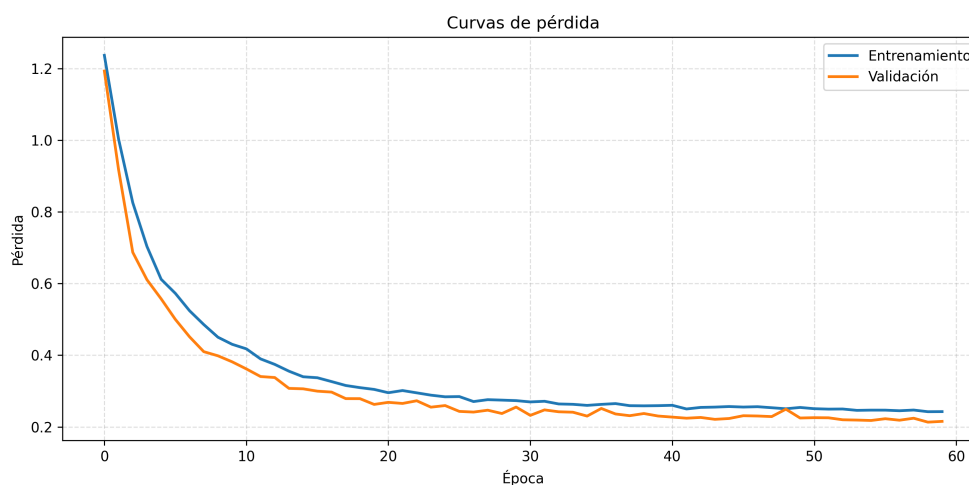


Fig. 3.19: Curvas de pérdida del modelo endógeno TFT para el pronóstico de la ley de cobre.

3.7.2. Razón hierro-sílice

En el caso de la razón hierro-sílice (Fe/SiO_2), los tres modelos aplicados bajo el mismo *pipeline* mostraron un comportamiento menos preciso debido a la elevada variabilidad de la serie. El modelo de regresión lineal obtuvo un desempeño de $MASE = 0,702$ y $R^2 = 0,939$, evidenciando que, si bien el modelo capta tendencias globales, se ve limitado frente a variaciones abruptas 3.20. El modelo LSTM, en cambio, logró mejorar la capacidad predictiva, alcanzando $MASE = 0,645$ y $R^2 = 0,955$, lo que confirma su

ventaja en el conjunto endógeno. Finalmente, el TFT mostró el desempeño más bajo ($MASE = 1,83$, $R^2 = 0,755$), con una mayor brecha entre entrenamiento y validación, lo cual revela sobreajuste 3.22.

Las Figuras 3.20, 3.21 y 3.22 ilustran las curvas de pérdida correspondientes a la regresión lineal, la LSTM y el modelo TFT.

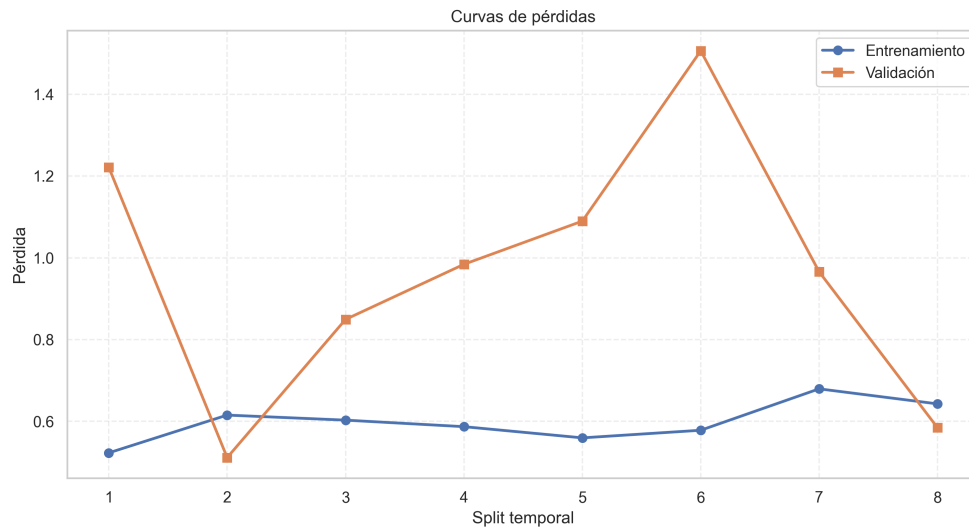


Fig. 3.20: Curvas de pérdida del modelo de regresión lineal para el pronóstico de la razón hierro-sílice (conjunto endógeno).

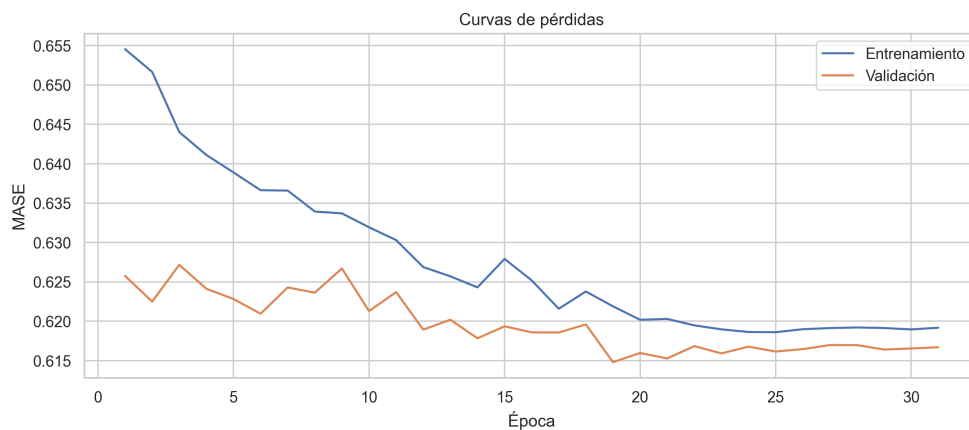


Fig. 3.21: Curvas de pérdida del modelo híbrido LSTM para el pronóstico de la razón hierro-sílice (conjunto endógeno).

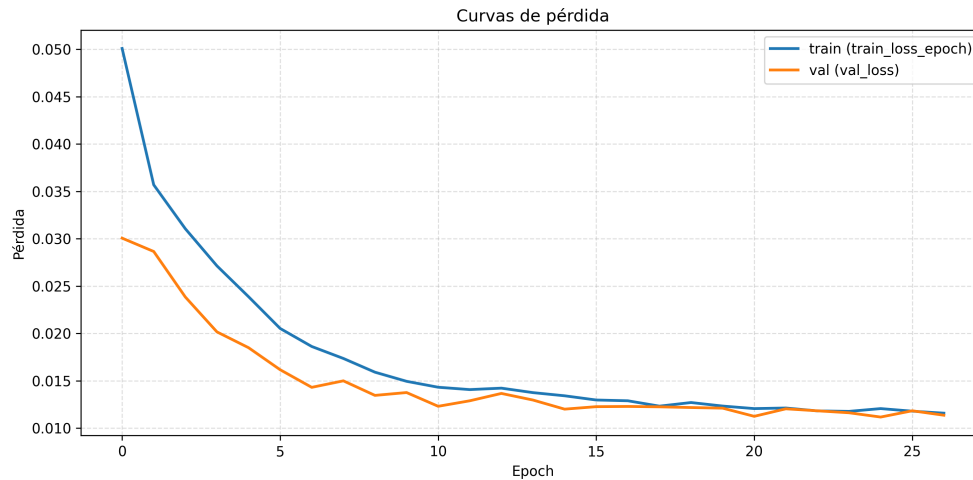


Fig. 3.22: Curvas de pérdida del modelo TFT para el pronóstico de la razón hierro-sílice (conjunto endógeno).

En síntesis, los resultados del conjunto endógeno muestran que:

- Para la ley de cobre, el modelo LSTM alcanzó el mejor desempeño, superando tanto a la regresión lineal como al TFT.
- Para la razón hierro-sílice, la LSTM también se consolidó como la mejor alternativa, aunque con mayores dificultades de generalización debido a la alta volatilidad de la serie.
- El TFT, pese a su sofisticación, presentó un desempeño más débil en ambas variables, lo que sugiere la necesidad de mayor ajuste de hiperparámetros o técnicas de regularización adicionales.

3.8. Modelos suavizados

Con el propósito de evaluar el impacto de la atenuación del ruido sobre el desempeño predictivo, se aplicó un suavizado previo a las series mediante una media móvil exponencial (EMA), de 3 períodos. En todos los casos el suavizado se aplicó tanto a las variables de entrada como a las variables objetivo, manteniendo luego el mismo *pipeline* de modelado.

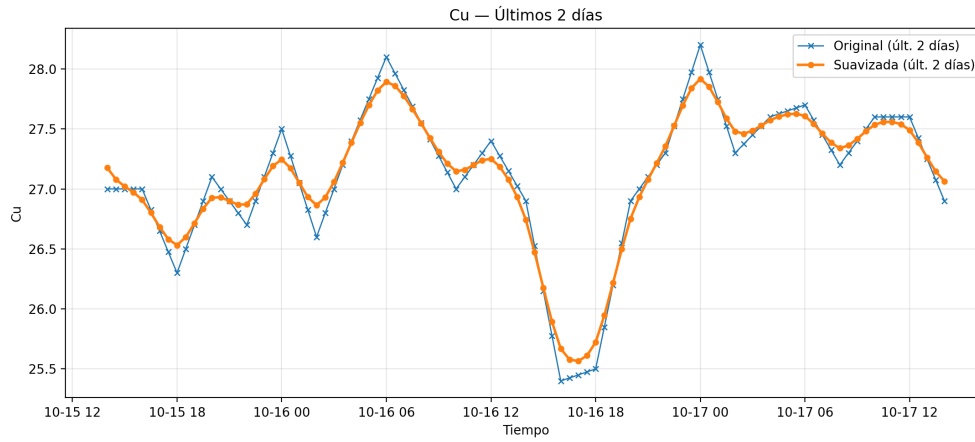


Fig. 3.23: Ley de cobre suavizado mediante EMA.

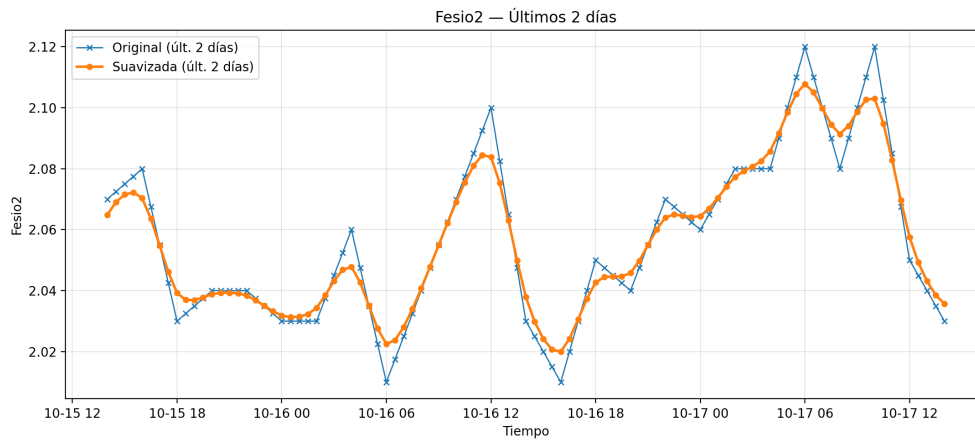


Fig. 3.24: Razón hierro-sílice suavizado mediante EMA.

3.8.1. Ley de cobre

En la ley de cobre, el **TFT** pero con errores elevados en el conjunto de prueba, obtuvo $MASE = 5,04$, $MAE = 0,589$ y $R^2 = 0,304$.

La **regresión lineal** presentó el peor comportamiento, con $MASE = 18,86$ y $R^2 = -1,53$ en prueba, lo que denota sobreajuste y escasa generalización bajo el suavizado.

El modelo **LSTM** también se vio afectado por el suavizado, con $MASE = 8,80$ y $R^2 = -0,80$ en prueba.

Las Figuras 3.25, 3.26 y 3.27 muestran curvas de pérdida más estables pero con mayores errores en validación y prueba respecto de los modelos no suavizados.

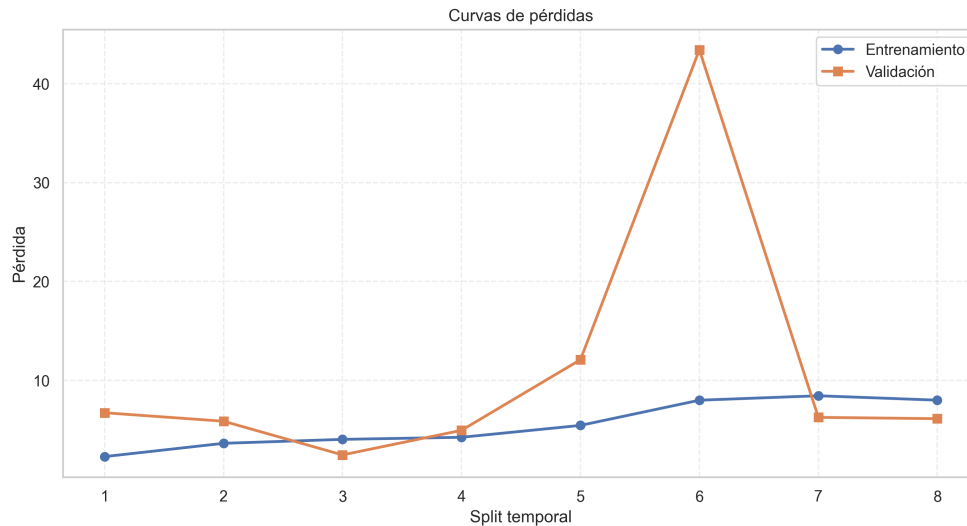


Fig. 3.25: Curvas de pérdida del modelo suavizado de regresión lineal para el pronóstico de la ley de cobre.

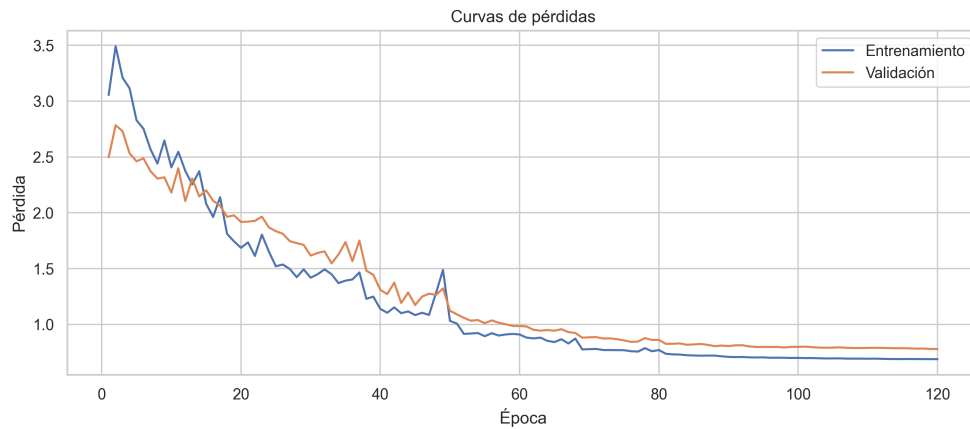


Fig. 3.26: Curvas de pérdida del modelo suavizado LSTM para el pronóstico de la ley de cobre.

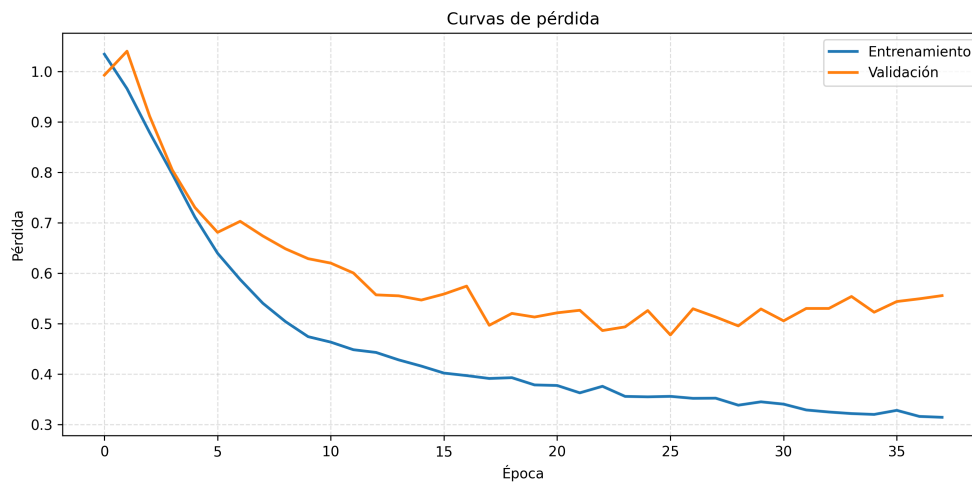


Fig. 3.27: Curvas de pérdida del modelo suavizado TFT para el pronóstico de la ley de cobre.

3.8.2. Razón hierro-sílice

Para la razón Fe/SiO_2 , el **TFT** logró el mejor desempeño relativo, pero con errores elevados en el conjunto de prueba, obtuvo $MASE = 5,19$ y $R^2 = 0,024$.

La **regresión lineal** presentó el peor comportamiento, con $MASE = 13,15$ y $R^2 = -0,62$ en el conjunto de prueba, lo que denota sobreajuste y escasa generalización bajo el suavizado.

El modelo **LSTM** también se vio afectado por el suavizado, con $MASE = 9,41$ y $R^2 = -1,83$ en prueba.

Las Figuras 3.28, 3.29 y 3.30 muestran una aparente estabilización de las curvas, no obstante, los errores en validación y prueba aumentan respecto a los modelos no suavizados.

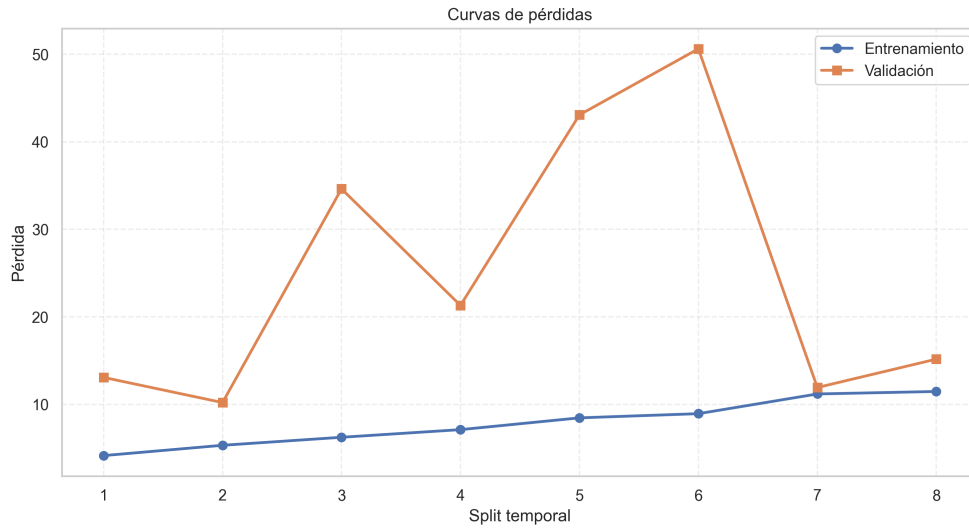


Fig. 3.28: Curvas de pérdida del modelo suavizado de regresión lineal para el pronóstico de la razón hierro-sílice.

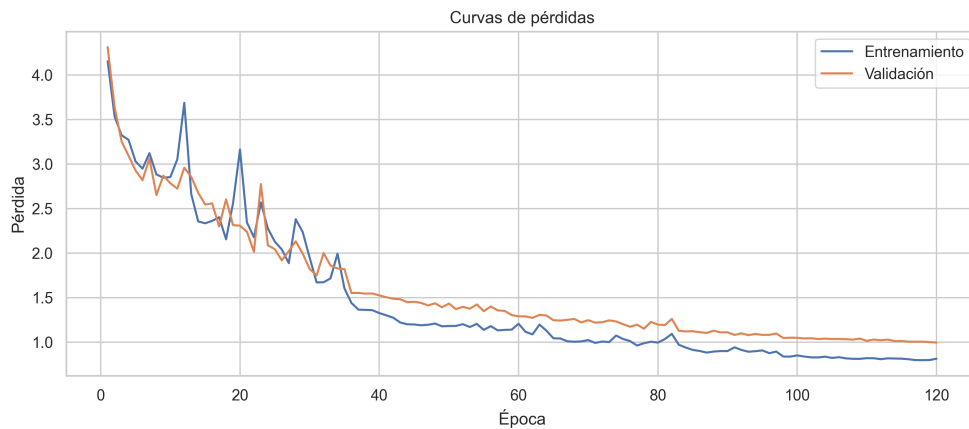


Fig. 3.29: Curvas de pérdida del modelo suavizado LSTM para el pronóstico de la razón hierro-sílice.

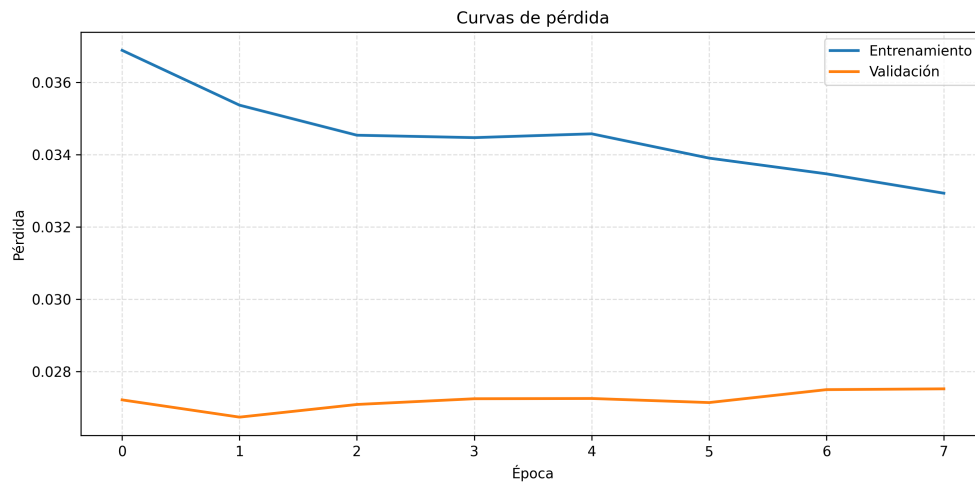


Fig. 3.30: Curvas de pérdida del modelo suavizado TFT para el pronóstico de la razón hierro-sílice.

Capítulo 4

Resultados

4.1. Resultados Regresión lineal para pronóstico de ley de cobre

Para el pronóstico de la ley de cobre (Cu) el mejor modelo correspondió a una regresión lineal con los siguientes parámetros:

- **Grado polinomial:** $degree = 1$ (modelo lineal).
- **Regularización:** $alpha = 0,3$.
- **Intercepto:** activado ($fit_intercept = True$).
- **Selección de variables:** $k = 40$ (tras expansión polinomial).
- **Número de parámetros efectivos:** 41 coeficientes (40 variables más intercepto).

En la Tabla 4.1 se presentan los principales resultados del modelo en el conjunto de prueba.

Tabla 4.1: Métricas de desempeño del modelo de regresión lineal en el conjunto de prueba para la ley de cobre.

Métrica	Valor
MAE	1.0486
RMSE	1.3369
R^2	-0.7059
MASE	6.0330
Sesgo de pronóstico (FB %)	-3.589 %
Coefficiente de correlación de Pearson	0.549
Información mutua normalizada (NMI)	0.722

4.1.1. Correlación cruzada

La Figura 4.1 muestra la correlación cruzada entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente $lag = -3$, aproximadamente -90 minutos, con una correlación máxima de 0.603, lo cual indica que las predicciones tienden a ir rezagadas respecto de la serie real.

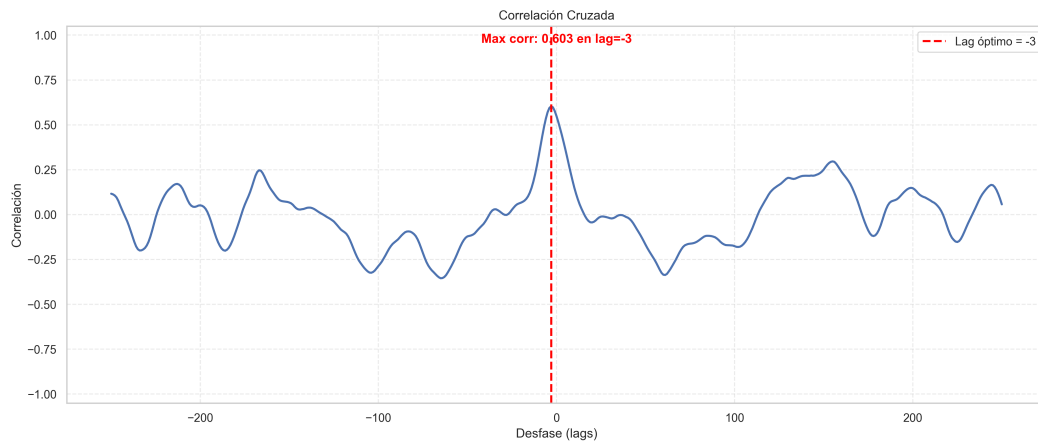


Fig. 4.1: Correlación cruzada para la ley de cobre utilizando regresión lineal.

La Figura 4.2 muestra la correlación cruzada después de un proceso de interpolación de 29 puntos entre cada valor, logrando así una aproximación más exacta de los minutos de desfase entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente de -88 minutos y máxima correlación de 0.587.

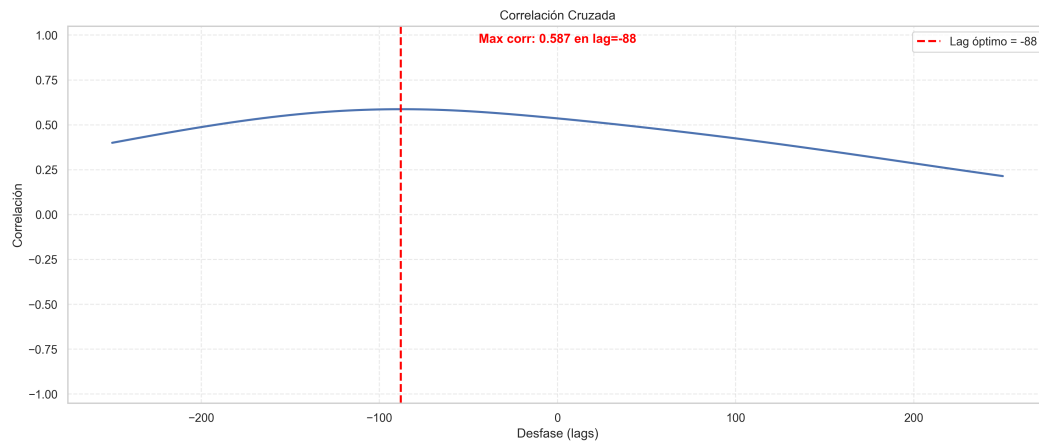


Fig. 4.2: Correlación cruzada interpolada para la ley de cobre utilizando regresión lineal

4.1.2. Diagrama de dispersión

La Figura 4.3 presenta el diagrama de dispersión entre los valores predichos y los objetivos. Se observa una correlación de Pearson moderada ($r = 0,549$) y un valor de NMI de 0.722. Sin embargo, el coeficiente de determinación negativo ($R^2 = -0,706$) evidencia que el modelo no logra explicar adecuadamente la variabilidad de la serie, mostrando además dispersión considerable alrededor de la recta identidad.

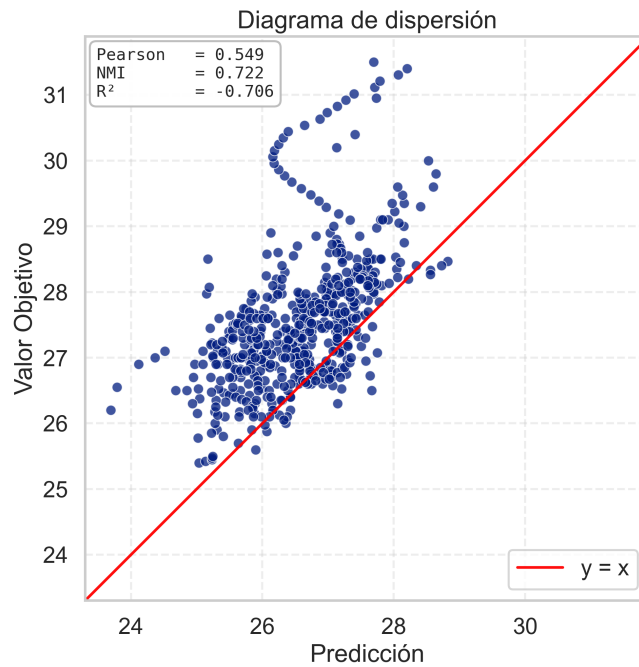


Fig. 4.3: Diagrama de dispersión entre valores objetivo y predichos de la ley de cobre utilizando el modelo de regresión lineal.

4.1.3. Serie temporal

Finalmente, en la Figura 4.4 se muestra la comparación entre la serie temporal de los valores objetivo de la ley de cobre y las predicciones del modelo. Se aprecia que el modelo reproduce parcialmente la tendencia general, aunque presenta rezagos temporales y dificultades para capturar picos y caídas abruptas, lo que limita su capacidad predictiva.

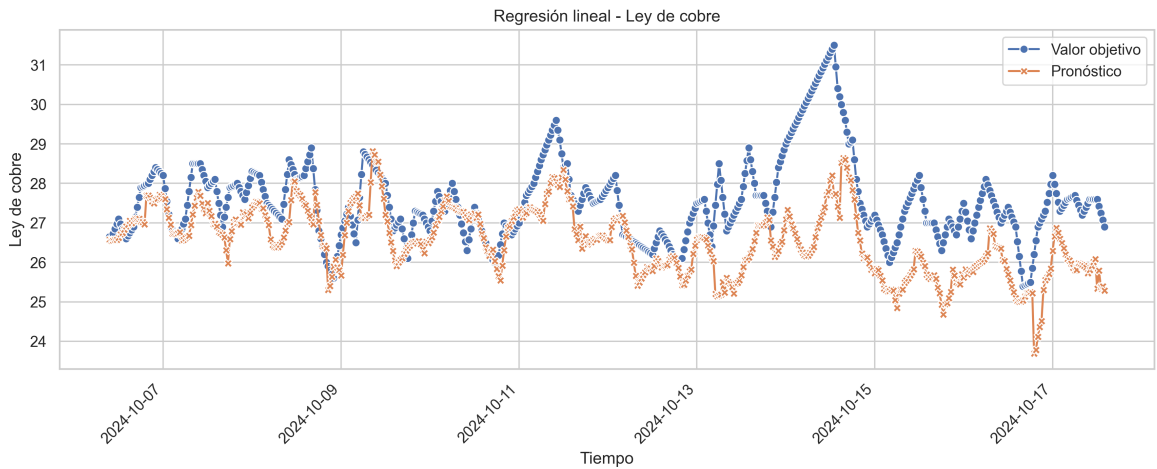


Fig. 4.4: Comparación entre valores objetivo y predicciones de la ley de cobre en el conjunto de prueba utilizando el modelo de regresión lineal.

4.2. Resultados Regresión lineal para pronóstico de la razón Fe/SiO_2

Para el pronóstico de la razón Fe/SiO_2 el mejor modelo correspondió a una regresión lineal con los siguientes parámetros:

- **Grado polinomial:** $degree = 2$ (modelo cuadrático).
- **Regularización:** $alpha = 10$.
- **Intercepto:** activado ($fit_intercept = True$).
- **Selección de variables:** $k = 8126$ (tras expansión polinomial).
- **Número de parámetros efectivos:** 8,127 coeficientes.

En la Tabla 4.2 se presentan los principales resultados del modelo en el conjunto de prueba.

Tabla 4.2: Métricas de desempeño del modelo de regresión lineal en el conjunto de prueba para la razón Fe/SiO₂.

Métrica	Valor
MAE	0.0281
RMSE	0.0353
R ²	0.0458
MASE	4.0818
Sesgo de pronóstico (FB %)	0.077 %
Coefficiente de correlación de Pearson	0.463
Información mutua normalizada (NMI)	0.525

4.2.1. Correlación cruzada

La Figura 4.5 muestra la correlación cruzada entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente $lag = -3$, aproximadamente -90 minutos, con una correlación máxima de 0.510, lo cual indica que las predicciones tienden a ir rezagadas respecto de la serie real.

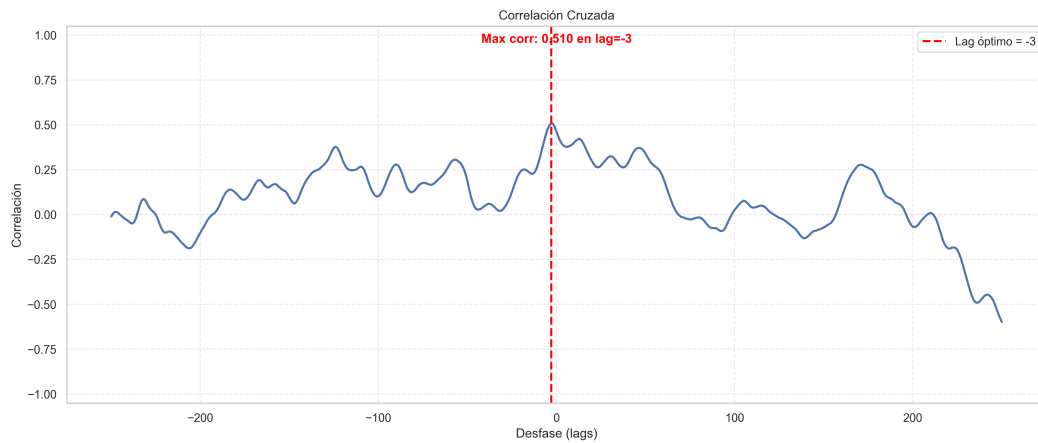


Fig. 4.5: Correlación cruzada para la razón Fe/SiO₂ utilizando regresión lineal.

La Figura 4.6 muestra la correlación cruzada después de un proceso de interpolación de 29 puntos entre cada valor, logrando así una aproximación más exacta de los minutos de desfase entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente de -83 minutos y máxima correlación de 0.519.

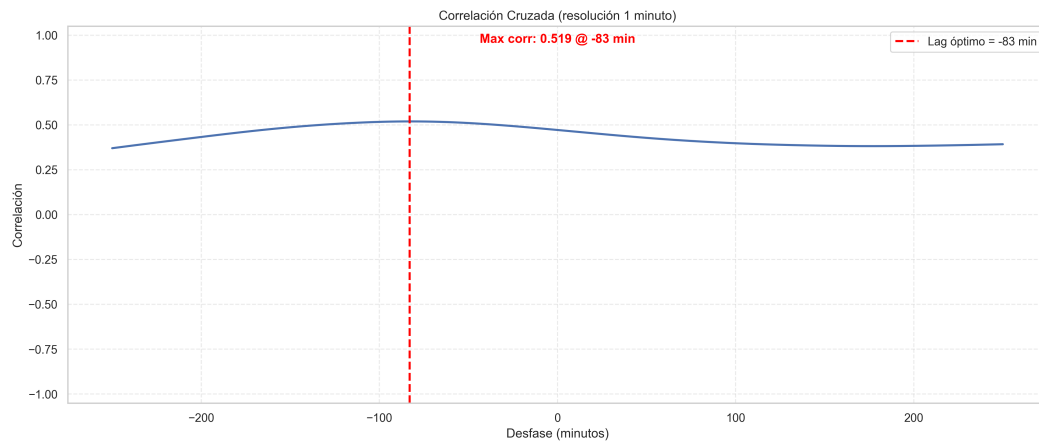


Fig. 4.6: Correlación cruzada interpolada para la razón Fe/SiO₂ utilizando regresión lineal

4.2.2. Diagrama de dispersión

La Figura 4.7 presenta el diagrama de dispersión entre los valores predichos y los objetivo. Se observa una correlación de Pearson moderada ($r = 0,463$) pero con dispersión considerable alrededor de la recta identidad, lo que refleja limitaciones del modelo para capturar completamente la variabilidad de la serie.

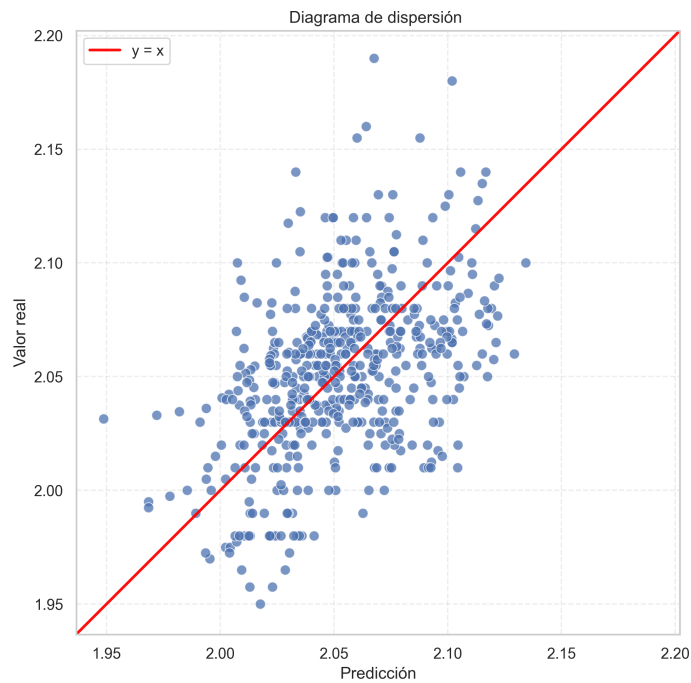


Fig. 4.7: Diagrama de dispersión entre valores objetivo y predichos de la razón Fe/SiO_2 utilizando el modelo de regresión lineal.

4.2.3. Serie temporal

Finalmente, en la Figura 4.8 se muestra la comparación entre la serie temporal de los valores objetivo de la razón Fe/SiO_2 y las predicciones del modelo. Se aprecia que el modelo reproduce parcialmente la tendencia general, aunque presenta rezagos y limitaciones en la captura de picos y caídas abruptas.

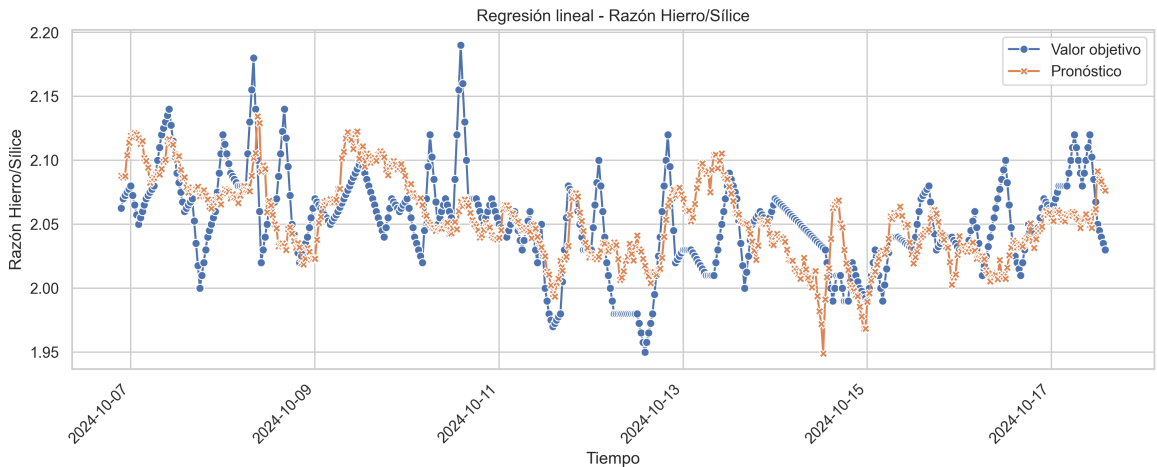


Fig. 4.8: Comparación entre valores objetivo y predicciones de la razón Fe/SiO_2 en el conjunto de prueba utilizando el modelo de regresión lineal.

4.3. Resultados LSTM para pronóstico de ley de cobre

Para el pronóstico de la ley de cobre (Cu) se implementó un modelo *Long Short-Term Memory* (LSTM) en combinación con un modelo lineal base que estima el valor objetivo a partir de variables exógenas, mientras que la red neuronal LSTM se encarga de modelar los residuos del modelo lineal. El mejor conjunto de hiperparámetros fue obtenido mediante *RandomizedSearch* optimizando la métrica MASE, resultando en la siguiente configuración:

- **Ventana de retroceso (lookback):** 12 pasos.
- **Tamaño de capa oculta:** 64 neuronas.
- **Número de capas:** 3.
- **Dropout:** 0.2.
- **Bidireccional:** activado.
- **Tasa de aprendizaje:** $lr = 0,001$.
- **Tamaño de batch:** 32.

- **Épocas:** 120 (con *early stopping*, paciencia de 12).
- **Regularización:** $weight_decay = 1e-4$.

En la Tabla 4.3 se presentan los principales resultados del modelo en el conjunto de prueba.

Tabla 4.3: Métricas de desempeño del modelo LSTM en el conjunto de prueba para la ley de cobre.

Métrica	Valor
MAE	1.1002
RMSE	1.4336
R ²	-0.963
MASE	6.322
Sesgo de pronóstico (FB %)	-3.517 %
Coefficiente de correlación de Pearson	0.430
Información mutua normalizada (NMI)	0.643

4.3.1. Correlación cruzada

La Figura 4.9 muestra la correlación cruzada entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente $lag = -3$, aproximadamente -90 minutos, con una correlación máxima de 0.483, lo que confirma que el modelo presenta rezago respecto de la serie real.

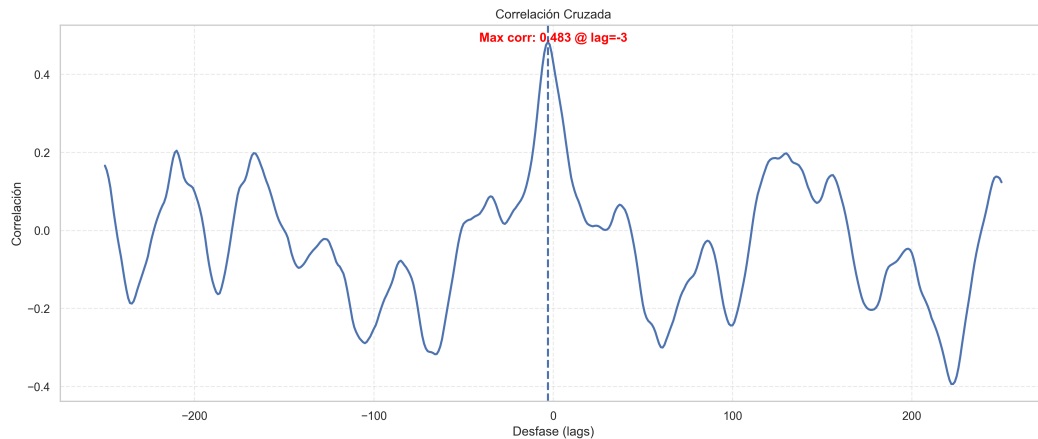


Fig. 4.9: Correlación cruzada para la ley de cobre utilizando el modelo LSTM.

La Figura 4.10 muestra la correlación cruzada después de un proceso de interpolación de 29 puntos entre cada valor, logrando así una aproximación más exacta de los minutos de desfase entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente de -89 minutos y máxima correlación de 0.488.

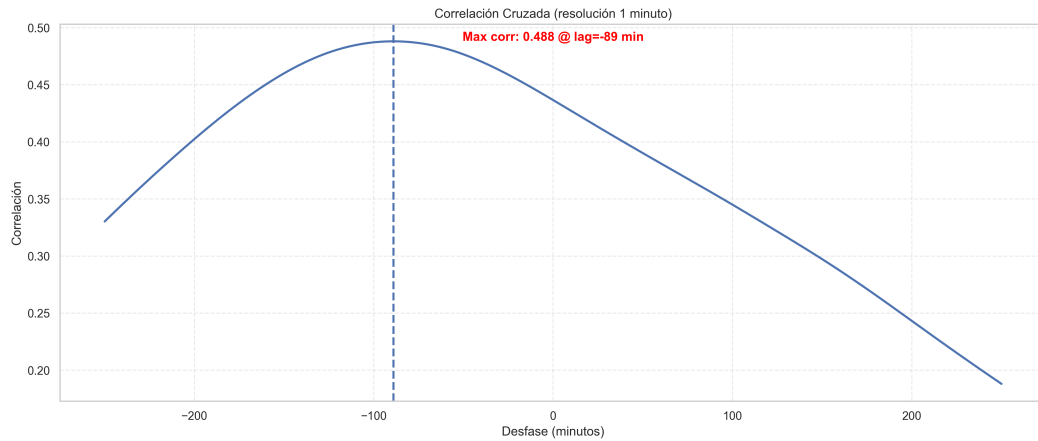


Fig. 4.10: Correlación cruzada interpolada para la ley de cobre utilizando el modelo LSTM.

4.3.2. Diagrama de dispersión

La Figura 4.11 presenta el diagrama de dispersión entre los valores objetivo y predichos. Se observa una correlación de Pearson moderada ($r = 0,430$) y un valor de NMI de 0.643, aunque la dispersión alrededor de la recta identidad sigue siendo elevada, indicando limitaciones en la capacidad de generalización del modelo.

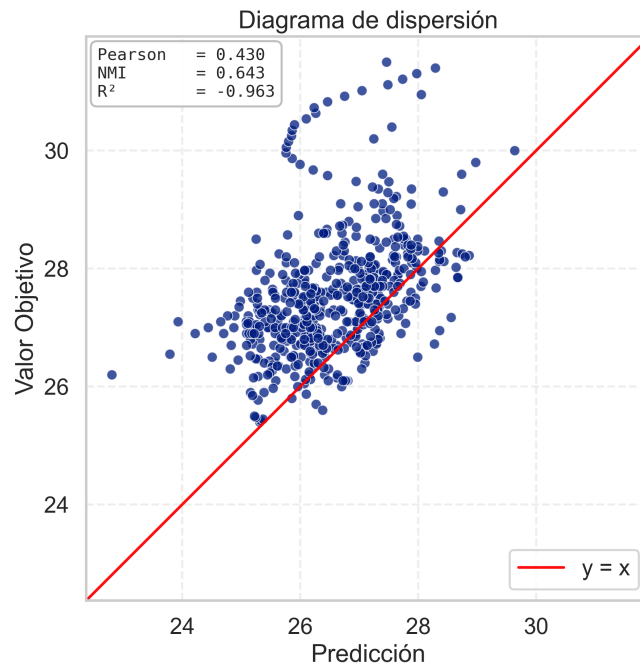


Fig. 4.11: Diagrama de dispersión entre valores objetivo y predichos de la ley de cobre para el modelo LSTM.

4.3.3. Serie temporal

Finalmente, en la Figura 4.12 se muestra la comparación entre la serie temporal de los valores objetivo y las predicciones generadas por el modelo. Si bien el LSTM captura parcialmente la tendencia de la serie, se aprecia que persisten rezagos y dificultades en la representación de picos y caídas abruptas, lo cual limita su desempeño predictivo.

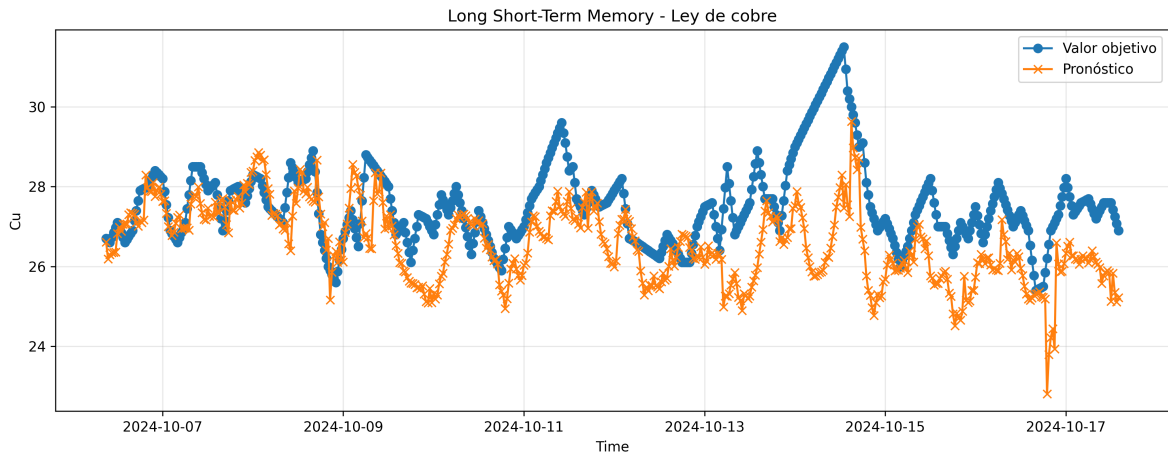


Fig. 4.12: Comparación entre valores objetivo y predicciones de la ley de cobre en el conjunto de prueba utilizando el modelo LSTM.

4.4. Resultados LSTM para pronóstico de la razón Fe/SiO_2

Para el pronóstico de la razón Fe/SiO_2 se implementó un modelo *Long Short-Term Memory* (LSTM) en combinación con un modelo lineal base, donde la red neuronal se entrenó sobre los residuos del modelo lineal. El mejor conjunto de hiperparámetros fue obtenido mediante *RandomizedSearch* optimizando la métrica MASE, resultando en la siguiente configuración:

- **Ventana de retroceso (lookback):** 12 pasos.
- **Tamaño de capa oculta:** 64 neuronas.
- **Número de capas:** 1.
- **Dropout:** 0.1.
- **Bidireccional:** activado.
- **Tasa de aprendizaje:** $lr = 0,001$.
- **Tamaño de batch:** 96.
- **Épocas:** 120 (con *early stopping*, paciencia de 12).

- **Regularización:** $weight_decay = 0,0$.

En la Tabla 4.4 se presentan los principales resultados del modelo en el conjunto de prueba.

Tabla 4.4: Métricas de desempeño del modelo híbrido LSTM en el conjunto de prueba para la razón Fe/SiO₂.

Métrica	Valor
MAE	0.0340
RMSE	0.0435
R ²	-0.475
MASE	4.846
Sesgo de pronóstico (FB %)	-0.674 %
Coefficiente de correlación de Pearson	0.276
Información mutua normalizada (NMI)	0.455

4.4.1. Correlación cruzada

La Figura 4.13 muestra la correlación cruzada entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente $lag = -3$, aproximadamente -90 minutos, con una correlación máxima de 0.309, lo cual indica un rezago en las predicciones respecto de la serie real.

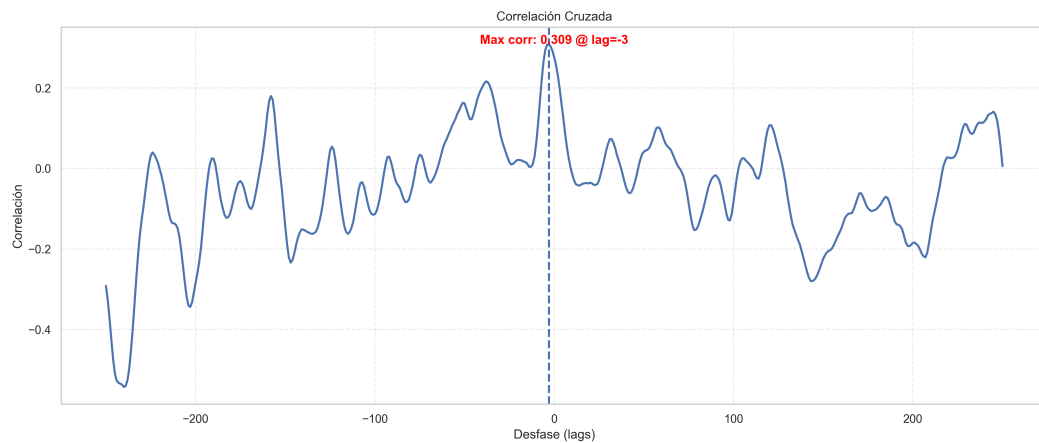


Fig. 4.13: Correlación cruzada para la razón Fe/SiO₂ utilizando el modelo LSTM.

La Figura 4.14 muestra la correlación cruzada después de un proceso de interpolación de 29 puntos entre cada valor, logrando así una aproximación más exacta de los minutos de desfase entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente de -97 minutos y máxima correlación de 0.314.

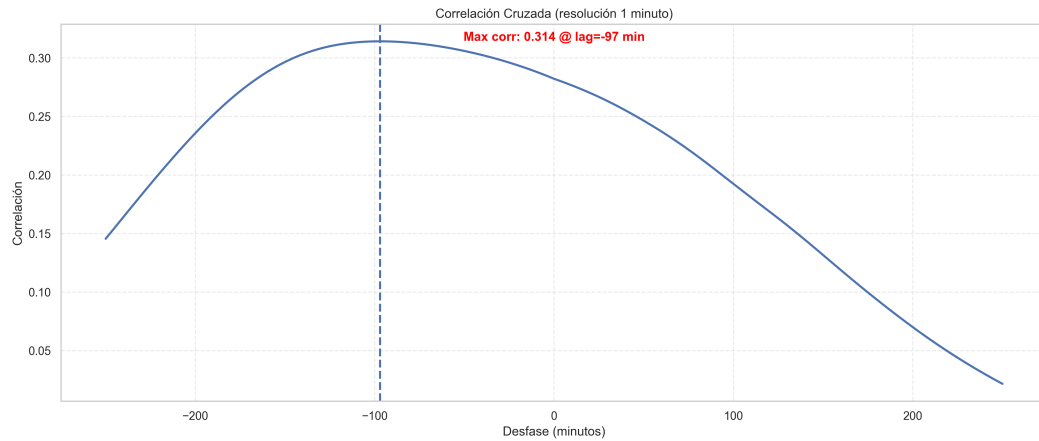


Fig. 4.14: Correlación cruzada interpolada para la razón Fe/SiO₂ utilizando el modelo LSTM.

4.4.2. Diagrama de dispersión

La Figura 4.15 presenta el diagrama de dispersión entre los valores objetivo y predichos. La correlación de Pearson es baja ($r = 0,276$) y el valor de NMI es de 0.455, lo que refleja una capacidad limitada del modelo para capturar la variabilidad de la serie, con amplia dispersión en torno a la recta identidad.

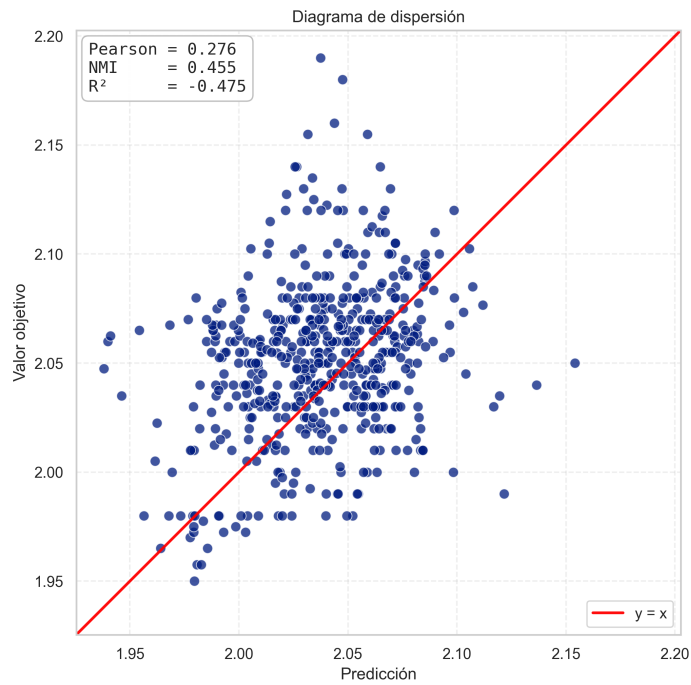


Fig. 4.15: Diagrama de dispersión entre valores objetivo y predichos de la razón Fe/SiO₂.

4.4.3. Serie temporal

Finalmente, en la Figura 4.16 se muestra la comparación entre la serie temporal de los valores objetivo de la razón Fe/SiO₂ y las predicciones del modelo. Aunque el LSTM reproduce parcialmente las tendencias generales, persisten rezagos y dificultades para capturar con precisión picos y caídas abruptas.

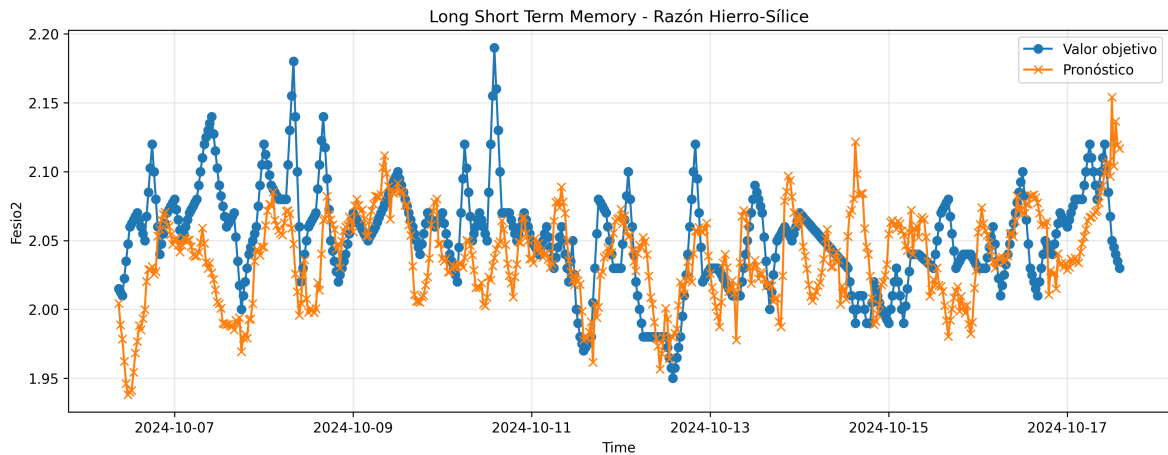


Fig. 4.16: Comparación entre valores objetivo y predicciones de la razón Fe/SiO_2 en el conjunto de prueba utilizando el modelo LSTM.

4.5. Resultados TFT para pronóstico de ley de cobre

Para el pronóstico de la ley de cobre (Cu) se implementó el modelo *Temporal Fusion Transformer* (TFT), el cual combina mecanismos de atención y estructuras recurrentes para modelar series temporales multivariadas. El mejor conjunto de hiperparámetros se obtuvo mediante *RandomSearch* optimizando la métrica MASE, resultando en la siguiente configuración:

- **Tamaño de capa oculta:** 32 neuronas.
- **Número de cabezas de atención:** 8.
- **Dropout:** 0.2.
- **Tasa de aprendizaje:** $lr = 0,0007$.
- **Longitud del codificador:** 72 pasos.
- **Número de parámetros efectivos:** 124,579.

En la Tabla 4.5 se presentan los principales resultados del modelo en el conjunto de prueba.

Tabla 4.5: Métricas de desempeño del modelo TFT en el conjunto de prueba para la ley de cobre.

Métrica	Valor
MAE	0.6791
RMSE	0.8787
R ²	0.262
MASE	3.9073
Sesgo de pronóstico (FB %)	0.891 %
Coefficiente de correlación de Pearson	0.593
Información mutua normalizada (NMI)	0.831

4.5.1. Correlación cruzada

La Figura 4.17 muestra la correlación cruzada entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente $lag = -3$, aproximadamente -90 minutos, con una correlación máxima de 0.623, lo cual indica que las predicciones presentan cierto rezago respecto de la serie real.

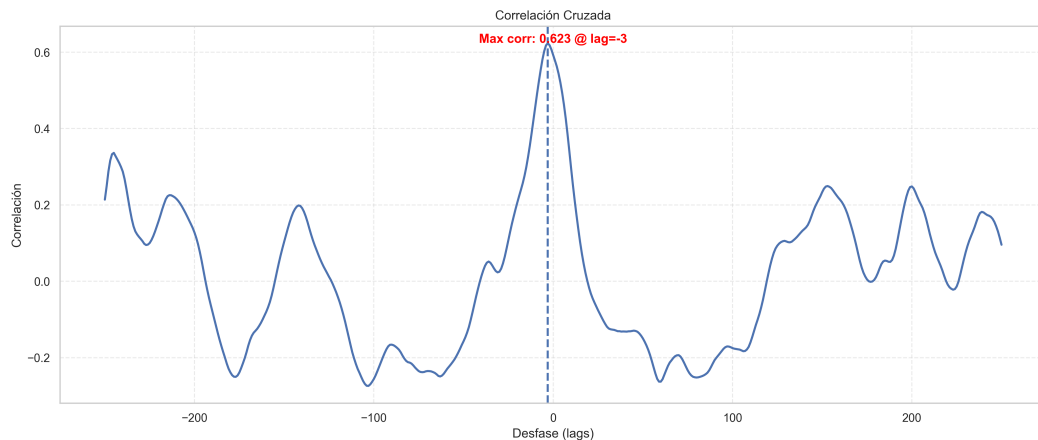


Fig. 4.17: Correlación cruzada para la ley de cobre utilizando el modelo TFT.

La Figura 4.18 muestra la correlación cruzada después de un proceso de interpolación de 29 puntos entre cada valor, logrando así una aproximación más exacta de los minutos de desfase entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente de -89 minutos y máxima correlación de 0.627.

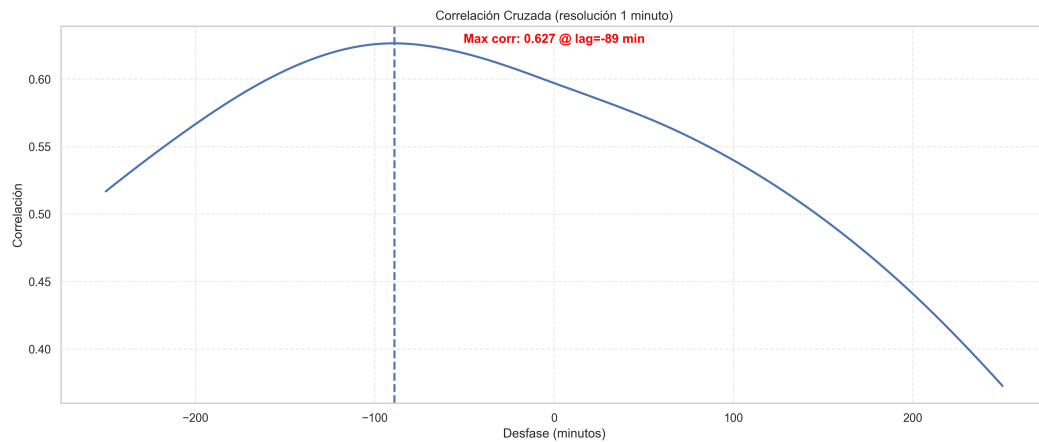


Fig. 4.18: Correlación cruzada interpolada para la ley de cobre utilizando el modelo TFT.

4.5.2. Diagrama de dispersión

La Figura 4.19 presenta el diagrama de dispersión entre los valores objetivo y predichos en el conjunto de prueba. Se observa una correlación de Pearson moderada-alta ($r = 0,593$) y un valor de NMI de 0.831, lo que evidencia que el modelo captura de manera más robusta la relación entre predicciones y valores objetivo en comparación con los modelos lineales y LSTM.



Fig. 4.19: Diagrama de dispersión entre valores objetivo y predichos de la ley de cobre utilizando el modelo TFT.

4.5.3. Serie temporal

Finalmente, en la Figura 4.20 se muestra la comparación entre la serie temporal de los valores objetivo de la ley de cobre y las predicciones del modelo. Se aprecia que el TFT logra una mejor representación de la tendencia general, con capacidad de seguir de manera más ajustada los patrones de variación, aunque todavía con dificultades para capturar todos los picos y caídas abruptas.

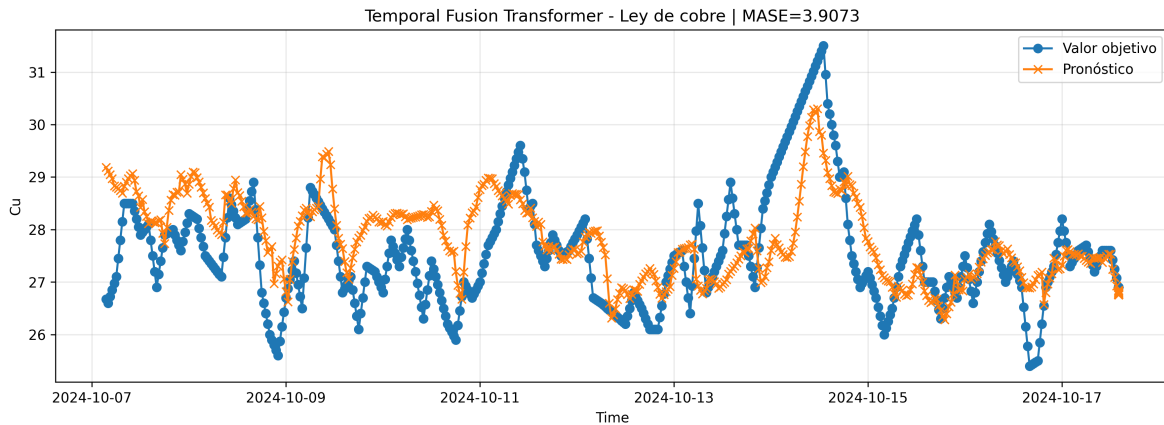


Fig. 4.20: Comparación entre valores objetivo y predicciones de la ley de cobre en el conjunto de prueba utilizando el modelo TFT.

4.6. Resultados TFT para pronóstico de la razón Fe/SiO_2

Para el pronóstico de la razón Fe/SiO_2 se implementó el modelo *Temporal Fusion Transformer* (TFT), el cual utiliza mecanismos de atención para capturar dependencias temporales de corto y largo plazo. El mejor conjunto de hiperparámetros se obtuvo mediante *RandomSearch* optimizando la métrica MASE, resultando en la siguiente configuración:

- **Tamaño de capa oculta:** 32 neuronas.
- **Número de cabezas de atención:** 8.
- **Dropout:** 0.25.
- **Tasa de aprendizaje:** $lr = 0,001$.
- **Longitud del codificador:** 72 pasos.
- **Número de parámetros efectivos:** 105,011.

En la Tabla 4.6 se presentan los principales resultados del modelo en el conjunto de prueba.

Tabla 4.6: Métricas de desempeño del modelo TFT en el conjunto de prueba para la razón Fe/SiO₂.

Métrica	Valor
MAE	0.0276
RMSE	0.0361
R ²	-0.025
MASE	3.943
Sesgo de pronóstico (FB %)	-0.107 %
Coefficiente de correlación de Pearson	0.331
Información mutua normalizada (NMI)	0.687

4.6.1. Correlación cruzada

La Figura 4.21 muestra la correlación cruzada entre los valores objetivo y las predicciones. Se observa una correlación máxima de 0.331 en $lag = 0$, lo que indica que, a diferencia de otros modelos evaluados, las predicciones del TFT no presentan rezago temporal significativo respecto de la serie real.

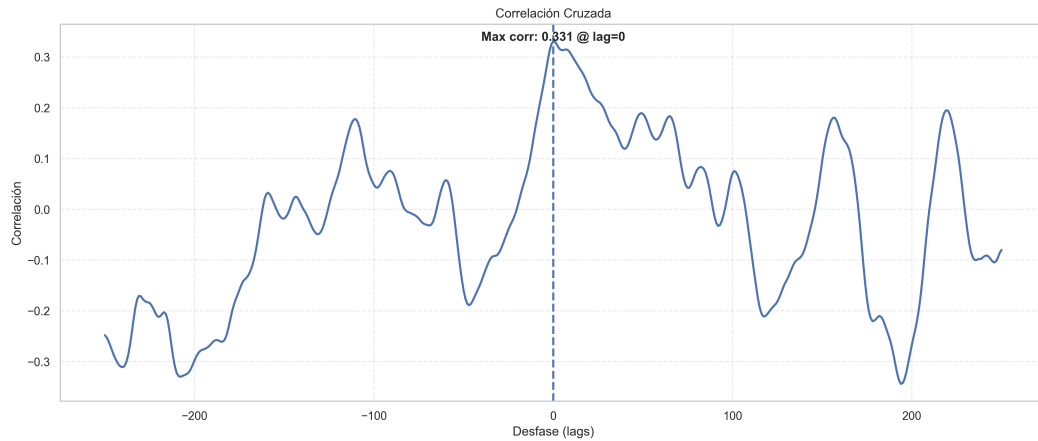


Fig. 4.21: Correlación cruzada para la razón Fe/SiO₂ utilizando el modelo TFT.

La Figura 4.22 muestra la correlación cruzada después de un proceso de interpolación de 29 puntos entre cada valor, logrando así una aproximación más exacta de los minutos de desfase entre los valores objetivo y las predicciones. Se observa un desfase temporal de aproximadamente de 11 minutos y máxima correlación de 0.334

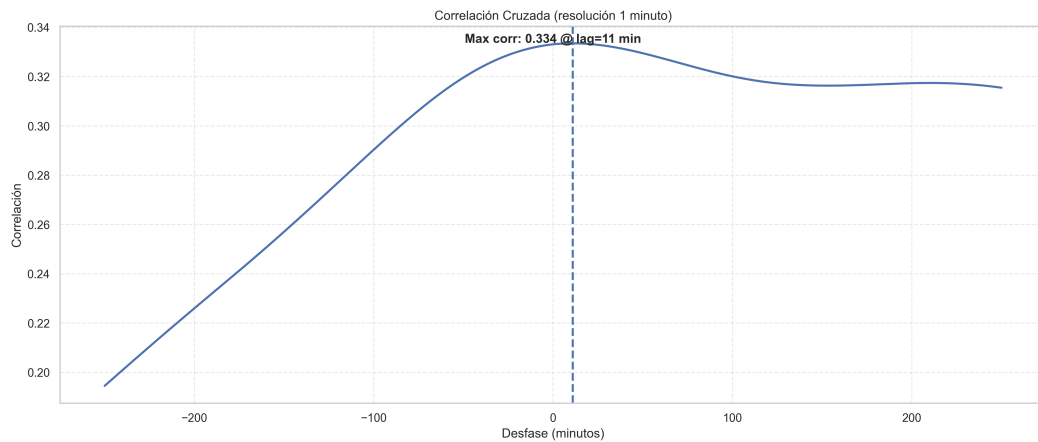


Fig. 4.22: Correlación cruzada interpolada para la razón Fe/SiO₂ utilizando el modelo TFT.

4.6.2. Diagrama de dispersión

La Figura 4.23 presenta el diagrama de dispersión entre valores objetivo y predichos en el conjunto de prueba. Se observa una correlación de Pearson baja ($r = 0,331$) y un valor de NMI de 0.687. Aunque estos valores reflejan que el modelo logra capturar parte de la relación entre las variables, también evidencian dispersión significativa alrededor de la recta identidad.

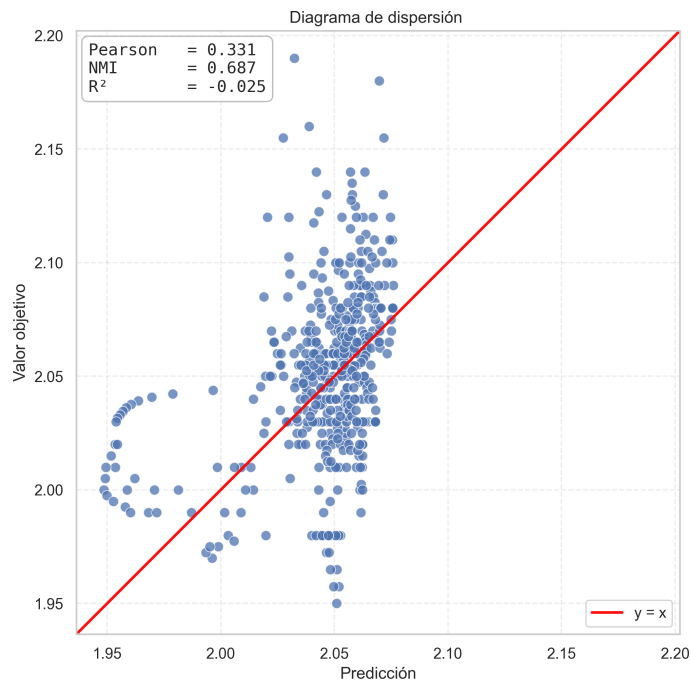


Fig. 4.23: Diagrama de dispersión entre valores objetivo y predichos de la razón Fe/SiO₂ utilizando el modelo TFT.

4.6.3. Serie temporal

Finalmente, en la Figura 4.24 se muestra la comparación entre la serie temporal de los valores objetivo de la razón Fe/SiO₂ y las predicciones del modelo. El TFT logra reproducir parcialmente la tendencia de la serie, con un mejor alineamiento temporal (sin rezagos aparentes), aunque con dificultades para capturar variaciones abruptas.

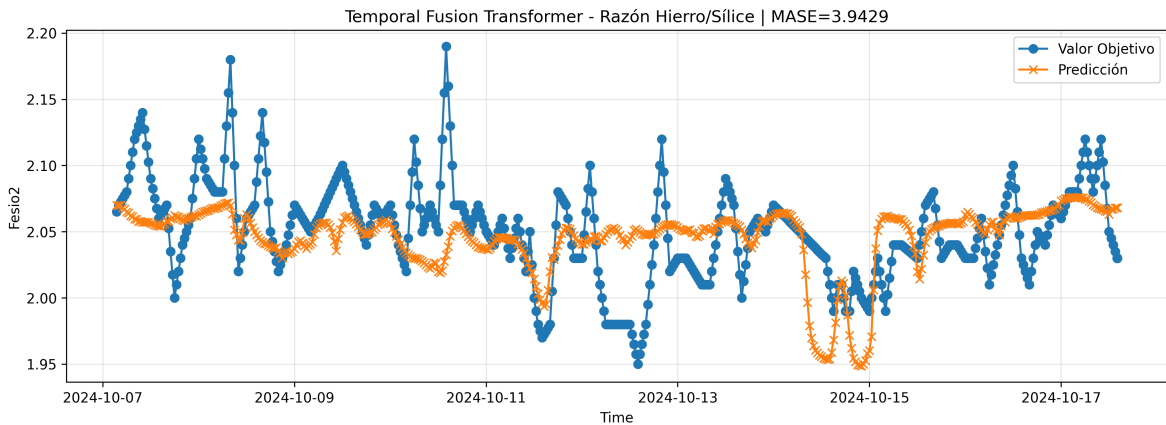


Fig. 4.24: Comparación entre valores objetivo y predicciones de la razón Fe/SiO_2 en el conjunto de prueba utilizando el modelo TFT.

4.7. Discusión de resultados

En la ley del cobre Cu , el **TFT** ofrece el mejor compromiso entre error y asociación: alcanza el menor MASE (3.9073), el único R^2 positivo (0.262) y las mayores dependencias lineal y no lineal ($r = 0,593$, $NMI = 0,831$), superando a la regresión lineal y al LSTM, cuyos MASE rondan 6 con R^2 negativos. Aunque estos últimos capturan parte de la forma de la señal presentan errores de escala y, sobre todo, de sincronía: la correlación cruzada muestra un rezago cercano a -90 minutos, también observable en el TFT para Cu . En sesgo, el TFT se mantiene próximo a cero ($FB = 0,891\%$), mientras que la regresión y el LSTM tienden a subestimar ($FB -3,5\%$).

Para la razón Fe/SiO_2 , el problema es más exigente y ningún modelo explica adecuadamente la varianza de prueba. Aun así, el **TFT** vuelve a ser el más competitivo en error y estructura (MASE = 3,943; $NMI = 0,687$), seguido de la **regresión lineal** (MASE = 4,0818; $R^2 = 0,0458$). La correlación cruzada muestra rezagos ~ -90 minutos en regresión y LSTM, mientras que el TFT se alinea mejor (pico cercano a $lag = 0$ y $\sim +11$ minutos tras interpolación), indicando que el principal déficit no es de nivel (FB entre $-0,674\%$ y $0,077\%$), sino de *temporalidad y forma*: los modelos siguen la tendencia, pero no predicen con precisión picos y caídas.

4.8. Resultados conjunto endógeno

En esta sección se presentan los resultados de los modelos entrenados bajo el enfoque endógeno, es decir, utilizando el pasado de las variables objetivo (Cu y Fe/SiO_2).

4.8.1. Ley de cobre

Para el pronóstico de la ley de cobre (Cu) el mejor desempeño se obtuvo con un modelo LSTM, alcanzando valores de $MASE = 0,459$, $MAE = 0,0797$, y $R^2 = 0,984$, con una correlación de Pearson de 0.992. En comparación, la regresión lineal mostró un desempeño de ($MASE = 0,545$, $R^2 = 0,982$), mientras que el modelo TFT presentó mayores dificultades de generalización ($MASE = 1,20$, $R^2 = 0,925$).

En la Tabla 4.7 se resumen las métricas de los modelos evaluados.

Tabla 4.7: Métricas de desempeño en el conjunto endógeno para la ley de cobre.

Modelo	MAE	R^2	MASE	r (Pearson)
Regresión Ridge	0.095	0.982	0.545	0.992
Híbrido LSTM	0.080	0.984	0.459	0.992
TFT	0.209	0.925	1.200	0.962

Las Figuras 4.25–4.27 muestran ejemplos representativos de las evaluaciones gráficas: correlación cruzada, diagrama de dispersión y comparación de series temporales del modelo LSTM. Se aprecia que el modelo LSTM logra la mayor proximidad a la recta identidad y un ajuste más robusto frente a las oscilaciones de la serie.

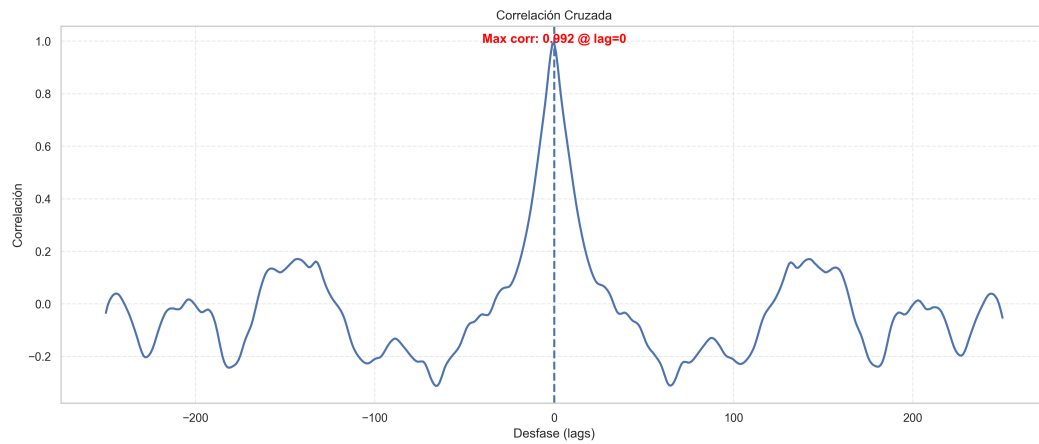


Fig. 4.25: Correlación cruzada entre valores objetivo y predicciones para la ley de cobre (Modelo LSTM).

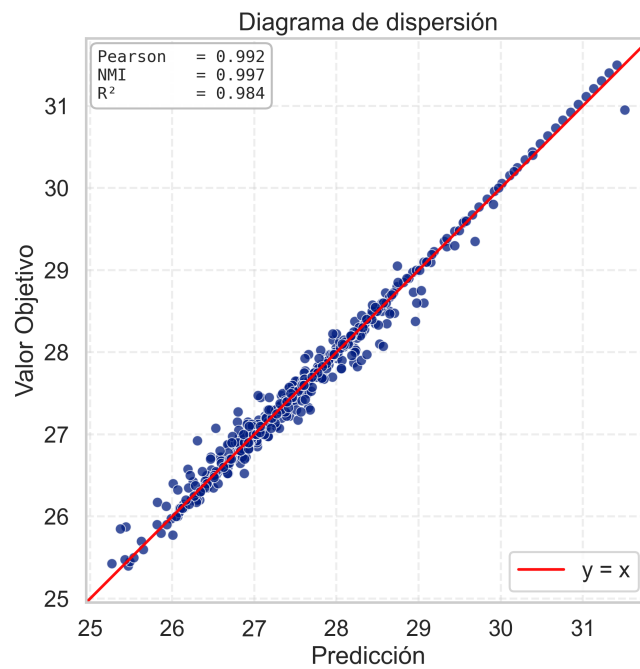


Fig. 4.26: Diagrama de dispersión de valores objetivo vs. predicciones de la ley de cobre (Modelo LSTM).

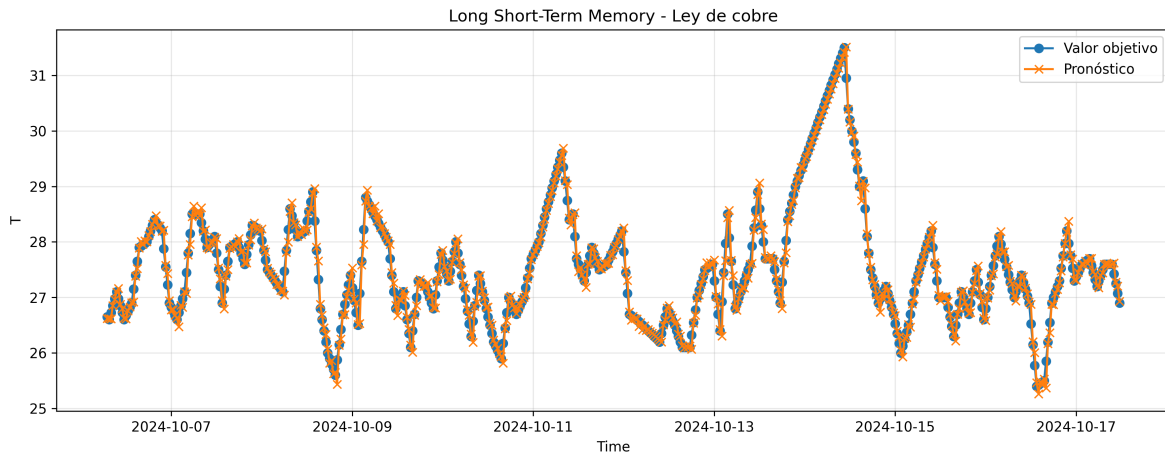


Fig. 4.27: Serie temporal entre valores objetivo y predicciones de la ley de cobre (Modelo LSTM).

4.8.2. Razón hierro–sílice

En el caso de la razón hierro–sílice (Fe/SiO_2), los resultados muestran mayor complejidad debido a la elevada variabilidad de la serie. El modelo LSTM fue nuevamente el que entregó el mejor equilibrio entre precisión y generalización, alcanzando $MASE = 0,645$, $MAE = 0,0045$ y $R^2 = 0,955$, con correlación de Pearson de 0.978. La regresión lineal robusta obtuvo $MASE = 0,702$, $R^2 = 0,939$, mientras que el TFT presentó menor capacidad predictiva ($MASE = 1,83$, $R^2 = 0,755$).

La Tabla 4.8 resume los resultados obtenidos.

Tabla 4.8: Métricas de desempeño en el conjunto endógeno para la razón hierro–sílice.

Modelo	MAE	R^2	MASE	r (Pearson)
Regresión Ridge	0.0048	0.939	0.702	0.974
Híbrido LSTM	0.0045	0.955	0.645	0.978
TFT	0.0128	0.755	1.835	0.870

Las Figuras 4.28–4.30 muestran las evaluaciones gráficas. Al igual que en el caso del cobre, la LSTM se acerca más a la dinámica de la serie.

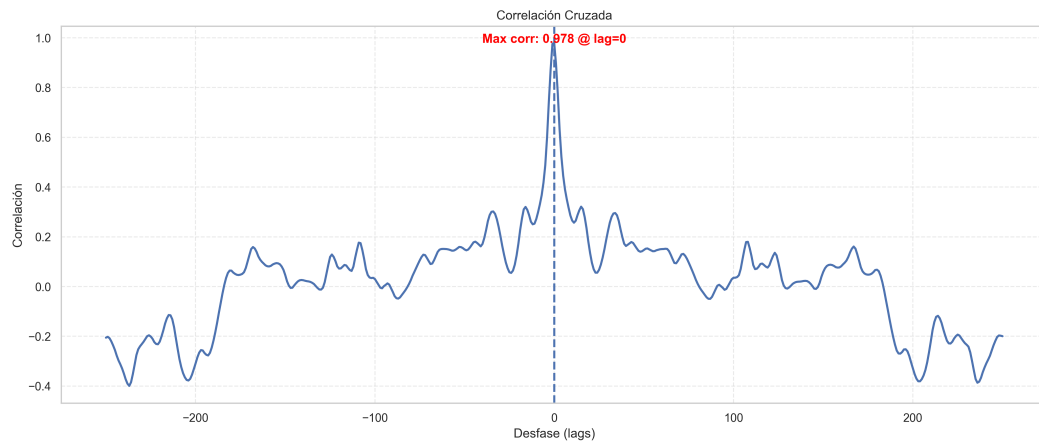


Fig. 4.28: Correlación cruzada entre valores objetivo y predicciones para la razón hierro-sílice.

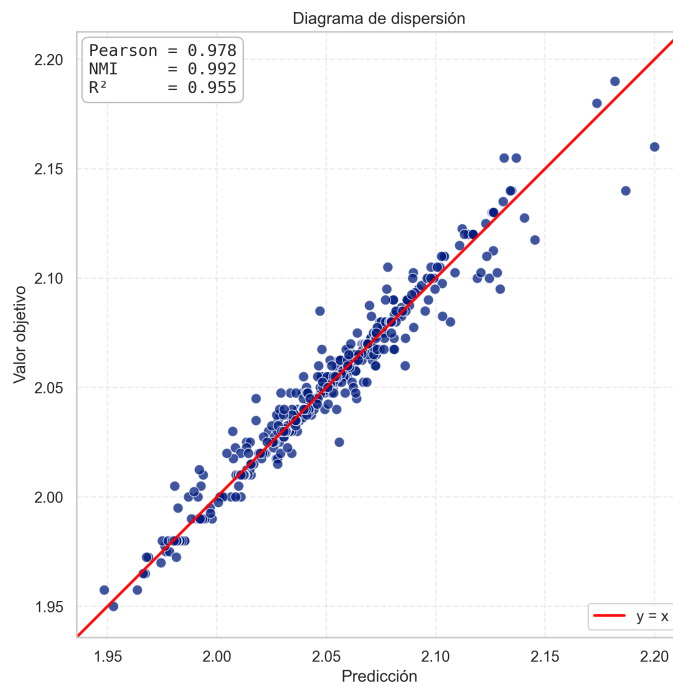


Fig. 4.29: Diagrama de dispersión de valores objetivo vs. predichos de la razón hierro-sílice.

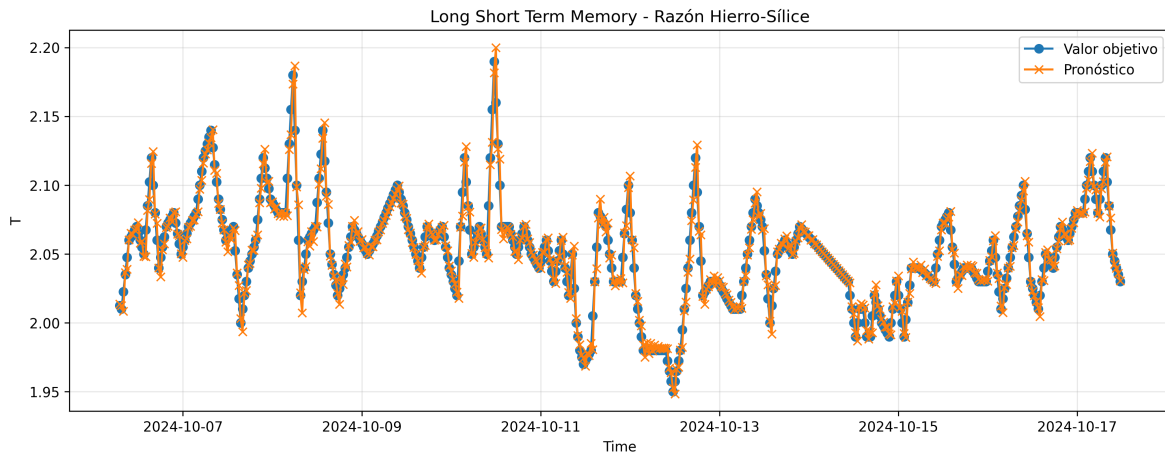


Fig. 4.30: Comparación temporal entre valores objetivo y predicciones de la razón hierro-sílice.

4.8.3. Discusión de resultados modelo endógeno

En el enfoque endógeno para la ley de Cu , los resultados evidencian que la **LSTM** ofrece el mejor equilibrio entre precisión y capacidad de generalización, con $MASE = 0,459$, $MAE = 0,0797$ y $R^2 = 0,984$, manteniendo además una asociación casi perfecta con la serie real ($r = 0,992$). La regresión lineal se aproxima en R^2 pero con mayor error escalado ($MASE = 0,545$), mientras que el **TFT** muestra una generalización peor en comparación. Las evaluaciones gráficas confirman que la LSTM se adhiere con mayor fidelidad a la recta identidad y replica mejor la dinámica de la serie.

Para Fe/SiO_2 , se observa un patrón con mayor complejidad de la señal: la **LSTM** vuelve a liderar con $MASE = 0,645$, $MAE = 0,0045$ y $R^2 = 0,955$ ($r = 0,978$), seguida de la regresión lineal ($MASE = 0,702$, $R^2 = 0,939$) y con un **TFT** de menor desempeño ($MASE = 1,835$, $R^2 = 0,755$) (Tabla 4.8). Los gráficos muestran que la LSTM captura con mayor estabilidad los ciclos y cambios de nivel, mientras que el TFT suaviza en exceso y pierde amplitud ante variaciones más abruptas. En conjunto, estos hallazgos refuerzan que, bajo un esquema endógeno, la inclusión del pasado del objetivo y la capacidad de memoria de LSTM son determinantes para reducir el error y mejorar la sincronía de las predicciones.

4.9. Resultados modelo suavizado

En esta sección se presentan los resultados de los modelos entrenados mediante un conjunto de datos previamente suavizado, es decir, utilizando una EMA de 3 periodos simétrica.

4.9.1. Ley de cobre

En el caso de la ley de cobre (*Cu*), el mejor resultado se obtuvo con el modelo *Temporal Fusion Transformer* (TFT), alcanzando $MASE = 5,038$, $MAE = 0,589$ y $R^2 = 0,304$, con correlación de Pearson $r = 0,570$. La *regresión lineal* y el modelo *LSTM* presentaron un deterioro de la capacidad predictiva, con R^2 negativos y aumentos del MASE en comparación a los modelos sin suavizado. En particular, la regresión alcanzó $MASE = 18,864$ y $R^2 = -1,531$, mientras que la LSTM tuvo un $MASE = 8,803$ y $R^2 = -0,803$.

Tabla 4.9: Métricas de desempeño (suavizado) para la ley de cobre.

Modelo	MAE	R^2	MASE	r (Pearson)
Regresión Ridge	1.071	-1.531	18.864	0.597
Híbrido LSTM	1.030	-0.803	8.803	0.516
TFT	0.589	0.304	5.038	0.570

En las Figuras 4.31–4.32–4.33 se observa mayor error en validación y prueba, consistente con la pérdida de precisión reportada.

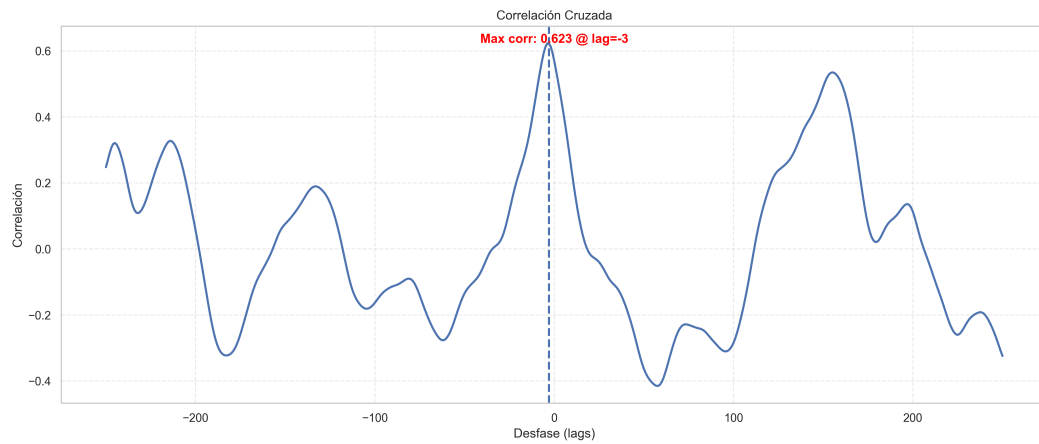


Fig. 4.31: Correlación cruzada para la ley de cobre modelo TFT suavizado.

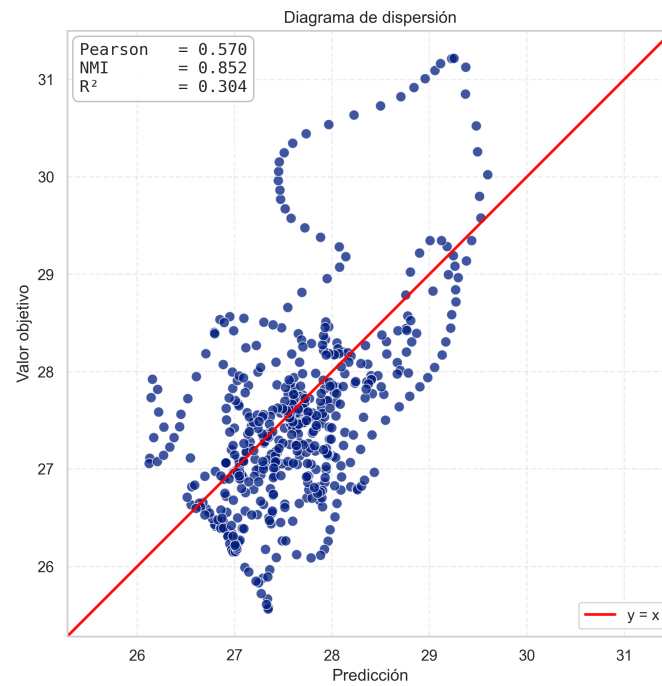


Fig. 4.32: Diagrama de dispersión para la ley de cobre modelo TFT suavizado

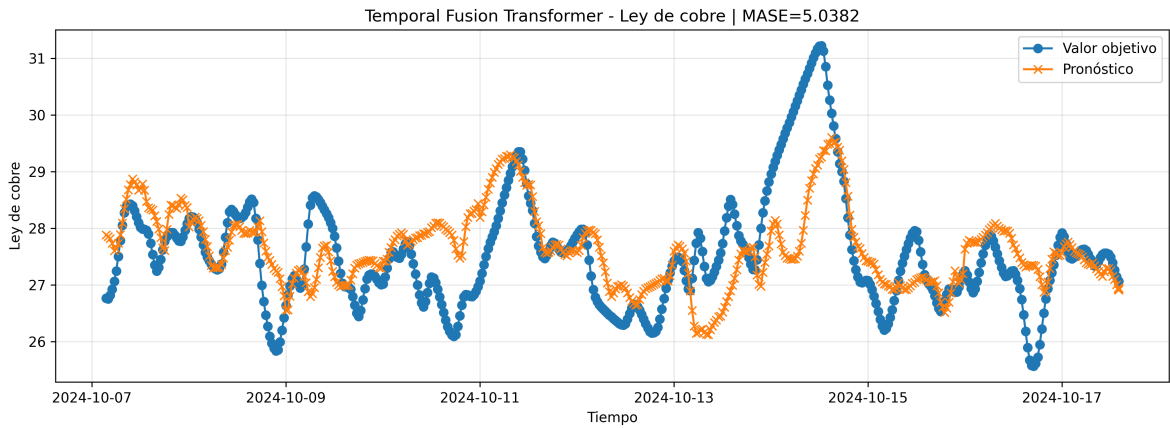


Fig. 4.33: Serie temporal para la ley de cobre modelo TFT suavizado

4.9.2. Razón hierro–sílice

Para la razón Fe/SiO_2 , el mejor desempeño también se obtuvo con el TFT, con $MASE = 5,189$, $MAE = 0,025$ y $R^2 = 0,024$ ($r = 0,391$). La *regresión lineal* y el *LSTM* mostraron $R^2 < 0$ y MASE elevados.

Tabla 4.10: Métricas de desempeño (suavizado) para la razón hierro–sílice.

Modelo	MAE	R^2	MASE	r (Pearson)
Regresión Ridge	0.027	-0.624	13.155	0.277
Híbrido LSTM	0.045	-1.827	9.411	0.052
TFT	0.025	0.024	5.189	0.391

Las Figuras 4.34–4.35–4.36 muestran las evaluaciones gráficas para Fe/SiO_2 utilizando el modelo TFT suavizado.

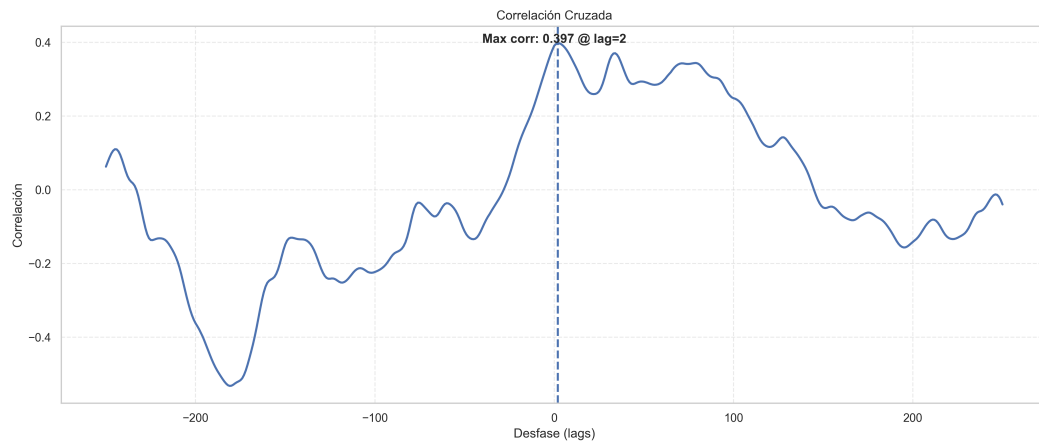


Fig. 4.34: Correlación cruzada para la razón Fe/SiO_2 modelo TFT suavizado.

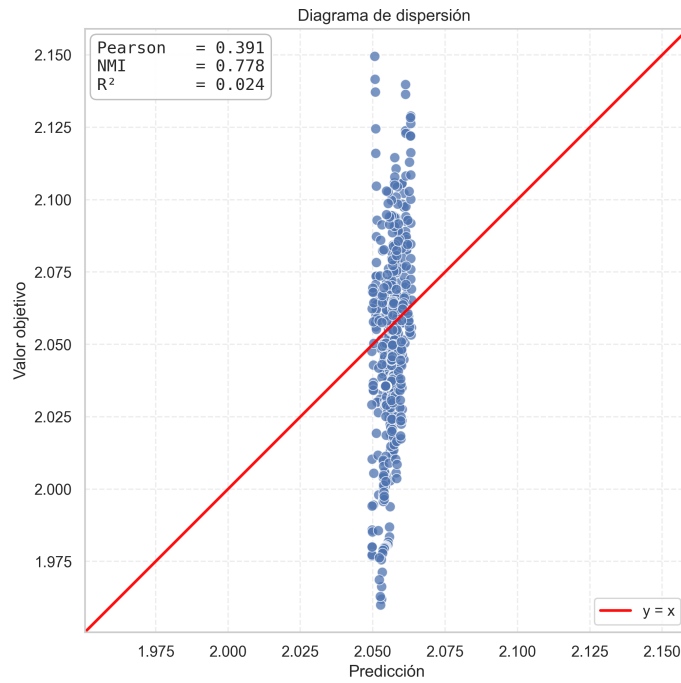


Fig. 4.35: Diagrama de dispersión para la razón Fe/SiO_2 modelo TFT suavizado

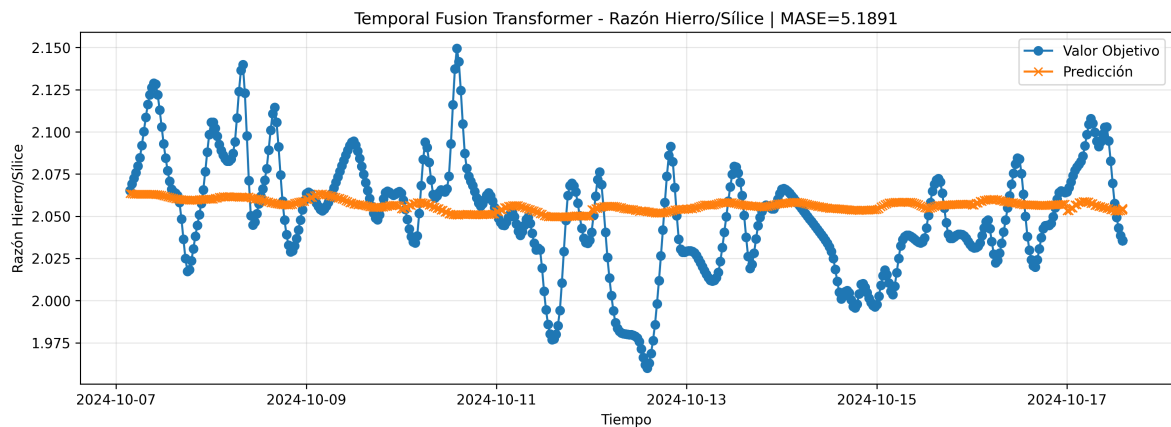


Fig. 4.36: Serie temporal para la razón Fe/SiO_2 modelo TFT suavizado

4.9.3. Discusión de resultados modelo suavizado

Los resultados utilizando el *suavizado simétrico* (EMA de 3 periodos) no mejoró el desempeño predictivo: se observó una **pérdida de precisión** y de **capacidad de generalización** en los modelos para ambas variables objetivo. En consecuencia, la hipótesis de que “variables de proceso con variaciones más suaves serían más fáciles de estimar” no se sostiene en este estudio; bajo la metodología aplicada en esta memoria de título el suavizado perjudicó las predicciones.

Capítulo 5

Conclusiones

El presente trabajo tuvo como objetivo desarrollar una herramienta de monitoreo no invasiva para estimar en tiempo real el contenido de cobre y la razón hierro-sílice en la escoria de un horno de conversión flash, utilizando datos de proceso y modelos de *Deep Learning* implementados en Python. A partir del trabajo realizado, se obtuvieron los siguientes hallazgos principales:

- **Viabilidad del monitoreo no invasivo:** Se demostró que es posible estimar las variables de interés (Cu y Fe/SiO_2) a partir de datos de proceso, reduciendo la dependencia de muestreos fisicoquímicos invasivos. Los modelos de tipo endógeno poseen un elevado error en las predicciones, pese a ello, se concluye que el presente trabajo se puede implementar como una herramienta de monitoreo para seguir la tendencia de las variables objetivo.
- **Comparación de conjuntos de datos:** El uso del conjunto endógeno, que utiliza el pasado variables objetivo, mejoró de manera consistente el desempeño de los modelos respecto al conjunto exógeno. No obstante, el conjunto exógeno conserva un valor práctico al permitir generar predicciones sin necesidad de mediciones previas de las variable objetivo, lo que favorece su aplicación en escenarios de monitoreo en línea.
- **Desempeño de los modelos:**
 - La **regresión lineal** entregó un punto de partida sólido, evidenciando que buena parte de la dinámica puede explicarse mediante relaciones lineales y

variables rezagadas, aunque con limitaciones en escenarios de alta volatilidad.

- El **modelo LSTM** resultó ser el más robusto, alcanzando los mejores valores de MASE y coeficientes de correlación, tanto en cobre como en razón hierro-sílice, lo que confirma su capacidad para modelar dependencias no lineales y patrones complejos en el conjunto de datos endógeno.
- El **TFT** mostró un comportamiento diferenciado dependiendo del conjunto de datos utilizado. En el caso exógeno, el modelo logró capturar de manera más efectiva las relaciones no lineales entre las variables de proceso y las variables objetivo, alcanzando métricas mejores frente a los otros enfoques. Esto puede atribuirse a que, al no incluir el pasado de las variables objetivo, el TFT aprovecha con mayor ventaja sus mecanismos de atención y compuertas para identificar patrones complejos. En contraste, en el conjunto endógeno el modelo no logró superar a la LSTM y presentó mayores dificultades de generalización.
- **Efecto del suavizado de las series:** Si bien, en principio, al suavizar una serie temporal esperamos que los modelos predigan mejor al reducir la variabilidad y los picos/caídas abruptos, en el presente trabajo esto no se verificó bajo la metodología aplicada, todos los enfoques evaluados se comportaron peor. Una explicación posible es que el suavizado elimina justamente parte de la información de alta frecuencia que se utiliza como datos operacionales relevantes (cambios de régimen, perturbaciones), reduciendo el contraste temporal que los modelos utilizan para generar pronósticos. Desde la perspectiva del sesgo-varianza, el suavizado reduce la varianza pero aumenta el sesgo, favoreciendo el subajuste cuando predominan patrones no lineales de corta duración. Por lo tanto, en nuestro caso el suavizado no aporta mejoras en la precisión y, de considerarse su uso, debería plantearse un *pipeline* específico por arquitectura (filtros específicos, realineación de rezagos, ingeniería de variables y re-calibración de hiperparámetros) antes de reentrenar los modelos.
- **Métrica principal:** El uso de MASE como métrica de referencia permitió evaluar los modelos en un marco comparable e independiente de la escala de las variables, facilitando la interpretación y comparación entre distintos

enfoques.

- **Impacto industrial:** Los resultados confirman que los modelos propuestos pueden integrarse en sistemas de monitoreo en línea, contribuyendo a mejorar la estabilidad del proceso, reducir la variabilidad operativa y optimizar la toma de decisiones en el horno de conversión flash.

En conclusión, la investigación valida la factibilidad de construir un sistema de monitoreo no invasivo basado en datos de proceso y modelos de *Deep Learning*. Dicho sistema representa un avance hacia la digitalización y optimización de la metalurgia extractiva, aportando tanto desde el punto de vista operativo como estratégico a la industria del cobre.

5.1. Consideraciones en aplicación a la industria

La implementación de los modelos desarrollados en este trabajo requiere evaluar no sólo su precisión estadística, sino también su factibilidad de aplicación en un entorno industrial real. En este sentido, se analizan tres aspectos centrales: los costes computacionales asociados a cada modelo, la aceptabilidad de sus tasas de error en un contexto productivo, y los retrasos observados en las predicciones (delays).

5.1.1. Costes computacionales

El modelo de *regresión lineal* constituye una alternativa de muy bajo coste computacional, pudiendo ejecutarse en hardware sencillo como un PLC o incluso microcontroladores, lo que lo hace atractivo para tareas de monitoreo en tiempo real con recursos limitados. Por su parte, la arquitectura *LSTM* requiere de un mayor poder de cómputo, siendo recomendable ejecutarla en un computador industrial o un servidor local con GPU modesta. Finalmente, el modelo *TFT* demanda un alto rendimiento computacional, especialmente en su etapa de entrenamiento, siendo poco viable en un PLC y más apropiado para estaciones de trabajo de alto desempeño o servidores especializados.

5.1.2. Tasas de error

Las métricas obtenidas muestran que los modelos presentan errores relativamente bajos, especialmente en el caso del conjunto endógeno con LSTM, donde los valores de MASE y MAE son suficientemente reducidos como para ser considerados aceptables en aplicaciones industriales. Sin embargo, en escenarios exógenos, los errores aumentan, lo que implica que, si bien los resultados son útiles como referencia en línea, no reemplazan completamente los análisis químicos de laboratorio, sino que los complementan al ofrecer un monitoreo continuo de la tendencia de las variables objetivo.

5.1.3. Retrasos en las predicciones

Un aspecto relevante identificado corresponde a los retrasos (lags) observados en ciertas predicciones. Estos no se originan en la arquitectura del modelo, sino en la naturaleza del conjunto de datos, donde la interpolación de largos intervalos sin mediciones de ley de cobre introdujo un desfase aparente en las curvas de correlación cruzada. Aun así, los modelos mantienen valor práctico, ya que permiten detectar tendencias y variaciones significativas en el proceso, funcionando como una herramienta de alerta temprana que complementa el muestreo químico tradicional.

En síntesis, cada modelo ofrece un balance diferente entre coste, precisión y desfase temporal. La regresión lineal es la más simple y de bajo costo, la LSTM equilibra precisión y recursos, y el TFT apunta a escenarios donde se disponga de hardware avanzado. Esta evaluación permite dimensionar de manera realista su potencial integración a sistemas industriales de monitoreo y control.

5.2. Trabajos futuros

Si bien los resultados obtenidos validan la factibilidad de desarrollar una herramienta de monitoreo no invasiva basada en datos de proceso y modelos de *Deep Learning*, existen diversas oportunidades de mejora y líneas de investigación que podrían potenciar este trabajo en el futuro:

- **Optimización de arquitecturas avanzadas:** Explorar variantes del *Temporal Fusion Transformer* (TFT) con técnicas de regularización adicionales, mayor ajuste de hiperparámetros y estrategias de calibración robusta, con el fin de superar las limitaciones observadas en la generalización, especialmente en la razón hierro-sílice.
- **Modelos híbridos más sofisticados:** Ampliar el enfoque híbrido combinando componentes lineales, LSTM y mecanismos de atención, lo que permitiría integrar la fortaleza de diferentes paradigmas de modelado y mejorar la capacidad predictiva bajo condiciones de alta variabilidad.
- **Integración de variables adicionales:** Incorporar nuevas fuentes de información del proceso (por ejemplo, sensores complementarios o parámetros operacionales no considerados en esta memoria de título, con el fin de enriquecer el espacio de predictores y capturar efectos aún no modelados.
- **Implementación en línea:** Desplegar los modelos en entornos de monitoreo en tiempo real, integrándolos a sistemas industriales, para evaluar su desempeño bajo condiciones operacionales dinámicas y en interacción con los operadores del proceso.
- **Generalización a otros procesos metalúrgicos:** Extender la metodología propuesta a diferentes etapas de la cadena de producción de cobre o a otros procesos pirometalúrgicos, de manera de evaluar la escalabilidad del enfoque y sus beneficios a nivel industrial.

En definitiva, el trabajo presente sienta una base sólida para la construcción de sistemas de monitoreo inteligentes en la metalurgia extractiva. La combinación de datos de proceso y técnicas de *Deep Learning* abre un camino hacia la operación optimizada de los hornos de conversión flash, con proyección de aplicación en otros procesos metalúrgicos de relevancia para la industria minera.

Bibliografía

- [1] “Historia metalúrgica chilena - introducción,” [Online]. Available: <https://construmine.webnode.cl/1/historia-metalurgica-chilena-introduccion/>. Accessed: Oct. 15, 2024.
- [2] “Chile y el cobre: una historia entrelazada,” [Online]. Available: <https://www.codelco.com/chile-y-el-cobre-una-historia-entrelazada>. Accessed: Oct. 15, 2024.
- [3] L. Arias, S. Torres, C. Toro, E. Balladares, R. Parra, C. Loeza, C. Villagrán, and P. Coelho, “Flash smelting copper concentrates spectral emission measurements,” *Sensors (Switzerland)*, vol. 18, no. 7, 2018.
- [4] M. Joseph, *Modern Time Series Forecasting with Python: Explore Industry-Ready Time Series Forecasting Using Modern Machine Learning and Deep Learning*, 1st ed. Birmingham, U.K.: Packt Publishing, 2022.
- [5] B. A. R. Morales, “Machine learning, sensores ópticos en hornos de fusión de baño,” Master’s thesis, Universidad de Concepción, 2022.
- [6] J. L. T. Sanhueza, “Cálculo de temperatura y emisividad en datos de espectrometría de procesos pirometalúrgicos por método inverso,” Master’s thesis, Universidad de Concepción, 2022.
- [7] J. Torres-Sanhueza, F. Rivas, F. G. Pérez, R. Parra, S. N. Torres, and S. E. Godoy, “Radiometric methods for high temperature flame estimation: Robustness analysis and application to an industrial flash smelting furnace,” *IEEE Trans. Instrum. Meas.*, vol. 72, no. 8, pp. 1–20, Aug. 2023.

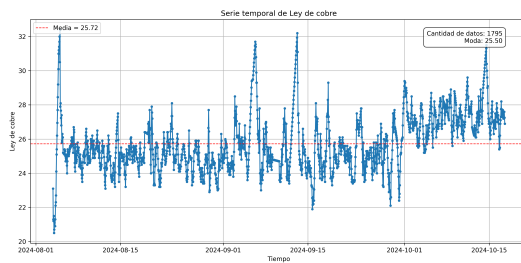
-
- [8] Y. A. Çengel, *Transferencia de Calor y Masa: Un Enfoque Práctico*, 3rd ed. México: McGraw-Hill Interamericana, 2007.
- [9] R. N. B. Ramírez, “Caracterización espectral de la combustión flash a nivel de laboratorio de especies mineralógicas individuales y un concentrado real,” Master’s thesis, Universidad de Concepción, 2018.
- [10] Codelco Educa, “Glosario de conceptos mineros: L (lamas, latón, levantamiento estratigráfico, ley de cobre, ...),” [Online]. Available: <https://www.codelcoeduca.cl/codelcoeduca/site/edic/base/port/glosario.html?letra=l>. Accessed: Jul. 15, 2025.
- [11] B. Wang, H. Yang, Z. Jin, Z. Liu, and M. Zou, “Effect of fe/sio₂ ratio and fe₂o₃ on the viscosity and slag structure of copper-smelting slags,” *Metals*, vol. 12, no. 1, p. 24, 2022, [Online]. Available: <https://doi.org/10.3390/met12010024>. Accessed: Jul. 15, 2025.
- [12] F. P. V., “Predicción de series de tiempo,” departamento de Ingeniería Eléctrica, Universidad de Concepción. Presentación de la asignatura Aprendizaje de Máquinas, 2024.
- [13] scikit-learn developers, “Cross-validation: Evaluating estimator performance,” User Guide, *scikit-learn: Machine Learning in Python*, online; accessed August 3, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/cross_validation.html
- [14] —, “sklearn.model_selection.TimeSeriesSplit,” API documentation, *scikit-learn: Machine Learning in Python*, online; accessed August 3, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html
- [15] —, “sklearn.model_selection.GridSearchCV,” API documentation, *scikit-learn: Machine Learning in Python*, online; accessed August 3, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
- [16] P. Schober, *Coeficiente de correlación lineal de Pearson*. Documento académico, s.f., material de estadística sobre el coeficiente de correlación lineal de Pearson.

-
- [17] P. Laarne, M. A. Zaidan, and T. Nieminen, “ennemi: Non-linear correlation detection with mutual information,” *SoftwareX*, vol. 14, p. 100686, 2021.
- [18] GeeksforGeeks, “Cross-Correlation Analysis in Python,” Tutorial, *GeeksforGeeks*, online; accessed August 4, 2025. [Online]. Available: <https://www.geeksforgeeks.org/machine-learning/cross-correlation-analysis-in-python/>
- [19] scikit-learn developers, “sklearn.linear_model.LinearRegression,” API documentation, *scikit-learn: Machine Learning in Python*, online; accessed August 20, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
- [20] GeeksforGeeks, “Introduction to Long Short-Term Memory (LSTM),” Tutorial, *GeeksforGeeks*, online; accessed August 20, 2025. [Online]. Available: <https://www.geeksforgeeks.org/deep-learning/deep-learning-introduction-to-long-short-term-memory/>
- [21] Dataness AI, “Understanding Temporal Fusion Transformer,” Blog post, *Medium*, online; accessed August 20, 2025. [Online]. Available: <https://medium.com/dataness-ai/understanding-temporal-fusion-transformer-9a7a4fcde74b>
- [22] P. Mahamulkar. (2023, December 17) Feature scaling using standardization, normalization and robust scaling. <https://medium.com/@prasadmahamulkar/feature-scaling-using-standardization-normalization-and-robust-scaling-cb333f943f70>. Medium Article.
- [23] T. von Oertzen and S. M. Boker, “Time delay embedding increases estimation precision of models of intraindividual variability,” *Psychometrika*, vol. 75, no. 1, pp. 158–175, 2010.
- [24] W. G. Davenport, M. King, M. Schlesinger, and A. K. Biswas, *Extractive Metallurgy of Copper*, 4th ed. Oxford, UK: Elsevier Science Ltd, 2002.
- [25] F. Mattioli, “Enhancing urban air quality predictions with temporal fusion transformers: A methodological evaluation,” Master Thesis, Universidad Politécnica de Madrid, Madrid, Spain, July 2024, supervised by Aurora Pérez Pérez.

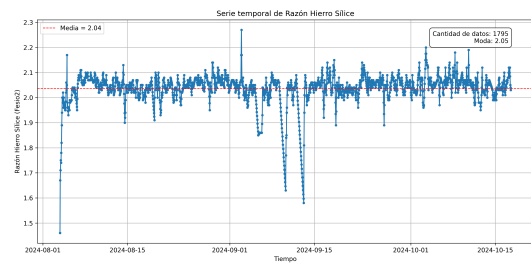
Appendices

Apéndice A

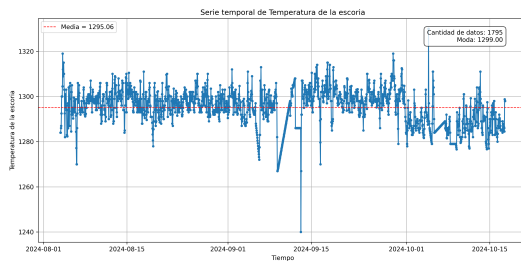
ANEXO: Figuras



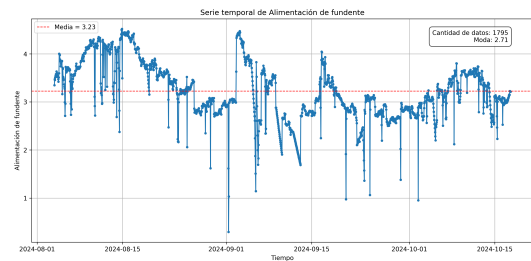
(a) Serie temporal de Ley de cobre.



(b) Serie temporal de Razón Hierro Sílice.

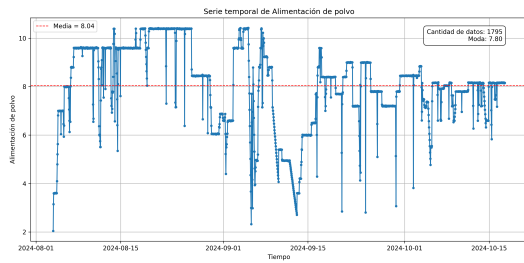


(c) Serie temporal de Temperatura de la escoria.

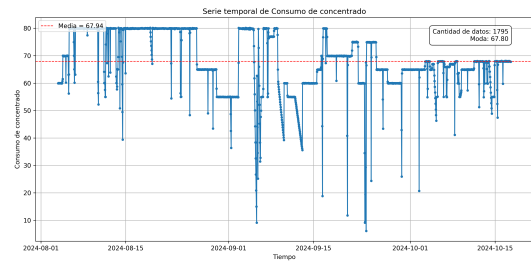


(d) Serie temporal de Alimentación de fundente.

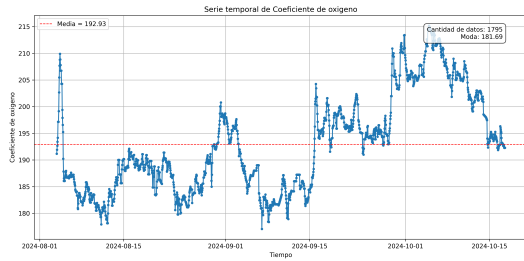
Fig. A.1: Series temporales con frecuencia de 1 hora.



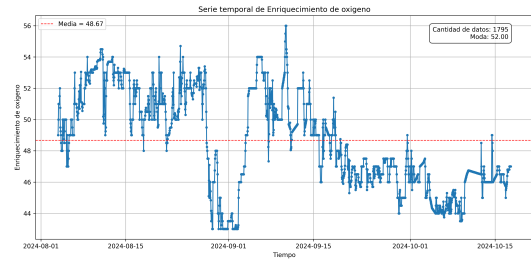
(a) Serie temporal de Alimentación de polvo.



(b) Serie temporal de Consumo de concentrado.

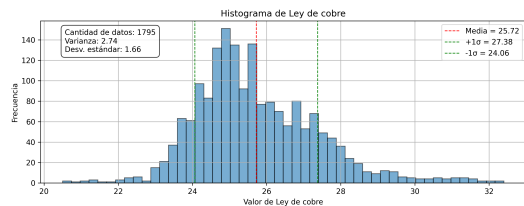


(c) Serie temporal de Coeficiente de oxígeno.

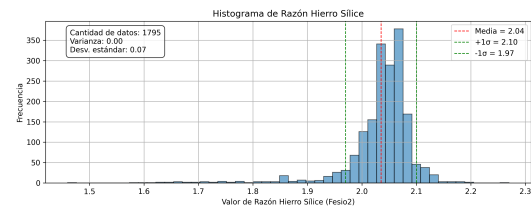


(d) Serie temporal de Enriquecimiento de oxígeno.

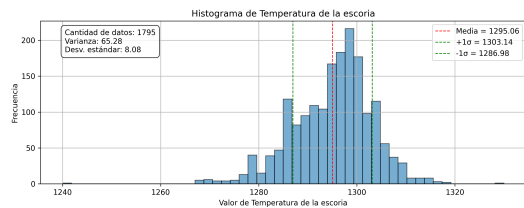
Fig. A.2: Series temporales con frecuencia de 1 hora.



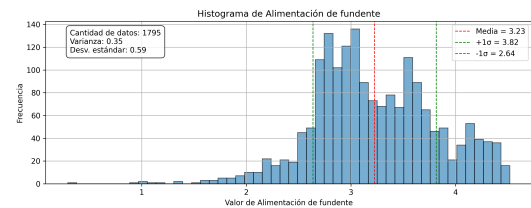
(a) Histograma de Ley de cobre.



(b) Histograma de Razón Hierro Sílice.

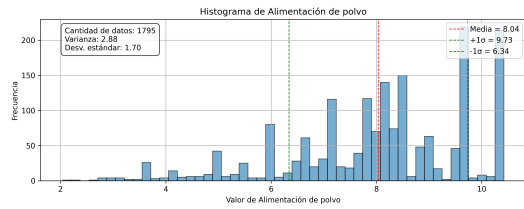


(c) Histograma de Temperatura de la escoria.



(d) Histograma de Alimentación de fundente.

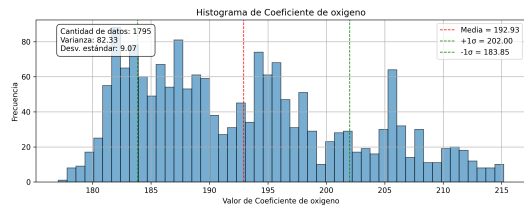
Fig. A.3: Histogramas con frecuencia de 1 hora.



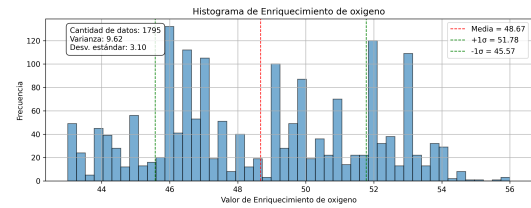
(a) Histograma de Alimentación de polvo.



(b) Histograma de Consumo de concentrado.

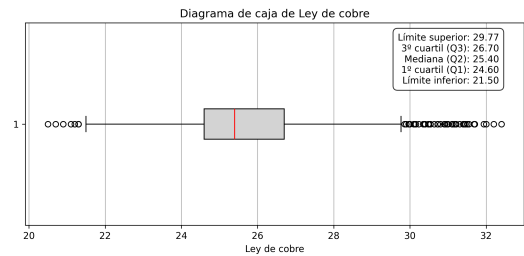


(c) Histograma de Coeficiente de oxígeno.

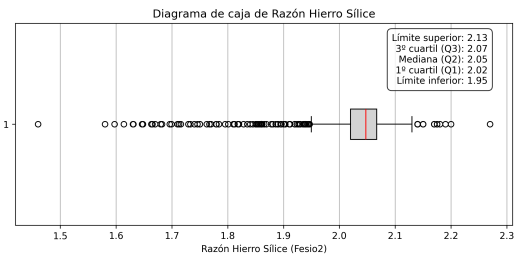


(d) Histograma de Enriquecimiento de oxígeno.

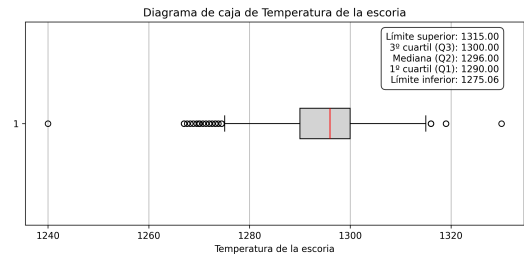
Fig. A.4: Histogramas con frecuencia de 1 hora.



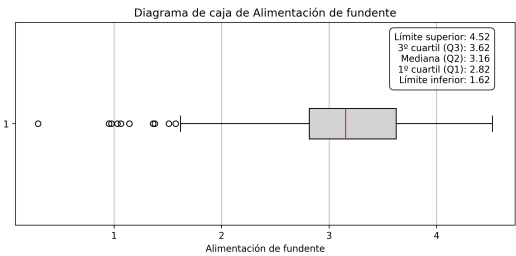
(a) Diagrama de caja de Ley de cobre.



(b) Diagrama de caja de Razón Hierro Sílice.

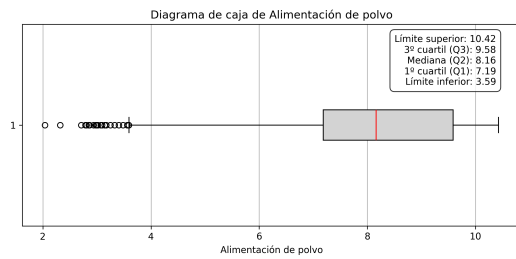


(c) Diagrama de caja de Temperatura de la escoria.

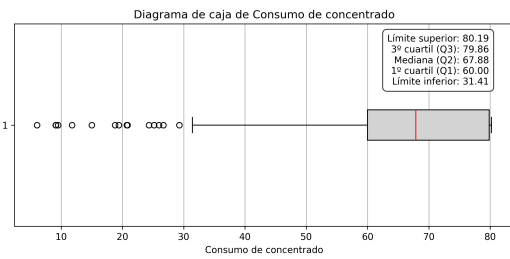


(d) Diagrama de caja de Alimentación de fundente.

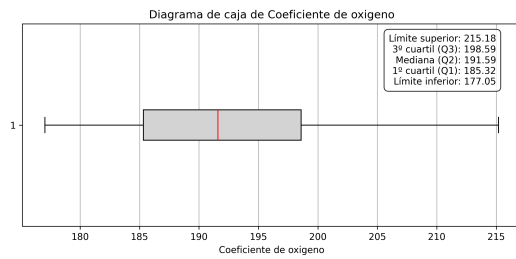
Fig. A.5: Diagramas de caja con frecuencia de 1 hora.



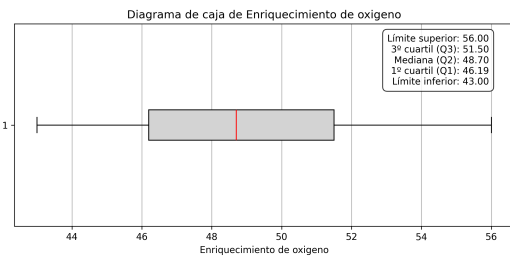
(a) Diagrama de caja de Alimentación de polvo.



(b) Diagrama de caja de Consumo de concentrado.



(c) Diagrama de caja de Coeficiente de oxígeno.



(d) Diagrama de caja de Enriquecimiento de oxígeno.

Fig. A.6: Diagramas de caja con frecuencia de 1 hora.

Apéndice B

ANEXO: Código

```
1 import pandas as pd
2
3 archivos = [
4     'ConsFeed ',
5     'Cu ',
6     'DustFeed ',
7     'Fesio2 ',
8     'FluxFeed ',
9     'OxygenCoeff ',
10    'OxygenEnrich ',
11    'SlagTemperature '
12 ]
13
14 # Rango temporal uniforme
15 inicio = pd.Timestamp("2024-08-03 20:00:00")
16 fin = pd.Timestamp("2024-10-17 14:00:00")
17 frecuencia = '1h'
18 indice_uniforme = pd.date_range(start=inicio, end=fin, freq=frecuencia
19     )
20
21 # Procesar cada archivo
22 for nombre in archivos:
23     archivo_excel = f"{nombre}.xlsx"
24     columna_dato = nombre # Se asume que la columna se llama igual al
25         archivo
```

```
25     print(f"Procesando: {archivo_excel}")
26
27     try:
28         # 1. Leer archivo
29         df = pd.read_excel(archivo_excel)
30
31         # 2. Convertir columna Time a datetime
32         df['Time'] = pd.to_datetime(df['Time'], dayfirst=True, errors
33             = 'coerce')
34         df = df.dropna(subset=['Time'])
35
36         # 3. Eliminar valores cero
37         df = df[df[columna_dato] != 0]
38
39         # 4. Filtrar por rango deseado
40         df = df[(df['Time'] >= inicio) & (df['Time'] <= fin)]
41
42         # 5. Agrupar por hora exacta con promedio
43         df = df.set_index('Time')
44         df_resampled = df.resample('1h').mean()
45
46         # 6. Reindexar sobre la malla uniforme
47         df_final = df_resampled.reindex(indice_uniforme)
48
49         # 7. Interpolación lineal
50         df_final[columna_dato] = df_final[columna_dato].interpolate(
51             method='linear', limit_direction='both')
52
53         # 8. Guardar archivo final
54         df_final.reset_index(inplace=True)
55         df_final.rename(columns={'index': 'Time'}, inplace=True)
56         df_final.to_excel(f"{nombre}_1H.xlsx", index=False)
57
58         print(f"{nombre} procesado correctamente.\n")
59
60     except Exception as e:
61         print(f"Error procesando {archivo_excel}: {e}\n")
```

Listing B.1: Preprocesamiento de los datos