

La investigación reportada en esta tesis es parte de los programas de investigación del CICESE (Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California).

La investigación fue financiada por el CONAHCYT (Consejo Nacional de Humanidades, Ciencias y Tecnologías).

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México). El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo o titular de los Derechos de Autor.

CICESE © 2023, Todos los Derechos Reservados, CICESE

Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California



Maestría en Ciencias en Ciencias de la Computación

Medición del campo magnético estelar con aprendizaje profundo

Tesis

para cubrir parcialmente los requisitos necesarios para obtener el grado de
Maestro en Ciencias

Presenta:

Joan Manuel Raygoza Romero

Ensenada, Baja California, México

2023

Tesis defendida por

Joan Manuel Raygoza Romero

y aprobada por el siguiente Comité

Dr. Irvin Hussein López Nava

Codirector de tesis

Dr. Julio César Ramírez Veléz

Codirector de tesis

Dr. Carlos Alberto Brizuela Rodríguez

Dra. Paulina Segovia Olvera



Dr. Pedro Gilberto López Mariscal

Coordinador del Posgrado en Ciencias de la Computación

Dra. Ana Denise Re Araujo

Directora de Estudios de Posgrado

Resumen de la tesis que presenta Joan Manuel Raygoza Romero como requisito parcial para la obtención del grado de Maestro en Ciencias en Ciencias de la Computación.

Medición del campo magnético estelar con aprendizaje profundo

Resumen aprobado por:

Dr. Irvin Hussein López Nava

Codirector de tesis

Dr. Julio César Ramírez Veléz

Codirector de tesis

Los campos magnéticos juegan un papel crucial en la evolución estelar. Para comprender esta evolución, es fundamental medir y trazar los campos magnéticos en la superficie estelar. Estas mediciones se logran a través de observaciones espectropolarimétricas, utilizando la Ecuación de Radiación y Transferencia de Luz Polarizada, que describe la propagación de la radiación electromagnética polarizada. Las propiedades magnéticas se infieren ajustando los parámetros de Stokes I, Q, U, V (donde I representa la intensidad, Q y U la polarización lineal y V la polarización circular). En este estudio, proponemos un enfoque de aprendizaje profundo utilizando una red neuronal densa para estimar los parámetros de Stokes basados en ocho parámetros de entrada que describen la configuración del campo magnético. Para lograr esto, realizamos experimentos de escalado de los datos, diferentes configuraciones de la arquitectura de la red neuronal y la comparación de dos enfoques, uno donde la red neuronal densa estima simultáneamente los cuatro parámetros de Stokes y otro donde se entrena un modelo por cada parámetro, encontramos que la red neuronal densa es capaz de estimar con precisión los parámetros de Stokes I y V, siendo este último el más relevante. Sin embargo, enfrentamos dificultades al estimar los parámetros de Stokes Q y U en casos donde su amplitud era baja. Posteriormente, utilizamos esta red neuronal densa como parte de la función objetivo en el algoritmo *Particle Swarm Optimization* para recuperar el campo magnético. El modelo mostró un buen rendimiento en la recuperación de los valores, destacando nuestra metodología como una técnica novedosa y general para el estudio de campos magnéticos en estrellas, superando las limitaciones de los métodos tradicionales y logrando abarcar una amplia gama de intensidades de campos magnéticos. Además, realizamos una transferencia de aprendizaje y re-entrenamiento del modelo de estimación de los parámetros de Stokes para una longitud de onda diferente. Esto permitió obtener resultados similares en la recuperación del campo magnético utilizando menos datos y reduciendo el tiempo de entrenamiento. Por lo tanto, nuestro trabajo presenta una metodología que podría aplicarse a cualquier estrella, partiendo del mismo modelo base.

Palabras clave: campo magnético, parámetros de Stokes, regresión de parámetros, aprendizaje profundo, optimización

Abstract of the thesis presented by Joan Manuel Raygoza Romero as a partial requirement to obtain the Master of Science degree in Computer Science.

Measuring the stellar magnetic field with deep learning

Abstract approved by:

Dr. Irvin Hussein López Nava

Thesis Co-Director

Dr. Julio César Ramírez Veléz

Thesis Co-Director

Magnetic fields play a crucial role in stellar evolution. To understand this evolution, it is essential to measure and map the magnetic fields on the stellar surface. These measurements are achieved through spectropolarimetric observations, using the polarized radiative transfer equation, which describes the propagation of polarized electromagnetic radiation. Magnetic properties are inferred by adjusting the Stokes parameters I, Q, U, V (where I represents intensity, Q and U represent linear polarization, and V represents circular polarization). In this study, we propose a deep learning approach using a dense neural network to estimate the Stokes parameters based on eight input parameters that describe the magnetic field configuration. To achieve this, we conducted scaling experiments on the data, explored different configurations of the neural network architecture, and compared two approaches: one where the dense neural network simultaneously estimates all four Stokes parameters, and another where a separate model is trained for each parameter. We found that the dense neural network is capable of accurately estimating Stokes parameters I and V, with the latter being the most relevant. However, we encountered difficulties in estimating Stokes parameters Q and U in cases where their amplitudes were low. Subsequently, we used this dense neural network as part of the objective function in the *Particle Swarm Optimization* algorithm to recover the magnetic field. The model performed well in recovering the values, highlighting our methodology as a novel and general technique for studying magnetic fields in stars, surpassing the limitations of traditional methods and covering a wide range of magnetic field intensities. Furthermore, we conducted transfer learning and retraining of the Stokes parameter estimation model for a different wavelength. This allowed us to achieve similar results in magnetic field recovery using less data and reducing training time. Therefore, our work presents a methodology that could be applied to any star, starting from the same base model.

Keywords: magnetic fields, Stokes parameters, parameters regression, deep learning, optimization

Dedicatoria

A mis amados padres, Zindy María Guadalupe Romero González y Manuel Raygoza Flores, cuya inquebrantable fe en mí y apoyo incondicional han sido pilares fundamentales a lo largo de mi tesis y mi vida. Asimismo, deseo extender esta dedicatoria a dos amigas excepcionales que se cruzaron en mi camino durante mis estudios de posgrado: Alejandra De La Peña Castro y Karla Lorena Martínez Mauricio. Su respaldo incondicional, tanto en lo académico como en lo personal, ha dejado una huella imborrable en mi camino.

Agradecimientos

Agradezco sinceramente al Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE) por proporcionarme el entorno propicio para llevar a cabo este estudio de investigación.

Extiendo mi gratitud al Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT) por brindarme el apoyo económico para realizar mis estudios de maestría.

Deseo expresar mi más sincero agradecimiento a mis asesores de tesis, el Dr. Irvin Hussein López Nava y el Dr. Julio César Ramírez Velázquez, cuya guía experta, valiosos comentarios y sugerencias resultaron fundamentales en la formación de esta tesis. También agradezco al Dr. Carlos Alberto Brizuela Rodríguez y a la Dra. Paulina Segovia Olvera, quienes formaron parte del comité y estuvieron disponibles para revisar el progreso y la redacción de la tesis. A mi familia y amigos, les agradezco su constante apoyo y compañía a lo largo de este camino.

Tabla de contenido

	Página
Resumen en español	ii
Resumen en inglés	iii
Dedicatoria	iv
Agradecimientos	v
Lista de figuras	viii
Lista de tablas	ix
Capítulo 1. Introducción	
1.1. Preguntas de investigación	2
1.2. Objetivos	2
1.2.1. Objetivo general	3
1.2.2. Objetivos específicos	3
1.2.3. Justificación	3
1.3. Síntesis de la metodología	4
1.4. Estructura de los capítulos	5
Capítulo 2. Marco Conceptual	
2.1. Dominio del problema	6
2.1.1. Espectropolarimetría	6
2.1.2. Aproximación de campo débil	8
2.1.3. Método de centro de gravedad	8
2.1.4. Medición de los valores atmosféricos y campo magnético	9
2.2. Trabajo previo relacionado	9
2.3. Métodos computacionales	14
2.3.1. Red neuronal densa	14
2.3.2. Algoritmo de optimización	17
2.3.3. Particle swarm optimization	18
2.3.4. Transferencia de aprendizaje	20
2.4. Resumen	21
Capítulo 3. Metodología	
3.1. Generación de datos sintéticos	23
3.2. Preprocesamiento de los datos	24
3.3. Modelo de estimación de parámetros de Stokes	26
3.4. Modelo de recuperación del campo magnético	28
3.5. Métricas de evaluación	30
3.6. Resumen	31
Capítulo 4. Resultados	
4.1. Selección del método de escalado	32
4.1.1. Rango alto	33

4.1.2.	Rango bajo	34
4.1.3.	Rango muy bajo	35
4.1.4.	Resumen	36
4.2.	Selección de la arquitectura	37
4.3.	Modelo general vs modelo especializado	38
4.4.	Efecto del tamaño del conjunto de datos	39
4.5.	Modelo de estimación de parámetros de Stokes	40
4.5.1.	Análisis del error en función de la amplitud del parámetro de Stokes	41
4.5.2.	Análisis de los percentiles	45
4.5.3.	Resumen	47
4.6.	Selección del algoritmo para la recuperación del campo magnético	48
4.7.	Recuperación del campo magnético	50
4.8.	Transferencia de aprendizaje	51
4.8.1.	Re-entrenamiento del modelo de estimación de parámetros de Stokes	52
4.8.2.	Recuperación del campo magnético con el modelo re-entrenado	53
4.8.3.	Resumen	54
Capítulo 5.	Conclusiones	
5.1.	Trabajo a futuro	56
Literatura citada	58

Lista de figuras

Figura	Página
1. Polarización lineal (izquierda) y polarización circular (derecha).	7
2. Arquitectura de una red neuronal densa.	14
3. Diagrama del proceso de entrenamiento de la red neuronal densa (lado izquierdo) y la optimización (lado derecho).	22
4. Ejemplo de los parámetros de Stokes generado por COSSAM.	24
5. Diagrama del flujo de la red neuronal densa.	26
6. Flujo de optimización para la recuperación del campo magnético.	28
7. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas combinaciones de los métodos de escalado para la entrada y la salida de los datos para el rango alto.	33
8. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas combinaciones de los métodos de escalado para la entrada y la salida de los datos para el rango bajo.	34
9. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas combinaciones de los métodos de escalado para la entrada y la salida de los datos para el rango muy bajo.	36
10. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas arquitecturas de redes neuronales densas agrupados por la cantidad de neuronas por capa para el rango alto.	37
11. MSE de cada parámetro de Stokes del modelo general y especializado para el rango alto.	38
12. WMAPE de cada parámetro de Stokes del modelo para la estimación de parámetros de Stokes agrupado por rangos de amplitud para el rango alto.	42
13. WMAPE de cada parámetro de Stokes del modelo para la estimación de parámetros de Stokes agrupado por rangos de amplitud para el rango bajo.	43
14. WMAPE de cada parámetro de Stokes del modelo para la estimación de parámetros de Stokes agrupado por rangos de amplitud para el rango muy bajo.	44
15. Estimación de los parámetros de Stokes de los percentiles 80 (arriba), 90 (en medio) y 99 (abajo) del rango alto.	45
16. Estimación de los parámetros de Stokes de los percentiles 80 (arriba), 90 (en medio) y 99 (abajo) del rango bajo.	46
17. Estimación de los parámetros de Stokes de los percentiles 80 (arriba), 90 (en medio) y 99 (abajo) del rango muy bajo.	47
18. MSE promedio en el espacio escalado de la mejor solución encontrada para el rango alto.	49
19. Gráficas de línea del MSE por cada parámetro de Stokes junto con su desviación estándar (barras verticales) para el rango alto.	51
20. Gráficas de línea del MSE en cada parámetro de Stokes por la cantidad de capas congeladas en el re-entrenamiento para el rango alto.	52

Lista de tablas

Tabla		Página
1.	Intervalos para los 8 atributos que describen la geometría magnética de la configuración dipolar.	23
2.	Parámetros del entrenamiento de la red neuronal para la estimación de los parámetros de Stokes.	27
3.	Parámetros del entrenamiento de la red neuronal para la estimación de los parámetros de Stokes.	29
4.	Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil de combinaciones de escalado de entrada y salida para el rango alto.	34
5.	Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil de combinaciones de escalado de entrada y salida para el rango bajo.	35
6.	Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil de combinaciones de escalado de entrada y salida para el rango muy bajo.	35
7.	Tabla con el error MSE promedio por parámetro de Stokes y el WMAPE promedio, junto con el tercer cuartil de distintas arquitecturas de red neuronal densa para el rango alto.	37
8.	Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil del modelo general y especializado para el rango alto	39
9.	Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil del modelo entrenado con distinta cantidad de datos para el rango alto.	39
10.	Tabla con el WMAPE por parámetro de Stokes y el WMAPE promedio, junto con el tercer cuartil para los distintos rangos.	40
11.	Tabla con el WMAPE por atributo del campo magnético del modelo en cada rango.	50
12.	Tabla con el WMAPE por parámetro de Stokes y el WMAPE promedio, junto con el tercer cuartil para los distintos rangos para el modelo base y el re-entrenado. . .	52
13.	Tabla con el WMAPE por atributo del campo magnético del modelo en cada rango para los distintos rangos para el modelo base y el re-entrenado.	53

Capítulo 1. Introducción

Los campos magnéticos son el origen de la actividad solar y estelar, y se acepta ampliamente que pueden desempeñar un papel muy importante en la evolución estelar: desde las estrellas jóvenes hasta los objetos compactos. Además, la fuerza y la topología del campo magnético varían según la etapa evolutiva, es decir, la edad de la estrella. Por lo tanto, es importante comprender cómo evolucionan los campos magnéticos estelares (Ramírez, J. et al. (2018)).

El trazado y la medición de los campos magnéticos estelares presentan varios desafíos debido a la naturaleza invisible de estos y a las complejas interacciones que ocurren dentro de las estrellas. Sin embargo, a lo largo de los años, se han desarrollado diferentes técnicas y herramientas para abordar esta tarea. Una de las técnicas más utilizadas es la espectropolarimetría, a partir de la Ecuación de Radiación y Transferencia de luz Polarizada (ERTP) que describe la propagación de la radiación electromagnética polarizada a través de medios astrofísicos, la cual aprovecha las variaciones en la polarización de la luz emitida por una estrella para inferir la presencia y las propiedades del campo magnético. Es a través del análisis de datos espectropolarimétricos que se pueden caracterizar adecuadamente los campos magnéticos (Degl'Innocenti & Landolfi (2004)). Esta técnica ha sido aplicada con éxito en estrellas de diferentes tipos y edades, proporcionando información valiosa sobre la distribución y la evolución de los campos magnéticos (Mathys (1995)).

Para ajustar las observaciones espectropolarimétricas se requiere el uso de códigos teóricos, estos códigos utilizan la ERTP para sintetizar los parámetros de Stokes a partir de una configuración del campo magnético sobre la estrella. En este trabajo, se considerará una configuración magnética dipolar descentrada generalizada. Se utiliza esta configuración para cubrir casos más complejos donde el campo magnético no es uniforme. Para obtener espectros sintéticos que se ajusten a las observaciones, se empleó el código COSSAM (Stift & Alecian (2012)), en el cual la superficie de la estrella se divide en 80 áreas, y en cada área se calculan los cuatro parámetros de Stokes mediante 8 atributos que describen la geometría magnética de la configuración dipolar.

Luego se integran sobre toda la superficie para obtener los perfiles de Stokes resultantes que se compararán con las observaciones. Esto requiere mucho tiempo de cómputo para modelar en el proceso de ajuste. Utilizando COSSAM estaríamos utilizando un enfoque teórico completo basado en la ERTP además permite generar un amplio conjunto de datos. Por esta razón, en este trabajo, se desarrolló un modelo basado en aprendizaje profundo para obtener una herramienta confiable que pueda sintetizar una gran cantidad de configuraciones magnéticas para los parámetros de Stokes de una manera asequible.

En la literatura científica, se han desarrollado diversos modelos destinados a realizar regresiones de las características atmosféricas y configuración magnética a partir de los parámetros de Stokes del Sol, nuestra estrella. Algunos ejemplos destacados en este campo de investigación son los artículos de Gafeira et al. (2021), Liu et al. (2020) y Knyazeva et al. (2022). Estos estudios se centran en obtener los datos espectropolarimétricos de regiones sumamente reducidas del disco solar, lo que permite una resolución espacial altamente detallada.

Mientras en los artículos de Ramírez, J. et al. (2018) y Córdova, J. et al. (2018), que abordan estrellas más allá del Sol, los datos espectropolarimétricos se obtienen del disco en su totalidad, lo que resulta en una falta de resolución espacial en dichas observaciones, ya que todas las estrellas se consideran objetos puntuales en los telescopios. En los enfoques en estos artículos, se emplea un modelo computacional como puede ser una red neuronal densa o una red neuronal convolucional (CNN) para obtener las características atmosféricas y/o configuración magnética en lugar de recurrir al método de optimización clásico.

1.1. Preguntas de investigación

- ¿Cuál es el error cuadrático medio (MSE) y el error porcentual medio absoluto ponderado (WMAPE) en la estimación de los parámetros de Stokes con una red neuronal densa en comparación al utilizar la E RTP calculada por COSSAM?
- ¿Cuál es el WMAPE al utilizar Particle Swarm Optimization (PSO) para recuperar la configuración magnética utilizando la red neuronal densa como parte de la función objetivo para estimar los parámetros de Stokes?
- ¿Cuál es la cantidad de datos que se requiere para realizar una transferencia de aprendizaje y re-entrenar el modelo con datos de otra estrella con la expectativa de un resultado similar?

1.2. Objetivos

En las siguientes subsecciones se exponen los objetivos del presente trabajo de investigación.

1.2.1. Objetivo general

Implementar y evaluar una red neuronal densa con distintas arquitecturas, escalados y enfoques para estimar la configuración magnética estelar para una configuración magnética dipolar descentrada generalizada.

1.2.2. Objetivos específicos

- Realizar una comparación entre las distintas arquitecturas de la red neuronal densa en la estimación de los parámetros de Stokes, con el objetivo de obtener la arquitectura que minimiza el error.
- Analizar los resultados utilizando distintos tratamientos a los atributos que describen la configuración magnética estelar y los parámetros de Stokes para la estimación de los mismos, con el fin de obtener resultados con menor error.
- Desarrollar y optimizar el entrenamiento del modelo seleccionado, utilizando una cantidad de 3.3 millones de datos, con el propósito de mejorar su capacidad para estimar con precisión los parámetros de Stokes.
- Integrar el modelo entrenado de estimación de parámetros de Stokes con un algoritmo de optimización robusto y eficiente, con el fin de lograr una recuperación precisa del valor del campo magnético.
- Adaptar y re-entrenar el modelo desarrollado para su aplicación en la estimación de los parámetros de Stokes de otras estrellas, con el propósito de demostrar su generalidad.

1.2.3. Justificación

La presente investigación tiene como propósito utilizar un enfoque teórico completo, a través de la técnica de espectropolarimetría en combinación con la E RTP para analizar datos espectropolarimétricos estelares. Este enfoque innovador nos permitirá obtener una comprensión más profunda de los procesos de formación de campos magnéticos en las estrellas y su influencia en la evolución estelar. La espectropolarimetría es una técnica que combina la espectroscopia y la polarimetría, y su aplicación en el estudio

de objetos celestes ofrece una visión detallada y exhaustiva de la luz emitida por las estrellas (European Southern Observatory (ESO) (SF)).

Los campos magnéticos estelares desempeñan un papel fundamental en la dinámica y evolución de las estrellas, afectando su estructura interna, actividad superficial e interacción con el entorno. Mediante la obtención de una comprensión más precisa de los procesos de formación de campos magnéticos estelares y su impacto en la evolución estelar, esta investigación abrirá nuevas puertas en el campo de la astrofísica. Los resultados obtenidos contribuirán a avanzar en áreas clave como la física estelar, la formación y evolución de estrellas y su campo magnético. Esta tesis de maestría representa una contribución significativa al conocimiento existente en el campo de la espectropolarimetría estelar y su aplicación en el análisis de datos espectropolarimétricos estelares. Los hallazgos y conclusiones obtenidos a partir de esta investigación podrán ser utilizados como base para futuras investigaciones y para ampliar nuestra comprensión del universo y su funcionamiento.

1.3. Síntesis de la metodología

El proceso se inicia mediante la generación de un conjunto de datos sintéticos a partir del programa COSSAM. Con este conjunto de datos, se procede a entrenar una red neuronal densa diseñada para estimar los datos espectropolarimétricos. Antes del entrenamiento, se realiza un escalado tanto de los datos de entrada como de los datos de salida. Una vez que la red neuronal densa ha sido debidamente entrenada, se integra con un algoritmo de optimización para formar la función objetivo, cuyo propósito es estimar los datos espectropolarimétricos para cada solución candidata. Cada solución candidata representa una configuración magnética en diferentes fases de rotación de una estrella. Para realizar la comparación, se calcula el error cuadrático medio entre la solución estimada por la red neuronal densa y los datos observados en cada fase de rotación de la estrella. Al concluir el proceso, se obtiene la configuración del campo magnético correspondiente para dicha estrella.

Cuando se necesita aplicar este proceso a una estrella diferente o en otra longitud de onda, se realiza una transferencia de aprendizaje en la red neuronal densa. Esta transferencia de aprendizaje permite re-entrenar la red con tan solo 35,000 casos. Una vez completado el re-entrenamiento, la red neuronal puede utilizarse en combinación con el algoritmo de optimización para obtener la configuración del campo magnético para la nueva estrella o longitud de onda específica.

Nuestra metodología se destaca como una técnica general y novedosa para el estudio de campos magnéticos en estrellas, superando las limitaciones de la aproximación de campo débil (WFA) y el método del centro de gravedad (CoG). Mediante la combinación de datos espectropolarimétricos, el uso de redes neuronales y posteriormente la aplicación de un algoritmo de optimización, logramos abarcar una amplia gama de intensidades de campos magnéticos. Además, al emplear transferencia de aprendizaje y re-entrenar la red neuronal, esta metodología podría ser aplicada a diferentes longitudes de onda o estrellas.

1.4. Estructura de los capítulos

La tesis se organiza en los siguientes capítulos, cada uno abordando aspectos específicos de la investigación:

Capítulo 1: Introducción. En este capítulo se establece el contexto y la relevancia del estudio. Se describe el problema de investigación, se plantea el objetivo general y se presentan las preguntas de investigación que guían el desarrollo de la tesis.

Capítulo 2: Marco Conceptual. Este capítulo presenta el marco conceptual que fundamenta la investigación. Se revisa la literatura existente y se discuten los estudios previos relacionados con el tema. Además, se establecen las bases teóricas necesarias para el entendimiento de la tesis.

Capítulo 3: Metodología. En este capítulo se detallan los enfoques y métodos utilizados en la investigación. Se describe la metodología empleada, incluyendo el diseño de los modelos, la generación del conjunto de datos, los escalados de datos y las métricas de evaluación.

Capítulo 4: Resultados. Se presentan y analizan los resultados obtenidos. Se identifican los hallazgos más relevantes y se establecen las relaciones entre los datos y los modelos propuestos de acuerdo a los objetivos de investigación planteados.

Capítulo 5: Conclusiones. En este último capítulo se presentan las conclusiones finales de la tesis. Se resumen los principales hallazgos, se destacan las contribuciones específicas realizadas en el estudio y se responden las preguntas de investigación planteadas en la introducción. Además, se discuten las implicaciones más amplias de los resultados y se plantean posibles áreas de desarrollo y trabajo futuro.

Capítulo 2. Marco Conceptual

2.1. Dominio del problema

El campo magnético desempeña un papel crucial en la evolución y actividad estelar, teniendo un impacto significativo en las estrellas. Es el impulsor detrás de la actividad estelar, como las manchas solares en el Sol. Las estrellas con campos magnéticos fuertes exhiben una variedad de características, como manchas estelares, flares (llamaradas) y eyecciones de masa coronal. Estos eventos son el resultado de interacciones dinámicas entre el campo magnético y el plasma estelar. Además, el campo magnético desempeña un papel importante en la pérdida de masa estelar. Los campos magnéticos fuertes interactúan con el viento estelar, causando efectos de aceleración y canalización que conducen a la pérdida de masa de la estrella.

Este proceso es particularmente relevante en estrellas masivas, donde la pérdida de masa puede tener un impacto significativo en su evolución y destino final (Nariai (1969)). Asimismo, el campo magnético influye en la evolución estelar al afectar procesos como la rotación estelar y la transferencia de masa en sistemas binarios. La interacción entre el campo magnético y el material estelar afecta la distribución de masa, la transferencia de momento angular y la velocidad de rotación de la estrella. Estos factores son cruciales para comprender la evolución y el comportamiento de las estrellas en el universo.

2.1.1. Espectropolarimetría

La espectropolarimetría ha surgido y se ha desarrollado predominantemente en el campo de la astrofísica. A lo largo del siglo XX, se ha ido reconociendo cada vez más la importancia de los campos magnéticos estelares. La mayoría, si no todas, las manifestaciones observables del magnetismo solar han sido analizadas mediante técnicas espectropolarimétricas. Como resultado, se han realizado grandes esfuerzos para mejorar la precisión de las mediciones y desarrollar una teoría que permita el análisis de estas mediciones espectropolarimétricas. No solo nos ayuda a desvelar los parámetros físicos del Sol y otras estrellas, sino que también se considera un pilar fundamental en nuestra comprensión de la interacción entre la radiación y la materia en el universo (Degl'Innocenti & Landolfi (2004)).

La ETRP desempeña un papel crucial en la interpretación de la información codificada en la radiación polarizada, en términos de las propiedades físicas de la fuente de luz. Es a través de la ETRP que

podemos inferir los parámetros astronómicos relevantes. La luz, o radiación electromagnética, constituye la principal fuente de información para los astrónomos. Aunque puede parecer sorprendente afirmar que no podemos medir directamente los parámetros físicos de la atmósfera solar, esta afirmación resalta el hecho de que la astrofísica es una ciencia observacional y no experimental (del Toro (2003)). En lugar de medir directamente temperaturas, velocidades o campos magnéticos de las estrellas, solo podemos medir la luz que llega hasta nosotros. Los parámetros astronómicos se deducen a partir de estas mediciones. Por lo tanto, la confiabilidad de las inferencias astronómicas depende en gran medida de la precisión de las mediciones de luz. En este sentido amplio, la espectropolarimetría abarca todas las mediciones reales realizadas por los astrónomos, ya que analiza la luz en función de sus dos características fundamentales: la longitud de onda y el estado de polarización (del Toro (2003)).

La polarización de la luz se describe mediante los parámetros de Stokes. Estos surgieron de la teoría matemática de la polarización de la luz y fueron propuestos en el siglo XIX por el físico británico George Gabriel Stokes. Consisten en cuatro valores numéricos que brindan una descripción completa del estado de polarización de la luz en un punto determinado del cielo. Estos parámetros se identifican comúnmente como I, Q, U y V. El parámetro I representa la intensidad de la luz sin polarizar, mientras que los parámetros Q y U representan la polarización lineal (lado izquierdo de la Figura 1) y el parámetro V representa la polarización circular (lado derecho de la Figura 1). Estos parámetros varían en respuesta a la presencia y las propiedades del campo magnético en la estrella.

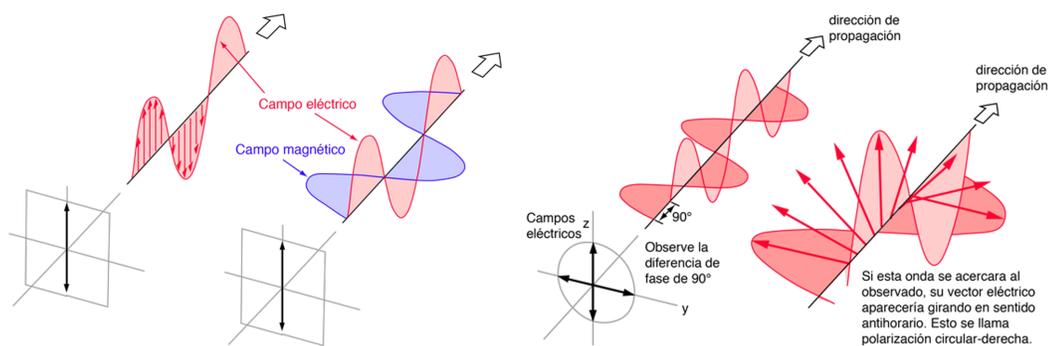


Figura 1. Polarización lineal (izquierda) y polarización circular (derecha).

La medición de la polarización de la luz mediante los parámetros de Stokes se realiza mediante espectropolarímetros. Estos instrumentos están diseñados para cuantificar la polarización de la luz. Estos instrumentos son fundamentales para el estudio de fenómenos astronómicos complejos y para profundizar en nuestro conocimiento del universo y sus objetos celestes.

2.1.2. Aproximación de campo débil

La aproximación de campo débil (WFA) es una técnica utilizada para medir el campo magnético proyectado a lo largo de la línea de visión. Esta aproximación se aplica cuando el ensanchamiento de la línea espectral que se utiliza para medir el campo es insignificante en comparación con otros procesos físicos de ensanchamiento, como la rotación, el ensanchamiento Doppler y la turbulencia micro y macro (Mathys & Stenflo (1987)). Estas condiciones imponen restricciones en la aplicación de esta técnica para medir campos magnéticos, especialmente cuando las intensidades son altas, del orden de kilo Gauss. La utilización de modelos teóricos nos permite abordar mediciones de campos magnéticos en estrellas con intensidades más allá del alcance de la WFA. Esto nos brinda la capacidad de investigar campos magnéticos de mayor intensidad y explorar su influencia en diversos fenómenos astrofísicos.

2.1.3. Método de centro de gravedad

A diferencia a la WFA, existe otra técnica conocida como el método del centro de gravedad (CoG), que ha demostrado tener menos limitaciones en el estudio de los campos magnéticos solares y estelares (Rees & Semel (1979), Mathys (1991)). Sin embargo, el CoG no utiliza modelos teóricos para el análisis de los datos. En cambio, se basa en los perfiles de Stokes observados para medir la intensidad del campo magnético. Esto implica que el método del CoG no es general y está limitado al estudio de campos magnéticos débiles (Ramírez (2020)).

En este contexto, la innovadora técnica que presentamos en este trabajo destaca como una opción única y general, ya que no está restringida a intensidades de campos magnéticos débiles. Esto se debe a que, para entrenar la red neuronal, hemos utilizado perfiles de Stokes teóricos obtenidos a partir de la resolución de la E RTP. Al emplear perfiles de Stokes teóricos en lugar de observados, podemos superar las limitaciones asociadas con la debilidad de los campos magnéticos. Nuestra metodología aprovecha la información proporcionada por los modelos teóricos, lo que nos permite abordar una amplia gama de intensidades de campos magnéticos estelares y solares.

2.1.4. Medición de los valores atmosféricos y campo magnético

La utilización de los métodos de WFA y CoG es de gran utilidad porque permiten medir el campo magnético efectivo sin recurrir a cálculos de transferencia radiativa. Actualmente, el método CoG es la técnica preferida utilizada para determinar el campo magnético efectivo. Sin embargo, ambos enfoques tienen limitaciones debido a la suposición de auto-similitud de las líneas, debiendo cumplir ambas condiciones simultáneamente: el uso de líneas sin mezclar y la medición de campos magnéticos débiles. Por lo tanto, determinar con precisión el rango de aplicación en el análisis de datos resulta difícil. Para medir y caracterizar los campos magnéticos estelares, se utilizan códigos teóricos que sintetizan los parámetros de Stokes a partir de una configuración del campo magnético sobre la estrella, para posteriormente comparar con los parámetros de Stokes observados.

A este proceso se le llama inversión, tiene lugar cuando se infiere los valores atmosféricos y/o del campo magnético a partir de perfiles observados de Stokes. La inversión se logra mediante la síntesis iterativa de perfiles de Stokes y su comparación con las observaciones. Se lleva a cabo un proceso de optimización de las condiciones físicas con el objetivo de minimizar la diferencia entre el perfil real y el sintetizado, hasta alcanzar un umbral definido. Las magnitudes utilizadas para sintetizar el perfil con el mayor acuerdo constituirán las mejores estimaciones del estado de la atmósfera que generó el perfil observado. En resumen, el proceso de inversión es una optimización que permite obtener información detallada sobre el campo magnético estelar a partir de la comparación y ajuste de los perfiles observados y sintetizados de Stokes.

2.2. Trabajo previo relacionado

En la actualidad, los estudios sobre magnetismo estelar se basan en el análisis de los parámetros de Stokes observados en estrellas. Por medio de la Ecuación de Radiación y Transferencia de luz Polarizada (ERTP), es que teóricamente se pueden modelar las distintas configuraciones del campo magnético estelar. Al ajustar los perfiles observados con aquellos obtenidos usando la ERTP, es que es posible obtener la configuración del campo magnético y/o las características atmosféricas de la estrella. Sin embargo, el proceso de ajuste de los parámetros de Stokes calculados teóricamente es computacionalmente costoso (Gafeira et al. (2021)), debido al tiempo requerido por los métodos de resolución para resolver las ecuaciones diferenciales de la ERTP para simular un modelo magnético dado. Ante este desafío, el

uso de algoritmos de aprendizaje automático, en particular el aprendizaje profundo, ha surgido como una posible solución, y varios estudios han comenzado a emplearlos para obtener las características atmosféricas y/o la configuración magnética a partir de los parámetros de Stokes (Ramírez, J. et al. (2018); Córdova, J. et al. (2018); Liu et al. (2020); Gafeira et al. (2021); Knyazeva et al. (2022)).

Las amplitudes de los perfiles de los parámetros de Stokes linealmente polarizados Q y U suelen ser más pequeñas que las de la polarización circular dada por el Stokes V y esta a su vez mas pequeña que la amplitud de la intensidad dada por el Stokes I. Como resultado, es común que en muchas mediciones de los parámetros de Stokes polarizados, Q y U, no se detecten con certeza debido a su baja intensidad y al nivel de ruido de las observaciones. Por lo tanto, típicamente solo se utilizan los parámetros de Stokes I y V para medir el campo magnético. Sin embargo, el parámetro de Stokes V también puede estar por debajo del nivel de ruido. Para abordar este problema, se han desarrollado técnicas multi-línea que suman múltiples líneas individuales en el dominio de velocidades (espacio Doppler), lo que resulta en un perfil conocido como firmas de Zeeman múltiples (MZS). Estas MZS contienen toda la información de los espectros polarizados, reduciendo el nivel de ruido y logrando una reducción de la dimensionalidad (Ramírez, J. et al. (2018); Córdova, J. et al. (2018)). Además, en los parámetros de Stokes se requiere una normalización o escalamiento para reducir las diferencias de magnitud entre ellos (Liu et al. (2020); Gafeira et al. (2021); Knyazeva et al. (2022)).

Debido a que aún son pocos los trabajos en esta área de investigación, a continuación haremos una revisión de ellos. Primero, empezamos por el estudio realizado por Córdova, J. et al. (2018), en el que se generaron los perfiles de Stokes V a partir de tres parámetros libres: temperatura efectiva (temperatura de la superficie visible), la fuerza de la gravedad y velocidad macro turbulenta (turbulencia causada por el medio interestelar). Se utilizó un modelo magnético centrado dipolar y se consideró una metalicidad (abundancia relativa de elementos más pesados que el helio) similar a la solar. Los perfiles de Stokes V fueron sometidos a Análisis de Componentes Principales (PCA) para obtener las firmas de Zeeman múltiples promedio y para eliminar el ruido, consistiendo cada firma de 279 puntos. Posteriormente, se agregó ruido gaussiano a las MZS para representar mejor los datos observados reales. Para el análisis, se empleó una red neuronal artificial de perceptrones multicapa para realizar la regresión del campo magnético efectivo. Los resultados obtenidos en el trabajo son regresiones dentro de un margen de error del 20 %, donde proponen como trabajo a futuro el uso de redes neuronales artificiales para datos reales y la expansión de la base de datos para incluir más variaciones con un número mayor de parámetros libres.

En el trabajo de Ramírez, J. et al. (2018), se llevó a cabo la síntesis de los perfiles de Stokes V a partir

de 7 parámetros libres. Estos parámetros incluyen los tres ángulos de Euler que describen la geometría magnética de la configuración dipolar, el ángulo de inclinación del eje de rotación estelar con respecto a la línea de visión, dos coordenadas que describen la posición del dipolo dentro de la estrella y la fuerza del momento dipolar. Siguiendo una metodología similar a la propuesta por Córdova, J. et al. (2018), se obtuvieron los MZS y se implementó un proceso de agregación de ruido gaussiano para obtener una disminución del del ruido. Con esto, se construyó un conjunto de datos que se utilizó para evaluar el rendimiento de tres modelos diferentes: Bayesian Ridge, Support Vector Machines y una red multicapa de perceptrones.

Los resultados obtenidos tienen un MAPE (error porcentual absoluto promedio) de alrededor del 10 %, y también se observó que era posible reducir el número de perfiles sintéticos necesarios para obtener un rendimiento satisfactorio. Además, se propuso un nuevo método para trabajar con los MZS, y se llevó a cabo una comparación de resultados utilizando datos observados de dos estrellas. Se encontró que las estimaciones obtenidas por el mejor modelo se aproximaban de manera casi idéntica a otras estimaciones previas. Cabe destacar que estos resultados son prometedores y abren nuevas perspectivas en el campo de la modelización del magnetismo estelar, permitiendo una mayor eficiencia en la obtención de información sobre los parámetros libres a partir de los perfiles de Stokes V sintéticos y observados.

En el trabajo de Liu et al. (2020) se analizaron los perfiles de Stokes observados con el instrumento GST/NIRIS (espectropolarímetro de imágenes en infrarrojo cercano en el telescopio de 1.6m Goode Solar), que proporciona vectores de magnetogramas de alta resolución y produce imágenes en escala de grises. Cada píxel de la imagen se trató como una muestra, a partir de la cual se obtuvieron los parámetros de Stokes I, Q, U y V. Utilizando el método de Milne-Eddington (ME), se calcularon la fuerza del campo magnético, el ángulo de inclinación y el ángulo de azimuth, que representan la configuración del campo magnético y son los valores que el modelo busca predecir. Los parámetros de Stokes Q, U y V fueron calibrados por I, normalizados y sometidos a un proceso de escalado.

El modelo de aprendizaje profundo utilizado en este estudio fue una red neuronal convolucional (CNN) con 3 canales de entrada correspondientes a los parámetros de Stokes Q, U y V. El modelo se entrenó para realizar la regresión de los tres valores de la configuración magnética para cada píxel. Posteriormente, este proceso se aplicó a todos los píxeles de una imagen y la suma de los resultados de cada píxel proporcionó los valores totales de la configuración magnética. El modelo computacional desarrollado demostró ser de 4 a 6 veces más rápido que el método ME tradicional. En términos de precisión, el modelo logró una alta correlación promedio de momento-producto de Pearson (PPMCC) para la fuerza del campo magnético, el ángulo de inclinación y el ángulo de azimuth. Específicamente, se obtuvieron

valores de 0.95, 0.88 y 0.83, respectivamente. Además, el error absoluto promedio (MAE) para la fuerza del campo magnético, el ángulo de inclinación y el ángulo de azimuth fue de aproximadamente 118, 5.3 y 11. Estos resultados indican un buen rendimiento del modelo en la predicción de los parámetros de configuración magnética a partir de los parámetros de Stokes sintetizados, lo que demuestra su utilidad en la obtención eficiente de información sobre el campo magnético estelar.

En el trabajo realizado por Gafeira et al. (2021), se utilizaron los datos del espectropolarímetro a bordo de la nave Hinode. Estos datos fueron sometidos a un proceso de optimización de los datos espectropolarimétricos observados con el cálculo de códigos teóricos para obtener la temperatura, velocidad, fuerza, azimuth e inclinación del campo magnético. Posteriormente, se normalizaron los parámetros de Stokes y los valores calculados previamente. Los cuatro parámetros de Stokes se utilizaron como entrada en un modelo computacional basado en una CNN para realizar la regresión de las estratificaciones de la configuración magnética. En este enfoque, cada muestra corresponde a un píxel de una imagen, similar al enfoque utilizado en el trabajo de Liu et al. (2020). El objetivo de este modelo computacional fue obtener de manera eficiente una estimación inicial de la configuración magnética, lo que permitiría reducir la cantidad de iteraciones necesarias para que el método de inversión converja.

Los resultados obtenidos mostraron una alta correlación de Pearson (0.99) para cada valor de la configuración magnética y las características atmosféricas. Sin embargo, los investigadores observaron una disminución en el rendimiento del modelo en áreas del mapa que no se utilizaron en el proceso de entrenamiento. Por lo tanto, recomiendan analizar con precaución los resultados obtenidos a partir del modelo CNN. En resumen, el enfoque propuesto por Gafeira et al. (2021) proporciona una forma eficiente de obtener una estimación inicial de la configuración magnética utilizando una red neuronal convolucional, lo cual puede mejorar la velocidad de convergencia del método de optimización. Además, este enfoque puede ser especialmente beneficioso en áreas que contienen umbras frías, lo que abre nuevas posibilidades de estudio en la caracterización del campo magnético estelar.

En el estudio llevado a cabo por Knyazeva et al. (2022), se empleó el modelo atmosférico con 11 parámetros para la generación de los perfiles de Stokes. Estos parámetros incluyeron la fuerza, inclinación y azimuth del campo magnético, ensanchamiento Doppler, amortiguamiento y fuerza de línea, continuo normalizado, gradiente del continuo normalizado de la función fuente, desplazamiento Doppler, factor de llenado magnético y desplazamiento Doppler. Utilizando estos parámetros, se generaron perfiles sintéticos de Stokes. Como parte del proceso de normalización, los parámetros de Stokes Q, U y V se multiplicaron por un factor de 1000, mientras que los 11 parámetros utilizados para generar los datos sintéticos se escalonaron. Además, se aplicó un proceso de agregación de ruido a los parámetros de Stokes.

Se utilizaron tres enfoques distintos para realizar la regresión de la configuración magnética. El primer enfoque consistió en entrenar un modelo computacional individual para cada uno de los 11 valores. El segundo enfoque involucró un único modelo computacional que realizara la regresión de los 11 valores simultáneamente. Por último, en el tercer enfoque, se utilizó el modelo computacional del segundo enfoque con sus pesos congelados, y se entrenó un modelo computacional individual para cada uno de los 11 valores. Todos estos modelos se basaron en una arquitectura de una red de perceptrones multicapa (MLP). Según los resultados obtenidos en los tres enfoques, se determinó que el mejor rendimiento se logró utilizando un bloque compartido de MLP y, posteriormente, un modelo computacional individual para cada valor de la configuración magnética, es decir, el tercer enfoque. Se obtuvieron los siguientes resultados para el error cuadrático medio (MSE) de la fuerza, inclinación y azimuth del campo magnético: 0.0019, 0.0025 y 0.0266, respectivamente. Los autores concluyen que este enfoque permite obtener resultados positivos que son aplicables en la práctica, además de presentar una alta velocidad de cálculo.

En la literatura científica, se han desarrollado varios modelos para realizar regresiones de las características atmosféricas y/o configuración magnética a partir de los parámetros de Stokes de nuestra estrella, el Sol. Los artículos de Gafeira et al. (2021), Liu et al. (2020) y Knyazeva et al. (2022) son ejemplos destacados en este campo de investigación. Estos estudios se enfocan en obtener los parámetros de Stokes de una región muy pequeña del disco solar, lo que permite una resolución espacial detallada. En contraste, en los artículos de Ramírez, J. et al. (2018) y Córdova, J. et al. (2018), que se centran en estrellas más allá del Sol, los parámetros de Stokes se obtienen del disco total (integrado), lo que resulta en una falta de resolución espacial en estas observaciones, ya que todas las estrellas se consideran objetos puntuales en los telescopios. Sin embargo, en el caso del Sol, es posible obtener información espacialmente resuelta, lo que permite observar diferentes regiones en su superficie.

En los enfoques presentados en estos artículos, se utiliza un modelo computacional para obtener las características atmosféricas y/o la configuración magnética en lugar de recurrir al método de optimización clásico. Este enfoque proporciona una alternativa eficiente y precisa para estimar estas propiedades, sin la necesidad de realizar iteraciones complejas de optimización. Además, algunos estudios utilizan el modelo computacional como una primera predicción de los valores, que luego pueden ser refinados utilizando métodos de optimización tradicionales.

2.3. Métodos computacionales

2.3.1. Red neuronal densa

Las redes neuronales densas son un componente fundamental de las arquitecturas de aprendizaje profundo y han demostrado ser altamente efectivas en muchas aplicaciones del mundo real (Paliwal & Kumar (2009)). Consisten en una capa de entrada, una o más capas ocultas y una capa de salida, donde los datos fluyen en una sola dirección desde la entrada hacia la salida. Cada neurona en la red recibe entradas de la capa anterior, las procesa utilizando un conjunto de pesos y sesgos, y pasa el resultado a la siguiente capa. Las capas ocultas de la red son responsables de extraer características complejas de los datos de entrada sin procesar y transformarlos en una representación adecuada para la tarea en cuestión. La capa de salida proporciona la predicción o decisión final. Un ejemplo de una arquitectura de una red neuronal se puede ver en la Figura 2.

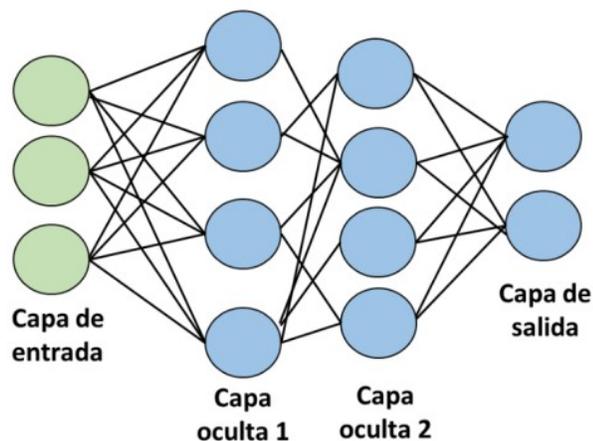


Figura 2. Arquitectura de una red neuronal densa.

El entrenamiento de una red neuronal densa es un problema de optimización, donde el objetivo es encontrar el conjunto de pesos y sesgos que minimizan el error de predicción en un conjunto de datos de entrenamiento. Esto se logra típicamente utilizando un algoritmo de optimización como el descenso de gradiente estocástico o una variante del mismo, y se utiliza el algoritmo de retropropagación para calcular los gradientes del error con respecto a los pesos y sesgos (Svozil et al. (1997)). La retropropagación es un algoritmo crucial en el entrenamiento de las redes neuronales densas. Consiste en calcular y ajustar los gradientes de error a lo largo de la red, comenzando desde la capa de salida y retrocediendo hacia las

capas ocultas. Esto se realiza utilizando la regla de la cadena y la derivada de la función de activación de cada neurona. Durante el entrenamiento, se calcula el error de predicción en cada paso y se ajustan los pesos y sesgos de la red para minimizar ese error. Este proceso se repite iterativamente hasta que la red converge a un punto donde el error de predicción es lo suficientemente bajo (Rumelhart et al. (1986)).

A continuación se detallarán cómo ciertos parámetros del entrenamiento de una red neuronal afectan a este proceso. Primero tenemos el tamaño del conjunto de datos que tiene un impacto en el entrenamiento de una red neuronal. Un conjunto más grande proporciona más información para que la red neuronal aprenda y puede ayudar a mejorar el rendimiento del modelo. Sin embargo, también implica un mayor costo computacional y puede requerir más tiempo de entrenamiento debido al mayor número de ejemplos a procesar. La función de activación juega un papel crucial en una red neuronal, ya que se utiliza para introducir no linealidades en las salidas de las neuronas. Diferentes funciones de activación, como la función sigmoide, la función ReLU o la función tangente hiperbólica, tienen diferentes propiedades y pueden ser más adecuadas para diferentes tipos de problemas. La elección de la función de activación puede afectar tanto al rendimiento del modelo como a la velocidad de convergencia durante el entrenamiento. Es importante seleccionar una función de activación que sea apropiada para el problema específico que se está abordando (Sharma et al. (2020)).

Una época representa una iteración a través de todo el conjunto de datos durante el entrenamiento de la red neuronal. El número de épocas indica cuántas veces se recorrerá el conjunto de datos completo durante el entrenamiento. Un número mayor de épocas permite que el modelo vea más ejemplos y puede ayudar a mejorar el rendimiento al permitir que la red neuronal se ajuste mejor a los datos. Sin embargo, es importante tener en cuenta que un número excesivamente alto de épocas puede llevar al sobreajuste del modelo a los datos de entrenamiento, lo que significa que el modelo se ajusta demasiado específicamente a los ejemplos de entrenamiento y no generaliza bien a nuevos datos. Para evitar el sobreajuste durante el entrenamiento se aplica una técnica llamada *early stopping*, la cual consiste en detener el entrenamiento cuando el rendimiento del modelo en un conjunto de validación deja de mejorar. Esto se logra mediante el seguimiento de una métrica de evaluación, como la precisión o el error, y detener el entrenamiento cuando dicha métrica empeora o deja de mejorar durante un número determinado de épocas consecutivas. El *early stopping* actúa como regularizador y ayuda a encontrar el punto óptimo en el que el modelo tiene un buen rendimiento tanto en los datos de entrenamiento como en los datos no vistos (Aggarwal (2018)).

El optimizador es el algoritmo utilizado para ajustar los pesos de la red neuronal durante el entrenamiento, basado en los gradientes calculados. Hay varios optimizadores disponibles, como el descenso de gradiente

estocástico (SGD) (Amari (1993)), Adam (Kingma & Ba (2014)), RMSProp (Hinton (2012)), entre otros. Cada optimizador tiene diferentes propiedades y puede influir en la velocidad de convergencia y el rendimiento del modelo. Este optimizador puede tener un momentum, el cual ayuda a acelerar el entrenamiento mediante la acumulación de gradientes de pasos anteriores y suaviza las actualizaciones de los pesos. Un valor de momentum alto permite que la red “salte” los mínimos locales y converja más rápidamente hacia el mínimo global, mientras que un valor bajo puede ayudar a evitar oscilaciones en la convergencia (Aggarwal (2018)).

La tasa de aprendizaje es un parámetro crítico que determina qué tan rápido se ajustan los pesos de la red neuronal durante el entrenamiento. Influye en la velocidad de convergencia y la estabilidad del entrenamiento. Una tasa de aprendizaje alta puede llevar a oscilaciones o divergencia en el entrenamiento, mientras que una tasa de aprendizaje baja puede hacer que el entrenamiento sea lento y se quede atrapado en mínimos locales. Encontrar la tasa de aprendizaje adecuada es importante para garantizar un entrenamiento estable y eficiente. Para esto se puede utilizar el *decay* o decaimiento de la tasa de aprendizaje, en la cual se va ajustando la tasa de aprendizaje a medida que avanza el entrenamiento. Se reduce gradualmente a lo largo del tiempo, permitiendo que el modelo realice actualizaciones más pequeñas a medida que se acerca a la convergencia (Patterson & Gibson (2017)).

El tamaño del lote (batch size) se refiere al número de ejemplos de entrenamiento que se utilizan en cada iteración durante el entrenamiento. Un tamaño de lote grande puede acelerar el entrenamiento al aprovechar la paralelización y el rendimiento del hardware, pero también puede requerir más memoria. Por otro lado, un tamaño de lote pequeño puede ayudar a evitar mínimos locales y generalizar mejor, pero puede ralentizar el entrenamiento debido a la mayor frecuencia de actualizaciones de los pesos (Bishop (1995)).

El proceso de inferencia en una red neuronal se refiere a la etapa en la cual se utiliza el modelo entrenado para hacer predicciones sobre nuevos datos de entrada. Antes de realizar la inferencia, es necesario realizar el mismo preprocesamiento de datos que se utilizó durante el entrenamiento. Esto puede incluir pasos como normalización, escalado, transformación de datos, entre otros, con el fin de preparar los datos de entrada de la misma manera en que se hicieron durante el entrenamiento. Después se realiza una propagación hacia adelante, durante este proceso los datos de entrada se pasan a través de la red neuronal capa por capa, desde la capa de entrada hasta la capa de salida. Cada neurona en una capa recibe las salidas de las neuronas de la capa anterior y realiza una operación lineal (producto punto entre los pesos y las entradas) seguida de una función de activación.

Este proceso se repite para todas las capas ocultas hasta llegar a la capa de salida. Una vez que los datos de entrada han sido propagados hacia adelante a través de la red neuronal, se obtienen las salidas en la capa de salida. Estas salidas pueden ser valores numéricos en el caso de problemas de regresión o probabilidades en el caso de problemas de clasificación. Dependiendo del tipo de problema, se puede aplicar una función de activación final en la capa de salida para obtener las predicciones finales.

Las salidas obtenidas en el paso anterior representan las predicciones o clasificaciones realizadas por la red neuronal. Es importante destacar que durante la inferencia, los pesos y los parámetros de la red neuronal se mantienen fijos, ya que la red está utilizando los conocimientos adquiridos durante el entrenamiento para hacer predicciones sobre nuevos datos. La inferencia es un proceso rápido y eficiente, una vez que el modelo ha sido entrenado. En resumen, el proceso de inferencia en una red neuronal implica el preprocesamiento de los datos de entrada, la propagación hacia adelante a través de las capas de la red, la obtención de las salidas y la interpretación de estas salidas para obtener las predicciones deseadas.

2.3.2. Algoritmo de optimización

Un algoritmo de optimización es una herramienta fundamental en varios campos, cuyo objetivo principal es encontrar la mejor solución posible dentro de un espacio de búsqueda determinado, ya sea minimizando o maximizando una función objetivo. El funcionamiento de un algoritmo de optimización se puede describir en pasos generales. En primer lugar, se define el problema de optimización, lo que implica identificar la función objetivo que se desea optimizar y las restricciones asociadas. Por ejemplo, al buscar el mínimo de una función, la función objetivo sería esa función y las restricciones podrían incluir límites en los valores de las variables (Kochenderfer & Wheeler (2019)).

A continuación, se selecciona un algoritmo de optimización adecuado para abordar el problema. Existen diversos algoritmos disponibles, como el descenso de gradiente, el algoritmo genético, el recocido simulado y la optimización por enjambres de partículas, entre otros. Cada algoritmo tiene características y enfoques propios, por lo que la elección depende del tipo de problema y las necesidades específicas. Una vez seleccionado el algoritmo, se inicia el proceso de optimización, que implica iterar a través de pasos con el objetivo de encontrar la solución óptima o una aproximación cercana. Estos pasos pueden variar según el algoritmo, pero generalmente incluyen evaluar la función objetivo en un conjunto de soluciones candidatas, calcular una medida de calidad (como el valor de la función objetivo) y actualizar las soluciones según reglas específicas.

Durante las iteraciones, el algoritmo de optimización busca mejorar gradualmente las soluciones utilizando la información proporcionada por la función objetivo y, en algunos casos, por restricciones adicionales. Esto se logra mediante la aplicación de operadores de búsqueda o técnicas específicas del algoritmo, como mutaciones, cruzamientos o ajustes de parámetros. El proceso continúa hasta que se cumpla un criterio de parada predefinido, que puede ser el número máximo de iteraciones, la convergencia de la solución o la obtención de una solución aceptable. En cada iteración, el algoritmo busca explorar y explotar el espacio de búsqueda de manera eficiente, con el objetivo de encontrar la solución óptima o una aproximación lo más cercana posible.

En resumen, los algoritmos de optimización son herramientas esenciales para encontrar soluciones óptimas o aproximadas en problemas de optimización. Funcionan mediante la iteración a través de un conjunto de soluciones candidatas, evaluando su calidad mediante una función objetivo y actualizando las soluciones según reglas específicas. La elección del algoritmo y su aplicación dependen del tipo de problema y las restricciones asociadas.

2.3.3. Particle swarm optimization

Particle Swarm Optimization (PSO) es un algoritmo de optimización basado en la simulación del comportamiento social de un enjambre de partículas. Se inspira en el comportamiento de los pájaros o peces en busca de comida, donde cada individuo del enjambre ajusta su posición en función de su propia experiencia y de la información compartida por otros individuos cercanos (Kennedy & Eberhart (1995)). El funcionamiento del algoritmo PSO se puede describir en los siguientes pasos. En primer lugar, se inicializa un enjambre de partículas en un espacio de búsqueda. Cada partícula representa una posible solución al problema de optimización y tiene asociadas una posición y una velocidad inicial.

En cada iteración, las partículas actualizan su velocidad y posición siguiendo dos componentes principales: la mejor posición que han alcanzado ellas mismas (*pbest*) y la mejor posición alcanzada por cualquier partícula en el enjambre (*gbest*). Estos valores son utilizados para ajustar la velocidad de la partícula en función de su dirección actual y su experiencia pasada. La fórmula para actualizar la velocidad de la partícula es:

$$v_i(t + 1) = w \cdot v_i(t) + c_1 \cdot r_1 \cdot (pbest_i - x_i(t)) + c_2 \cdot r_2 \cdot (gbest - x_i(t)) \quad (1)$$

La fórmula para actualizar la posición de la partícula es:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (2)$$

donde $v_i(t)$ es la velocidad de la partícula i en la iteración t , $x_i(t)$ es la posición de la partícula i en la iteración t , $pbest_i$ es la mejor posición alcanzada por la partícula i hasta el momento, $gbest$ es la mejor posición alcanzada por cualquier partícula en el enjambre, w es el factor de inercia que controla la influencia de la velocidad anterior, c_1 y c_2 son coeficientes de aprendizaje que controlan la influencia de $pbest$ y $gbest$ respectivamente, y r_1 y r_2 son valores aleatorios entre 0 y 1.

En cada iteración, las partículas actualizan su velocidad y posición, y se evalúa la calidad de las soluciones correspondientes. Si una partícula encuentra una posición mejor que su $pbest$, actualiza su $pbest$ personal. De manera similar, si alguna partícula encuentra una posición mejor que el $gbest$ global, se actualiza el $gbest$ del enjambre. El proceso de actualización de velocidades y posiciones se repite durante un número determinado de iteraciones o hasta que se cumpla un criterio de convergencia predefinido, como alcanzar un umbral de precisión o un número máximo de evaluaciones de la función objetivo. A continuación, se muestra el pseudocódigo del PSO implementado en este trabajo:

Algorithm 1 Particle Swarm Optimizacion (PSO)

```

Inicializar población
for  $t \leftarrow 1$  : generación máxima do
  for  $i \leftarrow 1$  : tamaño de la población do
    if  $f(x_{i,d}(t)) < pbest_i$  then
       $pbest_i = x_{i,d}(t)$ 
       $gbest = \min_i pbest_i$ 
    end if
    for  $d \leftarrow 1$ : dimensión do
       $v_{i,d}(t+1) = w \cdot v_{i,d}(t) + c_1 \cdot r_1 \cdot (pbest_i - x_{i,d}(t)) + c_2 \cdot r_2 \cdot (gbest - x_{i,d}(t))$ 
       $x_i(t+1) = x_i(t) + v_i(t+1)$ 
      if  $x_{i,d}(t+1) > x_{max}$  then
         $x_{i,d}(t+1) = x_{max}$ 
      else if  $x_{i,d}(t+1) < x_{min}$  then
         $x_{i,d}(t+1) = x_{min}$ 
      end if
    end for
  end for
end for

```

En resumen, Particle Swarm Optimizacion (PSO) es un algoritmo de optimización que simula el comportamiento de un enjambre de partículas para buscar soluciones óptimas. Cada partícula ajusta su velocidad

y posición en función de su experiencia personal y de la información compartida por el enjambre. Este enfoque de búsqueda cooperativa permite que las partículas exploren el espacio de búsqueda de manera eficiente y encuentren soluciones prometedoras.

2.3.4. Transferencia de aprendizaje

La transferencia de aprendizaje en redes neuronales es un enfoque en el aprendizaje automático que implica aprovechar el conocimiento adquirido por una red neuronal entrenada en una tarea para mejorar el rendimiento en otra tarea relacionada. En lugar de entrenar una red neuronal desde cero para una tarea específica, se utiliza una red pre-entrenada en una tarea previa como punto de partida. La idea fundamental detrás de la transferencia de aprendizaje es que las representaciones aprendidas por una red neuronal en una tarea pueden ser útiles para resolver problemas relacionados. Al entrenar una red neuronal en una tarea compleja y luego aplicarla a una tarea similar, se puede acelerar el proceso de entrenamiento y lograr un mejor rendimiento con menos datos de entrenamiento, si las tareas tuvieran poco en común, la transferencia de aprendizaje podría no resultar beneficiosa (Zhuang et al. (2021)).

El proceso de transferencia de aprendizaje generalmente implica dos etapas: la etapa de pre-entrenamiento y la etapa de ajuste fino. En la etapa de pre-entrenamiento, se entrena una red neuronal en un conjunto de datos grande y diverso, generalmente utilizando tareas de clasificación de imágenes como el reconocimiento de objetos en imágenes. Durante este entrenamiento, la red neuronal aprende a extraer características relevantes de las imágenes y a reconocer patrones importantes. En la etapa de ajuste fino, la red pre-entrenada se toma como punto de partida y se realiza un entrenamiento adicional en un conjunto de datos más pequeño y específico para la tarea de interés. En esta etapa, las capas superiores de la red se adaptan para la tarea específica, mientras que las capas inferiores (que han aprendido características generales) se mantienen fijas o se ajustan de manera más conservadora. Esto permite que la red se especialice en la nueva tarea sin perder el conocimiento previo adquirido en la etapa de pre-entrenamiento.

La transferencia de aprendizaje se ha utilizado con éxito en una amplia gama de aplicaciones, como el reconocimiento de objetos, el procesamiento del lenguaje natural, la traducción automática y muchas otras (Weiss et al. (2016)). Al aprovechar el conocimiento previo de una red pre-entrenada, la transferencia de aprendizaje permite el desarrollo de modelos más eficientes y precisos, especialmente cuando los conjuntos de datos de entrenamiento son limitados.

2.4. Resumen

En la actualidad la investigación sobre el magnetismo estelar se basa en el análisis de los parámetros de Stokes observados en estrellas. Mediante la E RTP, es posible modelar teóricamente las distintas configuraciones del campo magnético estelar. Al ajustar los perfiles observados con los obtenidos mediante la E RTP, se puede determinar la configuración del campo magnético y/o las características atmosféricas de la estrella.

Aunque todavía hay pocos trabajos en esta área de investigación, se han desarrollado diversos modelos para realizar regresiones de las características atmosféricas y/o la configuración magnética a partir de los parámetros de Stokes del Sol, nuestra estrella. Estos estudios se enfocan en obtener los parámetros de Stokes de una región muy pequeña del disco solar, lo que permite una resolución espacial detallada. En contraste a los artículos que se centran en estrellas más allá del Sol, obtienen los parámetros de Stokes del disco solar completo (integrado), lo que resulta en una falta de resolución espacial en estas observaciones, ya que todas las estrellas se consideran objetos puntuales en los telescopios.

Capítulo 3. Metodología

En el contexto de esta tesis es importante recordar que el objetivo es recuperar la configuración magnética de una estrella dados los parámetros de Stokes observados, utilizando un algoritmo de optimización en conjunto con una red neuronal densa para la estimación de los parámetros de Stokes. En este flujo de trabajo se establece una secuencia clara de pasos, desde la generación de datos sintéticos hasta la obtención de la configuración magnética estimada. Por lo que resulta de vital importancia comprender el flujo de datos.

Como etapa inicial, se genera un conjunto de datos sintéticos, como se describe en la sección 3.1, el cual se utilizará para el entrenamiento de una red neuronal densa encargada de la estimación de los parámetros de Stokes. Una vez que la red neuronal densa ha sido entrenada, se empleará en conjunto con un algoritmo de optimización, con el objetivo de llevar a cabo la estimación de los parámetros de Stokes para cada solución candidata. Cada solución candidata representa una configuración magnética en diferentes fases de rotación de la estrella. Estas distintas fases de rotación de la estrella se conocerán posteriormente, lo que permitirá inferir únicamente los 7 atributos (fuerza del momento dipolar, tres ángulos de euler, ángulo de inclinación y las dos coordenadas de la posición del dipolo) de la configuración magnética. Para llevar a cabo la optimización, se comparará el error cuadrático medio (MSE) entre los parámetros de Stokes estimados por la red neuronal densa, previamente entrenada, en las diferentes fases de rotación, y los parámetros de Stokes observados en esas mismas fases de rotación de una estrella específica. Al finalizar este proceso, se obtendrá la configuración del campo magnético correspondiente. El flujo de este proceso se representa visualmente en la Figura 3.

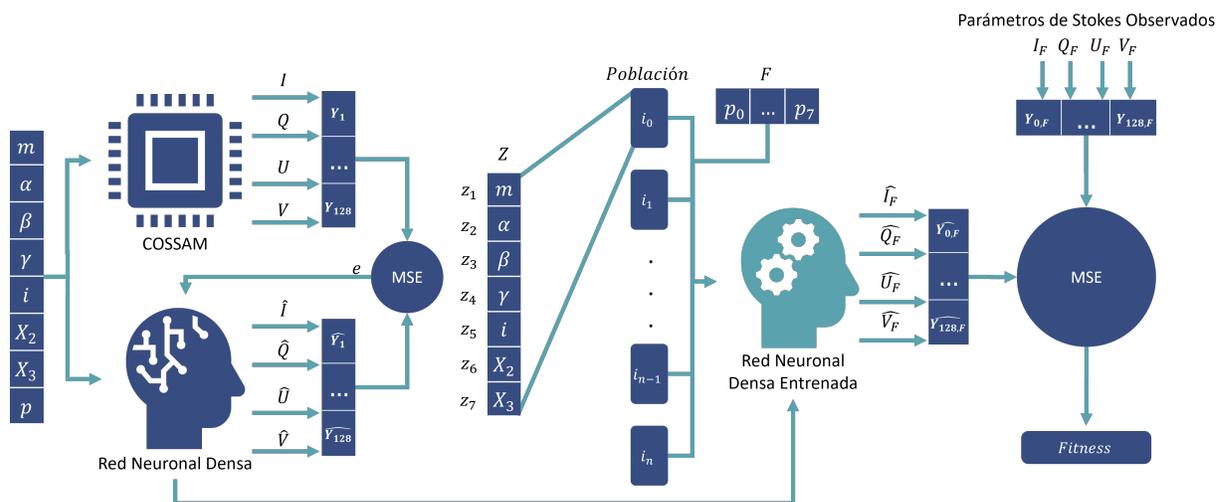


Figura 3. Diagrama del proceso de entrenamiento de la red neuronal densa (lado izquierdo) y la optimización (lado derecho).

3.1. Generación de datos sintéticos

En esta investigación se utilizó el programa COSSAM (Stift (2000)) para generar un conjunto de datos sintéticos con el objetivo de entrenar modelos para la estimación de los parámetros de Stokes utilizando este conjunto de datos sintéticos como los valores de verdad (ground truth). COSSAM emplea la E RTP a partir de la abundancia química y los parámetros atmosféricos del sol, el modelo del campo magnético es dipolar, y puede no estar centrado el dipolo (modelo dipolar descentrado) y una longitud de onda reducida para obtener una sola firma en la señal.

Se generaron un total de 3,300,000 de casos, cada una con 8 atributos que describen la configuración magnética de la configuración dipolar. La fuerza del momento dipolar (m), tres ángulos de Euler (α, β, γ) que describen la geometría de la estrella, el ángulo de inclinación (i) de la rotación de la estrella, dos coordenadas (X_2, X_3) que describen la posición del dipolo magnético dentro de la estrella, y la fase de rotación de la estrella (p) (Stift (1975)). Cada caso de los 8 atributos de entrada generó 128 atributos de salida, que conforman los parámetros de Stokes I, Q, U y V, con 32 atributos o puntos, estos puntos forman una señal discreta que representa cada parámetro de Stokes. Se detallan los rangos de los atributos de entrada en la Tabla 1, los cuales fueron utilizados para generar el conjunto de datos, y de los cuales se seleccionó una combinación aleatoria de 8 atributos para generar cada instancia. El conjunto de datos se ha dividido en tres rangos según la fuerza del momento dipolar: alto (100-5010), bajo (10-100) y muy bajo (0.1-10). Para más detalles en el capítulo 4.

Tabla 1. Intervalos para los 8 atributos que describen la geometría magnética de la configuración dipolar.

Atributo:	m	α	β	γ	i	X_2	X_3	p
Valor mínimo	0.1	-180.0	0.0	-180.0	0.0	0.0	0.0	0.0
Valor máximo	5010.0	180.0	180.0	180.0	180.0	0.2	0.2	1.0

En la Figura 4 se presenta un ejemplo de los perfiles de Stokes generados con los siguientes valores: $m = 891.3, \alpha = 71.5, \beta = 44.0, \gamma = -61.7, i = 54.9, X_2 = 0.0, X_3 = 0.0, p = 0.52$, y en la que cada perfil es un parámetro de Stokes, que serán la salida del primer modelo de estimación. Se puede observar una característica importante en los parámetros de Stokes, donde la amplitud de Stokes I es mucho mayor que la de Stokes V, y a su vez, la amplitud de Stokes V es mayor que la de los Stokes Q y U. Entre los cuatro parámetros de Stokes, el más importante es el Stokes V, ya que se trata del parámetro de Stokes polarizado con mayor amplitud (el Stokes I no es polarizado). Esto implica que la amplitud del Stokes V tiende a estar por encima del nivel de ruido en las observaciones astronómicas.

Por lo tanto, se considera que el parámetro de Stokes V es de particular importancia en el contexto de esta investigación, ya que su señal polarizada suele ser más pronunciada y detectable en comparación con los demás parámetros de Stokes, aun así existen observaciones astronómicas del Stokes Q y U, sobre todo en estrellas con un campo magnético fuerte.

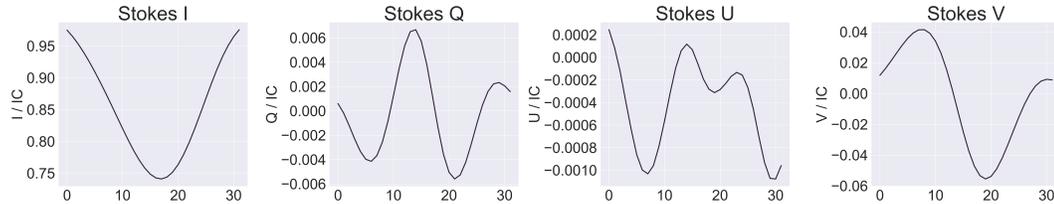


Figura 4. Ejemplo de los parámetros de Stokes generado por COSSAM.

En el caso de los atributos de la configuración del campo magnético, el más importante de ellos es la fuerza del momento dipolar m debido a su gran influencia en la amplitud de los parámetros de Stokes. Además, se consideran el ángulo de inclinación i y las coordenadas X_2 y X_3 , que describen la posición del dipolo dentro de la estrella, las coordenadas X_2 y X_3 serán analizadas como un vector al que llamaremos X_r . Por otro lado, los ángulos α , β y γ , que describen la geometría de la estrella, son menos relevantes, ya que diferentes combinaciones de estos tres ángulos podrían resultar en la misma geometría. Por último, la fase de rotación de la estrella, la cual se conoce de antemano en el proceso de recuperación del campo magnético, por lo que no se recuperara en el algoritmo de optimización. Por tanto, es esencial realizar un proceso de preprocesamiento de los datos para evitar que el modelo se sesgue solo hacia el Stokes I o para mejorar la estimación de los campos magnéticos grandes así como los campos magnéticos pequeños, con el objetivo de lograr un rendimiento óptimo en las distintas intensidades del campo magnético.

3.2. Preprocesamiento de los datos

En esta investigación, se emplearon cuatro métodos de escalamiento distintos tanto para los atributos de entrada (8 atributos) como para los datos de salida (32 puntos por cada parámetro de Stokes). Uno de estos métodos fue el escalamiento estándar (Pedregosa et al. (2011)), el cual consiste en normalizar o estandarizar de forma individual los atributos de la matriz de entrada X , donde cada renglón es una instancia y cada columna un atributo. El objetivo principal de este escalamiento es lograr que los atributos tengan un promedio de cero y una desviación estándar de uno. A continuación, se presenta la ecuación correspondiente a este procedimiento:

$$\hat{x}_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (3)$$

donde \hat{x}_{ij} representa el valor escalado del atributo x_{ij} en la matriz X , μ_j es el promedio del atributo j y σ_j es su desviación estándar.

Se empleó otro método de escalado denominado Min-Max (Pedregosa et al. (2011)), el cual tiene como objetivo escalar cada atributo dentro de un rango específico, en este caso, de 0 a 1. A continuación, se presenta la ecuación que representa este método:

$$\hat{x}_{ij} = \frac{x_{ij} - \text{mín}_j(X)}{\text{máx}_j(X) - \text{mín}_j(X)} \quad (4)$$

donde \hat{x}_{ij} representa el valor escalado del atributo x_{ij} en la matriz X , $\text{mín}_j(X)$ es el valor mínimo del atributo j en la matriz X , y $\text{máx}_j(X)$ es el valor máximo del atributo j en la matriz X .

Mediante esta ecuación, cada atributo se transforma de manera proporcional dentro del rango deseado, lo que facilita la comparación y el análisis de los datos en un contexto uniforme.

El tercer método de escalado utilizado es el Max-Abs (Pedregosa et al. (2011)), el cual tiene como objetivo escalar cada atributo en función del valor absoluto máximo. Este método se aplica para asegurar que todos los atributos estén dentro del rango $[-1, 1]$. A continuación, se muestra la ecuación que representa este método:

$$\hat{x}_{ij} = \frac{x_{ij}}{\text{máx}_j(|X|)} \quad (5)$$

donde \hat{x}_{ij} representa el valor escalado del atributo x_{ij} en la matriz X , y $\text{máx}_j(|X|)$ es el valor absoluto máximo de todos los atributos en la matriz X .

Al utilizar esta ecuación, cada atributo se escalamos de forma proporcional en relación con su valor absoluto máximo, lo que permite comparar y analizar los datos en un rango normalizado y consistente. Este enfoque resulta útil cuando se desea preservar la información relativa a la magnitud de los atributos, sin distorsionar sus proporciones originales.

Y por último el método de escalado utilizado es el transformador de cuantiles (Pedregosa et al. (2011)), el cual se utiliza para transformar los datos de manera que sigan una distribución uniforme. Esta técnica

es especialmente útil cuando los datos presentan desviaciones importantes de la distribución normal. El Transformador de Cuantiles asigna a cada valor de entrada un valor correspondiente en una distribución de probabilidad uniforme. Esto se logra calculando el rango de los cuantiles de los datos originales y luego asignando los valores transformados basados en estos cuantiles. A continuación se muestra la ecuación que representa este método:

$$\hat{x}_{ij} = F(x_{ij}) \quad (6)$$

donde \hat{x}_{ij} representa el valor escalado del atributo x_{ij} en la matriz X , y $F(x_{ij})$ es la función de distribución acumulativa (CDF) que asigna a x_{ij} su correspondiente valor en la distribución uniforme.

Al aplicar el transformador de cuantiles, se logra que los datos estén distribuidos de manera más uniforme.

3.3. Modelo de estimación de parámetros de Stokes

Para este problema, resulta imprescindible emplear un modelo computacional con capacidad para llevar a cabo una regresión multi-salida con el fin de obtener estimaciones de cada uno de los puntos de cada parámetro de Stokes. Por lo que se determinó que una red neuronal se ajusta para este problema. En el contexto más específico del flujo de la red neuronal densa para la estimación de parámetros de Stokes, representado en la Figura 5, se lleva a cabo un proceso dividido en etapas. Primero, se separa el conjunto de datos en tres conjuntos distintos: 75 % para entrenamiento, 15 % para validación y 10 % para pruebas.

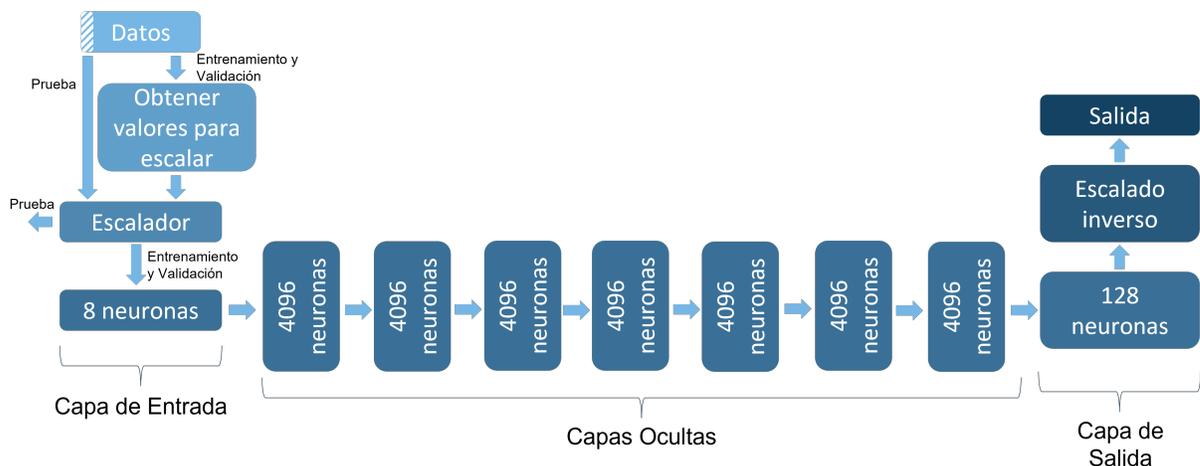


Figura 5. Diagrama del flujo de la red neuronal densa.

El conjunto de entrenamiento y validación se utiliza para obtener los valores necesarios para escalar los datos. Una vez escalados, se aplica este escalado a todo el conjunto de datos, tanto en la entrada como en la salida. Posteriormente, se utiliza el conjunto de entrenamiento y validación para entrenar la red neuronal densa. Una vez que ha sido entrenada, se realiza un escalado inverso para regresar al dominio original del conjunto de datos. Este último paso permite obtener la salida final del modelo. La implementación de la arquitectura y su entrenamiento se encuentra en el siguiente enlace: <https://github.com/DragonTaiko/stokes-parameters-magnetic-field>. La arquitectura de la red neuronal densa utilizada se muestra en la Figura 5.

Esta arquitectura consta de varias capas que han sido diseñadas de manera específica para abordar el problema de estimación del campo magnético. En la capa de entrada se han incluido 8 neuronas, ya que corresponden a los 8 atributos que definen el campo magnético. Además de 7 capas ocultas, cada una de ellas con 4096 neuronas y 128 neuronas de salida. La elección de la cantidad de capas y el número de neuronas por capa se realizó de manera empírica. Se tuvo en cuenta el error promedio en los cuatro parámetros de Stokes obtenido al comparar distintas arquitecturas de modelos de redes neuronales densas. Asimismo, se consideró especialmente el error en el Stokes V, dado su mayor importancia entre los parámetros de Stokes.

Tabla 2. Parámetros del entrenamiento de la red neuronal para la estimación de los parámetros de Stokes.

Atributo	Valor
Tamaño del Conjunto de Datos	50,000
Momentum	0.95
Función de Activación	ReLu
Epoas	1000
Optimizador	Descenso del Gradiente Estocástico
Early Stopping	25
Tasa de Aprendizaje*	$1e - 1$ (rango alto y bajo) y $1e - 2$ (rango muy bajo)
Batch Size	1024
Decaimiento	$1/(2 * dataset_size)$

* Las tasas de aprendizaje son variables en los experimentos de escalado

En la sección 4.2, se proporcionan más detalles sobre la selección de esta arquitectura específica. Se describen los criterios utilizados para determinar el número de capas ocultas, la cantidad de neuronas por capa y cómo se tuvieron en cuenta los errores promedio en los parámetros de Stokes durante este proceso de selección de la arquitectura. Los parámetros utilizados en los experimentos en el modelo de la estimación de parámetros de Stokes se encuentran detallados en la Tabla 2, la cual se aplicó a todos los experimentos, salvo algunas excepciones que se especifican en el pie de la tabla. Debido a restric-

ciones de tiempo, se decidió trabajar con un conjunto de datos de 50,000 casos para los experimentos que no impliquen el análisis del modelo cuando el conjunto de datos crece. Basándose en los resultados de experimentos empíricos, se implementó un enfoque de decaimiento de peso dinámico. Este enfoque involucra la variación del decaimiento de peso en función del tamaño del conjunto de datos. El decaimiento de peso desempeña un papel importante en el proceso de entrenamiento, y un enfoque dinámico permite la adaptación del decaimiento de peso a las características específicas del conjunto de datos. Se espera que estos parámetros mejoren la eficacia general del proceso de entrenamiento y la capacidad de generalización de la red.

3.4. Modelo de recuperación del campo magnético

En el modelo de optimización para la recuperación del campo magnético se presenta en la Figura 6. En primer lugar, se inicializan las soluciones candidatas asignándoles valores aleatorios dentro de los intervalos correspondientes vistos en la Tabla 1, cada solución candidata se conforma de 7 atributos de la configuración del campo magnético ($Z : [z_1 : m, z_2 : \alpha, z_3 : \beta, z_4 : \gamma, z_5 : i, z_6 : X_2, z_7 : X_3]$) y 7 fases de rotación conocidas de la estrella ($F : [p_0, \dots, p_7]$). Luego, se calcula la función objetivo ($f(Z)$) mediante la red a red neuronal entrenada para estimar los parámetros de Stokes de dicha configuración para cada fase de rotación, con la cual obtenemos el error entre los parámetros de Stokes de la solución candidata y los parámetros de Stokes de una estrella observada, de la cual se desea inferir el campo magnético.

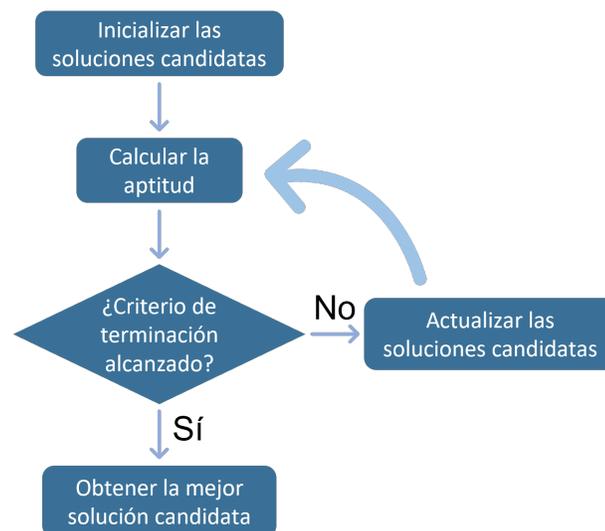


Figura 6. Flujo de optimización para la recuperación del campo magnético.

En este punto, la red neuronal entrenada juega un papel crucial, ya que se encarga de estimar los parámetros de Stokes de cada solución candidata. La red neuronal realiza esta tarea al estimar los valores de los parámetros de Stokes basándose en la configuración magnética asociada dada la solución candidata para cada fase de rotación. El proceso de optimización continúa hasta que se alcanza la condición de terminación predefinida. En ese momento, se obtiene la solución que mejor se asemeja en términos de los parámetros de Stokes y la configuración magnética subyacente.

Cabe destacar que para la recuperación del campo magnético de una estrella se le aportará los parámetros de Stokes de la estrella en distintas fases de rotación, esto para reducir la ambigüedad entre distintas configuraciones del campo magnético, estas rotaciones ya son conocidas previamente por lo que el algoritmo de optimización se encargará de recuperar los otros 7 valores que pertenecen a la configuración del campo magnético. En el modelo de la recuperación del campo magnético se seleccionó el algoritmo de optimización denominado Particle Swarm Optimization (PSO). En la sección 4.6, se llevará a cabo una discusión exhaustiva sobre las comparaciones realizadas con otros algoritmos, así como los fundamentos detrás de su elección.

Los detalles específicos de los parámetros utilizados para el algoritmo PSO se presentan en la Tabla 3 y la implementación del código se encuentra en <https://github.com/DragonTaiko/stokes-parameters-magnetic-field>. Para esta investigación, se optó por utilizar la versión clásica del PSO con inercia, considerando su capacidad para abordar problemas de optimización complejos. Además, se creó una muestra utilizando el método COSSAM, compuesta por 100 configuraciones del campo magnético. Cada configuración incluye 7 fases de rotación y los parámetros de Stokes correspondientes a cada fase. Se utilizan distintas fases de rotación de la estrella con el fin de obtener información más completa y precisa. Estas configuraciones del campo magnético se utilizaron como los valores de referencia (ground truth) a recuperar mediante el ajuste de los parámetros de Stokes de las 7 fases con las estimaciones proporcionadas por la red neuronal densa entrenada. Este proceso de recuperación del campo magnético se llevó a cabo mediante el uso de PSO y sus soluciones candidatas.

Tabla 3. Parámetros del entrenamiento de la red neuronal para la estimación de los parámetros de Stokes.

Atributo	Valor
Tamaño de la Población	2048
Inercia	0.7
Componente Personal (c_1)	1
Componente Social (c_2)	1
Inicialización	Aleatoria
Iteraciones	50
Condición de Terminación	Terminar las iteraciones

A continuación, se encuentra la formulación del problema de optimización:

$$\begin{aligned} & \text{Minimizar : } f(Z, F) \\ & f(Z, F) = \text{MSE}(Y_F, \hat{Y}_F) \\ & z_i \geq \text{mín}(X_{\text{cossam},i}) \quad \text{y} \quad z_i \leq \text{máx}(X_{\text{cossam},i}) \end{aligned}$$

donde Y_F es el vector que contiene el perfil de cada parámetro de Stokes de cada fase de rotación, \hat{Y}_F es la predicción de la estimación de los parámetros de Stokes de cada fase de rotación, $X_{\text{cossam},i}$ es el i -ésimo valor de la configuración del campo magnético y z_i es la predicción del valor de la configuración del campo magnético. La función objetivo depende de los 7 atributos del campo magnético de la solución candidata (Z) así como las 7 fases de rotación (F).

La función objetivo consiste en la minimización del MSE en el espacio escalado de cada parámetro de Stokes en cada fase de rotación. Además, se considera como restricciones en los valores mínimos y máximos para la configuración magnética, los cuales deben corresponder a los intervalos establecidos en la Tabla 1.

3.5. Métricas de evaluación

En este estudio, se emplearán dos métricas de rendimiento, el Error Cuadrático Medio (MSE) y el Error Porcentual Medio Absoluto Ponderado (WMAPE), para evaluar el rendimiento del modelo entre varias configuraciones y encontrar el mejor modelo. A continuación se presentan las ecuaciones para cada métrica:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad \text{y} \quad \text{WMAPE} = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{\sum_{i=1}^n |Y_i|} \quad (7)$$

donde n es el número de puntos o muestras, Y_i es el valor verdadero y \hat{Y}_i es la predicción del modelo para cada punto i .

MSE penaliza en mayor medida los errores grandes, facilitando la comparación de diferentes configuraciones del modelo en el mismo conjunto de datos. Por otro lado, WMAPE es una métrica útil para comparar las predicciones del modelo independientemente de la magnitud de los valores. Como métrica basada en porcentajes, WMAPE se utilizará para comparar la precisión de los valores predichos con los valores verdaderos entre parámetros de Stokes o cuando los datos tengan distinta escala.

Es importante recalcar que el cálculo del WMAPE promedio implica dividir el WMAPE de cada parámetro de Stokes entre la cantidad total de parámetros de Stokes. Esto se emplea para garantizar que los cambios de escala en los parámetros de Stokes no tengan un impacto desproporcionado y que todos los parámetros sean considerados con igual peso. A través del código *COSSAM*, se proporcionan combinaciones de valores de los 8 atributos del campo magnético (X_{cossam}), y a cambio, se obtienen los parámetros de Stokes I, Q, U y V (Y). La red neuronal densa se encarga de estimar los parámetros de Stokes (\hat{Y}), y se calcula el error entre Y y \hat{Y} utilizando las métricas previamente descritas para evaluar las estimaciones. En el caso de la recuperación del campo magnético mediante *COSSAM* se cuentan con los valores del campo magnético a inferir en las 7 fases distintas (X_{cossam}), así como los parámetros de Stokes correspondientes a cada fase (Y_F). El objetivo de PSO es ajustar los valores del campo magnético para reducir el error entre Y_F y \hat{Y}_F . Al finalizar, la salida de PSO será la predicción del campo magnético (Z).

3.6. Resumen

En resumen, primero se genera un conjunto de datos sintéticos que se utilizará para entrenar una red neuronal densa encargada de estimar los parámetros de Stokes. Una vez entrenada, esta red neuronal se empleará junto con un algoritmo de optimización como parte de la función objetivo para estimar los parámetros de Stokes de cada solución candidata, donde cada solución candidata representa una configuración magnética en diferentes fases de rotación de una estrella. Se compara el error cuadrático medio entre los parámetros de Stokes estimados por la red neuronal densa y los parámetros de Stokes observados en cada fase de rotación de la estrella y al finalizar el proceso, se obtiene la configuración del campo magnético correspondiente.

En la generación de datos sintéticos se realizó utilizando el programa *COSSAM*. Se generaron 3.3 millones de casos con 8 atributos que describen la geometría magnética de la configuración dipolar. Cada caso generó 128 atributos de salida, que representan los parámetros de Stokes I, Q, U y V. Se seleccionó un método de escalamiento tanto para los datos de entrada como para los de salida de entre cuatro métodos de escalamiento probados para evitar sesgos en los parámetros de Stokes.

Capítulo 4. Resultados

En este capítulo, se presentarán los resultados obtenidos en el presente estudio. Se comenzará con la selección del método de escalado que brinde un mejor rendimiento al modelo de estimación de parámetros de Stokes. A continuación, se mostrarán los experimentos preliminares realizados para la elección de la arquitectura de la red neuronal densa. Luego, se abordará la decisión sobre si la red neuronal densa debe estimar los cuatro parámetros de Stokes de manera conjunta (modelo general) o si se deben entrenar redes neuronales especializadas para cada parámetro (modelo especializado). Posteriormente, se llevará a cabo un análisis del tamaño del conjunto de datos y su impacto en el rendimiento del modelo. Finalmente, se presentarán los resultados finales del modelo de estimación de los parámetros de Stokes, así como un análisis exhaustivo de los mismos.

Posteriormente, la red neuronal densa entrenada para la estimación de los parámetros de Stokes se utilizará como parte de la función objetivo de un algoritmo de optimización. Para seleccionar el algoritmo más adecuado, se realizaron experimentos preliminares con diferentes opciones. Una vez seleccionado el algoritmo, se procedió a la recuperación del campo magnético y se analizaron los resultados obtenidos. Por último, se llevó a cabo una transferencia de aprendizaje y re-entrenamiento de la red neuronal densa para la estimación de los parámetros de Stokes en otra longitud de onda de la misma estrella. Se determinó el tamaño del nuevo conjunto de datos para el re-entrenamiento del modelo en un experimento, y en otro se determinó la cantidad de capas que debían ser re-entrenadas. Finalmente, se realizó un análisis comparativo de los resultados de la estimación de los parámetros de Stokes y la recuperación del campo magnético entre el modelo re-entrenado y el modelo base.

Durante los experimentos preliminares, se observó que el modelo presentaba un sesgo hacia fuerzas de momentos dipolares grandes (m). Como resultado, se tomó la decisión de segmentar los datos en tres categorías basadas en el rango de fuerzas de momentos dipolares: alto (100-5010), bajo (10-100) y muy bajo (0.1-10). Esta segmentación permitió obtener un mejor rendimiento del modelo en cada rango. El hardware utilizado para llevar a cabo los experimentos preliminares y los modelos finales contaba con una memoria RAM de 32 GB, un procesador Ryzen 5 5600X, y una tarjeta gráfica RTX 3070.

4.1. Selección del método de escalado

En este experimento se evaluaron los cuatro métodos de escalado (ver subsección 3.2) tanto para la entrada (fuerza del momento dipolar, tres ángulos de euler, ángulo de inclinación, dos coordenadas de la

posición del dipolo y la fase de rotación) como para la salida (32 puntos de cada parámetro de Stokes), lo que resultó en un total de 16 modelos diferentes. Estos modelos fueron entrenados y evaluados utilizando un subconjunto de los datos que constaba de 50,000 casos, con el propósito de determinar el método de escalado óptimo tanto para la entrada como para la salida del modelo de cada rango.

4.1.1. Rango alto

La Figura 7 muestra los resultados de combinar los distintos métodos de escalamiento a partir de un modelo de estimación base, ver subsección 3.3. Para mejorar el análisis y la visualización se removieron los datos atípicos. En el eje X se agrupan los métodos de escalamiento sobre los datos de salida, mientras que las series de colores corresponden a los escalados sobre los datos de entrada. En esta representación visual, se destaca claramente que el método Standard de escalado en la salida supera a los demás en términos del promedio de WMAPE de todos los parámetros de Stokes (extrema izquierda de la Figura 7). Por lo tanto, se recomienda utilizar este método de escalado para la salida del modelo.

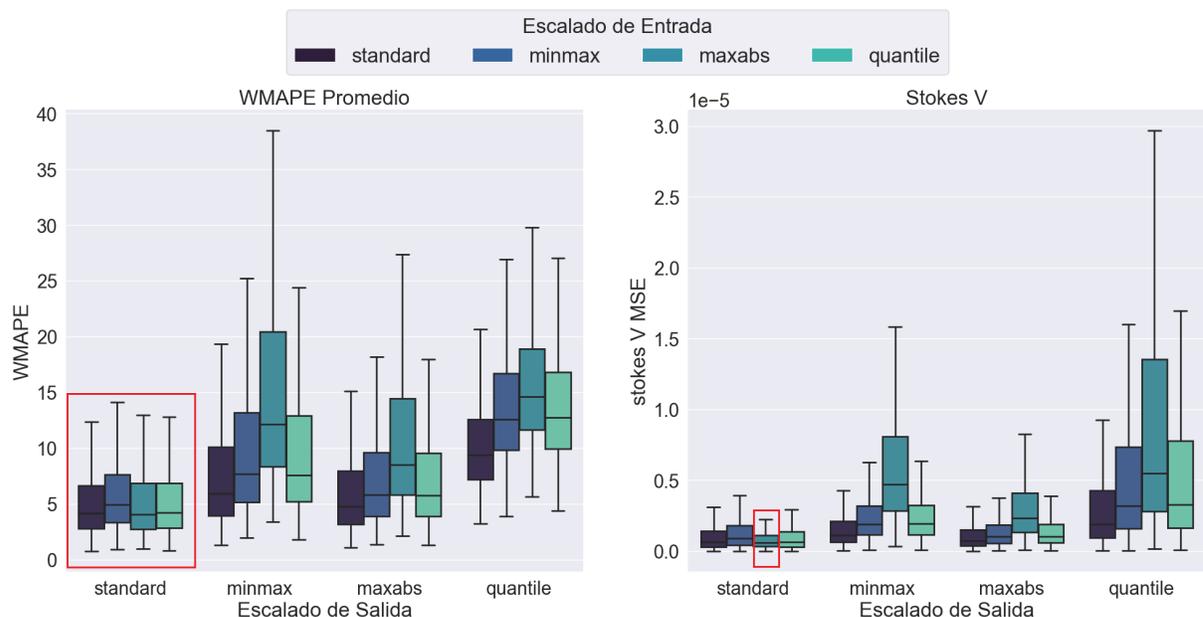


Figura 7. WMAPE promedio de los parámetros de Stokes y el MSE del Stokes V de las distintas combinaciones de los métodos de escalado para la entrada y la salida de los datos para el rango alto.

En cuanto al escalado en la entrada, no se observa una diferencia significativa entre los diferentes métodos. Sin embargo, para complementar esta información, la gráfica de la derecha en la Figura 7,

muestra el error cuadrático medio (MSE) para el parámetro de Stokes V, que es el parámetro más relevante en nuestra investigación. Una vez más, se observa una mejora similar con el método de escalado Standard en la salida. No obstante, en el caso del escalado de entrada, se aprecia una diferencia más notable con el método de escalado Max-Abs (resaltado en rojo).

Al examinar la Tabla 4, se puede observar que el modelo que utiliza el escalado de entrada Max-Abs junto con el escalado Standard en la salida obtiene un mejor rendimiento en términos del MSE de cada parámetro de Stokes, así como en el tercer cuartil; se utilizara continuamente el tercer cuartil en los errores debido a que hay ciertos casos donde el error es muy grande por lo que la desviación estándar no refleja adecuadamente la mayoría de los casos. Por lo tanto, con los resultados se recomienda seleccionar el método de escalado Max-Abs para la entrada y el método de escalado Standard para la salida.

Tabla 4. Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil de combinaciones de escalado de entrada y salida para el rango alto.

Escalado de Entrada	Escalado de Salida	Stokes I MSE (Q3)	Stokes Q MSE (Q3)	Stokes U MSE (Q3)	Stokes V MSE (Q3)	WMAPE promedio (Q3)
Standard	Standard	1.2e-06 (1.3e-06)	2.3e-07 (2.3e-07)	1.6e-07 (1.5e-07)	1.2e-06 (1.4e-06)	6.41 (6.64)
Standard	Max-Abs	1.2e-05 (1.6e-05)	2.4e-07 (2.3e-07)	1.9e-07 (1.7e-07)	1.2e-06 (1.4e-06)	9.52 (7.94)
Max-Abs	Standard	1.0e-06 (1.2e-06)	1.7e-07 (1.7e-07)	1.5e-07 (1.3e-07)	1.0e-06 (1.1e-06)	8.33 (6.83)
Quantile	Standard	1.1e-06 (1.3e-06)	2.1e-07 (2.2e-07)	1.5e-07 (1.4e-07)	1.1e-06 (1.3e-06)	6.85 (6.84)

4.1.2. Rango bajo

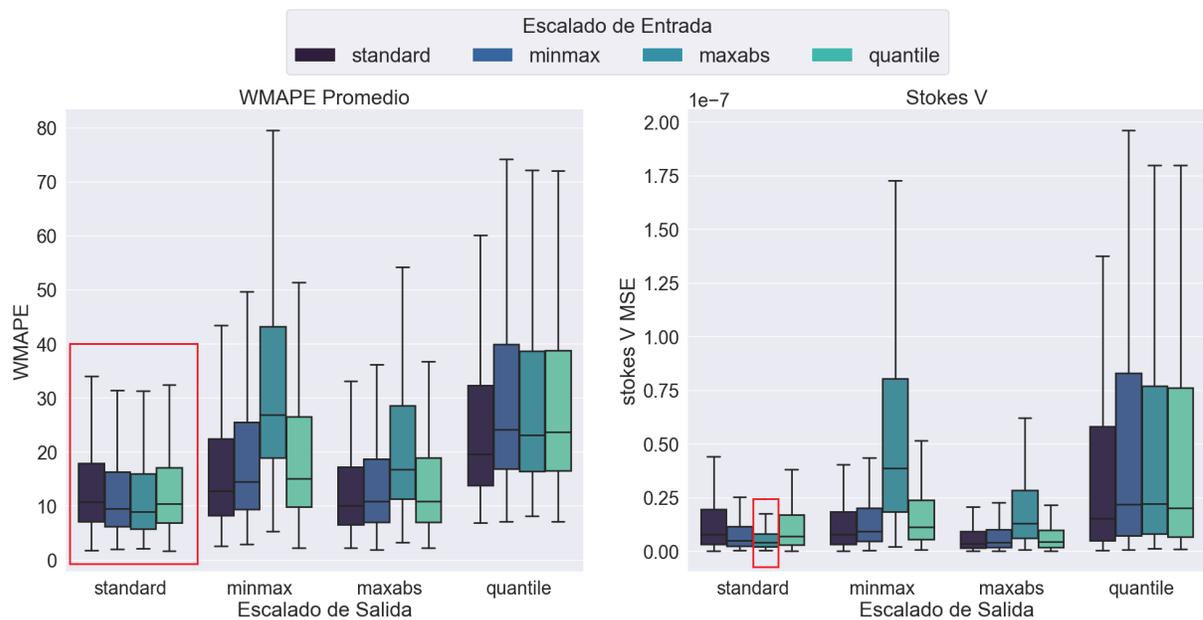


Figura 8. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas combinaciones de los métodos de escalado para la entrada y la salida de los datos para el rango bajo.

En el caso del rango bajo mostrado en la Figura 8, se observa un patrón similar al rango alto. Del lado izquierdo, se muestra que el escalado Standard en la salida es el mejor rendimiento en términos del WMAPE promedio de los parámetros de Stokes. Al examinar el MSE de Stokes V en el lado derecho de la Figura 8, se observa que no todas las combinaciones con el escalado Standard en la salida logran un error menor que otras opciones de escalado de salida. Sin embargo, destaca el escalado Max-Abs para la entrada y el escalado Standard para la salida (resaltado en rojo), especialmente al analizar la Tabla 5, donde se evidencia un menor MSE promedio por parámetro de Stokes, así como un menor promedio de WMAPE. Con base en estos resultados, se decidió seleccionar el escalado de entrada Max-Abs y el escalado Standard en la salida, de manera similar al rango alto.

Tabla 5. Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil de combinaciones de escalado de entrada y salida para el rango bajo.

Escalado de Entrada	Escalado de Salida	Stokes I MSE (Q3)	Stokes Q MSE (Q3)	Stokes U MSE (Q3)	Stokes V MSE (Q3)	WMAPE promedio (Q3)
Standard	Max-Abs	4.0e-06 (4.9e-06)	6.6e-12 (4.2e-12)	1.6e-12 (1.5e-12)	1.1e-08 (9.0e-09)	14.26 (17.14)
Min-Max	Standard	7.4e-10 (8.8e-10)	5.7e-12 (3.6e-12)	1.4e-12 (1.4e-12)	1.2e-08 (1.1e-08)	13.41 (16.27)
Min-Max	Max-Abs	7.0e-06 (8.7e-06)	6.9e-12 (4.7e-12)	1.8e-12 (1.7e-12)	1.1e-08 (1.0e-08)	15.99 (18.65)
Max-Abs	Standard	6.8e-10 (8.3e-10)	4.7e-12 (3.0e-12)	1.3e-12 (1.3e-12)	9.2e-09 (8.1e-09)	13.22 (15.98)

4.1.3. Rango muy bajo

Realizando un análisis similar para el rango muy bajo, se ha examinado el desempeño de distintas técnicas de escalado. En primer lugar, al observar el promedio de WMAPE de los parámetros de Stokes (gráfica de la izquierda) en la Figura 9, se aprecia que el escalado Quantile para la salida presenta un ligero margen de mejora. Sin embargo, al analizar el error cuadrático medio (MSE) de Stokes V (gráfica de la derecha) en la Figura 9, se constata que el escalado Quantile para la salida presenta problemas, mientras que otras combinaciones de escalado brindan resultados más favorables.

Tabla 6. Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil de combinaciones de escalado de entrada y salida para el rango muy bajo.

Escalado de Entrada	Escalado de Salida	Stokes I MSE (Q3)	Stokes Q MSE (Q3)	Stokes U MSE (Q3)	Stokes V MSE (Q3)	WMAPE promedio (Q3)
Standard	Max-Abs	1.3e-05 (1.7e-05)	9.5e-14 (1.1e-13)	8.1e-14 (1.0e-13)	3.5e-10 (3.1e-10)	1009.42 (67.46)
Standard	Quantile	1.2e-09 (1.2e-09)	2.9e-13 (2.1e-13)	1.7e-13 (1.4e-13)	4.1e-09 (1.3e-09)	67.87 (55.29)
Max-Abs	Standard	8.8e-10 (1.0e-09)	9.4e-14 (1.1e-13)	8.1e-14 (1.0e-13)	3.7e-10 (3.9e-10)	1048.75 (73.21)
Quantile	Max-Abs	1.0e-05 (1.3e-05)	9.6e-14 (1.1e-13)	8.4e-14 (1.0e-13)	3.9e-10 (3.6e-10)	750.30 (59.06)

En la Tabla 6, se muestran algunas de las combinaciones de escalado que funcionan mejor. Aunque no existe una combinación de escalado claramente superior, se ha decidido utilizar el escalado Standard en la entrada junto con el Max-Abs en la salida. Esta elección se basa en el mejor rendimiento observado

en el MSE del parámetro de Stokes V (resaltado en rojo en la gráfica de la derecha en la Figura 9). Es importante destacar que, en casos de campos magnéticos muy débiles, como el presente en esta situación, el parámetro de Stokes V es el único que proporcionará observaciones si se detecta un campo magnético en una estrella.

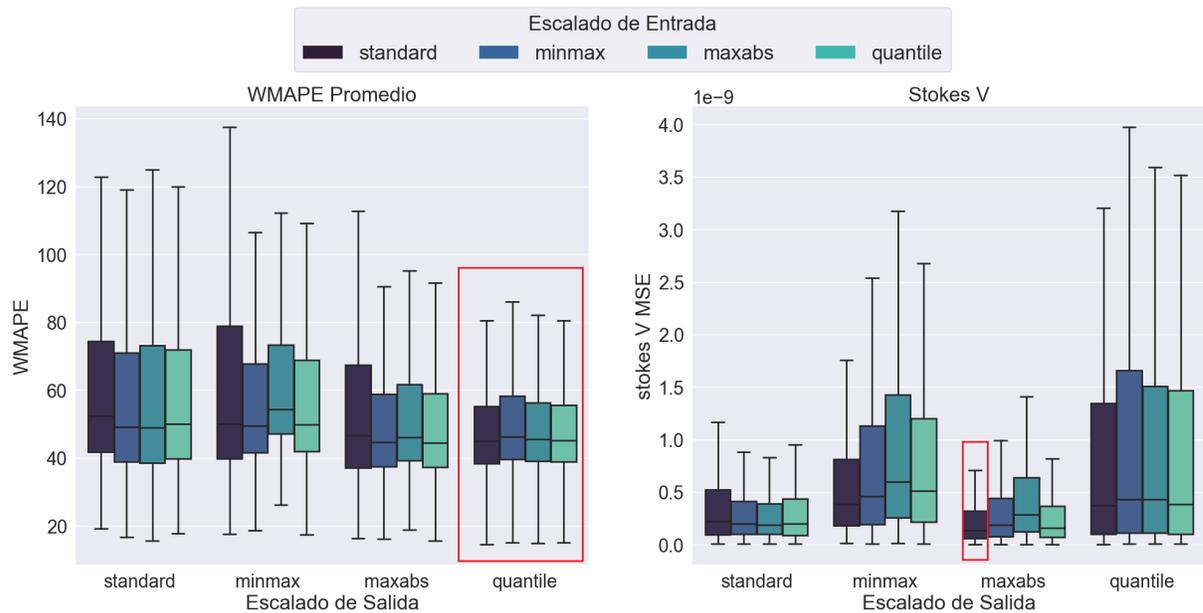


Figura 9. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas combinaciones de los métodos de escalado para la entrada y la salida de los datos para el rango muy bajo.

4.1.4. Resumen

En resumen, tras llevar a cabo un análisis minucioso en cada uno de los rangos (alto, bajo y muy bajo), se han elegido con cuidado los métodos de escalado más adecuados para optimizar el rendimiento del modelo. El objetivo principal de estos métodos de escalado es disminuir el sesgo inherente del modelo hacia los parámetros de Stokes con mayor amplitud, lo cual contribuirá a obtener resultados más precisos y confiables. En el caso del rango alto, se ha seleccionado el método de escalado Max-Abs para la entrada y el escalado Standard para la salida. Para el rango bajo, se ha utilizado la misma combinación de escalados que en el rango alto: Max-Abs para la entrada y Standard para la salida. Por otro lado, en el rango muy bajo se ha adoptado un enfoque ligeramente distinto, empleando el escalado Standard para la entrada y Max-Abs para la salida. Estas decisiones se han tomado en base al rendimiento observado en términos de WMAPE y MSE.

4.2. Selección de la arquitectura

Un aspecto crucial a tener en cuenta es la elección de la arquitectura de la red neuronal densa. Determinar el número óptimo de capas ocultas y neuronas por capa es fundamental para obtener el mejor rendimiento. En nuestro experimento, probamos diferentes configuraciones, incluyendo 2, 3, 4, 5, 6 y 7 capas ocultas con 512, 1024, 2048 y 4096 neuronas por capa, lo que nos lleva a un total de 24 modelos, este experimento se llevó a cabo en el rango alto y utilizando el escalado de entrada y salida seleccionado anteriormente (Max-Abs para la entrada y Standard para la salida).

En la Figura 10 se omiten los datos atípicos, así como los resultados de los modelos con dos capas para visualizar mejor estos resultados. En la gráfica izquierda de la Figura se muestra el WMAPE promedio de los parámetros de Stokes de estos modelos, y es evidente que a medida que aumentamos el número de capas ocultas, mejora el rendimiento del modelo. Además, al observar la gráfica de la derecha de la derecha en la que se presenta el MSE de Stokes V, podemos apreciar el mismo patrón, además de mejorar al aumentar el número de neuronas por capa oculta.

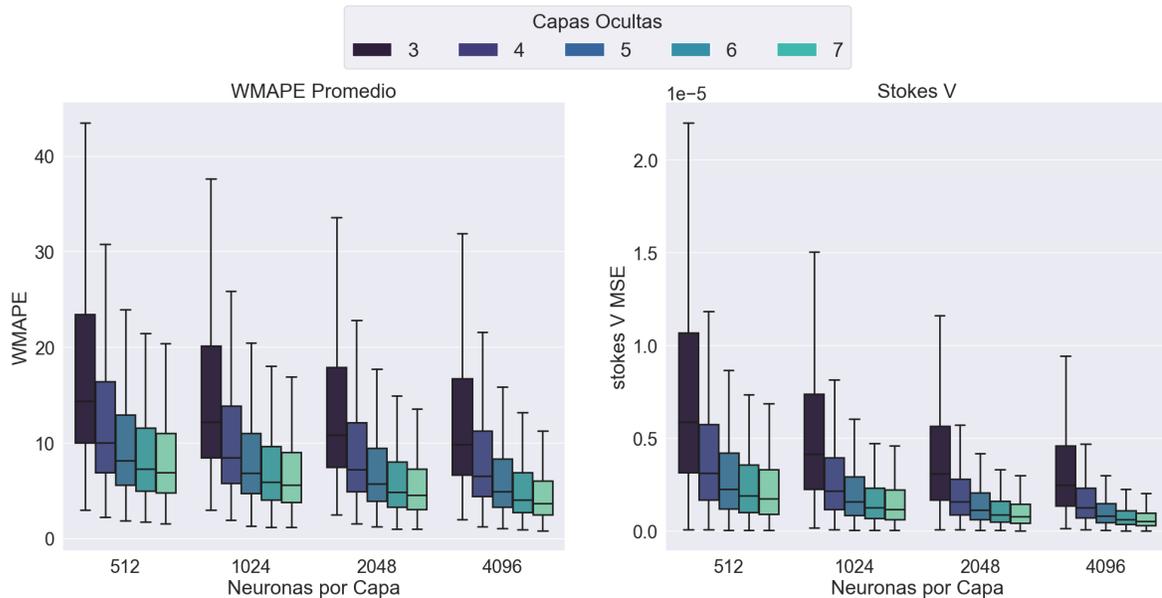


Figura 10. WMAPE promedio de los parámetros de Stokes y el MSE del Stoves V de las distintas arquitecturas de redes neuronales densas agrupados por la cantidad de neuronas por capa para el rango alto.

Tabla 7. Tabla con el error MSE promedio por parámetro de Stokes y el WMAPE promedio, junto con el tercer cuartil de distintas arquitecturas de red neuronal densa para el rango alto.

Num. Capas Ocultas	Neuronas por Capa	Stokes I MSE (Q3)	Stokes Q MSE (Q3)	Stokes U MSE (Q3)	Stokes V MSE (Q3)	WMAPE promedio (Q3)
5	4096	1.3e-06 (1.5e-06)	2.5e-07 (2.5e-07)	2.1e-07 (2.0e-07)	1.3e-06 (1.4e-06)	11.03 (8.32)
6	4096	1.0e-06 (1.2e-06)	1.7e-07 (1.7e-07)	1.5e-07 (1.3e-07)	1.0e-06 (1.1e-06)	8.33 (6.91)
7	2048	1.2e-06 (1.5e-06)	1.9e-07 (2.1e-07)	1.5e-07 (1.5e-07)	1.2e-06 (1.4e-06)	8.39 (7.26)
7	4096	9.3e-07 (1.0e-06)	1.5e-07 (1.4e-07)	1.2e-07 (1.1e-07)	8.8e-07 (9.8e-07)	7.09 (6.00)

En la Tabla 7, se presentan algunas de las mejores arquitecturas encontradas. El modelo más grande, con 7 capas ocultas y 4096 neuronas por capa oculta, muestra un menor MSE en cada parámetro de Stokes, así como un menor WMAPE promedio. En base a los resultados de este experimento, se selecciona la arquitectura con 7 capas ocultas y 4096 neuronas por capa oculta. Para los rangos bajo y muy bajo, utilizaremos la misma arquitectura que en el rango alto. Es importante destacar que por limitación del hardware no se pueden realizar experimentos con más neuronas y capas ocultas.

4.3. Modelo general vs modelo especializado

En la búsqueda del mejor modelo nos encontramos con dos enfoques, uno de ellos es la creación de un modelo para estimar cada parámetro de Stokes, lo que resultaría en un modelo especializado para cada uno de ellos. A estos modelos los llamaremos “modelos especializados”. Por otro lado, podríamos utilizar un único modelo capaz de predecir los cuatro parámetros de Stokes simultáneamente, al que denominaremos “modelo general”. Esto nos lleva a la siguiente pregunta: ¿Es mejor construir un modelo especializado para cada parámetro de Stokes o utilizar un único modelo capaz de predecir los cuatro parámetros simultáneamente? Para responder a esta interrogante, llevamos a cabo un experimento exhaustivo para decidir cuál enfoque consigue un modelo con un mejor rendimiento.

En la Figura 11 y en la Tabla 8, presentamos los resultados que muestran el rendimiento comparativo de estos enfoques. Es evidente que existe una diferencia notable entre ellos, siendo el modelo general el que sobresale en términos de rendimiento. El modelo general destaca debido a su capacidad para aprovechar los “pre-cálculos” de los parámetros de Stokes, es decir, las conexiones entre las neuronas entre los parámetros de Stokes durante el proceso de predicción. Al tener en cuenta los “pre-cálculos” de los demás parámetros, este modelo obtiene una mejora de la precisión y el rendimiento en general.

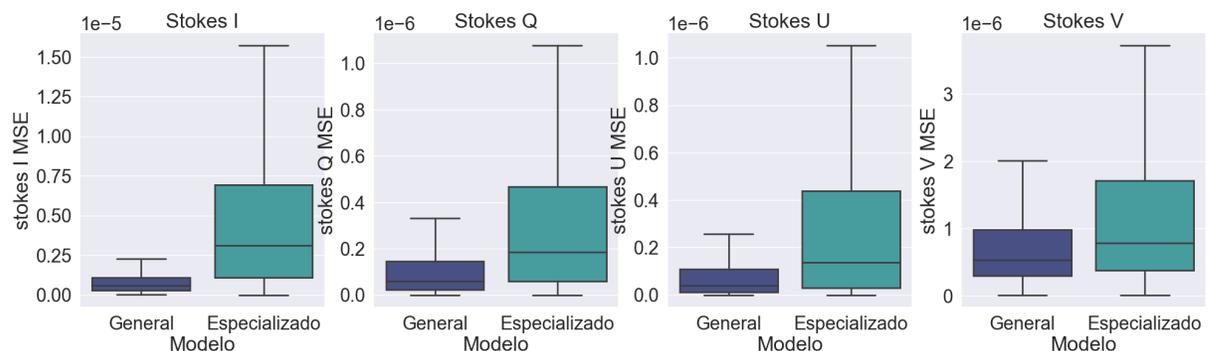


Figura 11. MSE de cada parámetro de Stokes del modelo general y especializado para el rango alto.

En contraste, los modelos especializados para cada parámetro de Stokes carecen del acceso a estos “pre-cálculos”, ya que su enfoque se limita a predecir un parámetro específico. Aunque estos modelos pueden capturar características únicas de cada parámetro de Stokes individual, no pueden aprovechar las conexiones con las neuronas que predicen los otros parámetros de Stokes. Esta limitación puede conducir a una menor precisión en la predicción de los parámetros de Stokes de manera aislada. Como resultado, el modelo general, al beneficiarse de la interacción entre los parámetros de Stokes, supera a los modelos especializados en términos de precisión y capacidad predictiva. Estos hallazgos respaldan la idea de utilizar el modelo general para predecir los parámetros de Stokes.

Tabla 8. Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil del modelo general y especializado para el rango alto

Modelo	Stokes I MSE (Q3)	Stokes Q MSE (Q3)	Stokes U MSE (Q3)	Stokes V MSE (Q3)	WMAPE promedio (Q3)
Especializado	5.8e-06 (6.9e-06)	4.4e-07 (4.6e-07)	4.8e-07 (4.3e-07)	1.5e-06 (1.7e-06)	10.25 (9.82)
General	9.3e-07 (1.0e-06)	1.5e-07 (1.4e-07)	1.2e-07 (1.1e-07)	8.8e-07 (9.8e-07)	7.09 (6.00)

4.4. Efecto del tamaño del conjunto de datos

Se llevó a cabo un análisis de la cantidad de datos de entrada para el modelo de estimación de parámetros de Stokes, utilizando el modelo de rango alto y el escalado de entrada Max-Abs y Standard para la salida. Los resultados se presentan en la Tabla 9, en ellos se observa que a medida que se aumentaba la cantidad de datos, los MSE de cada parámetro de Stokes mostraban una clara mejora en el rendimiento del modelo. Esto indica una mayor capacidad de generalización y una mejor capacidad para capturar las variaciones. Además, se observó una disminución constante en el WMAPE promedio a medida que se incrementaba el tamaño del conjunto de datos, lo cual confirma la robustez y precisión del modelo.

Tabla 9. Tabla con el MSE promedio por parámetro de Stokes y el WMAPE promedio de los parámetros de Stokes, junto con el tercer cuartil del modelo entrenado con distinta cantidad de datos para el rango alto.

Modelo	Stokes I MSE (Q3)	Stokes Q MSE (Q3)	Stokes U MSE (Q3)	Stokes V MSE (Q3)	WMAPE promedio (Q3)
50k	9.3e-07 (1.0e-06)	1.5e-07 (1.4e-07)	1.2e-07 (1.1e-07)	8.8e-07 (9.8e-07)	7.09 (6.00)
250k	2.3e-07 (2.9e-07)	3.1e-08 (3.6e-08)	2.4e-08 (2.6e-08)	2.0e-07 (2.4e-07)	3.79 (3.12)
500k	1.5e-07 (1.9e-07)	2.0e-08 (2.4e-08)	1.4e-08 (1.7e-08)	1.3e-07 (1.6e-07)	3.11 (2.55)
1.3M	1.0e-07 (1.3e-07)	1.3e-08 (1.6e-08)	9.5e-09 (1.1e-08)	9.1e-08 (1.1e-07)	2.70 (2.13)

Estos hallazgos respaldan la importancia de utilizar la mayor cantidad de datos disponibles para estimar los parámetros de Stokes en diferentes rangos. Con 1.3 millones de datos para el rango alto, un millón para el rango bajo y un millón para el rango muy bajo, se asegura un alto nivel de precisión en la

estimación de estos parámetros. Al igual que en la selección de la arquitectura, no se generó un conjunto de datos mayor por limitaciones de tiempo.

4.5. Modelo de estimación de parámetros de Stokes

A través de los experimentos preliminares mencionados anteriormente, logramos determinar el escalado de entrada y salida para cada rango, así como la arquitectura que tiene mejor rendimiento de las utilizadas de la red neuronal densa y la preferencia por un modelo que calcule los cuatro parámetros de Stokes simultáneamente. El flujo y la arquitectura del modelo de estimación de los parámetros de Stokes se presentan en la Figura 5.

A continuación, procederemos a evaluar el rendimiento del modelo en cada rango utilizando el conjunto de datos completo. En la Tabla 10 podemos observar el WMAPE de cada parámetro de Stokes así como el WMAPE promedio, junto con el valor para el tercer cuartil para los tres rangos (alto, bajo y muy bajo). Al utilizar el WMAPE para comparar los resultados de los modelos en sus respectivos rangos, se puede observar que en el caso del parámetro de Stokes I, el modelo que mejor lo estima es el del rango bajo. Sin embargo, el error en todos los modelos es tan bajo que resulta prácticamente despreciable. En general, todos los modelos estiman el parámetro de Stokes I de manera bastante precisa. Al analizar los demás parámetros de Stokes, se observa que el modelo del rango alto es el que mejor los estima, seguido por el modelo del rango bajo y, por último, el modelo del rango muy bajo. Esto concuerda con la amplitud de los parámetros de Stokes, ya que a medida que la fuerza del momento dipolar disminuye, se generan errores más grandes debido a la fuerte influencia de la fuerza del momento dipolar en la amplitud.

Tabla 10. Tabla con el WMAPE por parámetro de Stokes y el WMAPE promedio, junto con el tercer cuartil para los distintos rangos.

Rango del Modelo	Stokes I WMAPE (Q3)	Stokes Q WMAPE (Q3)	Stokes U WMAPE (Q3)	Stokes V WMAPE (Q3)	WMAPE promedio (Q3)
Alto	0.0279 (0.0348)	2.79 (2.35)	6.62 (4.76)	1.37 (1.59)	2.70 (2.13)
Bajo	0.0007 (0.0009)	11.27 (11.90)	24.93 (30.71)	1.82 (2.05)	9.51 (11.29)
Muy Bajo	0.1713 (0.2100)	687.17 (104.61)	2676.68 (1594.95)	6.48 (5.05)	842.62 (66.99)

En el caso de los parámetros de Stokes Q y U, que son los más complejos, se observa que el modelo del rango alto los estima bastante bien. Sin embargo, a medida que se pasa al rango bajo, estas estimaciones empeoran considerablemente, y en el caso extremo del rango muy bajo, el modelo no puede estimarlos

correctamente. Por lo tanto, no se recomienda utilizar las estimaciones para los parámetros de Stokes Q y U en el rango muy bajo. En el caso del rango bajo, se deben tener ciertas precauciones con las estimaciones, especialmente en el caso del parámetro de Stokes U, donde el promedio de error es cercano al 25 %. En cambio, en el rango alto, el modelo estima los parámetros de Stokes Q y U con muy poco error. En el caso del parámetro de Stokes V, se observa que tanto el modelo del rango alto como el del rango bajo obtienen un error promedio inferior al 2 %, lo cual indica que el modelo es robusto para estimar este parámetro de Stokes. En el caso del rango muy bajo, se nota una disminución en el rendimiento, pero el promedio de error no supera el 7 %, lo que aún hace que el modelo sea confiable para estimar el parámetro de Stokes V.

En resumen, los resultados demuestran que el modelo del rango alto es bastante robusto y confiable para todos los parámetros de Stokes, mientras que en el rango bajo, el modelo es confiable para los parámetros de Stokes I y V, y se tienen que tener ciertas precauciones para los parámetros de Stokes Q y U. Por último, en el rango muy bajo, el modelo es confiable para los parámetros de Stokes I y V, pero no es confiable para los parámetros de Stokes Q y U.

4.5.1. Análisis del error en función de la amplitud del parámetro de Stokes

A continuación, se llevó a cabo un análisis a profundidad del WMAPE en función de la amplitud del perfil del parámetro de Stokes. Sin embargo, antes de entrar en detalle, es necesario definir el cálculo de la amplitud para cada parámetro de Stokes. La amplitud de los parámetros de Stokes se define de la siguiente manera:

$$\begin{aligned}
 \textit{Amplitud}_{StokesI} &= \text{mín}(Y_I) \\
 \textit{Amplitud}_{StokesQ} &= \text{máx}(|Y_Q|) \\
 \textit{Amplitud}_{StokesU} &= \text{máx}(|Y_U|) \\
 \textit{Amplitud}_{StokesV} &= \text{máx}(|Y_V|)
 \end{aligned}
 \tag{8}$$

donde Y_I es el perfil de Stokes I, Y_Q es el perfil de Stokes Q, Y_U es el perfil de Stokes U y Y_V es el perfil de Stokes V. Dado a que los parámetros de Stokes Q, U y V atraviesan el cero, en la fórmula para definir su amplitud, se toma el valor máximo de los valores absolutos que pertenecen al parámetro de Stokes.

Utilizando estas definiciones de amplitudes de los parámetros de Stokes, se llevó a cabo un análisis del error a medida que la amplitud cambia en cada parámetro. Se generaron gráficas de barras donde cada barra representa un rango de amplitud. La altura de la barra corresponde al WMAPE promedio de los casos de prueba que están dentro en ese rango de amplitud, y se muestra un número encima de cada barra que representa el porcentaje de casos de prueba que se encuentran en ese rango de amplitud. Además, se trazaron líneas verticales en cada barra para representar la desviación estándar.

En la Figura 12 observamos los WMAPE de cada parámetro de Stokes por su rango de amplitud para el modelo de rango alto, en ella podemos observar a medida que la amplitud del Stokes I aumenta, también lo hace el error. Una mayor amplitud en el parámetro de Stokes I indica que la gaussiana correspondiente no es tan pronunciada, lo que significa que no hay una absorción significativa de la intensidad de la luz. Debido a esto, el perfil de Stokes I no varía mucho, lo que dificulta su predicción de manera precisa como en otros rangos de amplitud. Sin embargo, incluso en los peores casos, el error se mantiene por debajo del 0.06 %, lo que en general no presenta problemas significativos en su predicción.

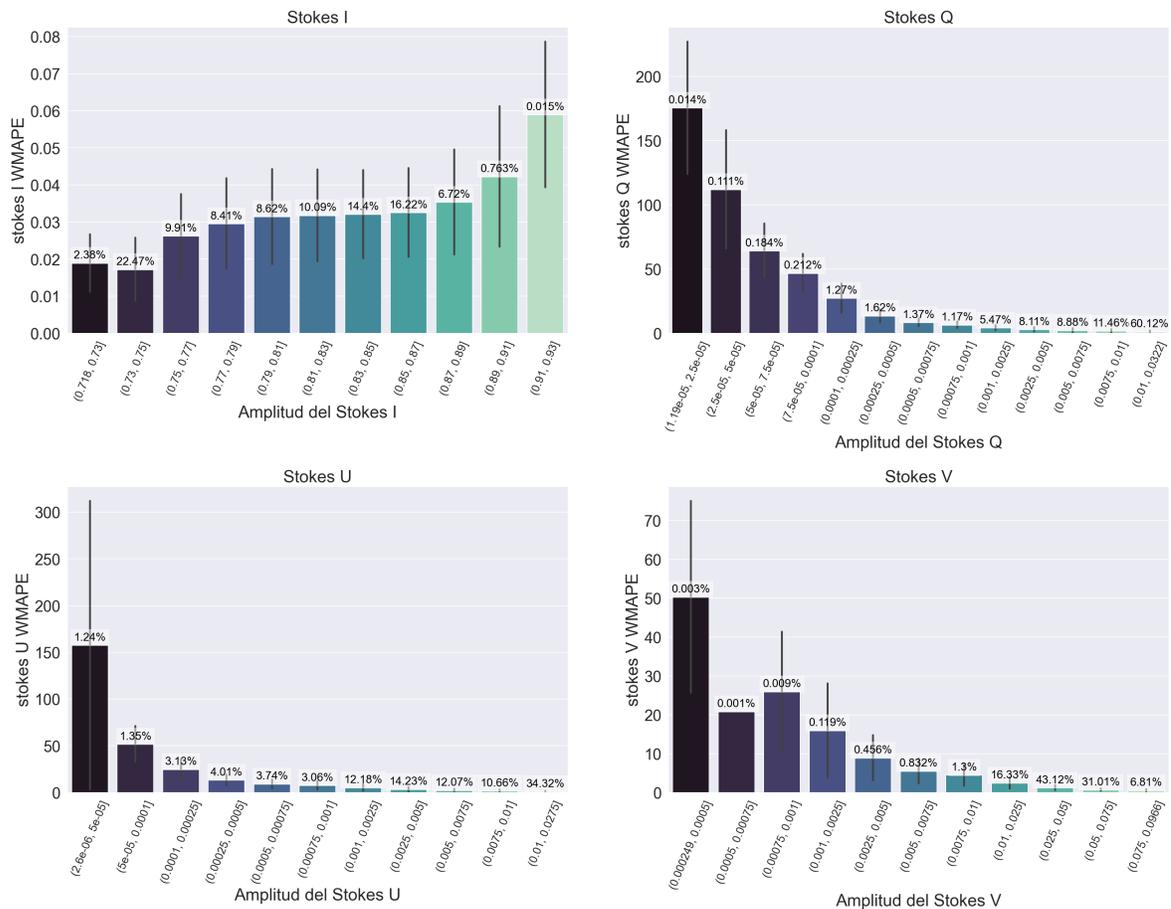


Figura 12. WMAPE de cada parámetro de Stokes del modelo para la estimación de parámetros de Stokes agrupado por rango de amplitud para el rango alto.

En el caso del parámetro de Stokes Q, a medida que la amplitud del parámetro disminuye, el error aumenta. Una menor amplitud en el Stokes Q indica una baja intensidad de luz polarizada, lo cual puede atribuirse a una estrella poco magnetizada o a un gran ángulo de inclinación de la estrella, lo que hace que el dipolo magnético no apunte directamente hacia nosotros, entre otros factores que podrían resultar en una baja amplitud. En el 96 % de los casos, el error es menor al 9 %. Esto indica que, en la gran mayoría de los casos, el modelo realiza una buena estimación del parámetro de Stokes Q.

Para el parámetro de Stokes U, se observa una tendencia similar al Stokes Q. El error aumenta a medida que la amplitud disminuye. Además, se observa que en el 90 % de los casos el error es menor al 10 %. Siendo los parámetros de Stokes Q y U polarizaciones lineales, se nota una diferencia considerable en la capacidad de predicción entre Stokes Q y U, siendo este último un poco más complicado de estimar para el modelo. Sin embargo, en la mayoría de los casos, el modelo realiza una buena estimación del perfil de Stokes U.

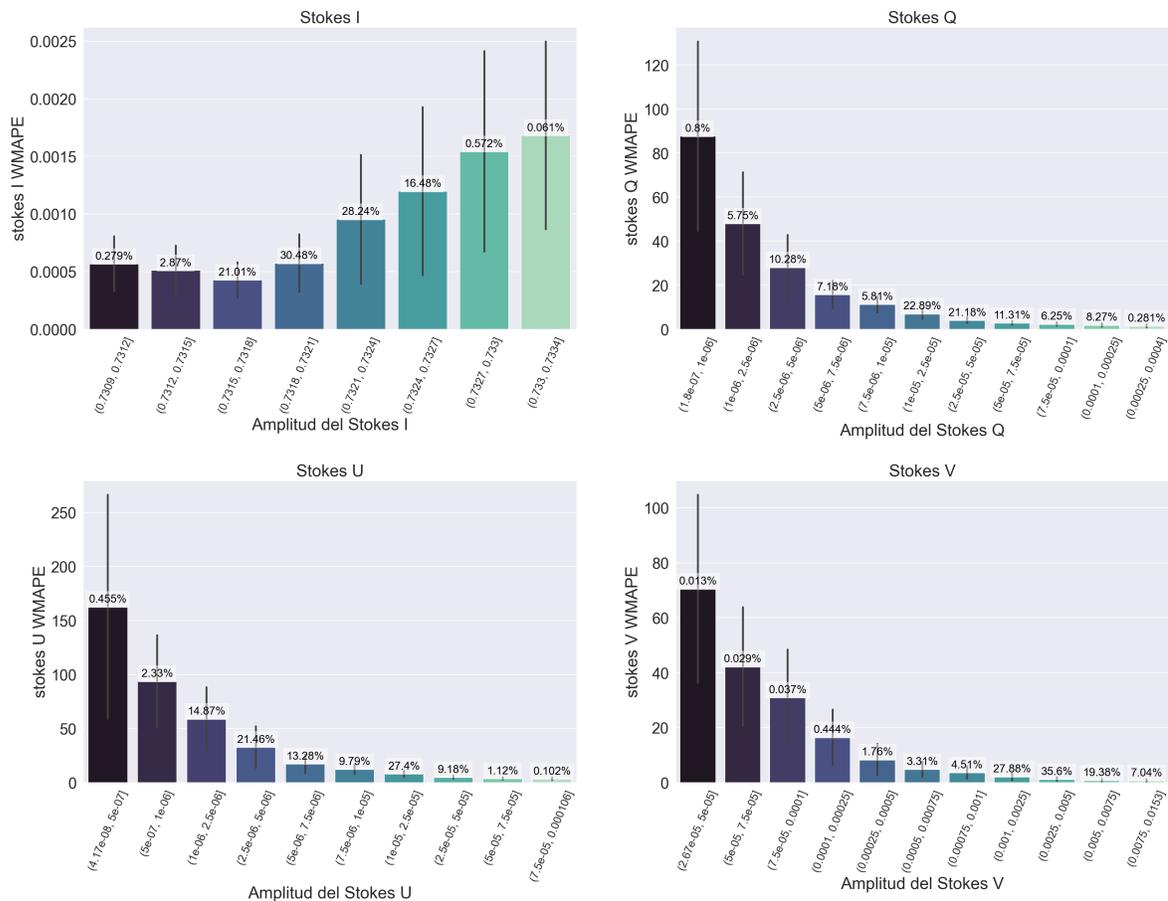


Figura 13. WMAPE de cada parámetro de Stokes del modelo para la estimación de parámetros de Stokes agrupado por rango de amplitud para el rango bajo.

En el Stokes V se observa una situación similar al Stokes Q y U. A medida que la amplitud disminuye, el error aumenta. Sin embargo, se obtiene que en el 99.5% de los casos, el error es inferior al 9%, incluso en el 97% de los casos se obtiene un error menor al 3%, por lo que el modelo es bastante robusto al estimarlo con un error muy bajo. El modelo en la mayoría de los casos puede estimar con un error bajo cada parámetro Stokes, y los errores grandes del modelo están claramente asociados con una menor amplitud para los parámetros de Stokes Q, U y V. El Stokes I es el más sencillo de estimar, seguido por el Stokes V, luego el Stokes Q y, por último, el Stokes U.

En el rango bajo, se muestra en la Figura 13, donde se observa que el Stokes I es bien estimado. En el 70% de los casos del Stokes Q, el error está por debajo del 7.5%. En el Stokes U, se presentan mayores dificultades, ya que el 48% de los casos tiene un error inferior al 12.5%. Por último, en el caso del parámetro de Stokes V, se observa que el 99% de los datos presenta un error inferior al 8.5%, e incluso el 90% tiene un error inferior al 2.5%. Esto demuestra que el modelo es confiable para estimar el parámetro de Stokes V, al igual que el modelo del rango alto.

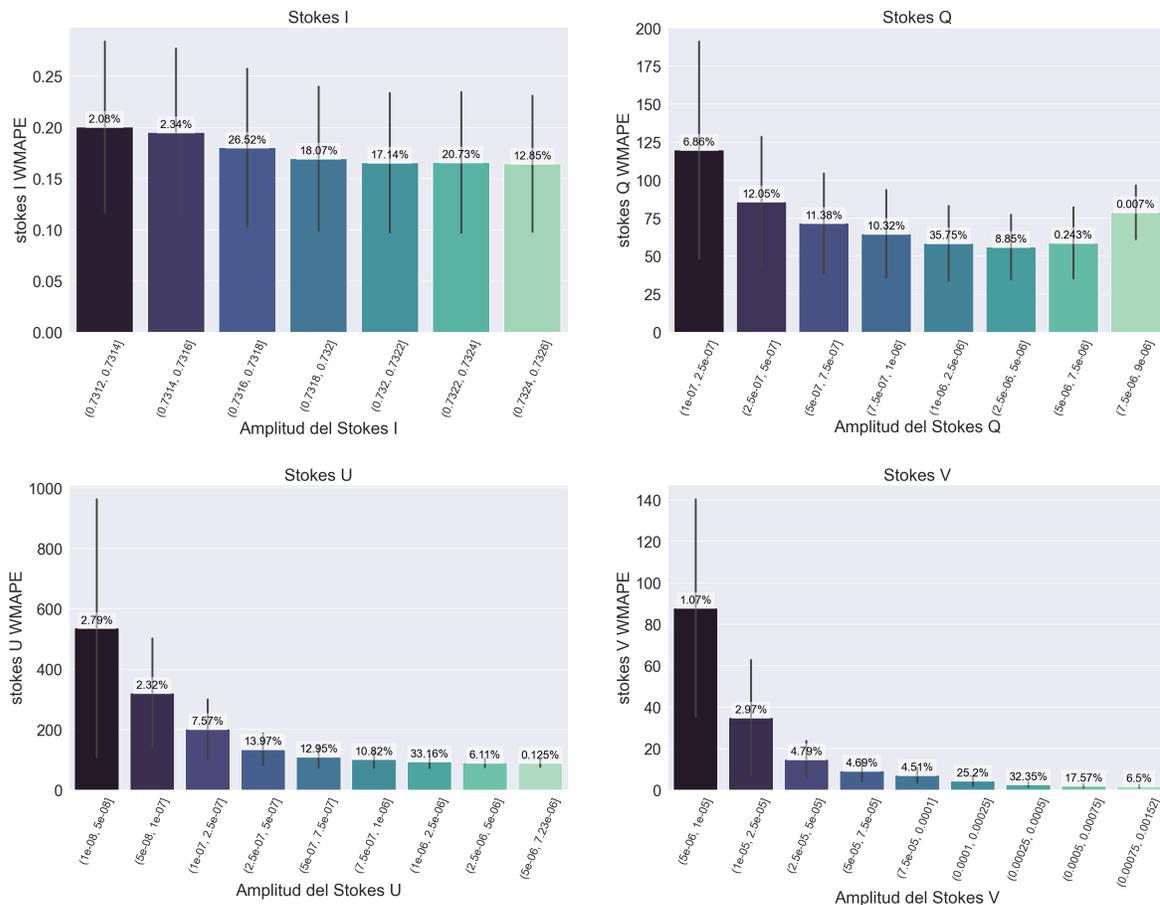


Figura 14. WMAPE de cada parámetro de Stokes del modelo para la estimación de parámetros de Stokes agrupado por rangos de amplitud para el rango muy bajo.

Por último, en la Figura 14, se muestra la misma información pero para el rango muy bajo, el cual ya habíamos identificado como el modelo que presenta más problemas. En el parámetro de Stokes I, nuevamente se observa un WMAPE muy bajo, lo que indica que el modelo es robusto al estimarlo. Sin embargo, en los casos de los parámetros de Stokes Q y U, se observa que los errores, incluso en los mejores rangos, están por encima del 55 % y 85 %, respectivamente. Por lo tanto, en estos dos parámetros de Stokes, el modelo no responde adecuadamente y no se recomienda su uso. Por otro lado, en el parámetro de Stokes V, se encuentra que en el 90 % de los casos, el error es menor al 9.5 %, e incluso en el 80 % de los casos, el error está por debajo del 4.5 %. Esto demuestra que el modelo es confiable al estimar el parámetro de Stokes V en el rango muy bajo, al igual que en los demás rangos.

4.5.2. Análisis de los percentiles

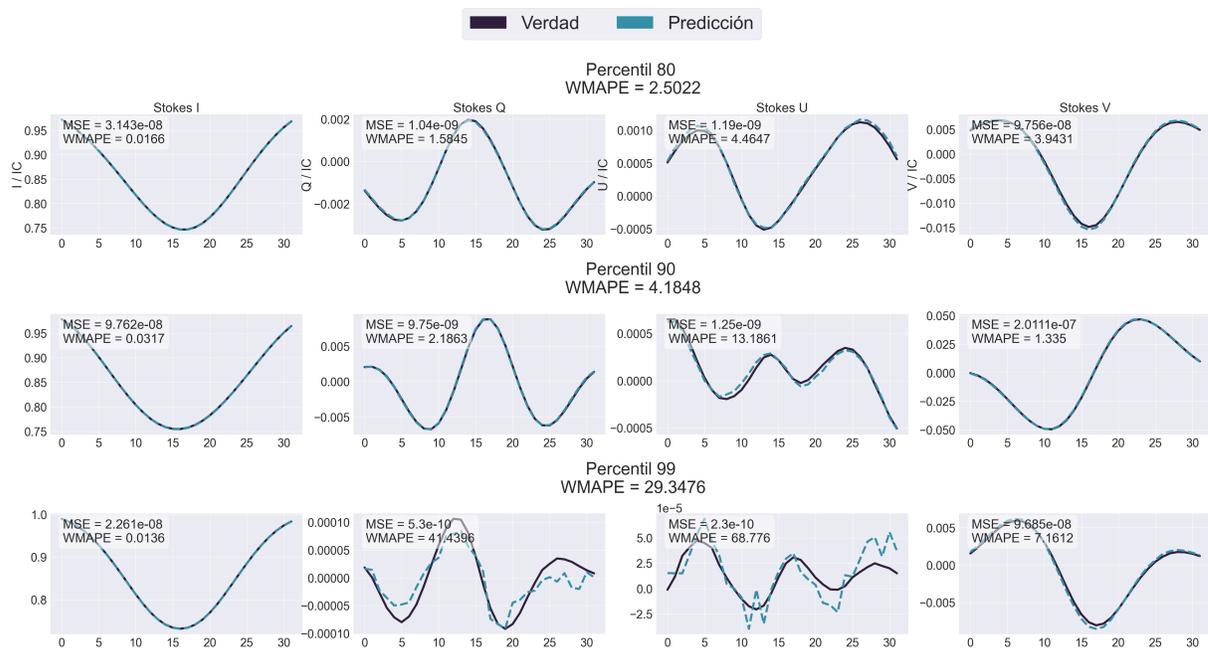


Figura 15. Estimación de los parámetros de Stokes de los percentiles 80 (arriba), 90 (en medio) y 99 (abajo) del rango alto.

En esta sección presentaremos la visualización de los percentiles 80, 90 y 99 de las estimaciones de los parámetros de Stokes en cada rango. Comenzando con el rango alto, como se muestra en la Figura 15. En el percentil 80, podemos observar que el modelo obtiene un WMAPE promedio del 2.5 %. El Stokes U es el peor estimado, con un error menor al 4.5 %, lo que indica que en el 80 % de los casos, el modelo

tendrá un error igual o menor a este valor. En el percentil 90, el WMAPE promedio es inferior al 4.2 %, y nuevamente el Stokes U es el más difícil de estimar, con un error ligeramente superior al 13 %. En el percentil 99, que representa uno de los peores casos, el WMAPE promedio es de casi el 30 %. En este caso, el Stokes U tiene un error de casi el 70 %.

Además, el Stokes Q también presenta un error grande, casi del 42 %. Por lo tanto, en algunos casos extremos, el modelo no responde adecuadamente. Sin embargo, el parámetro de Stokes V presenta un error del 7.1 %, lo que indica que incluso en los peores casos, la estimación del Stokes V sigue estando dentro de un rango manejable. En todos los casos, se observa que el Stokes I es muy bien predicho por el modelo.

Pasando al rango bajo, mostrado en la Figura 16, en el percentil 80 observamos algo similar, donde el parámetro de Stokes I es bien estimado junto con el parámetro de Stokes V. Sin embargo, el parámetro de Stokes Q presenta un error alrededor del 26 %, lo que ya es considerable, al igual que el parámetro de Stokes U, obteniendo un error del 29 %. En el percentil 90, el error en el parámetro de Stokes U se vuelve más pronunciado, alcanzando el 57 %. Aun así, el modelo estima correctamente los parámetros de Stokes I y V. Por último, en el percentil 99, observamos una situación similar al rango alto, donde se presentan grandes errores en los parámetros de Stokes Q y U.

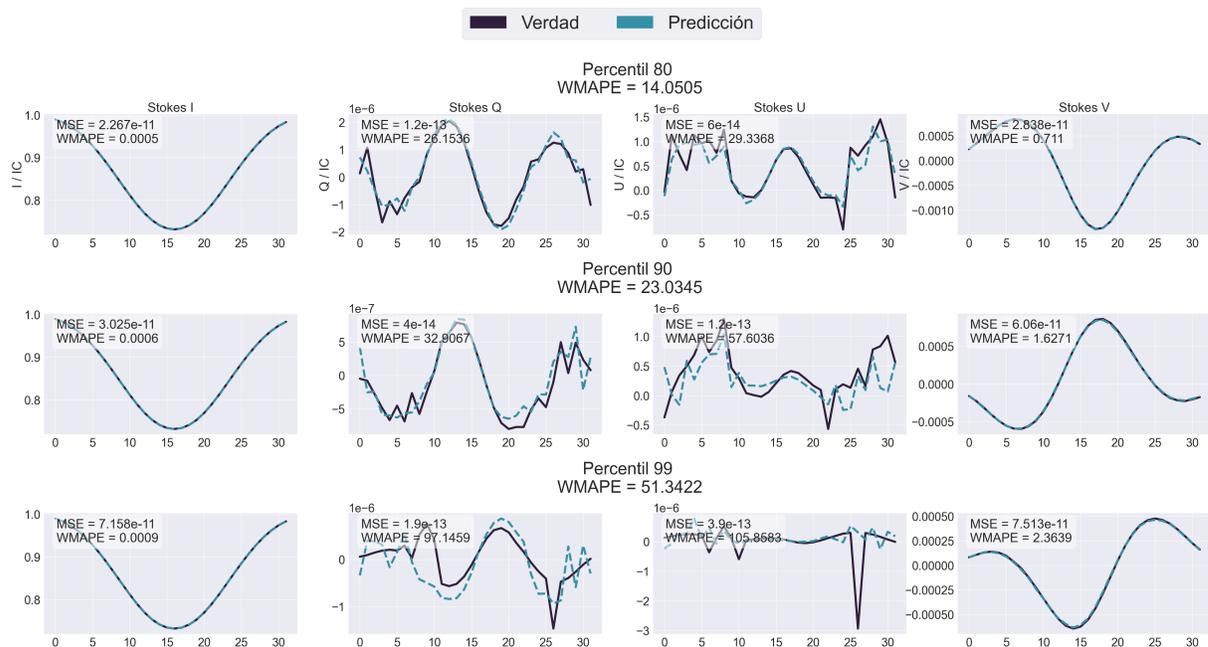


Figura 16. Estimación de los parámetros de Stokes de los percentiles 80 (arriba), 90 (en medio) y 99 (abajo) del rango bajo.

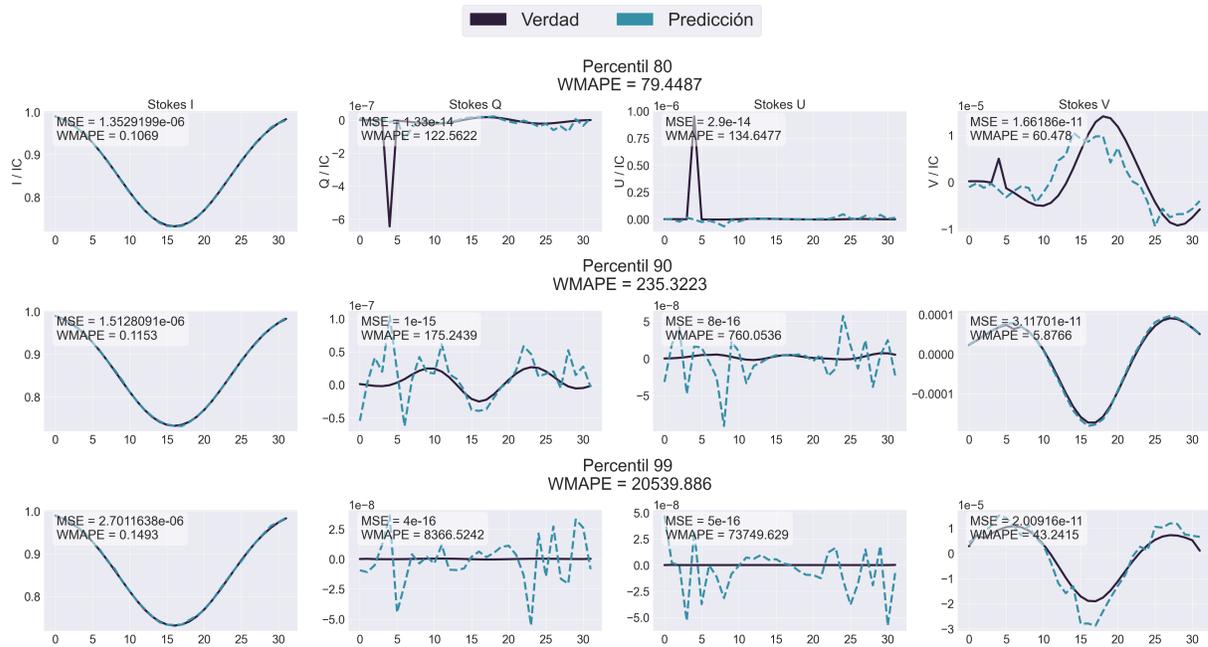


Figura 17. Estimación de los parámetros de Stokes de los percentiles 80 (arriba), 90 (en medio) y 99 (abajo) del rango muy bajo.

En el rango muy bajo, ilustrado en la Figura 17, podemos ver que el modelo no puede estimar adecuadamente los parámetros de Stokes Q y U en ninguno de los casos, obteniendo errores superiores al 120 %. Por lo tanto, no se recomienda utilizar las estimaciones de los parámetros de Stokes Q y U en el rango muy bajo. En cuanto al parámetro de Stokes I, se observa el mismo comportamiento que en los otros casos. En el parámetro de Stokes V, se observan errores grandes del 60 % y 43 % en los percentiles 80 y 99, respectivamente. Sin embargo, en el percentil 90, el error es bastante manejable y se encuentra dentro del margen de error considerado bueno. Por lo tanto, en el caso del parámetro de Stokes V, es necesario tomar algunas precauciones en ciertos casos en los que la estimación de este parámetro podría ser incorrecta en el rango muy bajo.

4.5.3. Resumen

En resumen, el modelo muestra un rendimiento muy bueno en la estimación de los parámetros de Stokes en el rango alto, con errores bajos en la mayoría de los casos. Sin embargo, se observan errores significativos superiores al 41 % en los parámetros de Stokes Q y U en el percentil 99, lo que indica una falta

de confiabilidad del modelo en situaciones extremas. En el rango bajo, el modelo continúa estimando muy bien los parámetros de Stokes I y V, obteniendo errores menores al 0.0009% y 2.5%, respectivamente. Aun así se presentan dificultades en la estimación de los parámetros Q y U, especialmente en los percentiles 90 y 99. En el rango muy bajo, el modelo no es capaz de estimar adecuadamente los parámetros de Stokes Q y U en muchos casos, lo que desaconseja su uso en estas condiciones. No obstante, el modelo sigue mostrando un buen rendimiento en los parámetros de Stokes I y V, excepto en ciertos casos extremos donde el Stokes V no se estima correctamente en el rango muy bajo. El WMAPE promedio del modelo son inferiores al 0.2% y 6.5% en los parámetros de Stokes I y V en todos los rangos, mientras que se presentan problemas en ciertos casos del rango alto y bajo, y en muchos casos del rango muy bajo en los parámetros de Stokes Q y U.

4.6. Selección del algoritmo para la recuperación del campo magnético

Una vez que los parámetros de Stokes han sido estimados a partir de la red neuronal, el siguiente paso es la recuperación del campo magnético a partir de los parámetros de Stokes observados (ver Figura 3). En otras palabras, buscamos determinar qué combinación de los atributos del campo magnético produce ciertos parámetros de Stokes. Para lograr esto, es necesario seleccionar un algoritmo de optimización adecuado. Realizamos pruebas con varios algoritmos, incluyendo Mínimos Cuadrados (LS) (Moré (1978)), Dual Annealing (DA) (Tsallis & Stariolo (1996)), Dual Annealing + Mínimos Cuadrados (DA+LS), Particle Swarm Optimization (PSO) (Kennedy & Eberhart (1995)) y Black Hole Optimization (BHO) (Hatamlou (2013)), esto para el rango alto.

El enfoque Dual Annealing + Mínimos Cuadrados combina los dos algoritmos de la siguiente manera: se utiliza Dual Annealing como inicializador, ya que Mínimos Cuadrados es sensible a la inicialización. Luego, se realiza una última optimización con Mínimos Cuadrados partiendo del resultado final obtenido mediante Dual Annealing. Todos los algoritmos fueron configurados para tener aproximadamente el mismo tiempo de ejecución promedio y se ejecutó 10 veces la misma configuración del algoritmo de optimización por cada caso de prueba. Para LS, se ejecutó tres veces para que tardara el mismo tiempo que los otros algoritmos y se seleccionó el mejor resultado. En el caso de DA, se realizaron 330 evaluaciones de la función. En el enfoque DA+LS, se llevaron a cabo 170 evaluaciones de la función con DA y una ejecución de LS. Para PSO, se utilizó una población de 868 partículas y 10 iteraciones, mientras que para BHO se empleó una población de 1600 estrellas y 10 iteraciones. Las configuraciones de los

algoritmos han sido diseñadas con el objetivo de garantizar un tiempo de ejecución promedio igual en cada caso, a la vez que constituyen una prueba rápida para identificar el algoritmo de optimización que mejor se adapta a esta problemática.

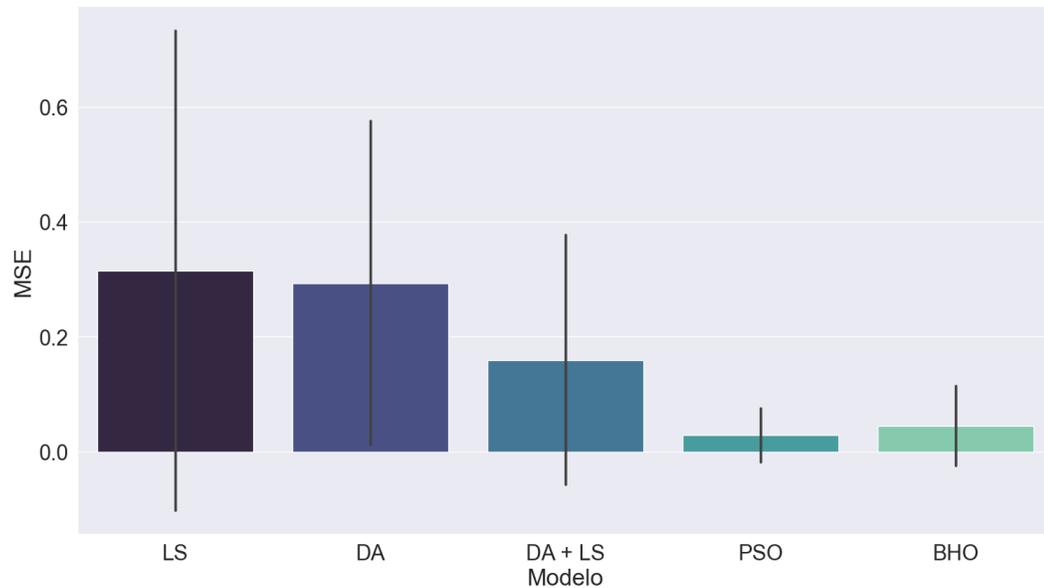


Figura 18. MSE promedio en el espacio escalado de la mejor solución encontrada para el rango alto.

En la Figura 18, se muestra el MSE en el espacio escalado de los parámetros de Stokes para la mejor solución encontrada por cada algoritmo de optimización en cada instancia de prueba. Además, se incluyen barras verticales que representan la desviación estándar. Los resultados revelan que el enfoque de combinación de búsqueda por DA y LS proporciona resultados significativamente superiores en comparación con el uso individual de estos algoritmos. Sin embargo, tanto PSO como BHO son los dos algoritmos que ofrecen los mejores resultados en términos de error. La ventaja de PSO y BHO radica en su capacidad para paralelizar el cálculo de la aptitud de cada solución candidata. Entre los dos, PSO muestra una ligera superioridad al lograr errores más bajos y una menor desviación estándar. Por consiguiente, se decidió utilizar el algoritmo PSO para llevar a cabo la recuperación del campo magnético en el rango alto, así como en los demás rangos. Estos hallazgos respaldan la elección del algoritmo PSO como el enfoque óptimo para la estimación precisa de los parámetros de Stokes, ya que ofrece un rendimiento mejorado en términos de precisión dado el MSE y estabilidad mediante la desviación estándar.

4.7. Recuperación del campo magnético

En la sección siguiente, analizaremos los resultados de la recuperación del campo magnético utilizando el algoritmo PSO en conjunto con la red neuronal densa entrenada para ajustar los parámetros de Stokes de las soluciones candidatas a los observados (generados por COSSAM). Para obtener resultados más estables, se ejecutó el algoritmo PSO 5 veces para cada instancia de prueba y se seleccionó la mejor solución de esas 5 ejecuciones como la solución final. Cada ejecución de PSO tomó un promedio de 16 segundos, por lo que las 5 ejecuciones del algoritmo tardaron en total 80 segundos.

En la Tabla 11, se presenta el WMAPE de los atributos más importantes a recuperar. Podemos observar que el WMAPE en la fuerza del momento dipolar (m) incrementa a medida que este valor disminuye, sin embargo, este error se mantiene por debajo del 3% en todos los rangos. Por otro lado, en el ángulo de inclinación (i) se obtienen errores más grandes y no siguen la misma tendencia que en el caso anterior, pero el error se mantiene por debajo del 9% en todos los rangos. Por último, en el vector X_r que representa la posición del dipolo, se observa una tendencia similar a la fuerza del momento dipolar, pero más pronunciada, con errores del 2%, 7% y 11.5% para los rangos alto, bajo y muy bajo, respectivamente. Es en el rango muy bajo donde la recuperación de este atributo se vuelve un poco más complicada.

A partir de estos resultados, podemos concluir que el conjunto de la red neuronal densa entrenada como parte de la función objetivo de PSO para recuperar los valores de la configuración del campo magnético obtiene buenos resultados con errores bajos. En la aplicación de esta metodología con datos reales, cada recuperación del campo magnético, es decir, cada caso de prueba representaría una estrella diferente. Esto significa que también se puede determinar la configuración del campo magnético en un tiempo adecuado.

Tabla 11. Tabla con el WMAPE por atributo del campo magnético del modelo en cada rango.

Rango del Modelo	Fuerza del Momento Dipolar (m)	Inclinación (i)	Posición del Dipolo (X_r)
Alto	0.34	8.55	1.74
Bajo	1.61	5.35	7.09
Muy Bajo	2.79	6.81	11.54

4.8. Transferencia de aprendizaje

En esta sección se aborda la optimización y recuperación del campo magnético mediante el modelo de estimación de parámetros de Stokes con transferencia de aprendizaje. Es importante destacar que este enfoque funciona de manera efectiva para una longitud de onda o estrella específica. Sin embargo, al aplicarlo a otra longitud de onda o estrella, el modelo no estimaría correctamente los parámetros de Stokes, lo cual afectaría la recuperación precisa del campo magnético. Para superar esta limitación, se exploró la transferencia de aprendizaje utilizando datos de una longitud de onda distinta, lo que permitiría reducir la cantidad de datos necesarios para el entrenamiento. Para determinar la cantidad adecuada de datos requeridos y obtener resultados precisos, se llevó a cabo un experimento preliminar, este experimento se realizó para el modelo del rango alto.

En la Figura 19 se presentan los resultados de este experimento, en ellos se pueden observar cómo el error disminuye y su desviación estándar se reduce a medida que se realiza el re-entrenamiento con más datos. Sin embargo, en este estudio, es crucial lograr una aplicación fácil y rápida a cualquier estrella. Por lo tanto, basándonos en los resultados obtenidos, se decidió seleccionar un conjunto de datos de 35,000 casos para el re-entrenamiento del modelo de estimación de parámetros de Stokes.

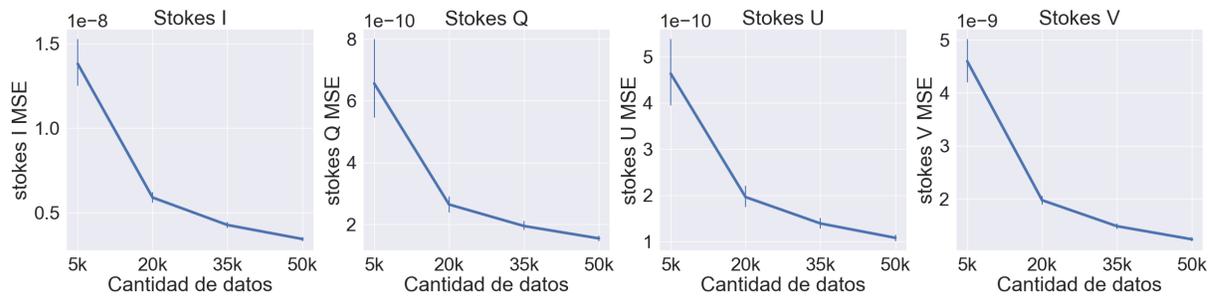


Figura 19. Gráficas de línea del MSE por cada parámetro de Stokes junto con su desviación estándar (barras verticales) para el rango alto.

Además, se llevó a cabo otro experimento para determinar qué capas deberían congelarse y así reducir el tiempo necesario para el re-entrenamiento. En la Figura 20, se muestra el MSE de cada parámetro de Stokes congelando desde cero hasta todas las capas ocultas. Es importante destacar que las capas congeladas se refiere desde la entrada hasta la salida. Los resultados revelaron que, aunque la desviación estándar es similar en todos los casos, el menor error promedio se obtuvo al congelar las primeras cinco capas y re-entrenar solo las tres capas finales. A partir de este punto, se observó un aumento en el error. Por lo tanto, basándonos en los resultados de este experimento, se decidió realizar la transferencia de

aprendizaje y re-entrenamiento utilizando los pesos de las primeras cinco capas congeladas.

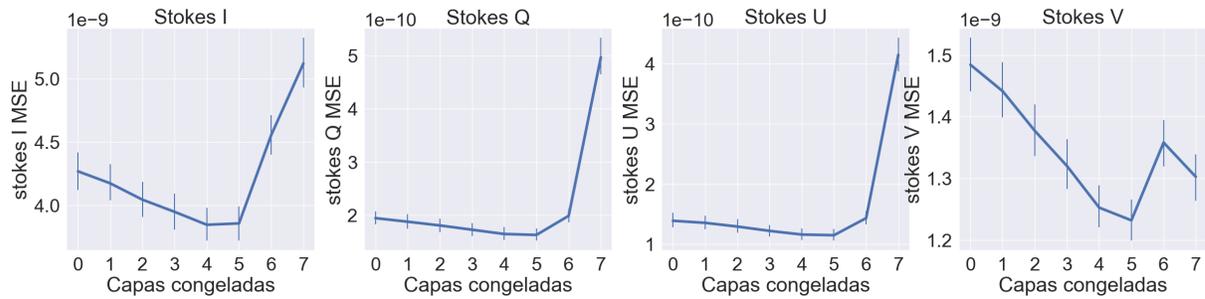


Figura 20. Gráficas de línea del MSE en cada parámetro de Stokes por la cantidad de capas congeladas en el re-entrenamiento para el rango alto.

Con base en los resultados obtenidos en ambos experimentos, se determinó que el conjunto de datos para la transferencia de aprendizaje y re-entrenamiento consistiría en 35,000 casos. Esta elección se basó en la reducción del tiempo necesario para generar estos datos. Además, se estableció qué capas se re-entrenarían y qué capas tendrían sus pesos congelados. Las primeras cinco capas se mantendrían con los pesos congelados, mientras que las últimas tres capas se re-entrenarían. Se espera que esta configuración logre un equilibrio entre el tiempo necesario para generar un nuevo conjunto de datos para una estrella diferente y el tiempo requerido para el re-entrenamiento, a fin de obtener resultados confiables.

4.8.1. Re-entrenamiento del modelo de estimación de parámetros de Stokes

Tabla 12. Tabla con el WMAPE por parámetro de Stokes y el WMAPE promedio, junto con el tercer cuartil para los distintos rangos para el modelo base y el re-entrenado.

Rango del Modelo	Modelo	Stokes I WMAPE (Q3)	Stokes Q WMAPE (Q3)	Stokes U WMAPE (Q3)	Stokes V WMAPE (Q3)	WMAPE promedio (Q3)
Alto	Base	0.0279 (0.0348)	2.79 (2.35)	6.62 (4.76)	1.37 (1.59)	2.70 (2.13)
	Re-entrenado	0.0044 (0.0054)	11.14 (5.60)	24.30 (13.27)	2.20 (2.49)	9.41 (5.23)
Bajo	Base	0.0007 (0.0009)	11.27 (11.90)	24.93 (30.71)	1.82 (2.05)	9.51 (11.29)
	Re-entrenado	0.0001 (0.0001)	34.07 (43.12)	56.51 (75.26)	5.47 (5.94)	24.01 (30.87)
Muy Bajo	Base	0.1713 (0.2100)	687.17 (104.61)	2676.68 (1594.95)	6.48 (5.05)	842.62 (66.99)
	Re-entrenado	0.3488 (0.4219)	1216.21 (145.68)	2565.84 (188.52)	7.45 (6.25)	947.46 (88.16)

En la Tabla 12, se presentan los resultados en términos de WMAPE para cada parámetro de Stokes, así como el WMAPE promedio de los parámetros de Stokes. Dado que las amplitudes son diferentes, se utiliza el WMAPE para una comparación adecuada. Al analizar los resultados obtenidos, se observa que el error tiende a aumentar en el modelo re-entrenado, a excepción del caso del parámetro Stokes I en los

rangos alto y bajo. En el caso de Stokes Q, el modelo re-entrenado muestra errores significativos en el rango alto, y no se recomienda su uso en los demás rangos debido a los altos niveles de error. En cuanto a Stokes U, la situación es aún más preocupante, ya que solo el modelo entrenado desde cero con todos los datos resulta confiable para realizar predicciones. Sin embargo, en todos los demás rangos, incluido el modelo re-entrenado, se obtienen errores considerablemente altos, lo que lo hace poco confiable para este parámetro de Stokes.

Por último, al analizar el rendimiento en el parámetro de Stokes V, se observa una degradación en el desempeño del modelo re-entrenado en comparación con el modelo base. A pesar de ello, el rendimiento general sigue siendo bueno, siendo el peor caso en el rango muy bajo, donde el modelo re-entrenado presenta un error promedio del 7.45 %. En conclusión, el modelo re-entrenado para los parámetros Stokes I y Stokes V sigue ofreciendo buenos resultados, mientras que los problemas persisten en los casos de Stokes Q y U, que ya eran los más complicados para el modelo base.

4.8.2. Recuperación del campo magnético con el modelo re-entrenado

Habiendo determinado la cantidad de datos para el re-entrenamiento del modelo de estimación de parámetros de Stokes (35,000) y la cantidad de capas congeladas (las primeras 5 capas), se llevaron a cabo las transferencias de aprendizaje y re-entrenamientos de los modelos para cada rango. Con estos modelos re-entrenados, se realizó la recuperación del campo magnético de la misma manera que con el modelo base, utilizando 5 ejecuciones de PSO para cada instancia. En la Tabla 13, se muestra el WMAPE de los atributos más importantes del campo magnético para el modelo base y el modelo re-entrenado en los distintos rangos. Para el rango alto, se observan resultados bastante similares, con diferencias de menos del 0.4 % en la fuerza del momento dipolar (m), incrementándose a casi el 1 % en la posición del dipolo (X_r), y alrededor del 3 % en el ángulo de inclinación (i). Es en este último atributo donde el modelo re-entrenado muestra un mejor rendimiento.

Tabla 13. Tabla con el WMAPE por atributo del campo magnético del modelo en cada rango para los distintos rangos para el modelo base y el re-entrenado.

Rango del Modelo	Modelo	Fuerza del Momento Dipolar (m)	Inclinación (i)	Posición del Dipolo (X_r)
Alto	Base	0.34	8.55	1.74
	Re-entrenado	0.67	5.64	2.72
Bajo	Base	1.61	5.35	7.09
	Re-entrenado	3.90	4.29	16.41
Muy Bajo	Base	2.79	6.81	11.54
	Re-entrenado	2.93	6.64	8.73

En el rango bajo, se aprecia una mayor diferencia en la fuerza del momento dipolar entre el modelo base y el re-entrenado, llegando casi al 2.5 %. En la inclinación, nuevamente se obtiene un error menor en el modelo re-entrenado, con una diferencia de casi el 1 %, mientras que en este rango se observa que el atributo más desafiante es la posición del dipolo, donde el modelo re-entrenado empeora en un 9.5 % en comparación del modelo base. En el rango muy bajo, se observa un error muy similar en la fuerza del momento dipolar y el ángulo de inclinación, y la diferencia notable se encuentra en la posición del dipolo, donde el modelo re-entrenado logra reducir el error en aproximadamente un 3 % en comparación con el modelo base.

4.8.3. Resumen

En resumen, los experimentos preliminares nos permitieron establecer parámetros para el re-entrenamiento del modelo de estimación de los parámetros de Stokes. Con estos parámetros fijados, obtuvimos los resultados del modelo para cada rango, tanto en la estimación de los parámetros de Stokes como en la recuperación del campo magnético. En cuanto a la estimación de los parámetros de Stokes, el modelo generalmente empeora en cada parámetro, pero sigue obteniendo un rendimiento muy bueno en los Stokes I y V. Sin embargo, los Stokes Q y U siguen siendo los más desafiantes, especialmente en el rango bajo y muy bajo, donde se observan errores considerables en muchos casos.

En cuanto a la recuperación de la configuración del campo magnético, en general, se obtiene un rendimiento similar al modelo base en la fuerza del momento dipolar, sin que el error supere el 4 % en ningún caso. En el ángulo de inclinación, se observan pequeñas diferencias, excepto en el rango alto, donde el modelo re-entrenado muestra una mejora considerable. Aun así, el error en todos los casos ronda el 8.5 %. En cuanto a la posición del dipolo, se obtienen mayores diferencias entre los modelos en el rango bajo y muy bajo. En el rango bajo, el modelo re-entrenado muestra un rendimiento considerablemente peor, con un error cercano al 16.5 %. En cambio, en el rango muy bajo, el modelo re-entrenado muestra un mejor rendimiento.

Capítulo 5. Conclusiones

A través de nuestra investigación se implementaron y evaluaron modelos de aprendizaje profundo para estimar los parámetros de Stokes y posteriormente, recuperar la configuración magnética estelar para una configuración dipolar descentrada generalizada. Para la estimación de los parámetros de Stokes, utilizamos una red neuronal densa debido a su capacidad de salida múltiple y su capacidad de adaptarse al problema. Además, realizamos experimentos preliminares para determinar la arquitectura adecuada de la red neuronal. Los resultados obtenidos fueron muy satisfactorios. Al estimar los parámetros de Stokes I y V, logramos un MSE de $1.0e-07$ y $9.1e-08$, respectivamente. En términos de WMAPE, obtuvimos 0.02 % y 1.37 % para los parámetros de Stokes I y V en el rango alto, y valores similares para el rango bajo, con un WMAPE de 0.0007 % y 1.82 % para los mismos parámetros. Incluso en el rango muy bajo, los resultados fueron aceptables, con un WMAPE de 0.17 % y 6.48 % para Stokes I y V, respectivamente.

En general, los resultados indican que el modelo logra un WMAPE menor al 0.2 % y 6.5 % para los parámetros de Stokes I y V en cualquier rango. Sin embargo, en el caso de los parámetros de Stokes Q y U, encontramos que su rendimiento se degradaba a medida que la fuerza dipolar disminuía. Específicamente, en el rango muy bajo, el modelo no puede estimar correctamente estos dos parámetros de Stokes. Además, notamos que el rendimiento de los parámetros de Stokes V, Q y U se vio más afectado en las amplitudes menores.

Se obtuvo un rendimiento muy bueno en la recuperación de la configuración del campo magnético mediante la combinación de la red neuronal densa entrenada y el algoritmo de optimización PSO. En particular, evaluamos el WMAPE en la estimación de la fuerza del momento dipolar en los tres rangos diferentes, y los resultados fueron notables: 0.34 %, 1.61 % y 2.79 % para los rangos alto, bajo y muy bajo, respectivamente. Asimismo, analizamos el WMAPE para el ángulo de inclinación en los mismos tres rangos, obteniendo valores de 8.55 %, 5.35 % y 6.81 %. En relación al parámetro de la posición del dipolo, obtuvimos un WMAPE de 1.74 %, 7.09 % y 11.54 % en los rangos alto, bajo y muy bajo, respectivamente. Sin embargo, se observó una degradación en el rendimiento en este último atributo a medida que la fuerza del momento dipolar disminuía. Estos resultados respaldan la efectividad de la metodología propuesta para la recuperación de la configuración magnética.

A través de los experimentos preliminares, logramos establecer los parámetros adecuados para el re-entrenamiento del modelo de estimación de los parámetros de Stokes. Determinamos que con 35,000 casos es suficiente para obtener un rendimiento satisfactorio al re-entrenar la red neuronal densa en una longitud de onda diferente, lo cual resulta en una reducción significativa del tiempo de entrenamiento.

Estos resultados respaldan la viabilidad de utilizar el modelo re-entrenado para la recuperación del campo magnético. En general, se obtuvieron resultados similares en la recuperación de la fuerza del momento dipolar y el ángulo de inclinación utilizando el modelo re-entrenado. Sin embargo, en el caso de la posición del dipolo se observaron algunas diferencias más pronunciadas en algunos casos, aunque en términos generales se obtuvo un rendimiento bueno. Estos hallazgos destacan la promesa y la utilidad del modelo re-entrenado en la recuperación del campo magnético y la aplicación para otras longitudes de onda u otras estrellas.

Nuestro trabajo presenta una metodología innovadora para la medición del campo magnético en estrellas utilizando la teoría completa de la E RTP, mediante el uso de una red neuronal densa y PSO. Aunque se identificaron limitaciones en la estimación de los parámetros Stokes Q y U en rangos bajos, nuestros resultados prometen un avance significativo en la comprensión de los campos magnéticos y la configuración estelar. Recomendamos ampliar el conjunto de datos de entrenamiento y realizar una limpieza exhaustiva de los casos con amplitudes muy bajas debido a que en la práctica no tienen la suficiente amplitud para poder ser capturadas en las mediciones astronómicas.

Estos resultados prometen un avance significativo en la comprensión de los campos magnéticos y la configuración estelar. Mediante la mejora continua del modelo, podremos perfeccionar la estimación de los parámetros de Stokes y la recuperación de la configuración del campo magnético para obtener valiosos conocimientos sobre el comportamiento complejo de los campos magnéticos estelares, ahorrando tiempo y recursos.

5.1. Trabajo a futuro

Aplicando la metodología desarrollada a datos reales obtenidos de observaciones astronómicas utilizando la red neuronal densa entrenada y el uso de PSO, podremos estimar los parámetros de Stokes y así recuperar la configuración del campo magnético de estrellas reales. El uso de datos reales nos permitirá evaluar el rendimiento y la precisión de nuestra metodología en situaciones del mundo real. Será interesante analizar cómo se comporta el modelo entrenado con diferentes tipos de estrellas, incluyendo aquellas con campos magnéticos más complejos o con amplitudes bajas. La aplicación de técnicas de limpieza y pre-procesamiento de datos en la amplitud de los perfiles de Stokes es un área de investigación relevante que puede mejorar la calidad de los datos. Con esto se podría aprovechar la sensibilidad de los instrumentos para filtrar los datos que pueden estar fuera del rango de captura de las mediciones astronómicas.

Es importante resaltar que en este estudio se realizó el análisis sin la incorporación de ruido en los datos. Sin embargo, resulta crucial considerar la introducción de ruido en los datos y evaluar cómo afecta al rendimiento del modelo. Esta estrategia permitiría simular situaciones más realistas, ya que en la práctica los datos astronómicos suelen contener ruido. La introducción de ruido en los datos puede tener un impacto significativo en la capacidad del modelo para realizar estimaciones precisas. Por lo tanto, resulta necesario evaluar cómo el ruido afecta la calidad de las predicciones y determinar si es posible aplicar técnicas de reducción de ruido para mitigar este efecto adverso.

Además, es interesante experimentar con otras arquitecturas de redes neuronales profundas, como redes recurrentes, convolucionales o *transformers*. Estas arquitecturas podrían ofrecer ventajas adicionales en términos de captura de patrones temporales, extracción de características espaciales o modelado de relaciones a largo plazo, respectivamente. La exploración de estas alternativas podría brindar nuevas perspectivas y mejorar aún más el desempeño del modelo.

En este estudio, se empleó el modelo de dipolo descentrado generalizado como enfoque principal. Sin embargo, también se podría considerar la utilización de un modelo multipolar para representar superficies de campo magnético más complejas. Esto permitiría realizar un mapeo del campo magnético en toda la estrella, abriendo la posibilidad de explorar configuraciones magnéticas más detalladas y sofisticadas. La inclusión de un modelo multipolar ampliaría el espectro de análisis y proporcionaría una representación más completa de la distribución del campo magnético en la estrella.

Literatura citada

- Aggarwal, C. (2018). *Neural networks and deep learning*. Springer. <https://doi.org/10.1007/978-3-319-94463-0>.
- Amari, S. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4), 185–196. [https://doi.org/10.1016/0925-2312\(93\)90006-0](https://doi.org/10.1016/0925-2312(93)90006-0).
- Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford university press. <https://global.oup.com/academic/product/neural-networks-for-pattern-recognition-9780198538646>.
- Córdova, J., Navarro, S., & Ramírez, J. (2018). Machine learning for stellar magnetic field determination. *International Journal of Computational Intelligence Systems*, 11, 608–615. <https://doi.org/10.2991/ijcis.11.1.146>.
- Degl'Innocenti, E. & Landolfi, M. (2004). *Polarization in spectral lines*, volume 307. Springer Science & Business Media. <https://doi.org/10.1007/1-4020-2415-0>.
- del Toro, J. (2003). *Introduction to spectropolarimetry*. Cambridge University Press. <https://doi.org/10.1017/CB09780511536250>.
- European Southern Observatory (ESO) (SF). Polarimetría. Recuperado el 28 de abril del 2023. <https://www.eso.org/public/spain/teles-instr/technology/polarimetry/>.
- Gafeira, R., Orozco, D., Milić, I., Quintero, C., Ruiz, B., & Uitenbroek, H. (2021). Machine learning initialization to accelerate Stokes profile inversions. *Astronomy & Astrophysics*, 651, A31. <https://doi.org/10.1051/0004-6361/201936910>.
- Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information Sciences*, 222, 175–184. <https://doi.org/10.1016/j.ins.2012.08.023>.
- Hinton, G. (2012). Neural networks for machine learning. Recuperado el 17 de abril del 2023. https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.
- Kennedy, J. & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, 1942–1948 vol.4. <https://doi.org/10.1109/ICNN.1995.488968>.
- Kingma, D. & Ba, J. (2014). Adam: a method for stochastic optimization. *International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.1412.6980>.
- Knyazeva, I., Plotnikov, A., Medvedeva, T., & Makarenko, N. (2022). Multi-output deep learning framework for solar atmospheric parameters inferring from Stokes profiles. In *Advances in Neural Computation, Machine Learning, and Cognitive Research V*, 299–307. Springer International Publishing. https://doi.org/10.1007/978-3-030-91581-0_40.
- Kochenderfer, M. & Wheeler, T. (2019). *Algorithms for optimization*. Mit Press. <https://mitpress.mit.edu/9780262039420/algorithms-for-optimization/>.
- Landolfi, M. & Degl'Innocenti, E. (1982). Magneto-optical effects and the determination of vector magnetic fields from Stokes profiles. *Solar Physics*, 78(2), 355–364. <https://doi.org/10.1007/BF00151615>.
- Liu, H., Xu, Y., Wang, J., Jing, J., Liu, C., Wang, J., & Wang, H. (2020). Inferring vector magnetic fields from Stokes profiles of GST/NIRIS using a convolutional neural network. *The Astrophysical Journal*, 894(1), 70. <https://dx.doi.org/10.3847/1538-4357/ab8818>.

- Mathys, G. (1991). Spectropolarimetry of magnetic stars. II. The mean longitudinal magnetic field. *Astronomy & Astrophysics*, 89, 121. <https://ui.adsabs.harvard.edu/abs/1991A&AS...89..121M>.
- Mathys, G. (1995). Spectropolarimetry of magnetic stars. V. The mean quadratic magnetic field. *Astronomy & Astrophysics*, 293, 746–763. <https://ui.adsabs.harvard.edu/abs/1995A&A...293..746M>.
- Mathys, G. & Stenflo, J. (1987). Anomalous Zeeman effect and its influence on the line absorption and dispersion coefficients. *Astronomy & Astrophysics*, 171(1-2), 368–377. <https://ui.adsabs.harvard.edu/abs/1987A&A...171..368M>.
- Moré, J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical Analysis*, 105–116. Springer Berlin Heidelberg. <https://doi.org/10.1007/BFb0067700>.
- Nariai, K. (1969). Mass loss from coroneae and its effect upon stellar rotation. *Astrophysics and Space Science*, 3(1), 150–159. <https://doi.org/10.1007/BF00649601>.
- Paliwal, M. & Kumar, U. (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1), 2–17. <https://doi.org/10.1016/j.eswa.2007.10.005>.
- Patterson, J. & Gibson, A. (2017). *Deep Learning: A Practitioner's Approach*. O'Reilly Media, Inc. <https://www.oreilly.com/library/view/deep-learning/9781491924570/>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>.
- Phuoc, G. & Pham, T. (2018). Measuring the Stokes polarization parameters. In *6th International Conference on the Development of Biomedical Engineering in Vietnam (BME6)*, 593–596. Springer Singapore. https://doi.org/10.1007/978-981-10-4361-1_101.
- Ramírez, J., Yáñez, C., & Córdoba, J. (2018). Using machine learning algorithms to measure stellar magnetic fields. *Astronomy & Astrophysics*, 619, A22. <https://doi.org/10.1051/0004-6361/201833016>.
- Ramírez, J. (2020). On the use of the first-order moment approach for measurements of H_{eff} from LSD profiles. *Monthly Notices of the Royal Astronomical Society*, 493(1), 1130–1138. <https://doi.org/10.1093/mnras/staa301>.
- Rees, D. & Semel, M. (1979). Line formation in an unresolved magnetic element: a test of the centre of gravity method. *Astronomy & Astrophysics*, 74(1), 1–5. <https://ui.adsabs.harvard.edu/abs/1979A&A...74...1R>.
- Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>.
- Sharma, S., Sharma, S., & Athaiya, A. (2020). Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, 04, 310–316. <http://dx.doi.org/10.33564/IJEAST.2020.v04i12.054>.
- Stift, M. (1975). A non-axisymmetric rigid rotator model for magnetic stars. *Monthly Notices of the Royal Astronomical Society*, 172(1), 133–139. <https://doi.org/10.1093/mnras/172.1.133>.

- Stift, M. (2000). COSSAM: Codice per la Sintesi Spettrale nelle Atmosfere Magnetiche. https://www.ada2012.eu/cossam_ff/index.html.
- Stift, M. & Alecian, G. (2012). Modelling ApBp star atmospheres with stratified abundances consistent with atomic diffusion. *Monthly Notices of the Royal Astronomical Society*, 425(4), 2715–2721. <https://doi.org/10.1111/j.1365-2966.2012.21681.x>.
- Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1), 43–62. [https://doi.org/10.1016/S0169-7439\(97\)00061-0](https://doi.org/10.1016/S0169-7439(97)00061-0).
- Tsallis, C. & Stariolo, D. (1996). Generalized simulated annealing. *Physica A: Statistical Mechanics and its Applications*, 233(1), 395–406. [https://doi.org/10.1016/S0378-4371\(96\)00271-3](https://doi.org/10.1016/S0378-4371(96)00271-3).
- Weiss, K., Khoshgoftaar, T., & Wang, D. (2016). A survey of transfer learning. *Journal of Big data*, 3(1), 1–40. <https://doi.org/10.1186/s40537-016-0043-6>.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43–76. <https://doi.org/10.1109/JPROC.2020.3004555>.