



Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y Tecnología

Máster Universitario en Astrofísica y Técnicas de Observación
en Astronomía

Medición directa de la pequeña escala del medio intergaláctico con parejas cercanas de QSOs

Trabajo fin de estudio presentado por:	Juan Antonio Castro Fernández.
Tipo de trabajo:	Iniciación a la investigación en astronomía y astrofísica.
Línea de trabajo:	Cosmología, evolución del Universo y astrofísica de altas energías.
Director/a:	José Oñorbe Bernis. Francisco Miguel Montenegro Montes.
Fecha:	16/07/2025

Resumen

El objetivo principal de este estudio es estudiar las limitaciones observacionales para caracterizar la estructura a pequeña escala del medio intergaláctico (MIG) a partir de espectros de cuásares cercanos de calidad media o baja. Para ello, se ha implementado un método computacional en Python basado en el análisis estadístico de las diferencias de fase entre las líneas de visión de pares de cuásares próximos. Este enfoque permite acceder a información sobre la coherencia del flujo transmitido en escalas transversales del orden de decenas a cientos de kilopársecs.

Con el fin de validar la metodología, se han empleado simulaciones hidrodinámicas cosmológicas, que permiten generar espectros sintéticos con propiedades controladas y reproducir diversas configuraciones observacionales e instrumentales. Estas simulaciones han sido fundamentales para estudiar la respuesta del método frente a variaciones en la resolución espectral, la relación señal-ruido y la separación angular entre líneas de visión.

Los resultados obtenidos permiten establecer un conjunto de condiciones mínimas que garantizan la viabilidad del método propuesto, proporcionando una herramienta útil para el análisis de datos observacionales basados en el estudio de la pequeña escala del medio intergaláctico con datos limitados por resolución o ruido, y que será aplicable a los catálogos que se obtendrán en los próximos años (p.e. WEAVE-QSO), cuyo volumen de datos no tiene precedentes en comparación con los catálogos actuales.

Palabras clave: medio intergaláctico, bosque Lyman- α , cuásar, cosmología, métodos numéricos

Abstract

The main objective of this study is to investigate the observational limitations in characterizing the small-scale structure of the intergalactic medium (IGM) using medium- or low-quality spectra of closely-spaced quasars. To this end, a computational method has been implemented in Python, based on the statistical analysis of phase differences between the lines of sight of nearby quasar pairs. This approach provides access to information on the coherence of the transmitted flux on transverse scales ranging from tens to hundreds of kiloparsecs.

To validate the methodology, cosmological hydrodynamical simulations have been employed. These simulations allow the generation of synthetic spectra with controlled properties and the reproduction of various observational and instrumental configurations. They have been essential for studying the method's response to changes in spectral resolution, signal-to-noise ratio, and angular separation between lines of sight.

The results obtained establish a set of minimum conditions that ensure the viability of the proposed method, offering a useful tool for the analysis of observational data aimed at probing the small-scale structure of the IGM under resolution- or noise-limited conditions. The method is particularly well-suited for application to upcoming data sets (e.g., WEAVE-QSO), whose volume will be unprecedented compared to current quasar catalogs.

Keywords: intergalactic medium, Lyman- α forest, quasi-stellar object, cosmology, numeric methods

Índice de contenidos

1.	Introducción	1
1.1.	Motivación	1
1.2.	Planteamiento del problema	2
1.3.	Estructura del documento	3
1.4.	Estado del arte	3
1.4.1.	Fundamentos de Cosmología: expansión del Universo y medida de distancias	4
1.4.2.	Cuásares: Quasi-stellar objects	6
1.4.3.	Bosque de Lyman- α	7
1.4.4.	Propiedades internas del gas en el MIG	9
1.4.5.	Revisión de avances en caracterización de la pequeña escala del MIG	15
2.	Objetivos	17
2.1.	Objetivo general	17
2.2.	Objetivos específicos	17
3.	Material y método	18
3.1.	Marco teórico del método	18
3.1.1.	Análisis estadístico en el espacio de Fourier	18
3.1.2.	Medida no degenerada de la pequeña escala del MIG	20
3.2.	Implementación y ejecución del método	25
3.2.1.	Datos y entorno tecnológico	25
3.2.2.	Implementación en Python	29
4.	Discusión y resultados	45
4.1.	Introducción a los resultados	45

4.2.	Resultados a alta resolución espacial: 30 km/s.....	47
4.2.1.	Resultados obtenidos con SNR=6.....	47
4.2.2.	Resultados obtenidos con SNR=4.....	49
4.2.3.	Resultados obtenidos con SNR=2.....	50
4.3.	Resultados a media resolución espacial: 60 km/s.....	53
4.3.1.	Resultados obtenidos con SNR=6.....	53
4.3.2.	Resultados obtenidos con SNR=4.....	54
4.3.3.	Resultados obtenidos con SNR=2.....	57
4.4.	Resultados a baja resolución espacial: 100 km/s.....	57
4.4.1.	Resultados obtenidos con SNR=6.....	57
4.4.2.	Resultados obtenidos con SNR=4.....	59
4.4.3.	Resultados obtenidos con SNR=2.....	61
4.5.	Degeneración con el flujo medio.....	62
4.5.1.	Nuevos factores de normalización	62
4.5.2.	Resultados obtenidos en el estudio de la degeneración	64
5.	Conclusiones y Trabajo Futuro.....	67
5.1.	Limitaciones del método y los estudios	67
5.2.	Conclusiones extraídas del estudio	68
5.3.	Futuras líneas de trabajo	69
	Referencias bibliográficas.....	71
Anexo A.	Índice de Acrónimos	76

Índice de figuras

Figura 1. Bosque de $Ly\alpha$ en el espectro de emisión de los cuásares 3C 273 y Q1422+2309	8
Figura 2. Espectro normalizado de un QSO obtenido de simulaciones hidrodinámicas a redshift $z=2.4$	9
Figura 3. Distribución de densidad de materia bariónica (izquierda) y materia oscura (derecha) para $z=3$, obtenida de simulaciones hidrodinámicas.	10
Figura 4. Relación densidad-temperatura en el gas a $z=2.4$ en simulaciones hidrodinámicas	11
Figura 5. Bosque de $Ly\alpha$ y parámetros fundamentales del gas en el medio intergaláctico, obtenidos de simulaciones hidrodinámicas	14
Figura 6. PDF de la fase de los campos de densidad a distinta separación $\mathbf{r} \perp$, modo $\mathbf{k}$ y escala de Jeans λ_J	23
Figura 7. Espectro de las parejas de QSOs SDSS J073522.43+295710.1, SDSS J073522.55+295705.0 (panel A) y SDSS J102116.98+111227.6, SDSS J102116.47+111227.8 (panel B).....	25
Figura 8. Estructura interna de los archivos de parejas de QSOs de simulación hidrodinámica en formato HDF5	30
Figura 9. Bosque de $Ly\alpha$ a partir del flujo normalizado ideal (Panel A) y bosque de $Ly\alpha$ a partir del flujo normalizado con ruido para una SNR=10 (Panel B) y SNR=4 (Panel C)	36
Figura 10. Diagrama de flujo simplificado del proceso de aplicación de la transformada de Fourier y generación de la PDF	42
Figura 11. PDF de la diferencia de fase para los archivos de simulación con modelo térmico T000 y los modelos teóricos T000, T002 y T003 agrupadas por distancia transversal y modos k	43
Figura 12. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=6 y 40 parejas de QSOs.....	48

Figura 13. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=4 y 110 parejas de QSOs.....	49
Figura 14. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=2 y 195 parejas de QSOs.....	51
Figura 15. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=2 y 500 parejas de QSOs.....	52
Figura 16. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=6 y 115 parejas de QSOs.....	54
Figura 17. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 170 parejas de QSOs.....	55
Figura 18. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 150 parejas de QSOs.....	56
Figura 19. Resultados de la caracterización de modelos térmicos con resolución espacial de 100 km/s, SNR=6 y 170 parejas de QSOs.....	58
Figura 20. Resultados de la caracterización de modelos térmicos con resolución espacial de 100 km/s, SNR=4 y 300 parejas de QSOs.....	60
Figura 21. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 500 parejas de QSOs.....	61
Figura 22. Resultados de la caracterización del modelo térmico T000 a distintos flujos medios con resolución espacial de 60 km/s, SNR=4 y 170 parejas de QSOs.....	64
Figura 23. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 170 parejas de QSOs y el rango de variación de T000 debido a la degeneración con el flujo medio.....	65

Índice de tablas

Tabla 1. <i>Modelos térmicos de las simulaciones hidrodinámicas</i>	28
Tabla 2. <i>Factores de normalización para los tres modelos térmicos obtenidos con los métodos de bisección y Brent</i>	34
Tabla 3. <i>Distribución de parejas de QSOs esperadas para $z=2.4$, $SNR \geq 3$, $SB=21$ y $seeing=2.6$ por bloque de observación</i>	40
Tabla 4. <i>Número aproximado de parejas por bloque de observación para WEAVE a $seeing$ 2.6 y $SB=21$</i>	47
Tabla 5. <i>Factores de normalización para los tres modelos térmicos obtenidos con el método de bisección para $F_1 = 0.80460$</i>	63
Tabla 6. <i>Factores de normalización para los tres modelos térmicos obtenidos con el método de bisección para $F_2 = 0.82260$</i>	63

1. Introducción

1.1. Motivación

El medio intergaláctico (MIG) es el gas extremadamente tenue que ocupa los vastos espacios entre las galaxias. Está compuesto en su mayoría por hidrógeno ionizado, aunque también contiene helio y trazas de otros elementos ligeros. A pesar de su baja densidad, el MIG alberga la mayor parte de la materia bariónica del Universo. Debido a su presencia ubicua y a su papel en los intercambios de materia y energía entre galaxias, el MIG es una pieza clave para comprender cómo evolucionan las estructuras a gran escala en el cosmos. Su estudio permite reconstruir etapas tempranas del Universo y entender los procesos físicos que han guiado la formación y distribución de la materia a lo largo del tiempo.

Aproximadamente 370.000 años después del Big Bang, la recombinación de la materia bariónica dio lugar a la formación de átomos neutros y a la liberación de los fotones que hoy observamos como el fondo cósmico de microondas (CMB; Wilson, 1980). Comenzaba entonces la denominada Edad Oscura, caracterizada por la ausencia de fuentes visibles de luz. Con la expansión del Universo, el hidrógeno y el helio primordiales se enfriaron hasta temperaturas del orden de $20K$, mientras que las fluctuaciones de materia oscura y la gravedad moldeaban la materia bariónica, iniciando la formación de las primeras estructuras luminosas, como galaxias y estrellas (McQuinn, 2016; Xu et al., 2013).

Las primeras estrellas y núcleos de galaxia activos (AGN) fueron responsables de emitir fotones de alta energía capaces de reionizar el gas intergaláctico. Este calentamiento dejó un rastro térmico en el MIG que aún puede ser detectado hoy en día. A pesar de que el gas se enfría posteriormente por expansión adiabática y efecto Compton, estos procesos son mucho más lentos que el calentamiento inicial, lo que hace que el rastro térmico de la reionización predomine a temperaturas del orden de 10^4K (Hui & Gnedin, 1997). Esta memoria térmica convierte al MIG en un archivo histórico de los eventos de reionización, por lo que su estudio proporciona información clave sobre la evolución térmica y la reionización del Universo (Hui & Haiman, 2003).

1.2. Planteamiento del problema

La caracterización del estado térmico del MIG no puede realizarse en escalas arbitrarias: mientras que a gran escala (orden de Mpc) las fluctuaciones bariónicas trazan bien la distribución de materia oscura, a escalas menores ($\sim 100\ kpc$) entran en juego efectos físicos como la presión térmica, que suprime las fluctuaciones bajo una escala crítica conocida como escala o suavizado de Jeans. Este concepto, análogo al clásico criterio de colapso gravitatorio, limita la capacidad del gas para aglomerarse y constituye una medida directa del historial térmico del MIG (Gnedin et al., 1998; Kulkarni et al., 2015).

La escala de Jeans depende del calor inyectado durante la reionización cósmica, y establece la masa mínima necesaria para el colapso de estructuras y, por tanto, desempeña un papel fundamental en la formación de galaxias. En este sentido, tiene profundas implicaciones cosmológicas, pues determina hasta qué punto el MIG, en función de su estado térmico, puede inducir la formación de estructuras en el Universo. De hecho, estudios recientes muestran que la huella térmica de la reionización del hidrógeno neutro (HI) podría aún persistir en el MIG hasta desplazamientos al rojo de hasta $z \sim 4 - 5$ (Furlanetto & Oh, 2009), siendo en consecuencia medible mediante técnicas apropiadas como el análisis espectral de las líneas de absorción en la línea de visión de cuásares, conocidas como bosque de Lyman- α .

El análisis de estos espectros de absorción permite estudiar las propiedades físicas del MIG (densidad, temperatura, estado de ionización), sin embargo, esta técnica es muy sensible a la resolución espectral: sólo una resolución suficientemente alta permite identificar las estructuras sutiles que revelan la variabilidad del MIG a pequeña escala (Lidz et al., 2010).

Una estrategia particularmente potente es el análisis del bosque de Ly α en pares cercanos de cuásares. Al estudiar espectros de líneas de visión separadas por distancias transversales de cientos de kilopársecs, se puede medir directamente la coherencia del bosque y, por tanto, obtener una estimación empírica de la escala característica del MIG (Rorai et al., 2017). Hasta ahora, estas medidas se han limitado a un número muy reducido de espectros de altísima resolución. Sin embargo, nuevos proyectos como WEAVE-QSO y DESI (Alexander et al., 2023;

Jin et al., 2024) están generando grandes catálogos de espectros de resolución media o baja, lo que abre la puerta al estudio de estas escalas con muestras de mayor tamaño.

A pesar de los avances mencionados, aún no se ha determinado con claridad cuáles son los requisitos técnicos mínimos (como resolución espectral, relación señal-ruido y tamaño de muestra) para poder estimar de manera robusta la escala característica del MIG mediante los nuevos catálogos de cuásares de resolución media o baja.

En este Trabajo de Fin de Estudios (TFE) se analizarán los métodos empleados para la caracterización de la pequeña escala del MIG y se aplicarán a espectros de resolución media y baja, con el objetivo de establecer los criterios necesarios para que estas medidas puedan realizarse a partir de espectros de baja calidad, simulando las condiciones observacionales futuras mediante el uso de simulaciones hidrodinámicas cosmológicas.

1.3. Estructura del documento

En este Capítulo se presenta el contexto teórico que permite comprender con precisión los estudios llevados a cabo en este TFE, así como la motivación y el planteamiento de la problemática abordada, ambos expuestos en los apartados anteriores. En el Capítulo 2 se presentan tanto los objetivos generales como específicos formulados. En el Capítulo 3 se detalla el método empleado en la caracterización de la pequeña escala del MIG y su marco teórico específico, así como los pasos necesarios para su reproducción a partir de los datos utilizados y la implementación de éste dentro del entorno tecnológico en el que se desarrollan estos estudios. En el Capítulo 4 se exponen los resultados obtenidos en la aplicación del método a los datos obtenidos de simulaciones hidrodinámicas y se desarrolla la discusión sobre éstos. En el Capítulo 5 se detallan las conclusiones extraídas de los análisis y estudios aplicados y se plantean las posibles líneas de trabajo futuras.

1.4. Estado del arte

A lo largo de esta sección se revisarán los conceptos fundamentales necesarios para comprender tanto los análisis desarrollados en los posteriores capítulos como los resultados obtenidos a partir de éstos. Asimismo, se presentarán los principales avances realizados hasta

la fecha en el estudio del medio intergaláctico, poniendo especial atención a los métodos desarrollados y empleados en la caracterización de la pequeña escala del MIG.

1.4.1. Fundamentos de Cosmología: expansión del Universo y medida de distancias

El análisis de los datos observacionales del MIG, en particular el que se lleva a cabo a partir de espectros de cuásares, requiere un marco teórico sólido en cosmología. En este contexto, resulta imprescindible revisar algunos conceptos fundamentales que permiten interpretar de forma rigurosa los resultados derivados de simulaciones y observaciones.

1.4.1.1. Expansión del Universo y desplazamiento al rojo

Uno de los pilares de la cosmología contemporánea es la constatación empírica de que el Universo se encuentra en expansión. Esta expansión fue evidenciada por Edwin Hubble en 1929 al observar la correlación entre la distancia de las galaxias y el desplazamiento hacia el rojo de sus espectros, y es actualmente conocida como la Ley de Hubble. Una de las principales consecuencias de la expansión es, por tanto, el conocido como corrimiento al rojo cosmológico o *redshift*, que refleja el alargamiento progresivo de las longitudes de onda de los fotones a medida que estos atraviesan un Universo en expansión.

El *redshift* puede cuantificarse mediante la siguiente relación:

$$\lambda_{obs} = (1 + z) \cdot \lambda_{emit} , \quad (1)$$

donde λ_{obs} es la longitud de onda observada, λ_{emit} es la longitud de onda emitida por la fuente, y z representa el *redshift* en el que se encuentra la fuente. A diferencia del efecto Doppler clásico, este fenómeno no implica movimiento relativo en un espacio estático, sino la dilatación del propio espacio-tiempo. La observación de líneas espectrales conocidas desplazadas al rojo constituye, por tanto, una prueba directa de la expansión cósmica.

Desde el punto de vista observacional, el *redshift* permite detectar en el rango del espectro visible líneas emitidas originalmente en el ultravioleta, si las fuentes que las emitieron se encuentran a suficiente distancia. En el caso que ocupa a estos estudios, la línea Ly α

(1215.67 Å en el vacío) puede observarse a $\lambda \approx 4133$ Å para una fuente situada en $z = 2.4$, permitiendo una observación de dicha línea con telescopios terrestres.

1.4.1.2. Definición de distancias y normalización de unidades

El concepto de distancia en cosmología difiere del utilizado en los contextos euclídeos de la física clásica ya que, en contextos cosmológicos, el espacio mismo se encuentra en expansión. Esto obliga a definir distintas nociones de distancia según el marco teórico adoptado. Se destacan dos definiciones de distancia:

Por un lado, la distancia propia, definida como la distancia física real entre un observador (normalmente en la Tierra) y un punto del Universo; y la distancia comóvil, que mide la separación entre dos puntos del Universo excluyendo el efecto de la expansión misma, y que permanece constante si no hay movimiento peculiar. Esta definición resulta especialmente útil en simulaciones cosmológicas, motivo por el cual se utilizará de forma habitual en los resultados de este TFE.

Las distancias comóviles se expresan habitualmente en megapársecs comóviles $cMpc/h$, donde h es la constante de Hubble normalizada, definida como:

$$h = \frac{H_0}{100 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1}} \quad (2)$$

siendo H_0 la constante de Hubble, cuyo valor actual se estima entre $67.4 \text{ km s}^{-1} \text{Mpc}^{-1}$, a partir del CMB (Planck Collaboration et al., 2020) y $73 \text{ km s}^{-1} \text{Mpc}^{-1}$, estimado a partir de observaciones de super novas tipo Ia (Riess et al., 2022).

El uso de unidades comóviles permite comparar resultados obtenidos bajo diferentes supuestos cosmológicos, al desacoplar la distancia del valor de H_0 , lo cual resulta especialmente relevante en el análisis del bosque de Lyman- α .

En los posteriores estudios, las escalas espaciales k derivadas de la transformada de Fourier del flujo de cuásares proporcionan la inversa de la escala espacial, y pueden darse tanto en unidades de s/km y unidades de $h/cMpc$, normalizando así las escalas asociadas a las fluctuaciones del MIG.

1.4.1.3. Velocidades de recesión en un Universo en expansión

Una implicación destacada del modelo de universo en expansión es que las velocidades de recesión derivadas de objetos suficientemente lejanos pueden exceder el valor de la velocidad de la luz en el vacío, sin que ello infrinja los principios de la relatividad. Esta aparente paradoja se resuelve al reconocer que estas velocidades no representan desplazamientos locales de objetos a través del espacio, sino el crecimiento de la métrica del espacio-tiempo entre dos puntos comóviles. En otras palabras, no existe una transmisión de información a velocidades supralumínicas, sino una expansión global del tejido del Universo.

En contextos observacionales, es habitual expresar la velocidad de recesión efectiva como:

$$v = c \cdot z \quad (3)$$

Esta expresión permite definir una distancia como una velocidad de recesión en, comúnmente, km/s . Si bien esta relación es una aproximación válida únicamente a *redshifts* bajos (a *redshifts* mayores la relación se vuelve no lineal y es necesario usar modelos relativistas), permite ilustrar cómo velocidades muy elevadas son una consecuencia natural del modelo cosmológico y no deben interpretarse como una violación de la física relativista.

1.4.2. Cuásares: Quasi-stellar objects

Un cuásar, u objeto cuasi estelar (*quasi-stellar object*, QSO), es un objeto astronómico extremadamente luminoso y compacto, cuya apariencia estelar contrasta con su verdadera naturaleza: la emisión energética generada por la acreción de materia en torno a agujeros negros supermasivos situados en el centro de galaxias distantes. Emiten en un amplio rango del espectro electromagnético, desde radio hasta rayos X, y a niveles de intensidad muy altos, lo que permite que sean observados incluso a distancias muy lejanas por el desplazamiento al rojo de sus emisiones. El primero en identificarse fue 3C 273, a partir de observaciones radio y espectros con $z \approx 0.16$ (Schmidt, 1963), lo que marcó un hito en la comprensión de estos objetos.

Los QSOs son especialmente relevantes para el estudio del MIG por tres razones clave: su gran distancia permite observar, con telescopios terrestres, el bosque de Ly α en el espectro visible

para $z \gtrsim 2$; su enorme luminosidad posibilita obtener espectros con alta relación señal/ruido y resolución espectral a pesar de sus distancias; y su continuo espectral facilita la detección de líneas de absorción del hidrógeno neutro. Estas propiedades los convierten en fuentes ideales para investigar la distribución del hidrógeno en el Universo a través de análisis espectrales.

En este TFE se emplearán espectros generados a partir de simulaciones hidrodinámicas, constituyendo la base sobre la cual se desarrollan los análisis estadísticos destinados a alcanzar los objetivos planteados.

1.4.3. Bosque de Lyman- α

Se conoce como bosque de Ly α al conjunto de líneas de absorción que aparecen en el espectro de un cuásar o una galaxia lejana, debido a que los fotones emitidos por el cuásar interaccionan con sucesivas nubes de hidrógeno neutro a lo largo de la línea de visión hasta su llegada al observador terrestre.

En 1971, durante las observaciones del cuásar 4C 05.34 se encontraron distintas líneas de absorción en su espectro y, tras su análisis, se determinó que pertenecían a la línea de Ly α (Lynds, 1971).

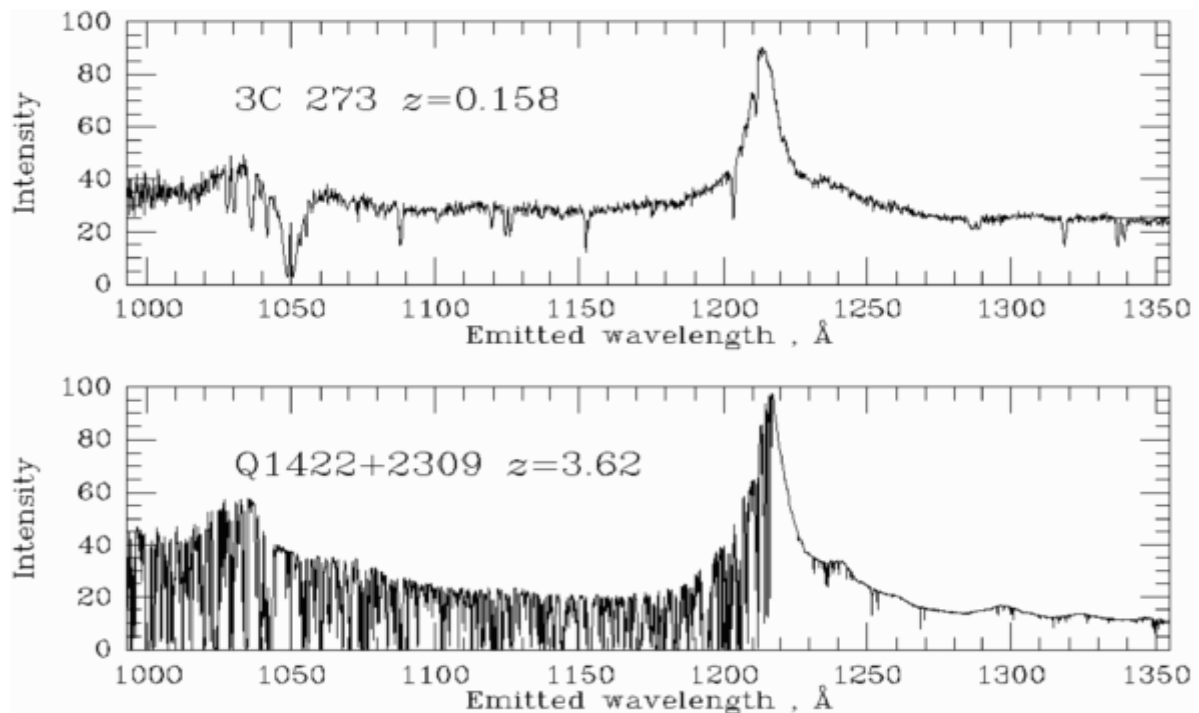
La absorción o línea de Lyman- α es la primera línea espectral del hidrógeno perteneciente a la Serie de Lyman, y que se produce cuando el electrón cae del orbital $n = 2$ al orbital $n = 1$ emitiendo radiación a una longitud de onda de 1215,67 Å en el vacío, lo que sitúa esta línea en el espectro ultravioleta lejano. Lo realmente interesante de esta banda de absorción es que, gracias al *redshift*, es visible en el óptico a determinadas distancias, posibilitando su detección con telescopios terrestres (Adelberger et al., 2004). El *redshift* al que la línea de Ly α es detectable en el óptico desde la superficie terrestre está en torno a $z \approx 2$.

En la Figura 1 puede verse el espectro de emisión de los cuásares 3C 273 y Q1422+2309, correspondientes al flujo continuo emitido, y en los cuales se aprecia un pico de emisión alrededor la longitud de onda correspondiente a la transición de Ly α , producida por la ionización del hidrógeno neutro circundante a éstos. En la Figura 2 se representa el bosque de Ly α para un QSO, obtenido a partir del flujo normalizado observado, es decir, la fracción del

flujo observado respecto al continuo emitido por el cuásar, que refleja cuánta absorción ha producido el medio intergaláctico. Estas absorciones corresponden todas a la transición $\text{Ly}\alpha$, pero se observan a diferentes longitudes de onda desplazadas al rojo, lo que refleja la distribución del hidrógeno neutro en distintas épocas del Universo.

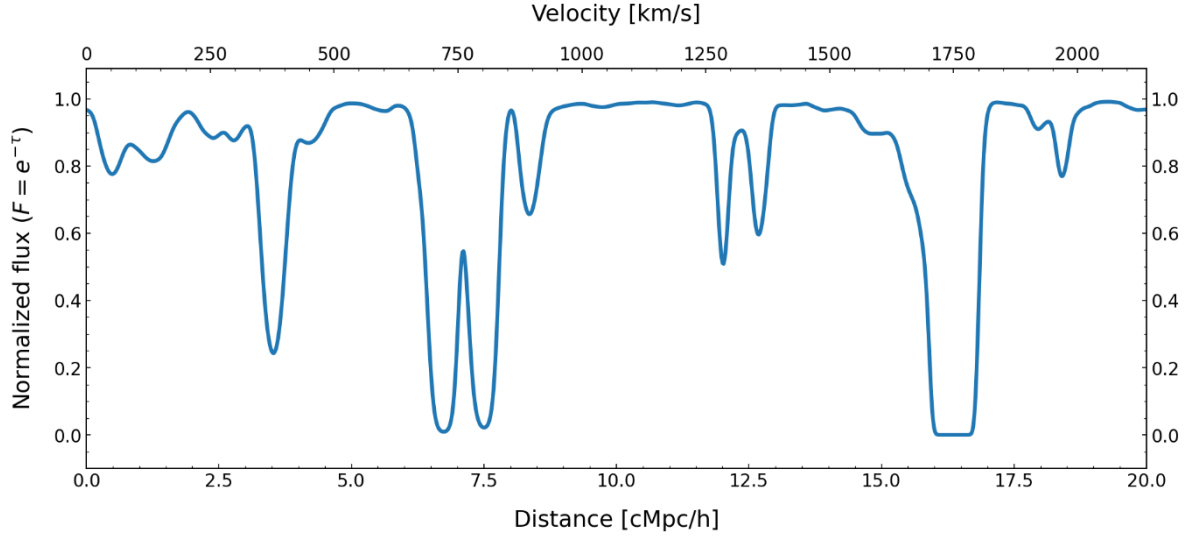
El descubrimiento del bosque de $\text{Ly}\alpha$ abrió una vía clave en el estudio del MIG a gran resolución y en escalas cosmológicas. Desde entonces, múltiples estudios han aprovechado estas líneas de absorción para caracterizar la morfología, evolución y dinámica del hidrógeno neutro en el MIG (Sargent et al., 1980; Wolfe, 1986; Foltz et al., 1984; Crofts, 1989). De forma análoga, el estudio de bosques de $\text{Ly}\alpha$ generados mediante simulaciones hidrodinámicas permitirán constreñir las propiedades a pequeña escala del MIG con alta precisión.

Figura 1. Bosque de $\text{Ly}\alpha$ en el espectro de emisión de los cuásares 3C 273 y Q1422+2309



Fuente: <https://astro.ucla.edu/~wright/Lyman-alpha-forest.html>

Figura 2. *Espectro normalizado de un QSO obtenido de simulaciones hidrodinámicas a redshift $z=2.4$*



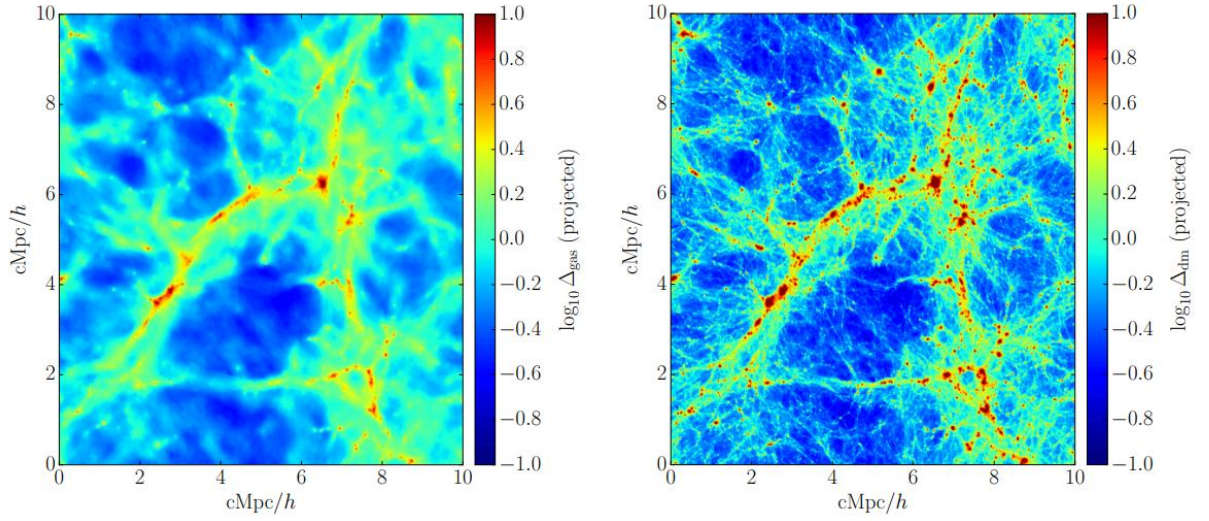
Fuente: elaboración propia.

1.4.4. Propiedades internas del gas en el MIG

Con el objetivo de comprender con profundidad los métodos empleados en la caracterización de la pequeña escala del MIG, es necesario analizar las propiedades termodinámicas del gas que lo conforma, así como las distintas variables que entran en juego en la regulación de su comportamiento. Partiendo de este contexto, la temperatura del gas y su relación con la densidad local es fundamental, ya que la presión generada por la temperatura interna del gas limita el crecimiento de las fluctuaciones de densidad inducidas por la materia oscura. En otras palabras, cuando se evalúa la distribución de la materia bariónica a grandes escalas (del orden de $\sim 5 \text{ Mpc}$), la materia bariónica tiende a seguir el potencial gravitatorio impuesto por la materia oscura, adoptando así ambas una distribución similar. Por el contrario, a escalas menores (inferiores a 1 Mpc), la materia bariónica deja de seguir la distribución filamentosa de la materia oscura debido a las diferencias en su comportamiento físico. Esto es, mientras que la materia oscura se comporta como un fluido acolisional cuya evolución está dominada por la gravedad, la materia bariónica, al ser un gas colisional, experimenta los efectos de la presión generada por la temperatura que suavizan las pequeñas fluctuaciones de densidad. A

este fenómeno se le conoce como filtrado de presión, e impide que la materia bariónica se aglutine con la misma eficacia a esas escalas, lo que introduce una diferencia clara en la distribución espacial de ambas materias. La escala característica a partir de la cual la presión térmica comienza a dominar sobre la gravedad y evita el colapso de pequeñas estructuras es la longitud de Jeans, que establece una masa mínima para la formación de objetos en el MIG y condiciona la evolución de las primeras estructuras bariónicas (McQuinn, 2016).

Figura 3. Distribución de densidad de materia bariónica (izquierda) y materia oscura (derecha) para $z=3$, obtenida de simulaciones hidrodinámicas.



Fuente: Kulkarni et al., 2015

Por tanto, la caracterización del estado térmico del gas resulta imprescindible para poder entender en detalle cómo evoluciona la distribución de la materia bariónica en el MIG.

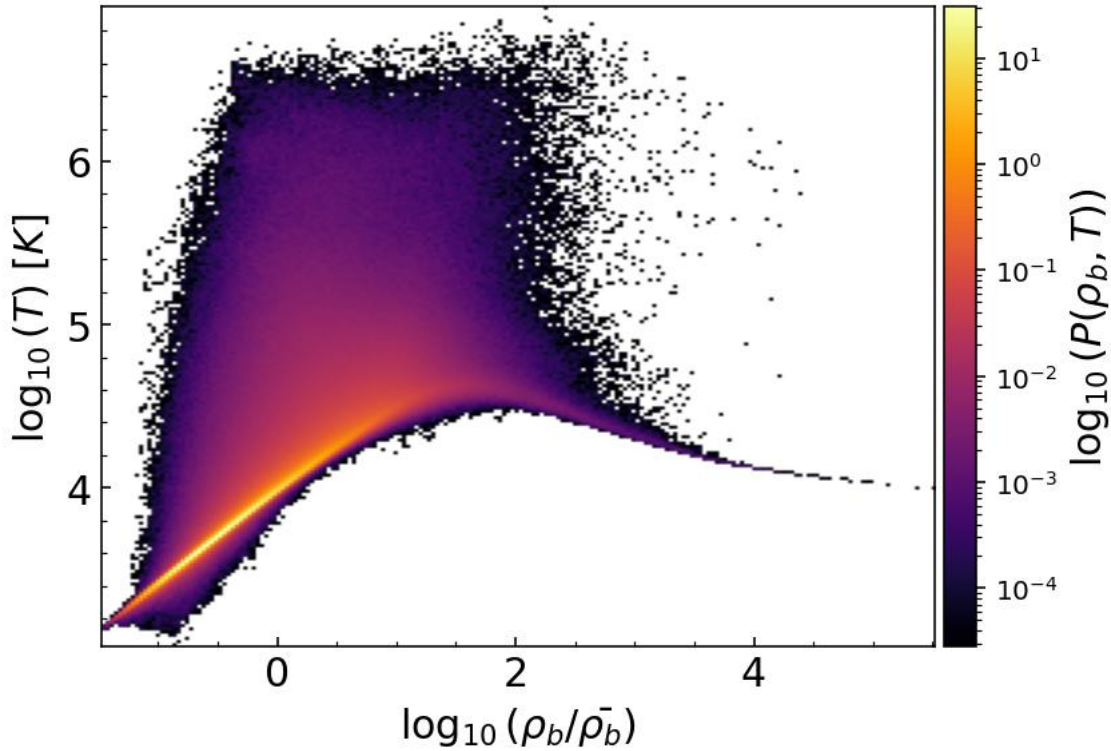
El estado térmico del gas en el MIG se suele describir con una ley de potencias tal que:

$$T = T_0 \left(\frac{\rho}{\bar{\rho}} \right)^{\gamma-1} = T_0 \cdot \Delta_b^{\gamma-1} \quad (4)$$

donde T_0 es la temperatura del gas a la densidad media $\bar{\rho}$, y $\gamma - 1$ es la pendiente de la relación densidad-temperatura, $T - \Delta_b$ (Lukić et al., 2015). De esta forma, T_0 representa un

parámetro característico de la historia térmica promedio del gas, mientras que γ indica la respuesta del medio a las variaciones locales de densidad.

Figura 4. Relación densidad-temperatura en el gas a $z=2.4$ en simulaciones hidrodinámicas



Fuente: elaboración propia

La Figura 4 muestra la función de densidad de probabilidad, PDF (*probability density function*), conjunta para la densidad y la temperatura de una matriz de datos correspondiente a una línea de visión aleatoria que cruza un volumen de $20 \times 20 \times 20 \text{ Mpc}$ en las simulaciones hidrodinámicas usadas en este TFE. Puede apreciarse la relación de ley de potencias entre densidad y temperatura en la zona del diagrama con mayor densidad de probabilidad (región amarilla) y que sigue una recta de pendiente $\gamma - 1$.

El valor de γ suele oscilar entre 1 y 1.6, aproximadamente (McQuinn, 2016). Así, un valor cercano a 1 corresponde a un gas casi isotérmico, donde la temperatura es prácticamente

independiente de la densidad local. Por otra parte, para valores de $\gamma > 1$ se tiene un calentamiento diferencial del gas, donde las regiones con mayor densidad tienen temperaturas mayores debido a la compresión del gas y otros procesos físicos como la reionización o la inyección de energía de otras fuentes (Hui & Gnedin, 1997).

El parámetro γ es crucial para comprender cómo la presión térmica regula las fluctuaciones de densidad en el MIG. Como se ha descrito anteriormente, una relación densidad-temperatura con $\gamma > 1$ significa que las regiones más densas son también las más calientes, lo que implica a su vez un aumento de la presión térmica local y, por tanto, una amortiguación de las perturbaciones presentes a escalas menores. La relación $T - \Delta_b$ tiene también influencia en otros parámetros fundamentales del gas como la densidad de hidrógeno neutro en el medio intergaláctico n_{HI} , y la profundidad óptica τ , a partir de la cual se obtiene el bosque de Ly α .

En el marco teórico en el que se encuadra este TFE, la absorción de la línea de Ly α se caracteriza por la profundidad óptica τ , que cuantifica la fracción de energía que ha sido absorbida a lo largo de la línea de visión. La profundidad óptica tiene una fuerte dependencia con la densidad de materia bariónica en unidades de la densidad media Δ_b y con la fracción de hidrógeno neutro x_{HI} , y se relaciona con la probabilidad de absorción a través de $P = \exp(-\tau_{Ly\alpha})$ (McQuinn, 2016). Por otra parte, cabe destacar que la fracción de hidrógeno neutro presente en el gas, dada por la Ecuación 5, resulta determinante en la absorción de Ly α y depende directamente del equilibrio entre la tasa de recombinación α_A , que escala con la temperatura según $\alpha_A \propto T^{-0.7}$ (McQuinn, 2016), la densidad de electrones n_e y la tasa de fotoionización de hidrógeno Γ_{HI} . Por tanto, una estimación precisa de estos parámetros es necesaria para la correcta caracterización del bosque de Ly α y la realización de análisis estadísticos con errores e incertidumbre mínimos.

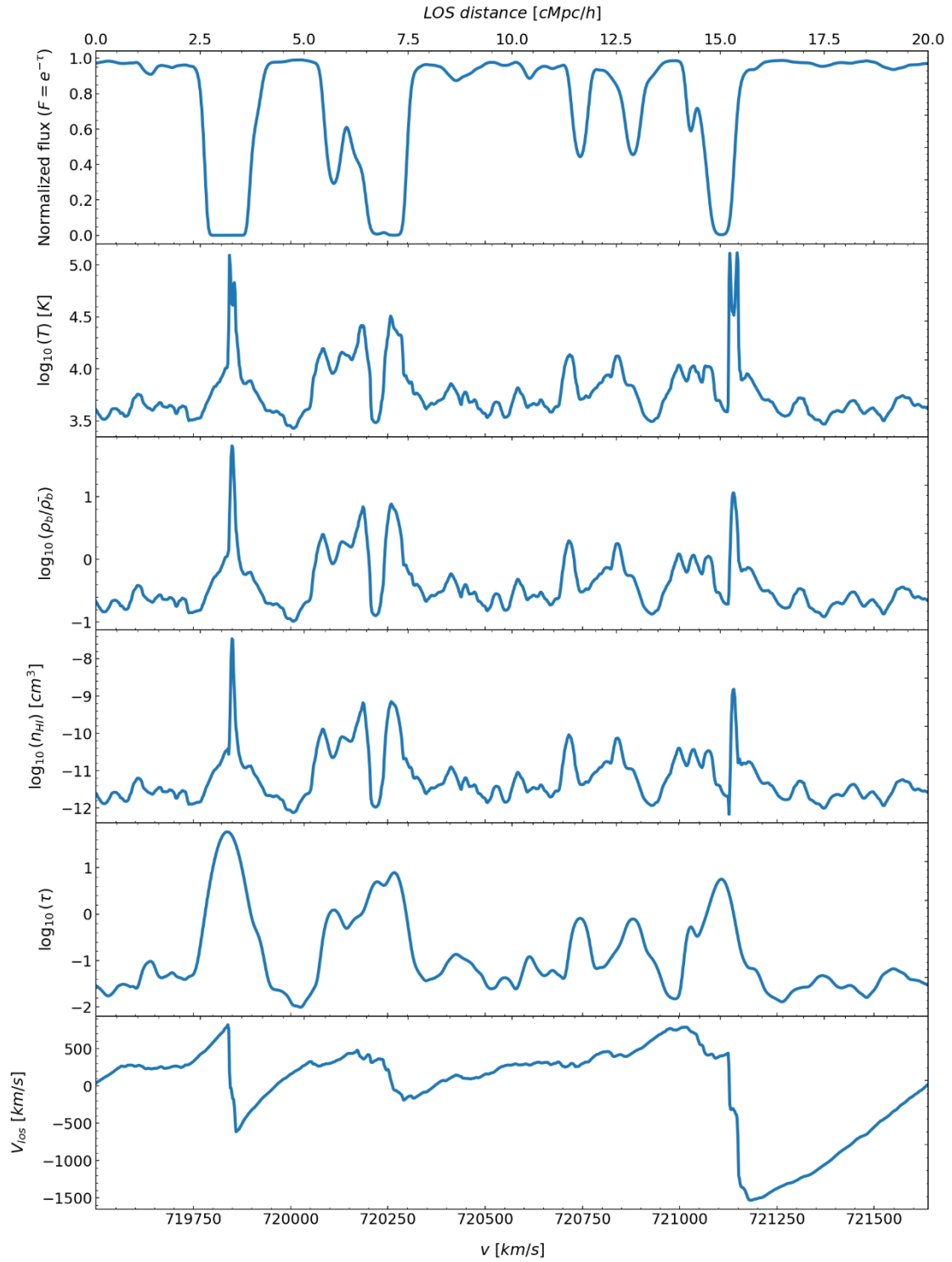
$$x_{HI} = \frac{\alpha_A n_e}{\Gamma_{HI}} \quad (5)$$

La relación existente entre los distintos parámetros del MIG y el bosque de Ly α puede verse con detalle en la Figura 5. En ella se representan en distintas gráficas, en orden descendente,

el bosque de Ly α obtenido a partir del flujo normalizado, el logaritmo de la temperatura del gas, el logaritmo de la variación de la densidad respecto a la densidad media, el logaritmo de la densidad de la columna de hidrógeno neutro, el logaritmo de la profundidad óptica y la velocidad de recesión en la línea de visión, frente a la distancia en la línea de visión del volumen de simulación (eje superior) y la velocidad real de recesión (eje inferior). Como se ha comentado y puede verse en la gráfica, un aumento de la densidad en regiones determinadas del gas implica a su vez un aumento de la temperatura y de la profundidad óptica. Esto se refleja en una disminución del flujo recibido, dando lugar a las líneas de absorción que conforman el bosque de Ly α . Como cabe esperar, una disminución del flujo recibido por la absorción en la línea de Ly α se corresponde con una mayor densidad de la columna de hidrógeno neutro. Por último, el aumento de densidad se refleja también en la componente de velocidad de recesión en la línea de visión, ya que el gas cae hacia los pozos de potencial desde ambos lados: desde el lado más cercano al observador (lo que genera velocidades peculiares positivas) y desde el lado más cercano al QSO, donde el gas se acerca al observador, produciendo velocidades peculiares negativas.

En definitiva, la caracterización de T_0 y γ es un paso imprescindible en la comprensión de la estructura del MIG a pequeña escala, ya que encapsulan la historia térmica del gas y su respuesta frente a los mecanismos de enfriamiento y calentamiento. Las observaciones, junto con las simulaciones que proporcionan resoluciones muy altas, no sólo permiten hacer una estimación de estos parámetros, sino que también sirven como vía para la validación de modelos cosmológicos y para el estudio del impacto de procesos como reionización y realimentación galáctica. En el contexto de este TFE, las restricciones que vienen impuestas por los valores de T_0 y γ son esenciales para la interpretación de los resultados derivados de los análisis estadísticos del bosque de Ly α y su relación con las propiedades internas del MIG.

Figura 5. *Bosque de Ly α y parámetros fundamentales del gas en el medio intergaláctico, obtenidos de simulaciones hidrodinámicas*



Fuente: elaboración propia

1.4.5. Revisión de avances en caracterización de la pequeña escala del MIG

Como se ha visto, el valor de T_0 en el interior del gas que compone el MIG es uno de los factores clave en el proceso de caracterización de la pequeña escala. Los métodos usados en la medida de temperatura del MIG son variados y quedan fuera del ámbito de estudio de este TFE, pero merece la pena mencionar alguno como, por ejemplo, la obtención de la temperatura a partir del ancho de las líneas de absorción en función de N_{HI} (Schaye et al., 1999) o la medida de la curvatura del espectro (Becker et al., 2011; Boera et al., 2014). Las medidas llevadas a cabo durante la segunda década del siglo XXI parecen coincidir en que la temperatura del MIG, para la densidad media y en un desplazamiento al rojo de $2 < z < 4$, está alrededor de $T_0 = (10 - 20) \times 10^3 K$ (Becker et al., 2011; Garzilli et al., 2012).

El estudio de la estructura del MIG a pequeña escala ha experimentado avances significativos gracias a nuevos enfoques basados en el análisis del flujo transmitido en el bosque de Ly α . Tradicionalmente, la principal herramienta para inferir el estado térmico del MIG ha sido el espectro de potencia del flujo longitudinal, sensible a parámetros clave como la temperatura del gas T_0 , el índice adiabático γ y la escala de suavizado de Jeans λ_J . Sin embargo, este método sufre una degeneración entre estos parámetros, ya que distintos modelos pueden producir espectros de potencia muy similares debido a la acción combinada del ensanchamiento térmico y el suavizado por presión, ambos responsables de la supresión de potencia en altas frecuencias (más detalle en §3.1.2).

Para superar esta limitación, Rorai et al. (2013) aplicaron un enfoque alternativo basado en el análisis de la diferencia de fases entre las transformadas de Fourier del flujo transmitido por parejas de cuásares cercanos. A diferencia del espectro de potencia, que solo mide la amplitud de las fluctuaciones, el análisis de fase revela información sobre la coherencia espacial de las estructuras que absorben la luz, y permite detectar correlaciones transversales entre líneas de visión que con los métodos tradicionales no era posible detectar sin la incertidumbre introducida por la degeneración entre los parámetros T_0 y γ . Este método es especialmente

sensible a λ_J , pero prácticamente insensible a T_0 y γ , acabando así con la degeneración del espectro longitudinal.

Otra ventaja clave de este enfoque es su robustez frente a los efectos de la resolución instrumental: el suavizado de Jeans actúa como un filtro de paso bajo, eliminando los modos de Fourier afectados por la resolución instrumental sin necesidad de un modelado detallado del instrumento. Además, el carácter no lineal del campo del flujo es lo suficientemente débil como para permitir aplicar estadística de Fourier con buena precisión.

Rorai et al. (2017) aplicaron esta técnica a una muestra de 25 pares de cuásares de alta resolución, alta SNR (15-20) y con separaciones transversales de $r_{\perp} = 100 - 500 \text{ kpc}$, extrayendo la distribución de diferencias de fase y ajustándola con una distribución de Cauchy. En sus resultados, obtuvieron una estimación de λ_J consistente con los modelos estándar de reionización, lo que valida empíricamente este método. Estos resultados consolidan el análisis de fases como una herramienta clave y no degenerada para explorar la física del MIG a escalas pequeñas, y sirve como punto de partida para los estudios que se desarrollan en los capítulos posteriores.

2. Objetivos

El objetivo principal de este TFE es llevar a cabo una investigación fundamentada en análisis estadísticos de datos procedentes de simulaciones hidrodinámicas que modelan la evolución del medio intergaláctico en el marco de los modelos cosmológicos actuales, para determinar los criterios de resolución de la pequeña escala del medio intergaláctico con espectros de resolución media y baja.

2.1. Objetivo general

Elaborar un proceso automatizado en Python que analice archivos de simulaciones hidrodinámicas en formato HDF5 y obtenga los criterios de resolución de la pequeña escala del medio intergaláctico para los modelos cosmológicos y térmicos en que están basados.

2.2. Objetivos específicos

Los objetivos específicos en los que se descompone el objetivo global de este TFE son:

- Analizar en profundidad las propiedades físicas del medio intergaláctico mediante el estudio de su evolución, composición y dinámica en el contexto cosmológico actual.
- Desarrollar automatismos para la extracción y procesamiento de datos contenidos en archivos HDF5 provenientes de simulaciones hidrodinámicas del MIG.
- Aplicar técnicas de análisis espectral, concretamente la transformada de Fourier, para caracterizar la estructura del MIG a distintas escalas.
- Evaluar el impacto de los siguientes parámetros observacionales en la recuperación de información física relevante:
 - Resolución espectral
 - Relación señal-ruido
 - Tamaño de la muestra
- Establecer criterios cuantitativos de resolución mínima para el estudio de la pequeña escala del MIG, identificando los límites de precisión alcanzables en función de los errores estadísticos y sistemáticos esperados.
- Determinar la existencia de posibles degeneraciones en el nuevo método.

3. Material y método

Los métodos de caracterización de la pequeña escala del MIG aplicados en este TFE se basan en estudios estadísticos aplicados sobre análisis frecuenciales con la transformada de Fourier de espectros de parejas de QSOs cercanos (Rorai et al., 2013). En esta sección se describirán los métodos de análisis espectral empleados, así como los automatismos implementados en la obtención de la estadística necesaria de los datos de simulaciones hidrodinámicas que permiten alcanzar los objetivos propuestos.

La ejecución de la investigación llevada a cabo se divide en dos grandes bloques: en primer lugar, estudio de las herramientas matemáticas y estadísticas usadas y, por último, preparación del entorno tecnológico, obtención de los datos e implementación del método de caracterización de la pequeña escala del MIG.

3.1. Marco teórico del método

3.1.1. Análisis estadístico en el espacio de Fourier

Como se comentaba al inicio de este Capítulo, los métodos empleados se basan en la transformada de Fourier y los análisis estadísticos aplicados a los espectros de las parejas de QSOs. Para poder procesar los datos y entender los resultados obtenidos es necesario comprender las herramientas matemáticas y estadísticas que se integran en los procesos de trabajo automatizados.

3.1.1.1. La transformada de Fourier aplicada a la caracterización del MIG

La transformada de Fourier de una función $f(x) \in R$ o C se define como:

$$\tilde{f}(k) = \int_{-\infty}^{\infty} f(x) e^{-2\pi i k x} dx, \quad (6)$$

donde:

- $f(x)$ es una función en el espacio real. En este caso, el flujo normalizado que conforma el bosque de Ly α .

- $\tilde{f}(k)$ es su representación en el espacio de frecuencias.
- k es el número de onda que, en este caso, representa la escala espacial inversa en unidades de s/km .

La transformada de Fourier aplicada al bosque de Ly α permite analizar las fluctuaciones en la transmisión del flujo a lo largo de la línea de visión hacia el QSO. Estas fluctuaciones del flujo están relacionadas con la distribución de densidad y temperatura del gas del MIG, por lo que aplicando la transformada de Fourier a un espectro de absorción se obtiene la distribución de la señal en el espacio de frecuencias que, en este caso, contiene la información sobre las escalas espaciales que dominan la señal. Esta transformada será la herramienta utilizada para extraer información de las estructuras del MIG a partir de los datos obtenidos de las simulaciones hidrodinámicas, analizando la correlación entre las señales de cuásares cercanos, cuantificando su grado de similitud y la coherencia entre sus estructuras.

En el espacio de Fourier, la correlación cruzada entre dos señales $f_1(x)$ y $f_2(x)$ se traduce en el producto entre $\tilde{f}_1(k)$ y $\tilde{f}_2^*(k)$, donde el segundo término es el conjugado complejo de la transformada obtenida para $f_2(x)$. Esta operación implica una resta de fases, ya que los exponentes complejos de cada señal se combinan como $e^{i\phi_1(k)} \cdot e^{-i\phi_2(k)} = e^{i[\phi_1(k)-\phi_2(k)]}$. La diferencia de fases obtenida proporciona una medida directa del desfase relativo que existe entre las fluctuaciones del flujo transmitido a lo largo de cada línea de visión y, por tanto, de la similitud entre las estructuras que atraviesan ambas señales. En otras palabras, cuanto menor sea esta diferencia, mayor será la coherencia de fase y más similares son las estructuras atravesadas por ambas líneas de visión.

3.1.1.2. Función de Densidad de Probabilidad: PDF

La función de distribución de probabilidad, PDF, es una herramienta estadística que describe la probabilidad de que una variable tome un determinado valor dentro de un intervalo dado. Esta función proporciona una visión completa de la distribución de los valores posibles, permitiendo analizar tanto la tendencia central como la dispersión y la asimetría de los datos. En el contexto del análisis del flujo transmitido en el MIG, la PDF se emplea para estudiar la distribución estadística de las diferencias de fase entre las transformadas de Fourier del flujo

transmitido en las parejas cercanas de QSOs. Como se verá a continuación, la PDF permitirá determinar la coherencia espacial de las estructuras del MIG en distintas escalas, identificando en qué modos de Fourier (escalas espaciales) persiste esta coherencia o se pierde, y relacionando esta coherencia directamente con la escala física de las fluctuaciones del gas.

A diferencia del espectro de potencias, la PDF de las diferencias de fase captura la dispersión angular entre modos equivalentes en líneas de visión distintas. Al ser esta estadística especialmente sensible a la distribución y alineamiento tanto de estructuras filamentosas como vacíos a lo largo del eje transversal, las variaciones en la distribución de las diferencias de fases pueden revelar la influencia de procesos y parámetros como la presión del gas o la dispersión peculiar de velocidades.

Esta herramienta estadística, junto a la transformada de Fourier, permitirá estudiar con precisión las características del MIG en aquellas escalas en las que el estado térmico del gas se impone a las fluctuaciones de densidad.

3.1.2. Medida no degenerada de la pequeña escala del MIG

Como se ha introducido en los apartados anteriores, el método de caracterización de la pequeña escala del MIG empleado está basado en el análisis de la diferencia de fases entre las transformadas de Fourier del flujo transmitido por parejas de QSOs cercanos, siguiendo la metodología introducida por Rorai et al. (2013). Este enfoque ofrece una vía alternativa a la vez que complementaria al espectro de potencia tradicional, adquiriendo una sensibilidad particular a la coherencia espacial de las estructuras responsables de la absorción. Además, supone una solución a la problemática de la degeneración del espectro de potencia del flujo.

3.1.2.1. Degeneración en el espectro de potencia en 1D

El espectro de potencia longitudinal del flujo transmitido ha sido una de las herramientas más utilizadas para inferir el estado térmico del MIG a través de sus parámetros T_0 , γ y la escala de suavizado de Jeans λ_J (Viel et al., 2009). Sin embargo, este método sufre una degeneración entre estos tres parámetros como demuestran Rorai et al. (2013), donde la distinción entre modelos con diferencias notables entre estos tres parámetros resulta extremadamente difícil al ajustarse todos con mucha precisión a espectros observacionales.

Esta degeneración surge porque tanto el suavizado térmico, que actúa en 1D en la línea de visión, y el suavizado por presión, que actúa en 3D sobre las estructuras formadas, generan un efecto similar sobre el espectro: un recorte de potencia en los modos k mayores, es decir, una supresión de las pequeñas escalas.

El ensanchamiento térmico Doppler introduce un corte en el espectro longitudinal a un valor de $k_{th} \sim 0.09 \text{ s/km}$, mientras que la escala de Jeans, que determina la longitud sobre la que la presión del gas suaviza las fluctuaciones de densidad, induce un corte espectral a lo largo de la línea de visión $k_J \sim 0.09 \text{ s/km}$ (Rorai et al., 2013). Dado que ambos procesos producen cortes espectrales similares, distintas combinaciones de T_0 , γ y λ_J pueden producir el mismo espectro de potencias.

En consecuencia, cualquier intento de caracterizar el estado térmico del MIG a partir del espectro longitudinal sin considerar esta degeneración lleva implícito el riesgo de llegar a conclusiones erróneas.

3.1.2.2. Ventajas del análisis de fase para parejas de QSOs en el espacio de Fourier

El cambio de paradigma introducido por el estudio de la fase en el espacio de Fourier viene motivado por una serie de ventajas que suponen una mejora significativa de los resultados que se pueden obtener de estos análisis.

Desde el punto de vista instrumental, una de las razones más relevantes por las que se decide trabajar en el espacio de Fourier viene dada por la resolución de los espectrógrafos empleados en las observaciones: la gran mayoría de parejas de QSOs cercanos son demasiado tenues como para poder observarse a alta resolución espacial ($FWHM \simeq 5 \text{ km s}^{-1}$) y resolver el bosque de Ly α . Esto implica que es necesario tener en cuenta de forma explícita los efectos de la resolución espectral de los instrumentos, lo que puede resultar complejo para espectros obtenidos por rendija (Rorai et al., 2013). Sin embargo, se puede considerar con una muy buena aproximación que el suavizado provocado por un límite de la resolución espectral actúa simplemente como un filtro de paso bajo sobre el flujo transmitido. Esto supone que la forma del espectro de potencia permanece inalterada para aquellos modos de Fourier k que son

menores que el valor de corte k_{res} impuesto por la resolución del instrumento. De esta forma, al trabajar en el espacio de Fourier, se pueden eliminar los modos $k \gtrsim k_{res}$ sin incurrir en un modelado preciso de los efectos de la resolución instrumental. Además, en las escalas en las que se producen las fluctuaciones de densidad éstas son débilmente no lineales, por lo que se conservan algunas propiedades de los campos Gaussianos aleatorios, como la independencia estadística de los modos de Fourier, simplificando el análisis de los errores y la interpretación estadística.

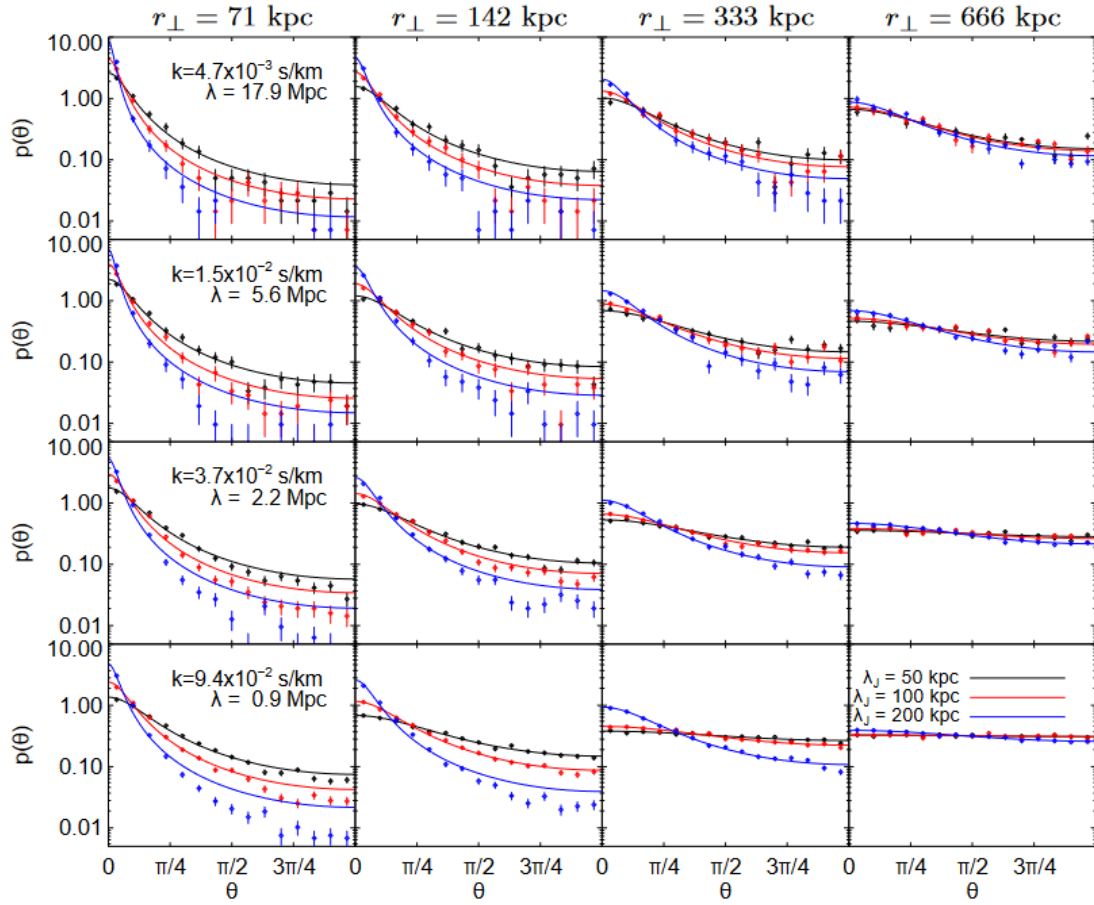
La ventaja más importante que presenta el análisis de la fase, sin embargo, radica en la rotura de la degeneración impuesta por el espectro de potencia longitudinal. La fase $\phi(k)$ de cada modo de Fourier proporciona información sobre la localización y la coherencia de las estructuras que modelan el flujo observado. Además, a diferencia del módulo de la transformada, su valor se ve significativamente menos afectado por errores de calibración, lo que constituye una ventaja significativa en el tratamiento de datos observacionales (Rorai et al., 2013).

El estudio de la fase es crucial ya que, dado que dos señales pueden tener el mismo espectro de potencia, pero fases distintas, el análisis exclusivo del módulo puede ignorar configuraciones espaciales específicas. En otras palabras: el módulo del espectro de Fourier otorga información sobre la amplitud de las fluctuaciones a distintas escalas, mientras que la fase contiene la información sobre la disposición espacial de dichas fluctuaciones. Por tanto, dos señales (flujos recibidos) pueden compartir exactamente el mismo espectro de potencia, pero diferir completamente en su distribución espacial si tienen fases distintas. Este hecho convierte al análisis de fase en una herramienta clave para poder acceder a configuraciones estructurales que el espectro de potencia por sí sólo no puede discernir.

El análisis de la fase permite ir más allá de la información contenida en una sola línea de visión al comparar cuán alineadas están estas estructuras entre dos líneas paralelas. Si las fases de un modo k son similares en ambas líneas, significa que la estructura que actúa sobre ese modo está presente en ambas trayectorias, lo que sugiere que el campo de absorción está transversalmente correlacionado. Además, mientras que esta coherencia es altamente

sensible a λ_J (al determinar la extensión tridimensional del gas), está exenta de los efectos provocados por el ensanchamiento térmico (que actúa localmente en la línea de visión), eliminando cualquier efecto que este tenga sobre la diferencia de fases.

Figura 6. PDF de la fase de los campos de densidad a distinta separación $r_\perp$, modo k y escala de Jeans λ_J



Fuente: Rorai et al., 2013

Estas características del análisis de la diferencia de fases entre distintos espectros de potencia se traducen en un comportamiento muy distinto en los estudios estadísticos derivados de él: tal y como demuestran Rorai et al. (2013), al construir la PDF de las diferencias de fase para distintos modos k se observa una fuerte dependencia con λ_J y una insensibilidad relativa a los

parámetros T_0 y γ (ver Figura 6), convirtiendo la estadística de fase en una medida no degenerada, capaz de extraer la escala de suavizado del MIG sin ambigüedad, incluso en estudios donde el espectro de potencias longitudinal no permite distinguir entre modelos térmicamente distintos. En esta Figura puede verse las PDF para la diferencia de fases $p(\theta)$ para distintas separaciones de distancia $r_\perp$, modo k y escala de Jeans λ_J , donde los puntos con barras de error representan la distribución de fases agrupadas del campo de densidad obtenidas de simulaciones hidrodinámicas y las líneas continuas corresponden al ajuste de máxima verosimilitud utilizando una distribución de Cauchy (Rorai et al., 2013).

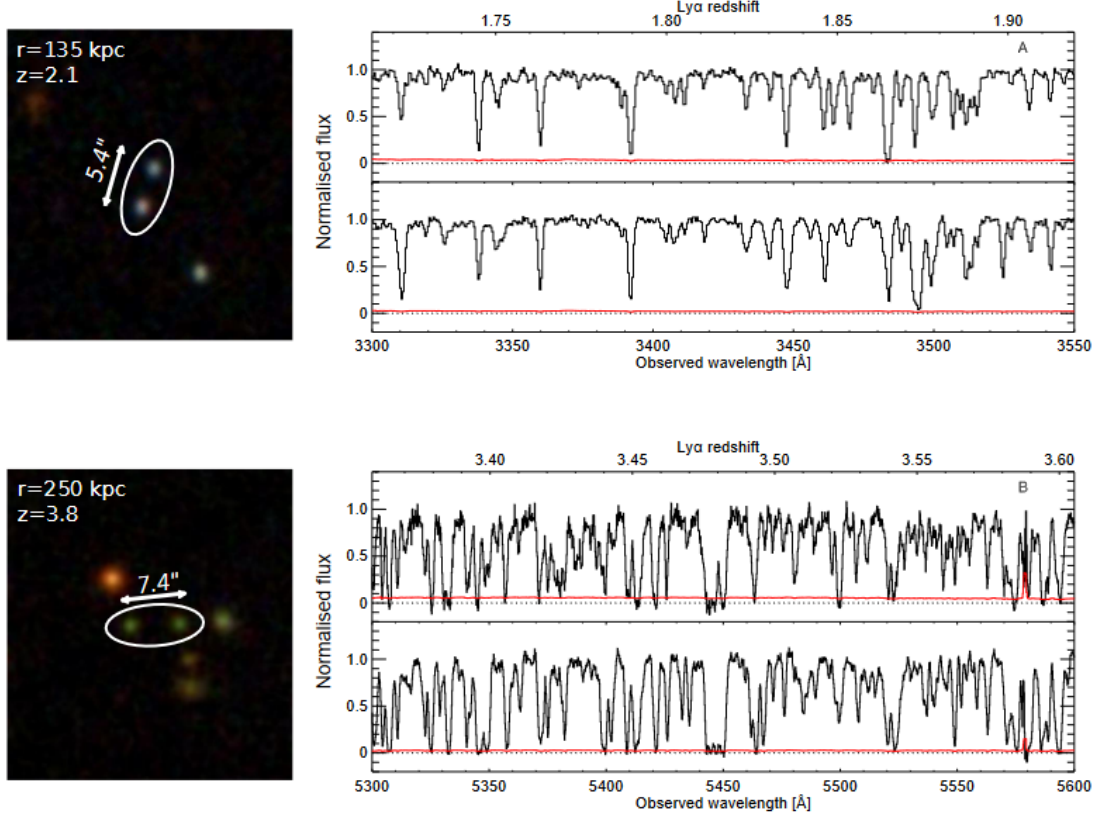
3.1.2.3. Obtención de la medida con parejas de QSOs de alta resolución

La combinación de los análisis de diferencia de fase y la PDF de éstas permitió a Rorai et al. (2017) obtener, a partir de parejas de QSOs cercanos de alta resolución, una medida de λ_J consistente con los modelos estándar de reionización del Universo por fotoionización ultravioleta.

Esta medida se obtuvo a partir del espectro de 25 parejas de QSOs con una separación transversal $r_\perp = 100 - 500 \text{ kpc}$ (Rorai et al., 2017). En la Figura 7 se muestra el flujo continuo normalizado del bosque de Ly α de dos parejas de QSOs empleadas en la obtención de esta medida, representado en función de la longitud de onda y el *redshift* correspondiente en el eje superior. Además, la línea de puntos señala el nivel cero de flujo y la línea continua roja la amplitud del nivel 1σ de ruido de cada píxel.

Como puede verse en la Figura 7, la resolución del bosque de Ly α para estos espectros es muy alta, permitiendo una clara distinción de las líneas de absorción en ambas parejas y garantizando unos resultados robustos. Estos resultados observacionales (ver Rorai et al., 2013), en contraste con las simulaciones hidrodinámicas, respaldan la elección del método basado en el análisis de la diferencia de fases para parejas de QSOs cercanos en los estudios realizados para la caracterización de la pequeña escala del MIG en este TFE.

Figura 7. Espectro de las parejas de QSOs SDSS J073522.43+295710.1, SDSS J073522.55+295705.0 (panel A) y SDSS J102116.98+111227.6, SDSS J102116.47+111227.8 (panel B)



Fuente: Rorai et al., 2017

3.2. Implementación y ejecución del método

3.2.1. Datos y entorno tecnológico

El primer paso para poder acometer los objetivos propuestos es disponer de una fuente de datos extensa, robusta y fiable, y un entorno tecnológico que permita su procesamiento y análisis de forma eficiente y reproducible.

3.2.1.1. Obtención de datos para el estudio del MIG: simulaciones NVB

Las simulaciones hidrodinámicas NVB, basadas en el código Nyx (Almgren et al., 2013), permiten modelar con alta precisión el comportamiento del medio intergaláctico en un

universo en expansión. Nyx combina la dinámica de partículas de la materia oscura con la hidrodinámica del gas bariónico, utilizando técnicas de refinamiento de malla adaptativo (AMR, *adaptive mesh refinement*) para mejorar la resolución en regiones de interés.

Estas simulaciones son especialmente útiles para el estudio del bosque de Ly α , ya que permiten reproducir con alta precisión la estadística del flujo de absorción observado en los espectros de cuásares lejanos. Gracias a la combinación de hidrodinámica con AMR y a la modelización precisa de la expansión cósmica, las simulaciones basadas en Nyx pueden capturar la evolución del gas bariónico a escalas cosmológicas. Ajustando parámetros como la resolución espacial, el tamaño de la caja de simulación y el modelo de ionización del medio intergaláctico, se logra una buena correspondencia con observables como la función de correlación del flujo transmitido, la distribución de anchura y profundidad de las líneas de absorción, y la dispersión entre líneas de visión cercanas. Esto las convierte en una herramienta de gran utilidad para interpretar la información obtenida del bosque Ly α y para comprender la distribución y dinámica del MIG durante los primeros miles de millones de años del Universo.

Las simulaciones hidrodinámicas usadas pertenecen al proyecto THERMAL (<http://thermal.joseonorbe.com/index.html>), y el autor de éstas es el también codirector de este TFE, José Oñorbe. El proyecto THERMAL consiste en un conjunto de más de 60 simulaciones hidrodinámicas con distintos modelos térmicos pero basadas en el mismo modelo cosmológico (consistente con las observaciones más recientes del CMB), y tiene como objetivo el estudio del estado térmico del MIG.

Para este TFE se han utilizado tres simulaciones concretas con las siguientes características:

- Modelo cosmológico: Λ CDM, consistente con observaciones (Planck Collaboration et al., 2020) con parámetros $\Omega_m = 0.3192$, $\Omega_L = 0.6808$, $\Omega_b = 0.04964$, $h = 0.6704$, $n_s = 0.96$, $\sigma_8 = 0.826$.
- Longitud de la caja de simulación: $20 \text{ Mpc}/h$.
- Resolución: 1024 celdas.
- *Redshift*: 2.4.

Para más detalle sobre los parámetros cosmológicos, véase Weinberg (1972). Cada una de las tres simulaciones comparte estas características y difieren únicamente en el modelo térmico empleado, donde tanto la temperatura T_0 como la escala de Jeans λ_J son distintas en cada caso. El motivo por el que se decide trabajar con simulaciones a $z = 2.4$ es que, en la práctica, la mayor densidad de QSOs observables se encuentra a este *redshift*, por lo que facilita la aplicación de los métodos implementados a los catálogos que están actualmente en desarrollo.

Adicionalmente, se utiliza un archivo HDF5 que contiene 50000 muestras de líneas de visión de la caja de simulación que sirven, entre otros objetivos, para caracterizar la relación entre los distintos parámetros del MIG (ver §1.1.4).

Todas las simulaciones se presentan en formato HDF5 (*Hierarchical Data Format 5*), que es un formato de archivo diseñado para manejar grandes volúmenes de datos y que es ampliamente usado en contextos de ciencia e ingeniería en el trabajo con estructuras de datos complejas. Cada archivo de simulación contiene un catálogo de parejas de QSOs que a su vez están organizadas según las distancias que las separan.

El uso de estas simulaciones permite aplicar análisis estadísticos a distintas escalas y con distintas muestras de parejas garantizando, además, la aleatoriedad de cada análisis, la homogeneidad de los datos y su correspondencia con las observaciones y modelos más recientes del Universo.

3.2.1.2. Modelos térmicos empleados

Las características de los tres modelos térmicos empleados en las simulaciones se resumen en la siguiente Tabla 1, donde se especifican el valor de T_0 , γ y λ_J :

Tabla 1. *Modelos térmicos de las simulaciones hidrodinámicas*

Modelo	T_0	γ	λ_J
T000	$1.024 \cdot 10^4 K$	1.592	75.6937 kpc
T002	$1.994 \cdot 10^4 K$	1.582	105.5019 kpc
T003	$6.727 \cdot 10^3 K$	1.604	61.6421 kpc

Fuente: elaboración propia

En adelante, se hará referencia a estos modelos térmicos por el nombre indicado en la columna “Modelo”.

3.2.1.3. Entorno tecnológico

Los flujos de trabajo automatizados (o *pipelines*) que extraen, procesan y analizan los datos de los archivos HDF5 están implementados en Python, en su versión 3.11. Los motivos por los que se han elegido esta tecnología son:

- Sintaxis clara: facilita el seguimiento y entendimiento del código.
- Facilidad de uso: es un lenguaje de programación que ofrece un aprendizaje rápido, permitiendo abordar desde el primer momento el tratamiento de datos y dejando a un lado la complejidad de la programación.
- Escalabilidad y eficiencia: Python permite manejar grandes volúmenes de datos e implementar automatismos que se adapten a variaciones en el tamaño de las estructuras de los datos tratados, garantizando la robustez y reusabilidad del código.
- Catálogo de librerías: la amplia gama de librerías disponibles en Python facilita enormemente la implementación de los flujos de trabajo ofreciendo, por ejemplo, librerías para el tratamiento de archivos HDF5, librerías científicas para los análisis de Fourier y estadísticos, o librerías que permitan una representación gráfica de los datos clara y concisa.

Por otra parte, como entorno de desarrollo integrado (IDE) se ha usado Spyder 6, ya que permite desarrollar y ejecutar código, inspeccionar las variables generadas, ejecutar comandos en la consola Python y visualizar las gráficas generadas, todo ello desde una misma interfaz. Estas características facilitan las labores de implementación y depuración de código y, además, permite la personalización y configuración del *kernel* usado, lo que facilita el cambio entre espacios de trabajo de forma ágil.

La combinación de Python 3.11 y Spyder 6 permiten disponer del entorno tecnológico apropiado para implementar los flujos de trabajo que llevan a la consecución de los objetivos planteados.

3.2.2. Implementación en Python

La implementación del método propuesto por Rorai et al. (2013) se lleva a cabo en distintas fases. En esta sección se abordarán las distintas etapas de la implementación en subapartados, detallando los procesos automatizados y su correspondencia con la teoría explicada en este mismo Capítulo y los anteriores.

3.2.2.1. Extracción de datos de archivos de simulación hidrodinámica en formato HDF5

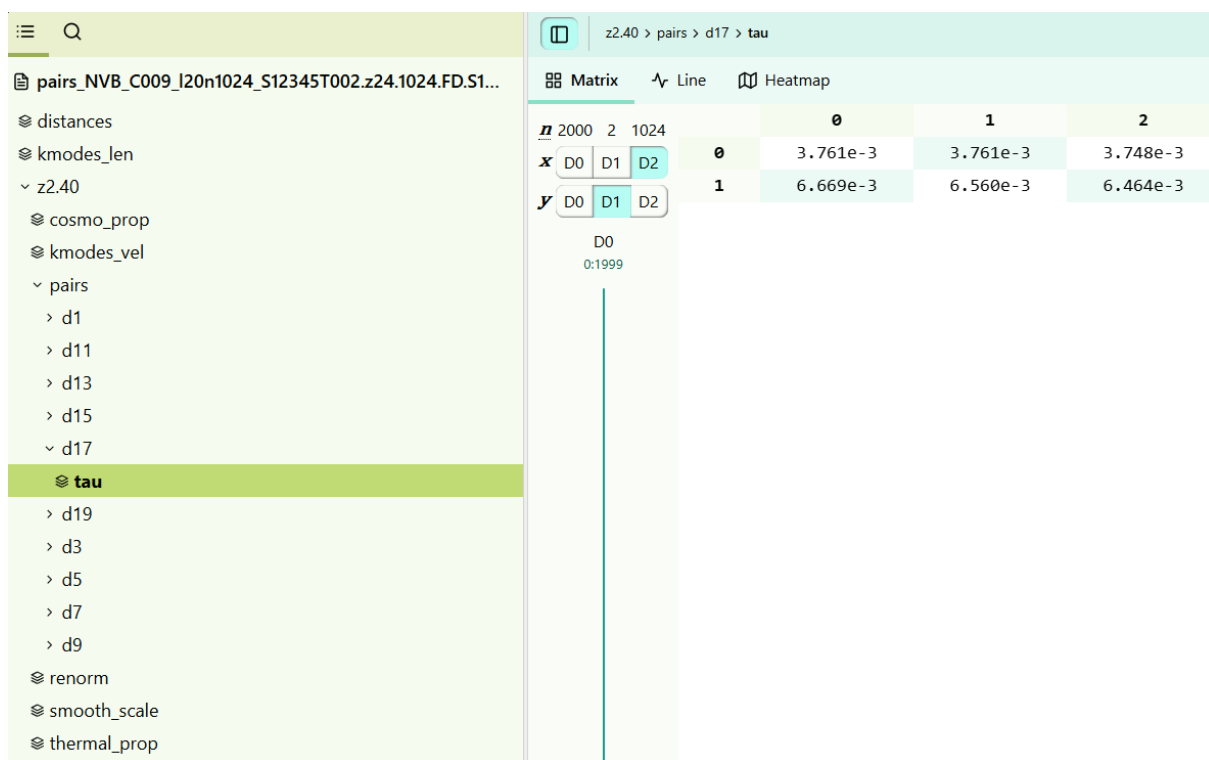
La extracción de datos de archivos HDF5 ha de hacerse con precisión y cuidado ya que su arquitectura jerárquica obliga a trabajar con estructuras de datos que mantengan la coherencia entre estos y faciliten su uso.

Cada archivo de parejas contiene la siguiente estructura, que puede verse en la Figura 8:

- `distances`: definición de distancias de separación entre QSOs.
- `kmodes_vel`: modos k en el espacio de Fourier en Mpc^{-1} .
- `z2.40`: estructura padre de toda la información relativa al estado térmico del gas en el MIG, las parejas formadas y las propiedades del modelo cosmológico:
 - `cosmo_prop`: parámetros del modelo cosmológico empleado en la simulación.
 - `kmodes_vel`: modos k en el espacio de Fourier en s/km .

- pairs: conjunto de 2000 parejas, agrupadas por distancias, donde cada pareja contiene la profundidad óptica de cada QSO muestreada en las 1024 celdas de resolución de la simulación.
- renorm: valor del factor de renormalización del flujo utilizado en la simulación.
- smooth_scale: valor de la escala de λ_f en kpc .
- thermal_prop: propiedades térmicas del gas del MIG en la simulación.

Figura 8. Estructura interna de los archivos de parejas de QSOs de simulación hidrodinámica en formato HDF5



Fuente: elaboración propia

En cuanto al archivo de 50000 muestras de líneas de visión del MIG, la estructura presentada tiene un único bloque principal donde se anidan bajo el mismo nodo todas las variables que contiene, destacando entre ellas la profundidad óptica τ , la variación de densidad sobre la densidad media Δ_b y la temperatura del gas T en cada punto muestreado de la línea de visión.

Para poder navegar por esta estructura jerárquica y extraer los datos necesarios se utiliza la librería de Python *h5py*, que permite trabajar con archivos HDF5 de forma rápida y sencilla accediendo mediante pares clave-valor a los distintos datos que contiene.

3.2.2.1. Preparación de las estructuras de datos

Una vez se han extraído los datos de los archivos HDF5 es necesario un procesamiento tanto estructural como matemático con el objetivo de disponer de estructuras reusables y de fácil acceso, y datos que coincidan en unidades y escalas con la realidad observable.

Las tareas de reestructuración de los datos, aunque breves y sencillas, son relevantes ya que garantizan la coherencia de los valores extraídos para cada parámetro. Esto ocurre porque, por ejemplo, el acceso a los datos de profundidad óptica de los archivos de 50000 líneas de visión devuelve una única matriz de dimensiones $[1, 5120000]$, correspondientes a 50000 líneas de visión con una resolución de 1024 celdas. Esto obliga a hacer una reestructuración de la matriz para que tenga dimensiones $[50000, 1024]$, de forma que cada fila de la matriz contenga los datos de una sola línea de visión, permitiendo elegir entre el trabajo con líneas de visión específicas, agrupaciones de éstas o con todas las líneas de visión disponibles, garantizando la coherencia de los datos y de los futuros análisis estadísticos. Esta misma reestructuración necesita ser aplicada para los demás parámetros del MIG, como la temperatura, la variación de densidad o la densidad de la columna de hidrógeno para poder representar con rigor los parámetros del MIG como se muestra en la Figura 5.

Por otro lado, los archivos que contienen las parejas de QSOs ya incluyen la estructura adecuada para el análisis estadístico, ya que cada conjunto de distancias (*distances*) se organiza como una matriz de dimensiones $[2000, 2, 1024]$, donde cada una de las 2000 entradas del primer índice representa una pareja de la línea de visión simulada. Cada pareja contiene dos vectores $[1, 1024]$, que corresponde a los valores de la profundidad óptica τ de cada uno de los QSO que componen la pareja. Esta estructura permite acceder fácilmente a los perfiles de absorción de cada pareja, asegurando tanto el aprovechamiento del código con estructuras análogas como de los propios datos usados.

3.2.2.2. Ajustes teóricos sobre los datos

El ajuste estructural no es el único que debe hacerse antes de abordar de lleno los análisis estadísticos de la fase en el dominio de Fourier: también es necesario aplicar ciertas correcciones matemáticas para conservar el rigor teórico entre los análisis realizados y las futuras observaciones.

3.2.2.2.1. Obtención de los ejes de representación

En primer lugar, es necesario obtener los datos de velocidad como representación de la distancia, junto con la distancia real de la caja de simulación reescalada según el parámetro de Hubble normalizado (ver §1.4.1.2), de forma que puedan representarse los datos de forma agnóstica al modelo cosmológico. Esto es sencillo de conseguir implementando los ajustes cosmológicos vistos en §1.4.1. Una vez se han obtenido estas distancias, y considerando el número de celdas de la caja, es sencillo obtener unos ejes de representación consistentes con las observaciones, sentando las bases para representar los parámetros del MIG como se muestran en la Figura 5.

3.2.2.2.2. Renormalización del flujo transmitido

Los cálculos realizados con el flujo normalizado que conforma el bosque de Ly α requiere ajustes más precisos y delicados. Este flujo normalizado viene dado por la siguiente expresión (Lukić et al., 2015):

$$F = e^{-\tau}, \quad (7)$$

donde τ es la profundidad óptica, obtenida de los archivos HDF5, por lo que la obtención del flujo normalizado que conforma el bosque de Ly α es directa.

Sin embargo, el análisis estadístico de la diferencia de fases entre pares de líneas de visión corre el riesgo de perder validez si se realiza directamente sobre los flujos simulados sin una normalización previa. Esto es debido a que cada archivo de simulación corresponde a un modelo térmico distinto, lo que afecta directamente a los parámetros que modelan la profundidad óptica. Como consecuencia, el flujo medio transmitido $\langle F \rangle = \langle e^{-\tau} \rangle$ difiere entre los modelos. Esta diferencia introduce un sesgo en la comparación de las fases ya que la relación entre la fase y la estructura espacial está modulada por el contraste del flujo. Es decir,

cambios abruptos en el flujo implican estructuras con bordes bien definidos, mientras que cambios suaves en el flujo implican estructuras más suaves y menos definidas, lo que afecta directamente a la fase de las señales. Por tanto, para que el análisis sea rigurosamente comparable desde el punto de vista estadístico y teórico, es necesario aplicar una normalización que iguale los flujos medios entre simulaciones. Esta normalización se hace a partir de la profundidad óptica de H_I en el bosque de $\text{Ly}\alpha$, que viene definida por la Ecuación 8:

$$\tau = A(\rho_b/\overline{\rho_b})^\beta, \quad (8)$$

donde la constante A se calcula comparando con observaciones y $\beta = 2 - 0.7(\gamma - 1)$ describe la dependencia de la profundidad óptica respecto de la densidad local (Lukić et al., 2015).

Por tanto, para poder comparar de forma coherente los distintos modelos, es imprescindible determinar previamente el valor de la constante A que normaliza el flujo transmitido. Esta normalización requiere seleccionar un flujo observacional de referencia, que funcione como punto de anclaje para los cálculos y garantice que las comparaciones entre las fases de los flujos obtenidos de distintos modelos sean consistentes y significativas.

El flujo normalizado medio de referencia elegido viene dado por las observaciones y estudios posteriores llevados a cabo por Becker et al. (2013), que proporcionan el flujo medio observacional en función del *redshift*, y cuyos cálculos se encuentran implementados en métodos de Python. Para el caso que ocupa a este TFE, el flujo medio normalizado en $z = 2.4$ corresponde a $\langle F \rangle = 0.81360$. Cabe destacar que los ficheros de simulación hidrodinámica proporcionan sus propios flujos de referencia y factores de normalización. No obstante, se opta por adoptar el flujo de referencia de Becker et al. (2013), debido a que está sólidamente contrastado mediante observaciones y resulta una base utilizada en la literatura reciente (ver Khan et al., 2024).

De esta forma, ya es posible calcular el factor de normalización A . Para los cálculos de A se adoptan dos métodos numéricos distintos: el método de la bisección y el método de Brent

(Press et al., 2007). El método de la bisección se implementa en Python manualmente, haciendo uso de bucles recursivos. Por otra parte, el método de Brent es accesible mediante la librería *scipy*, lo que simplifica y facilita su uso.

Para ambos métodos, la expresión que se pretende resolver viene dada por la Ecuación 9:

$$\frac{\sum_{i=1}^N e^{-A \cdot \tau_i}}{N - 1} - \langle F_{Becker} \rangle = 0, \quad (9)$$

donde el minuendo representa el flujo medio para la línea de visión de una matriz de profundidad óptica de dimensiones $1 \times N$, y el sustraendo es el flujo de referencia dado por Becker et al. (2013).

Los resultados obtenidos por cada método para los distintos modelos térmicos pueden consultarse en las Tablas 2 y 3:

Tabla 2. Factores de normalización para los tres modelos térmicos obtenidos con los métodos de bisección y Brent

Modelo térmico	Factor de normalización A con método de Bisección	Factor de normalización A con método de Brent
T_{000}	$A_{T_{000}} = 0.73608$	$A_{T_{000}} = 0.73611$
T_{002}	$A_{T_{002}} = 1.11129$	$A_{T_{002}} = 1.11133$
T_{003}	$A_{T_{003}} = 0.57693$	$A_{T_{003}} = 0.57688$

Fuente: elaboración propia

Como puede comprobarse, las diferencias en los factores de normalización son prácticamente despreciables en función del método de cálculo empleado (siendo todas inferiores al 0.01%), mientras que la diferencia entre los distintos modelos térmicos es muy significativa.

Adicionalmente, se calculan los factores A con el flujo de referencia intrínseco de las simulaciones, obteniendo un factor A igual al que proporcionan los archivos de simulación hidrodinámica, lo que corrobora el planteamiento teórico y garantiza la coherencia entre los posteriores análisis.

La consecución de esta meta permite, ahora sí, calcular un flujo normalizado a partir de la profundidad óptica, tal y como indica la Ecuación 10, y que es consistente entre distintos modelos y apto para su comparación en el espacio de Fourier:

$$F = e^{-A \cdot \tau} \quad (10)$$

3.2.2.2.3. Adición de ruido a las señales de flujo

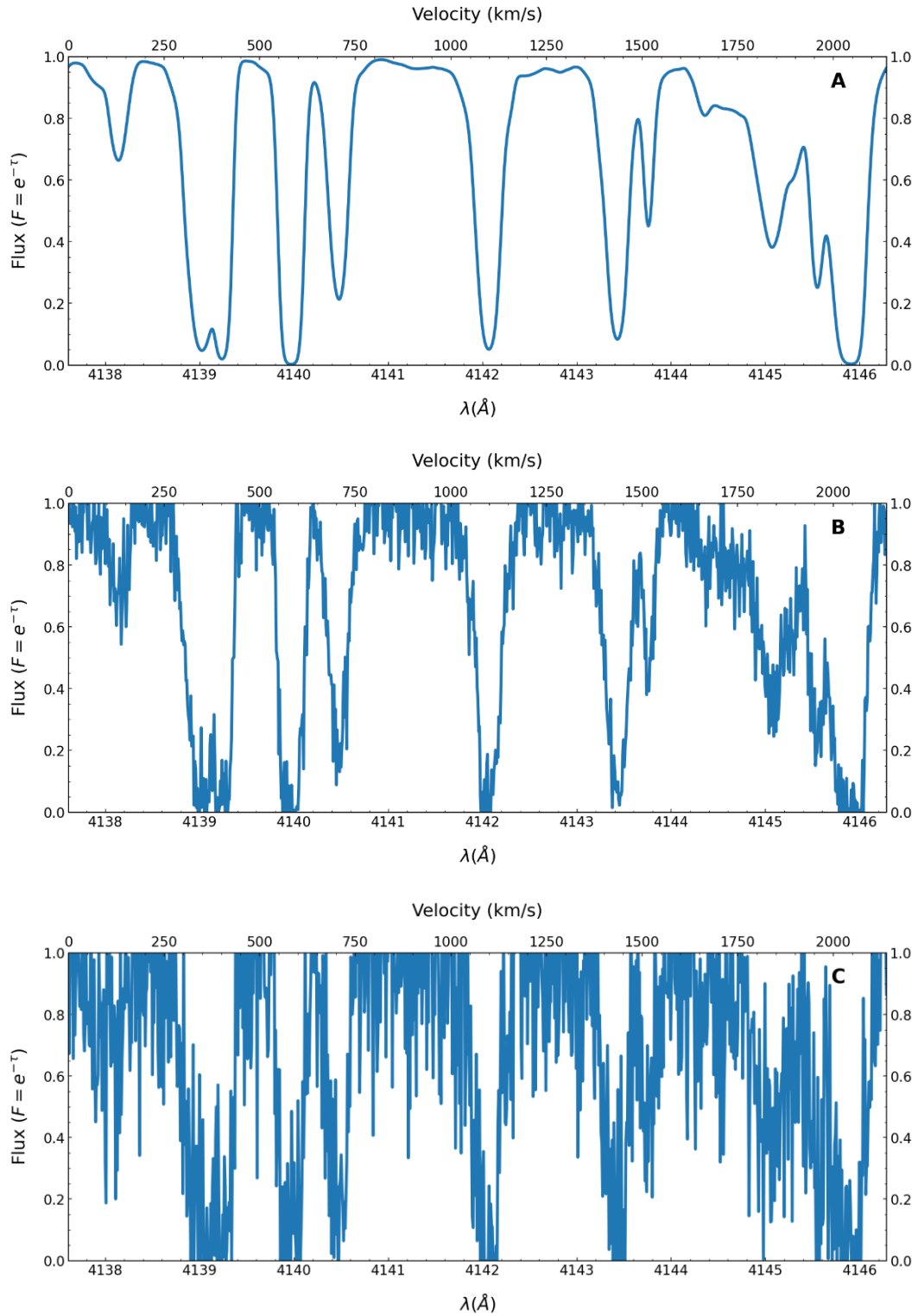
El flujo obtenido a partir de la Ecuación 10 dista de ser lo esperable de una observación en el contexto de los catálogos que están en desarrollo, ya que son señales ideales que carecen del ruido presente en cualquier observación. Por tanto, el siguiente paso en el ajuste de los datos es añadir ruido a las señales para que los posteriores análisis estadísticos se ajusten de la forma más precisa posible a la realidad.

El proceso de adición de ruido consiste en la suma de dos señales: la señal original del flujo y una señal de ruido blanco Gaussiano generada aleatoriamente. En este cálculo debe tenerse en cuenta la SNR de la señal. Estando ésta dada por la Ecuación 11, la SNR relaciona la potencia de la señal $P_{señal}$ y la potencia del ruido P_{ruido} :

$$SNR = \frac{P_{señal}}{P_{ruido}} \quad (11)$$

Así, la generación del ruido parte de una distribución normal estándar $N \sim \mathcal{N}(0,1)$ y se divide ésta por la SNR deseada para la simulación. Al dividir el ruido generado por la SNR, se ajusta la desviación estándar del ruido resultante a un factor $1/SNR$, modulando así su intensidad: cuanto mayor SNR, menor será el ruido añadido y mayor fidelidad de la señal simulada. Esta normalización es crucial para replicar condiciones observacionales reales y evaluar el impacto del ruido en los análisis posteriores.

Figura 9. Bosque de Ly α a partir del flujo normalizado ideal (Panel A) y bosque de Ly α a partir del flujo normalizado con ruido para una SNR=10 (Panel B) y SNR=4 (Panel C)



Fuente: elaboración propia

La diferencia entre la señal ideal obtenida a partir de la Ecuación 10 y la señal con el ruido añadido según una SNR dada puede verse en la Figura 9, donde la degradación de la señal complica notablemente la identificación de los picos de absorción presentes en el bosque de Ly α , además de modificar considerablemente la magnitud del flujo observado.

Esto finaliza los ajustes matemáticos necesarios en los datos para su aproximación máxima a las observaciones y al marco teórico en el que se realizan los estudios de este TFE, y permite avanzar a una fase siguiente de la implementación donde se aborden los ajustes entre los parámetros observacionales y los parámetros de simulación.

3.2.2.3. Ajustes entre observación y simulación

Los parámetros empleados en los ajustes de la simulación hidrodinámica, como la relación señal a ruido de la simulación SNR_{sim} y el número de parejas aleatorias generadas, necesitan considerar las diferencias intrínsecas entre la propia simulación y las observaciones reales que conforman los catálogos de resolución media y baja.

3.2.2.3.1. Ajuste de la SNR

Con el objetivo de asegurar que la relación señal a ruido de las señales analizadas sea coherente con la de las observaciones, se aplica una corrección sobre la SNR_{sim} en función de la SNR de la observación SNR_{obs} y las resoluciones de píxel de simulación y observación.

La señal total recibida se puede definir como $n \cdot S$, donde n es el número de píxeles y S es el valor medio de la señal por píxel. Por otra parte, el ruido total acumulado en la observación viene dado por $\sqrt{n} \cdot \sigma$. Esto permite definir SNR_{sim} tal que:

$$SNR_{sim} = \frac{n \cdot S}{\sqrt{n} \cdot \sigma} = \sqrt{n} \cdot \frac{S}{\sigma} = \sqrt{n} \cdot SNR_{obs}, \quad (12)$$

donde $n = \Delta v_{sim} / \Delta v_{obs}$, si se considera que la relación entre la distancia cubierta por un píxel de simulación Δv_{sim} y la cubierta por un píxel de observación es Δv_{obs} . Es decir, n es el número de píxeles originales promediados dentro de un píxel de simulación. De esta forma, la SNR_{sim} queda como:

$$SNR_{sim} = \sqrt{\frac{\Delta v_{sim}}{\Delta v_{obs}}} \cdot SNR_{obs} \quad (13)$$

Esto permite que, implementando un método en Python que haga este ajuste, se pueda trabajar con una SNR_{sim} que garantice la coherencia con las observaciones en los análisis estadísticos de la fase realizados.

Otra ajuste posible y más riguroso es la degradación de la propia simulación, obteniendo resoluciones propias similares a las resoluciones de las observaciones. Sin embargo, este método no sólo tiene una mayor complejidad computacional y teórica, sino que requiere la ejecución de una simulación hidrodinámica por cada valor concreto de resolución que se quiera obtener. Por estas razones, se adopta el ajuste explicado en este apartado, ya que ofrece una mayor flexibilidad y reutilización.

3.2.2.3.2. Cálculo y generación del número de parejas necesarias

Un aspecto de los parámetros de simulación que es imprescindible considerar es que los datos obtenidos de las simulaciones hidrodinámicas cubren una distancia determinada que viene dada por el tamaño de la caja, en este caso, $L = 20 \text{ Mpc}$. Por tanto, si se quiere trabajar con estos datos y compararlos con observaciones reales llevadas a cabo en un intervalo de z dado, es necesario reescalar el número de parejas simuladas para garantizar que la estadística aplicada una pareja de simulación sea también coherente con la aplicada a una pareja de observación.

En este caso, el ajuste entre simulación y observación es significativamente más sencillo ya que consiste simplemente en una división de la distancia comóvil de ese intervalo de z por la distancia cubierta por la caja. Este cociente proporciona el número de veces que es necesario replicar una pareja de simulación para abarcar las distancias reales de la observación. Para los estudios realizados en este TFE se asume que los QSOs observados se encuentran entre $z_{ini} = 2.2$ y $z_{end} = 2.6$, por lo que el número de veces que es necesario replicar una pareja de simulación para cubrir la distancia real de una pareja de observación se calcula con:

$$n_{\text{parejas-sim}} = \frac{dC_{z_{ini}, z_{fin}}}{L}, \quad (14)$$

donde $dC_{z_{ini}, z_{fin}}$ es la distancia comóvil para el modelo cosmológico empleado en el intervalo de z muestreado $[z_{ini}, z_{fin}]$, y L es el tamaño de la caja de simulación, ambos en Mpc .

Con este dato, por cada iteración se puede generar entonces un total de parejas de simulación que viene dado por $N_{total} = n_{\text{parejas-sim}} \cdot N_{obs}$, donde N_{obs} es el número de parejas de observación con el que se pretende trabajar. La implementación de este cálculo en el proceso automatizado de trabajo principal asegura que, independientemente de L , intervalo z y N_{obs} se obtenga en cada iteración el número necesario de parejas de simulación que garantice la coherencia entre simulaciones y observaciones.

Por último, para la generación de una muestra de parejas aleatorias con las que realizar la estadística de la diferencia de fase, debe considerarse la distribución real de parejas de QSOs en función de SNR_{sim} , *seeing* de las observaciones, distancia transversal que separa los QSOs y brillo superficial SB.

Los integrantes del proyecto WEAVE (Pieri et al., 2016) han pronosticado el número de parejas que cabe esperar en las observaciones en función de estos parámetros. Estos datos predictivos se muestran en la Tabla 3. Esta tabla muestra, para cada agrupación de distancias reales, la fracción de parejas que cabe esperar.

De esta manera, considerando estas distribuciones de parejas en función de la distancia que las separa, es posible generar una muestra aleatoria de parejas a partir de los datos de simulación que sea consistente con la distribución de parejas esperable cuando los catálogos se publiquen al término de sus observaciones.

Tabla 3. *Distribución de parejas de QSOs esperadas para $z=2.4$, $SNR \geq 3$, $SB=21$ y $seeing=2.6$ por bloque de observación*

Distancia transversal de separación real	Proporción de parejas de QSOs esperada
$0 < d < 100 \text{ kpc}$	5/23
$100 < d < 200 \text{ kpc}$	11/23
$200 < d < 300 \text{ kpc}$	7/23

Fuente: comunicación privada WEAVE-QSO Project

El algoritmo de análisis empleado comienza aplicando los ajustes de los parámetros de observación, como SNR_{sim} y N_{total} , y continúa generando, por cada iteración, muestras aleatorias de parejas a partir de los archivos de simulación hidrodinámica, haciendo uso de la distribución de la Tabla 3 para garantizar la máxima aleatoriedad de los datos y obtener una estadística robusta, fiable y acorde a las condiciones reales de observación.

3.2.2.4. Aplicación de la transformada de Fourier al flujo en el bosque de Ly α y generación de la PDF

El proceso de aplicación de la transformada de Fourier se integra en el algoritmo principal tras la generación de parejas aleatorias, una vez que se han preparado los datos y se han hecho los ajustes pertinentes. En esta sección se detallará el proceso de aplicación de la transformada de Fourier y los pasos necesarios para la generación de la PDF.

En primer lugar, es importante comprender cómo funciona el bucle que calcula los datos de la transformada de Fourier y de la PDF. Desde una visión de muy alto nivel, el algoritmo, con cada iteración, genera una muestra aleatoria de parejas, aplica la transformada de Fourier, obtiene las diferencias de fase, genera un histograma con los datos obtenidos y guarda los resultados en una estructura final, que es devuelta al terminar el bucle con los resultados

acumulados de todas las parejas generadas. Esto facilita el cálculo de las medias y desviaciones típicas de cada una de las agrupaciones de distancias y k .

El ángulo diferencia de fases de cada uno de los flujos transmitidos se calcula según la Ecuación 15:

$$\theta_{12}(k) = \arccos \left(\frac{\mathcal{R}[\delta \widetilde{F}_1^*(k) \delta \widetilde{F}_2(k)]}{\sqrt{|\delta \widetilde{F}_1(k)|^2 |\delta \widetilde{F}_2(k)|^2}} \right), \quad (15)$$

donde el numerador es el producto escalar de los flujos de cada uno de los QSOs de la pareja en el espacio de Fourier, y el denominador es el factor que normaliza cada componente obtenido del módulo de ambas transformadas (Rorai et al., 2013; Rorai et al., 2017). Así, puede calcularse con cada iteración la diferencia de fase de las señales $\theta_{12}(k)$ que permite posteriormente generar la PDF.

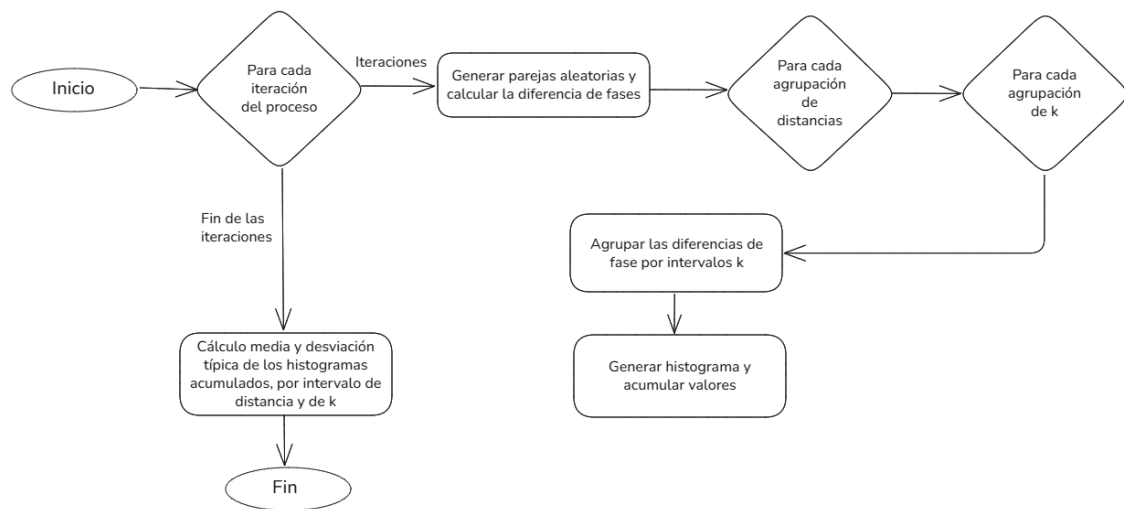
En la Figura 10 puede verse el diagrama de flujo simplificado correspondiente al proceso de generación de estos datos. Se ha optado por presentar un diagrama de flujo simplificado, sin indicar los flujos explícitos en los bucles anidados ya que el objetivo de este diagrama es comprender el proceso de generación de la PDF, y no profundizar en los subprocesos internos del mismo flujo. Por esta misma razón, se suprimen las entradas de los métodos y el uso de variables explícitas. El número de iteraciones máximas realizadas para este algoritmo se fija en 50.

Hay un paso concreto del proceso que merece la pena destacar, y es el paso de la agrupación de diferencias de fase en los intervalos de los modos k .

Para poder determinar con precisión la distribución de fases y generar la PDF, es imprescindible identificar qué fases corresponden a cada uno de los intervalos de k definidos. Esta asignación se lleva a cabo con el uso del método *digitize*, perteneciente a la librería *Numpy*. Este método compara cada valor de k con los bordes definidos en los intervalos k y devuelve el índice del intervalo- k correspondiente. Por ejemplo, si un valor de k se encuentra entre el segundo y el tercer borde, se le asigna el segundo intervalo. Esta indexación permite

agrupar las fases asociadas a cada valor de k según el intervalo k que les corresponde y calcular histogramas específicos para cada rango de k , lo cual es esencial para analizar la evolución de la distribución de fases en función de la escala física.

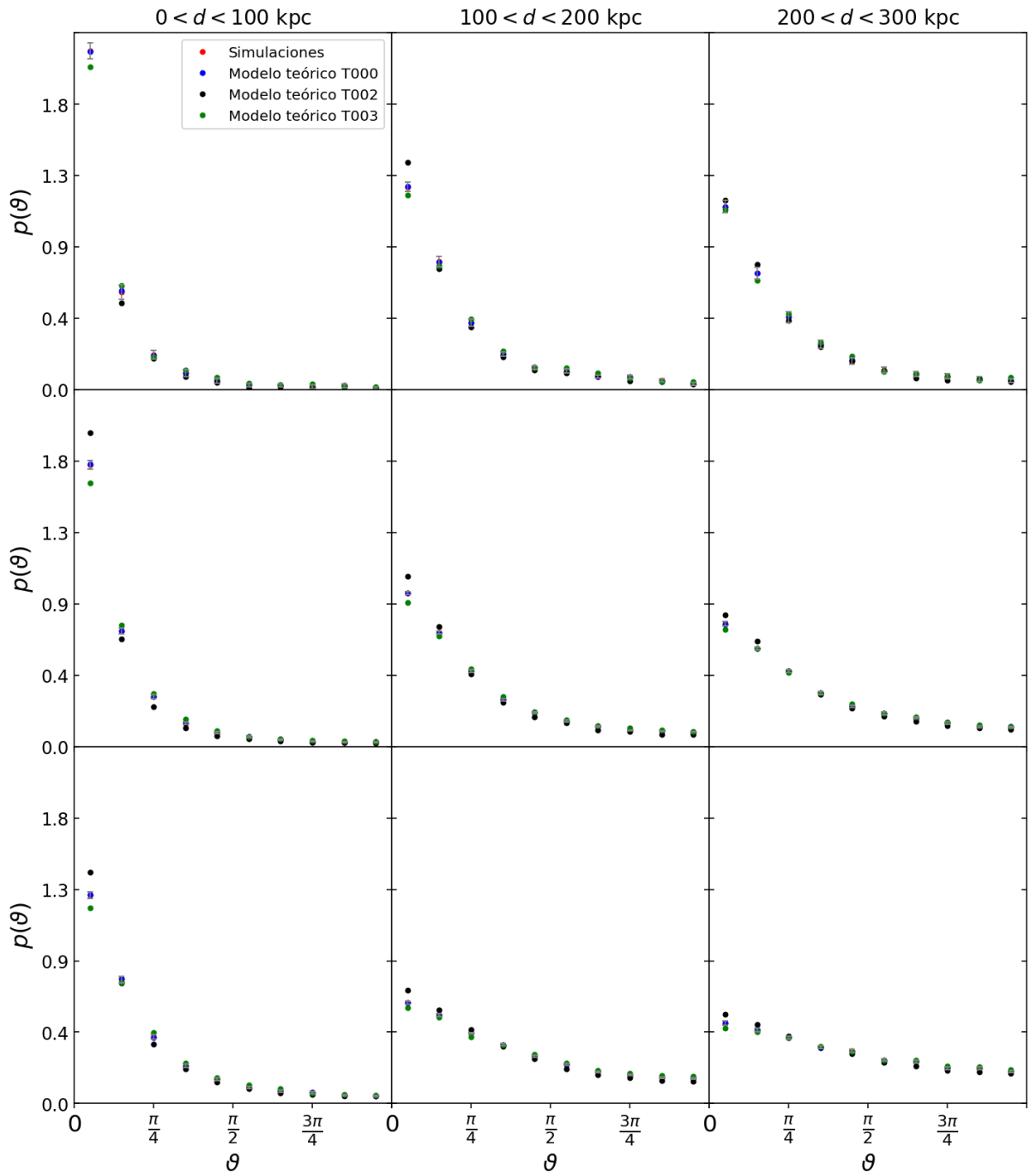
Figura 10. Diagrama de flujo simplificado del proceso de aplicación de la transformada de Fourier y generación de la PDF



Fuente: elaboración propia

La agrupación de los datos por intervalos de distancia transversal y en intervalos de modos k permiten construir una serie de gráficas conjuntas (ver Figura 11) donde, para cada columna, se muestran los datos de un mismo intervalo de distancia y, para cada fila, los datos de un mismo intervalo de modos de Fourier. Además, en cada gráfica se puede ver el ajuste de la simulación con las correspondientes barras de error y los tres modelos térmicos teóricos, que sirven para contrastar los resultados de la simulación y poder hacer un ajuste fino de los parámetros de ésta.

Figura 11. PDF de la diferencia de fase para los archivos de simulación con modelo térmico T000 y los modelos teóricos T000, T002 y T003 agrupadas por distancia transversal y modos k



Fuente: elaboración propia

Este proceso finalizaría la implementación del método de caracterización de la pequeña escala del MIG en Python. Como se ha visto a lo largo del capítulo, los ajustes teóricos y estructurales aplicados a los datos, así como una debida parametrización de los automatismos empleados, no solo garantizan el rigor científico de los resultados obtenidos, sino que, además, facilitan su reproducción y adaptación a distintos modelos, simulaciones y parámetros observacionales.

4. Discusión y resultados

En este capítulo se presentan los resultados obtenidos a partir del análisis de la estructura a pequeña escala del MIG, con especial atención a la influencia de la resolución espectral y la SNR de los flujos procesados en la caracterización de cada uno de los modelos analizados. Los datos han sido procesados y analizados siguiendo la metodología descrita en el Capítulo 3, lo que permite una comparación directa y coherente entre los distintos escenarios explorados.

4.1. Introducción a los resultados

Con el objetivo de determinar el impacto que tienen la resolución y la SNR en la diferenciación de los distintos modelos térmicos del MIG, se han considerado tres valores representativos de SNR de observación (6, 4 y 2), así como tres resoluciones espaciales distintas: 30, 60 y 100 km/s . Esta selección permite cubrir un rango de calidad de datos que abarca desde condiciones observacionales subóptimas hasta condiciones ligeramente superiores a las que cabe esperar con la instrumentación y tecnología actuales en condiciones normales, pero cuyo estudio permitirán establecer unos criterios de resolución tanto para los catálogos que llegarán en el futuro cercano como para aquellos que puedan aportar una calidad superior en el futuro lejano.

Los criterios de resolución instrumental se han tomado a partir de los modos de resolución de WEAVE (Jin et al., 2024), que dispone de un modo de alta resolución a 30 km/s y otro de media resolución de 60 km/s . Adicionalmente, se simula un tercer modo de 100 km/s para estudiar si es posible la caracterización de los modelos con una resolución baja. Aunque este TFE, como se ha señalado en capítulos anteriores, se centra en la medida de la pequeña escala del MIG a partir de parejas de cuásares de baja resolución, resulta igualmente interesante explorar el escenario de alta resolución utilizando la misma fuente de datos. Esta comparación permite ampliar el análisis y proporciona una visión más completa del impacto que tiene la variación de cada uno de los parámetros observacionales sobre los resultados obtenidos.

El impacto de la resolución y la SNR podrá analizarse a través del número de parejas necesarias para la caracterización de los modelos, y cuyo valor máximo se establece en 500, y en la

precisión de los resultados que se obtienen en las simulaciones. El procedimiento para determinar las condiciones de resolución consiste en fijar una resolución y una SRN y variar el número de parejas analizadas hasta poder resolver los modelos o, en caso de alcanzar el límite de parejas, determinar el mejor resultado esperable bajo esas condiciones.

Por otra parte, las parejas elegidas para todas las simulaciones y resultados se encuentran en un rango de distancias transversales comprendidas entre 100 y 200 $ckpc/h$, y los modos k de Fourier evaluados se varían en el intervalo comprendido entre 0.005 y 0.03 s/km . Este rango de operación se ha elegido después de diversas iteraciones con distintas distancias e intervalos de k , permitiendo elegir así unos valores tanto de distancias de separación como de k que sean consistentes con las escalas de Jeans λ_J de los modelos térmicos. Esto se debe a que el trabajo tanto con distancias muy superiores como inferiores a los valores de λ_J no permitiría determinar estadísticamente estas escalas en cada modelo a partir de las simulaciones, y éstas no aportarían valor estadístico ni interpretaciones físicas relevantes.

Además, para el total de parejas que conforman la muestra se especificará el número de bloques de observación (*observation block*, OB) necesarios para obtener el total de parejas que conforman la muestra, a una SNR y distancia transversal dadas. Un OB es una unidad estructurada de tiempo dentro de un programa astronómico que define una secuencia específica de observaciones. Incluye información detallada como el objeto que se va a observar, los instrumentos y configuraciones que utilizar, la duración, condiciones necesarias (como *seeing*) y cualquier restricción temporal. Su propósito es optimizar la planificación y ejecución eficiente de las observaciones en un telescopio.

En el intervalo de distancia de separación elegido para generar los resultados, la distribución de parejas esperables en 1 OB dada por la Tabla 4, donde se recoge el número aproximado de parejas con brillo superficial $SB = 21 \text{ mag/arcsec}^2$ y *seeing* 2.6, siendo éstas las condiciones donde se obtienen las distribuciones más favorables:

Tabla 4. *Número aproximado de parejas por bloque de observación para WEAVE a seeing 2.6 y SB=21*

	$SNR = 2$	$SNR = 4$	$SNR = 6$
Número de parejas	188	76	31

Fuente: comunicación privada WEAVE-QSO Project

En las siguientes secciones se presentan los resultados organizados jerárquicamente: primero según la resolución instrumental utilizada, y dentro de cada resolución, por el nivel de SNR correspondiente a las señales evaluadas. Además, cada figura muestra los modelos térmicos simulados junto con la escala de Jeans λ_J que le corresponde, y una escala en el eje de ordenadas ajustada a cada figura para facilitar la inspección visual de los resultados. En cuanto a la normalización del flujo necesaria para la consistencia de las comparativas, se han tomado los factores de normalización A obtenidos por el método de bisección, y dados por la Tabla 2.

4.2. Resultados a alta resolución espacial: 30 km/s

4.2.1. Resultados obtenidos con SNR=6

El primer escenario evaluado tiene sobre todo interés teórico desde el punto de vista de la relación entre la SNR y la resolución, ya que las condiciones en las que pueden obtenerse muestras de tamaño relevante de espectros con $SNR \sim 6$ son complejas de alcanzar con la tecnología actual, y por tanto el número de parejas disponibles con estas características son muy escasas, como indica la Tabla 4.

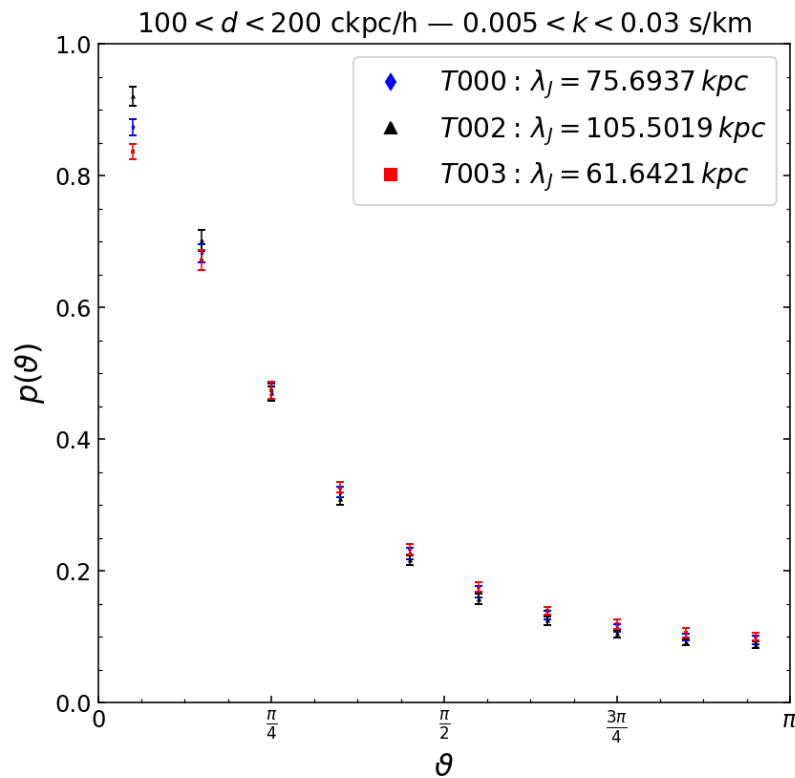
La combinación de una alta resolución con una alta SNR permite la caracterización de los modelos con tamaño de muestra de tan sólo 40 parejas, y siendo éste el mejor de los resultados en cuanto a tamaño de la muestra se refiere.

En la Figura 12 puede verse, atendiendo a los ángulos θ cercanos a 0, cómo la distinción entre los tres modelos térmicos es posible sin incertidumbre al no haber solape entre las barras de error y dada la alta densidad de probabilidad que presenta $P(\theta)$ para esos mismos valores. De

igual forma, el resto de los valores del histograma aportan estadística suficiente como para, en conjunto, garantizar la caracterización de los tres modelos sin incertidumbre.

Aun así, como se comentaba al inicio de este capítulo, la caracterización de los tres modelos a $SNR=6$ conlleva un elevado coste de obtención de la muestra, no tanto por el número de bloques de observación necesarios (2 en este caso) sino por la instrumentación necesaria. Además, es importante considerar que se están obteniendo estos resultados para unas condiciones de observación óptimas, lo que indica que no solo es necesario considerar resolución y SNR a la hora de hacer las observaciones y simulaciones, sino que, además, es importante tener en cuenta la disponibilidad de objetos observables con las características necesarias, condiciones de observación y viabilidad tecnológica para conformar la muestra de estudio con esos niveles de calidad.

Figura 12. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, $SNR=6$ y 40 parejas de QSOs

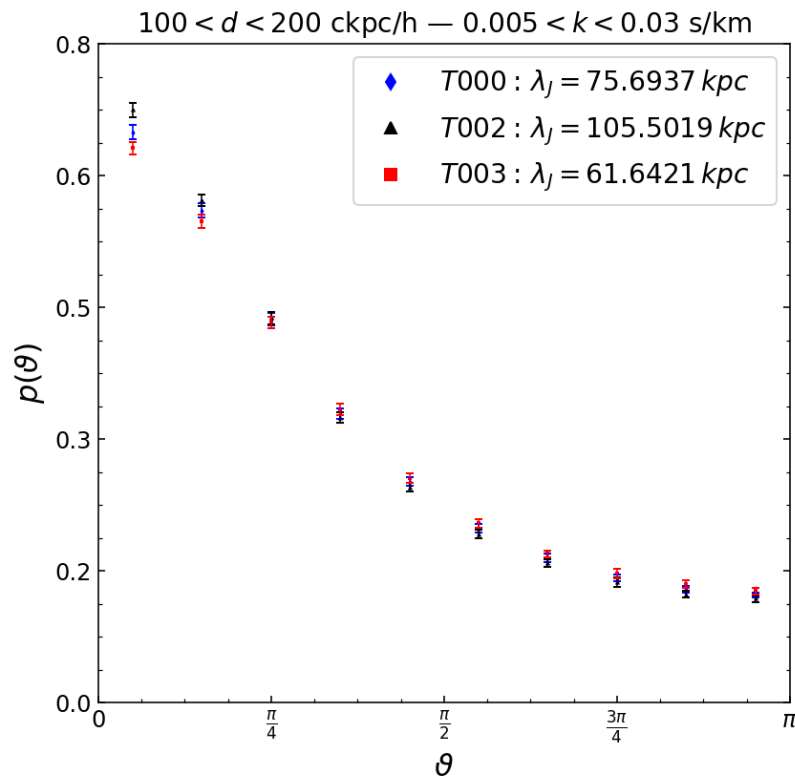


Fuente: elaboración propia

4.2.2. Resultados obtenidos con SNR=4

El siguiente caso estudiado en alta resolución dispone de una SNR menor, pero más cercana a los niveles esperables en las observaciones reales, y conllevando a su vez un aumento del tamaño de la muestra. En la Figura 13 se muestran los resultados de las simulaciones para SNR=4 y resolución espacial de 30 km/s.

Figura 13. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=4 y 110 parejas de QSOs



Fuente: elaboración propia

Para una resolución tan alta, el aumento de la SNR permite caracterizar los tres modelos para una muestra de tan solo 110 parejas de QSOs, obtenibles en 2 OBs. El elevado valor de $P(\theta)$ correspondiente a las diferencias de θ cercanas a 0 indica que las señales analizadas han viajado a través de las mismas estructuras del MIG (ver §3.1.2.2), conservando la coherencia

espectral y permitiendo resolver tres modelos térmicos con distinta escala de suavizado. Al igual que en el escenario anterior, el resto de los valores de $P(\theta)$, que también tienen relevancia estadística, se encuentran suficientemente separados y permiten, aun con errores, una mínima diferenciación entre los modelos, por lo que el conjunto total tiene la estadística suficiente como para asumir que los modelos se caracterizan sin incertidumbre significativa, tal y como ocurriera en §4.2.1.

De esta forma, se empieza a atisbar el impacto de la SNR, incluso para una misma resolución espacial, en la precisión de los resultados y el tamaño de la muestra necesario para obtenerlos sin incertidumbre.

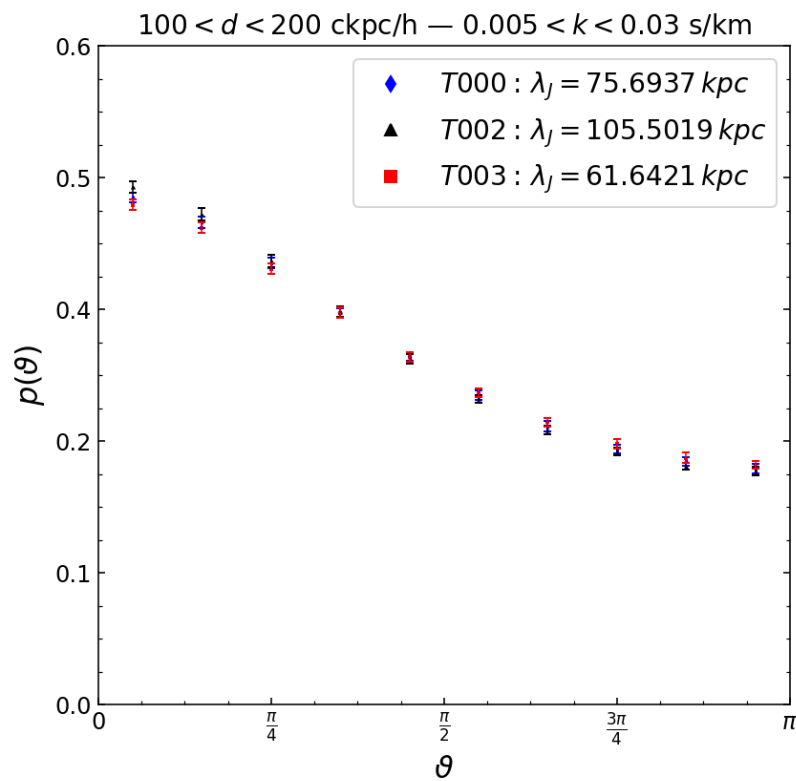
Estos resultados, aun para una resolución espacial alta, suponen un gran avance en la consecución de los objetivos iniciales de este TFE, ya que permitirían la caracterización con espectros de parejas de QSOs con SNRs muy bajas en comparación a las usadas por Rorai et al. (2013), como se explica en §1.4.5, ampliando las posibilidades de aplicación del método empleado en la caracterización de la pequeña escala del MIG.

4.2.3. Resultados obtenidos con SNR=2

Dentro de la caracterización de los modelos a alta resolución, este escenario es el más desfavorable al presentar la SNR más baja de las tres evaluadas. Para poder analizar este caso se han elegido dos resultados de entre todos los generados: los obtenidos con muestras de 195 y 500 parejas de QSOs. El objetivo de este planteamiento es ver el impacto que puede tener una variación significativa de la muestra a una SRN muy baja, incluso cuando se observa con una a alta resolución instrumental. En la Figura 14 puede verse el resultado del análisis para 195 parejas, donde las barras de error de los modelos analizados se solapan lo suficiente como para introducir una incertidumbre elevada en la distinción de éstos. Por otra parte, en la Figura 15 se exponen los resultados para las mismas condiciones de observación, pero usando en su lugar una muestra de 500 parejas. En este caso, es posible caracterizar sin errores el modelo T002, mientras que la distinción entre los modelos T001 y T003 es casi perfecta, afectada únicamente por un solape mínimo entre las barras de error.

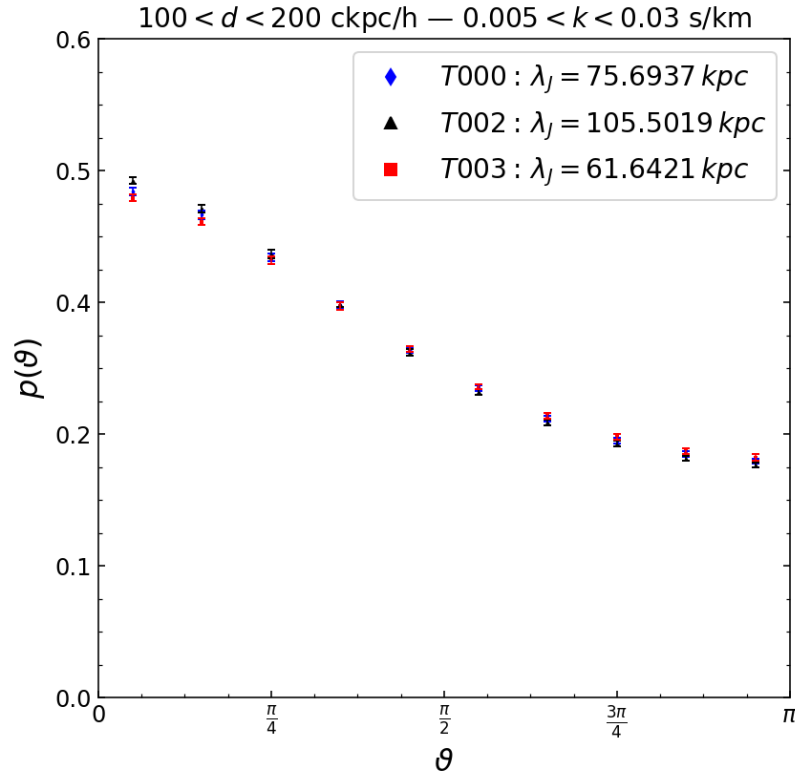
Hay que destacar el hecho de que este resultado se ha obtenido para el máximo número de parejas permitidas, 500, correspondientes a su vez a 5 OBs. Por tanto, en este caso extremo (en el que obtener una muestra de tal tamaño supone ya un desafío considerable) es posible caracterizar los tres modelos con una incertidumbre despreciable, aunque ello implique un alto coste asociado a la obtención de la muestra.

Figura 14. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=2 y 195 parejas de QSOs



Fuente: elaboración propia

Figura 15. Resultados de la caracterización de modelos térmicos con resolución espacial de 30 km/s, SNR=2 y 500 parejas de QSOs



Fuente: elaboración propia

En conclusión, ni siquiera una muy buena resolución espacial permite la distinción entre los tres modelos térmicos empleados en las simulaciones cuando la SNR es muy baja y el número de parejas no es extremadamente elevado. El impacto de una SNR baja en la caracterización de éstos, y en consecuencia de un nivel de ruido muy parecido al de la señal, es tan alto que, incluso en condiciones observacionales donde la muestra es considerable, hace prácticamente imposible la distinción entre los modelos térmicos sin introducir un elevado grado de incertidumbre, como puede apreciarse en las barras de error solapadas de los modelos en la Figura 14.

4.3. Resultados a media resolución espacial: 60 km/s

Tras estudiar los mejores escenarios posibles en cuanto a la resolución se refiere, en esta sección se presentan los resultados más relevantes de todo el capítulo, ya que se aproximan de forma más directa a los objetivos definidos en el Capítulo 2. En particular, se explora la caracterización del MIG a pequeña escala mediante simulaciones de espectros que no sólo presentan una baja SNR, sino que también han sido obtenidos a modos de media resolución instrumental.

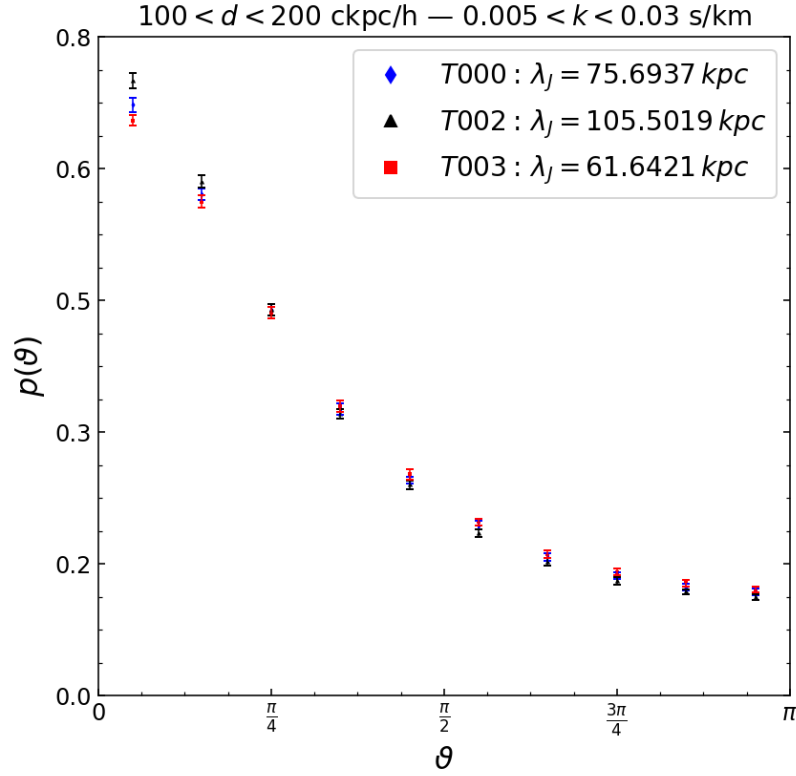
4.3.1. Resultados obtenidos con SNR=6

Teniendo como referencia los resultados expuestos en §4.2.1, la expectativa en cuanto a la caracterización a resolución baja para una SNR=6 es prometedora.

Así lo demuestra la Figura 16 donde, de nuevo, es viable una distinción de los tres modelos térmicos con SNR=6 y muestra de 115 parejas. Una vez más, hay que considerar la utilidad de estos análisis y las complejidades que conlleva su realización: en cuanto a su utilidad, permiten estudiar el impacto sobre el tamaño de la muestra, así como en la distribución de densidad de probabilidad que viene dada por $P(\theta)$, cuando se varía la SNR dentro de una misma resolución espacial; por otra parte, y como ya se ha comentado anteriormente, la cantidad de parejas de QSOs que puedan ofrecer una SNR=6 en 1 OB (para 115 parejas se necesitarían 4 OBs) es muy baja, por lo que el coste de elaboración de una muestra con estas características, por pequeña que sea ésta, podría ser considerablemente elevado o incluso inviable con la tecnología actual.

En cualquier caso, este estudio aporta un tamaño de muestra con el que sería viable la caracterización de los modelos en caso de disponer de suficientes datos con las características especificadas, y abre la puerta a futuros estudios cuando la tecnología permita las observaciones con estos niveles de SNR.

Figura 16. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=6 y 115 parejas de QSOs



Fuente: elaboración propia

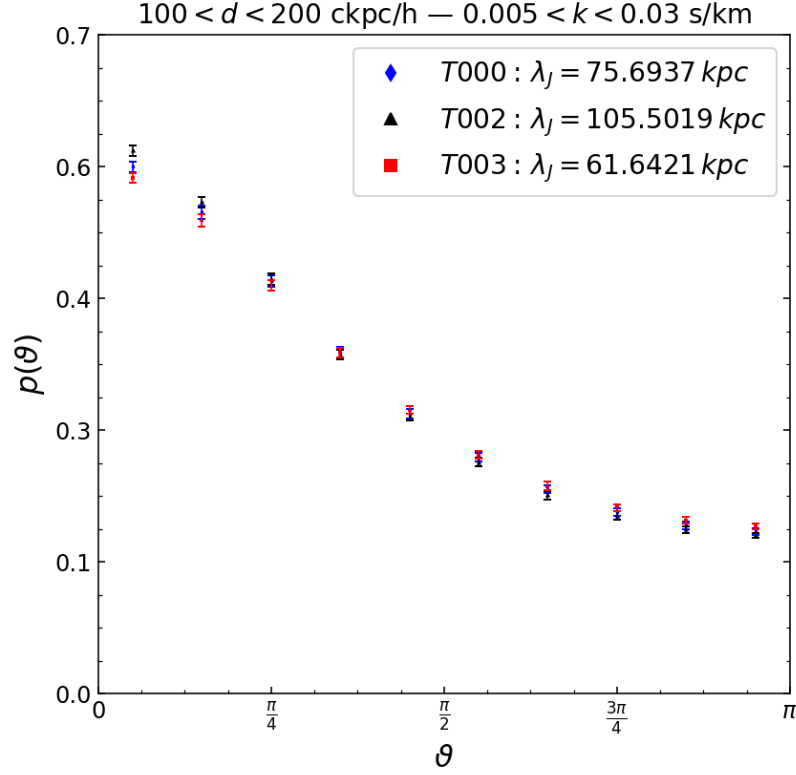
4.3.2. Resultados obtenidos con SNR=4

Estas simulaciones son, sin duda, las más interesantes de las que se presentan en esta sección: están realizadas con una resolución instrumental media real, y con una SNR que entra dentro de lo esperable en cuanto a obtención de espectros se refiere, contextualizada en el marco de los nuevos catálogos que están en desarrollo.

Considerando a los resultados anteriores, es de esperar que el número de parejas aumente con respecto al modo de alta resolución para esta misma SNR. Por tanto, para poder determinar las condiciones de tamaño de la muestra con precisión, se ha prestado especial énfasis en estas simulaciones, realizando numerosas iteraciones del método implementado con las mismas condiciones de simulación, asegurando así que la caracterización de los

modelos sea consistente y que calidad de los resultados obtenidos no sean fruto de la aleatoriedad de las muestras generadas.

Figura 17. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 170 parejas de QSOs



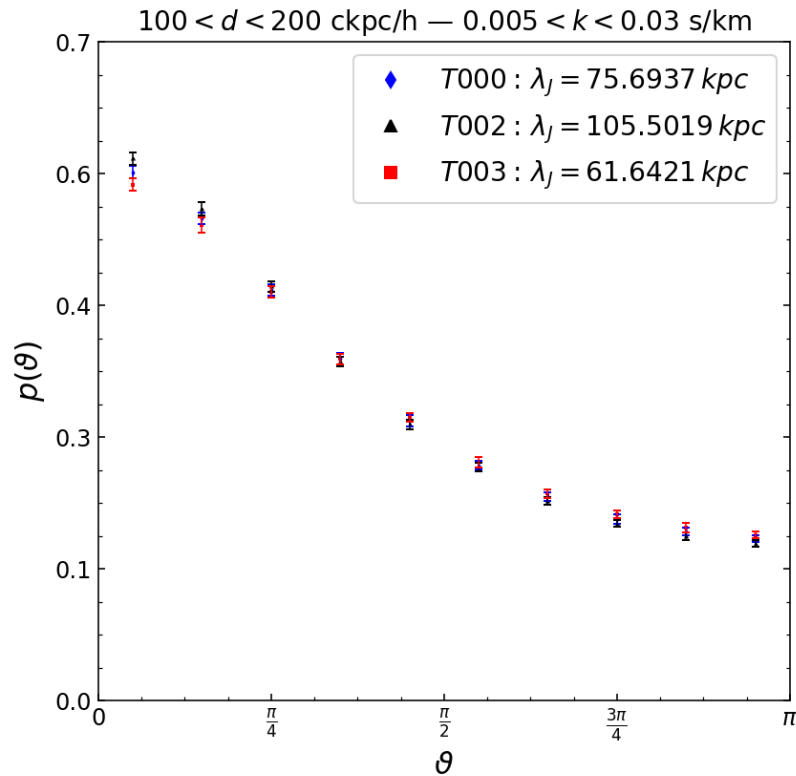
Fuente: elaboración propia

La Figura 17 muestra los resultados obtenidos en la caracterización de los modelos térmicos del MIG simulados: de nuevo, la función $P(\theta)$ muestra que es posible diferenciar los tres modelos para una muestra de 170 parejas, obtenibles en 3 OBs, introduciendo únicamente una incertidumbre despreciable entre los modelos T000 y T003, al igual que ocurre con los resultados expuestos en §4.2.3 para 500 parejas.

Este resultado es de una importancia superlativa, ya que implica que los objetivos propuestos para este TFE son alcanzables al permitir la caracterización de la pequeña escala del MIG con

espectros simulados de calidad media, con configuraciones equivalentes a observaciones con $SNR=4$ y una resolución instrumental real media/baja. Además, los resultados obtenidos en un intervalo de ± 20 parejas en la muestra no difieren en exceso de los obtenidos para 170, lo que permite disponer de cierto margen de operación a la hora de elaborar muestras de parejas de QSOs cercanos. De hecho, un ajuste con 150 parejas podría llevarse a cabo reduciendo a 2 el número de OBs necesarios, asumiendo tan solo una mínima incertidumbre.

Figura 18. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, $SNR=4$ y 150 parejas de QSOs



Fuente: elaboración propia

Un ejemplo de esta variación puede verse en la Figura 18, donde a pesar de no tener la diferenciación única del modelo T002 que se tiene en la Figura 17, la probabilidad de error en

la identificación de los modelos es lo suficientemente baja (por el grado de solape de las barras de error) como para asumir que la medida es correcta.

La caracterización de los tres modelos bajo estas condiciones de simulación supone que, a partir de los datos disponibles que se esperan de los nuevos catálogos en desarrollo, será posible determinar la escala de suavizado del MIG siempre que se disponga de un tamaño de muestra que se aproxime al indicado en este mismo apartado.

4.3.3. Resultados obtenidos con SNR=2

Para este apartado se han omitido los resultados correspondientes a la SNR y resolución espacial con las que se trabaja ya que, si bien con una resolución de 30 km/s ya resultaba difícil una caracterización clara de los modelos (por el elevado coste de obtención de la muestra) la situación se vuelve aún más desfavorable al operar con una menor resolución. Para el tamaño máximo de muestra considerado en este TFE, el nivel de incertidumbre asociado a resolución espacial de 60 km/s y SNR=2 hace inviable cualquier análisis con valor interpretativo, lo que justifica su exclusión de la presentación de resultados.

4.4. Resultados a baja resolución espacial: 100 km/s

Esta agrupación de resultados tiene como objetivo explorar los límites del método implementado, asumiendo los mismos valores de SNR que se han empleado para las resoluciones espaciales de WEAVE, pero llevando ésta a un valor extremadamente bajo.

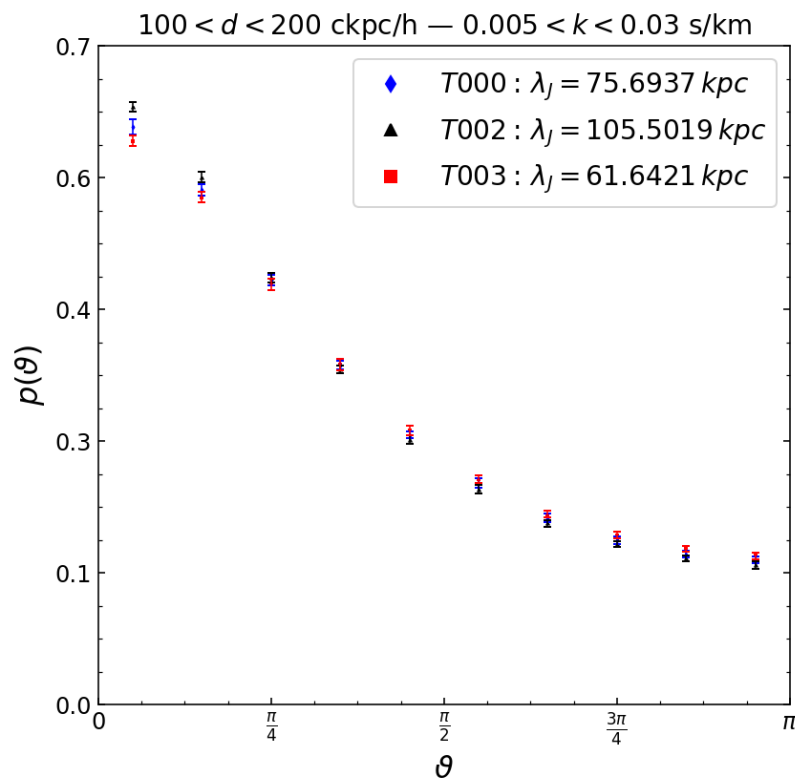
Aunque esta resolución instrumental sea inferior a las que ofrece, por ejemplo, WEAVE, resulta interesante determinar si sería posible caracterizar la pequeña escala del MIG con otros instrumentos que sí la ofrezcan o, considerando que la resolución afecta directamente a la SNR (ver §3.2.2.3.1), con datos que tengan una calidad extremadamente baja.

4.4.1. Resultados obtenidos con SNR=6

A estos niveles de resolución, la caracterización completa y con incertidumbre mínima es viable con una SNR lo suficientemente elevada en detrimento del tamaño de la muestra. En la Figura 19 se aprecia la distinción perfecta del modelo T002, y un mínimo solape entre los modelos T002 y T003, para una muestra de 170 parejas, correspondientes a 6 OBs.

Cabe esperar que, para muestras mayores, se puedan obtener resultados mejores. Sin embargo, considerando las ya mencionadas complejidades que conlleva la obtención de una muestra a $SNR=6$, se elige un valor de operación en el que el resultado puede considerarse excelente dadas las condiciones de resolución, donde sólo es necesario asumir una mínima incertidumbre entre dos de los modelos simulados.

Figura 19. Resultados de la caracterización de modelos térmicos con resolución espacial de 100 km/s, $SNR=6$ y 170 parejas de QSOs



Fuente: elaboración propia

Este resultado demuestra que una SNR lo suficientemente elevada permite operar con un rango de resoluciones que abarcan niveles que varían de muy alta (30 km/s) a muy baja calidad (100 km/s), por lo que refuerza la importancia que tiene en la caracterización de la

pequeña escala del MIG, como se viene demostrando en los resultados de los apartados anteriores.

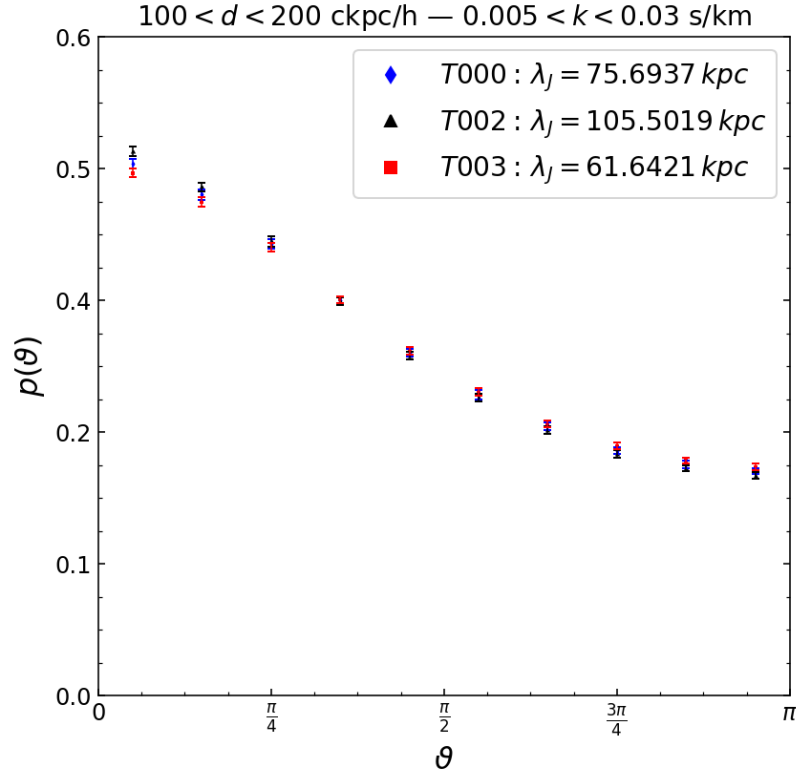
Analizar estos casos extremos permite saber cuáles son las condiciones de estudio y de calidad de los datos mínimas para afrontar análisis de caracterización de la pequeña escala del MIG con resoluciones muy bajas, e incluso analizar hasta dónde es posible llegar con un nivel de resolución muy bajo, pero con niveles de SNR ligeramente altos, ampliando el abanico de escenarios viables en los que llevar a cabo estudios de esta naturaleza.

4.4.2. Resultados obtenidos con SNR=4

La caracterización de modelos con una resolución muy baja y una SNR=4 conllevan un aumento de la muestra sustancioso, comparado con los resultados vistos a lo largo de este capítulo.

Una de las principales limitaciones que presenta este escenario es que, a medida que se aumenta el número de la muestra, la mejoría deja de ser tan proporcional a ésta como era para resoluciones mayores. Este fenómeno se puede apreciar en la comparativa entre las Figuras 20 y 21. En la Figura 20 puede verse que la caracterización de los modelos con un nivel de incertidumbre aceptablemente bajo supone disponer de una muestra de 300 parejas. El impacto de una baja resolución se hace notable tanto en el tamaño de la muestra como en los resultados que se obtienen, si se compara con los escenarios anteriores de SNR=4. Mientras que para una resolución espacial de 60 km/s y SNR=4 el tamaño de la muestra era de 170 ± 20 parejas (2-3 OBs), al disminuir la resolución espacial a 100 km/s , el tamaño de la muestra que permite la caracterización de los modelos, incluso con un nivel de incertidumbre ligeramente superior para los modelos T000 y T003, está alrededor de las 300 parejas, correspondientes a 4 OBs. Nótese que, con el aumento de la muestra, el número de OBs necesarios se ha duplicado.

Figura 20. Resultados de la caracterización de modelos térmicos con resolución espacial de 100 km/s, SNR=4 y 300 parejas de QSOs

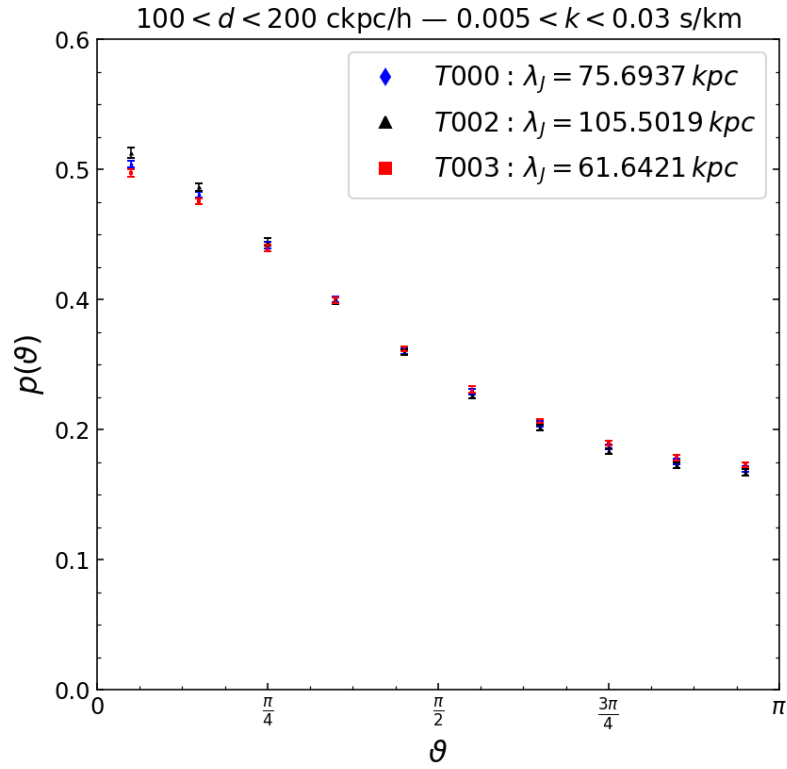


Fuente: elaboración propia

Por otra parte, un aumento de la muestra hasta 500 parejas de QSOs podría invitar a pensar que se obtiene una diferenciación total entre los tres modelos. Sin embargo, y como se ve en la Figura 21, una muestra tan elevada sólo permite diferenciar con claridad el modelo T002, mientras que se introduce una pequeña incertidumbre entre los modelos T000 y T003, demostrando que a resoluciones tan bajas el impacto del tamaño de la muestra, para una SNR fija, alcanza límites en los que deja de aportar una mejoría tan significativa como aportaba a resoluciones más altas. También es importante considerar los valores máximos de $P(\theta)$, donde la densidad de probabilidad de los valores de $\theta \approx 0$ disminuye respecto de los resultados obtenidos, por ejemplo, en §4.3.2 y §4.2.2 para esta misma SNR a otras resoluciones espaciales. Asimismo, dadas las dificultades inherentes a la obtención de una

muestra tan extensa (se necesitan 7 OBs), resulta justificado, considerando el equilibrio entre coste muestral y resolución de los tres modelos, optar por un tamaño de muestra más reducido, ya que las discrepancias entre modelos no inducen errores significativamente mayores.

Figura 21. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 500 parejas de QSOs



Fuente: elaboración propia

4.4.3. Resultados obtenidos con SNR=2

El escenario de simulación con SNR = 2 a muy baja resolución presenta las mismas limitaciones interpretativas que una SNR equivalente a resolución de 30 km/s, con el agravante de que la menor resolución provoca una degradación aún mayor de los resultados, como se detalla en §3.2.2.3.1.

Por esta razón, se omiten los resultados en este apartado, concluyendo que no es posible caracterizar los distintos modelos con una SNR tan baja a esta resolución.

4.5. Degeneración con el flujo medio

A lo largo del documento, y con especial énfasis en el Capítulo 3, se ha hablado de la degeneración que presentan métodos más tradicionales y cómo ha sido posible acabar con esta degeneración usando nuevos métodos. Sin embargo, el método de análisis de diferencia de fases en el dominio de Fourier puede no estar exento de otras degeneraciones como, por ejemplo, el flujo medio de referencia utilizado en las simulaciones (ver §3.2.2.2.2).

En esta sección se estudiará el impacto que tiene en la caracterización de los modelos térmicos una variación del flujo medio de referencia, con el objetivo de ratificar la validez de los resultados vistos en las secciones anteriores o, en caso contrario, constatar que tanto el método empleado como los resultados obtenidos necesitan una revisión adicional que permita desprenderse de esta degeneración.

Para conseguir tal objetivo se trabajará con la desviación del flujo medio obtenida por Becker et al. (2013), donde éste toma el valor $\langle F \rangle = 0.81360 \pm 0.009$ a $z = 2.4$, y se aplicarán a los modelos T000, T002 y T003. Partiendo de este flujo, ya usado en §3.2.2.2.2, se llevarán a cabo nuevos análisis considerando el margen de error comentado, siendo los nuevos valores del flujo medio $\langle F_1 \rangle = 0.8046$ y $\langle F_2 \rangle = 0.8226$.

4.5.1. Nuevos factores de normalización

Antes de hacer los nuevos análisis a distinto flujo medio, es necesario calcular los nuevos factores de normalización que se aplicarán en el cálculo del flujo normalizado a partir de la profundidad óptica, tal y como se detalla en §3.2.2.2.2.

En este caso, y por consistencia con los resultados anteriores, se usará únicamente el método de la bisección, de manera que el cálculo de los factores sea análogo al aplicado para los factores obtenidos en la segunda columna de la Tabla 2. En la Tablas 5 y 6 se muestran los nuevos factores de normalización obtenidos para F_1 y F_2 , respectivamente.

Tabla 5. Factores de normalización para los tres modelos térmicos obtenidos con el método de bisección para $F_1 = 0.80460$

Método de bisección	
Modelo térmico	Factor de normalización A_1
T000	$A_{1,T000} = 0.80413$
T002	$A_{1,T002} = 1.20727$
T003	$A_{1,T003} = 0.63217$

Fuente: elaboración propia

Tabla 6. Factores de normalización para los tres modelos térmicos obtenidos con el método de bisección para $F_2 = 0.82260$

Método de bisección	
Modelo térmico	Factor de normalización A_2
T000	$A_{2,T000} = 0.67169$
T002	$A_{2,T002} = 1.02005$
T003	$A_{2,T003} = 0.52467$

Fuente: elaboración propia

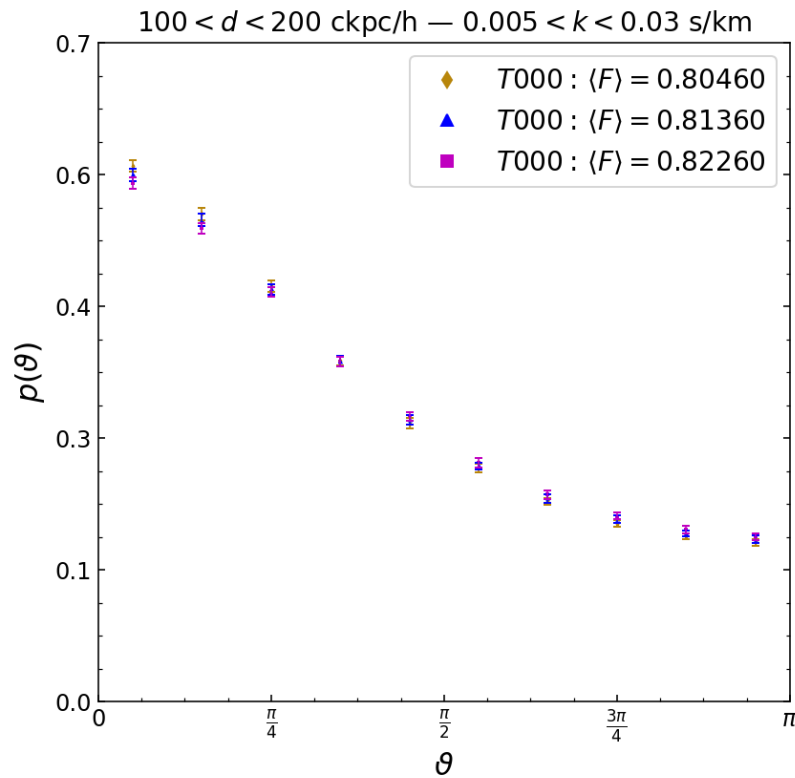
El procedimiento para detectar estas degeneraciones consiste en aplicar los mismos análisis que se han expuesto a lo largo de este capítulo, pero considerando, en lugar de tres modelos distintos, un mismo modelo con los tres flujos normalizados calculados con los nuevos factores A . De esta forma, se analizan los resultados para el modelo T000 con los factores A_{T000} ,

$A_{1,T000}$ y $A_{2,T000}$, obtenidos de las Tablas 2, 5 y 6, respectivamente, y así para los dos modelos restantes en todos los escenarios estudiados en los apartados anteriores.

4.5.2. Resultados obtenidos en el estudio de la degeneración

Para simplificar la representación de los resultados y facilitar el seguimiento del apartado, se presentará la caracterización obtenida tras aplicar el método para el modelo T000 en las condiciones más realistas, descritas en §4.3.2 (SNR=4, resolución espacial 60 km/s y muestra de 170 parejas). Dado que la caracterización de los demás modelos y escenarios considerados en este capítulo es análoga a la del caso seleccionado, se omite su representación explícita para favorecer una exposición más clara. No obstante, todos ellos han sido analizados con el mismo procedimiento, obteniendo resultados de la naturaleza similar a la que se exponen a continuación

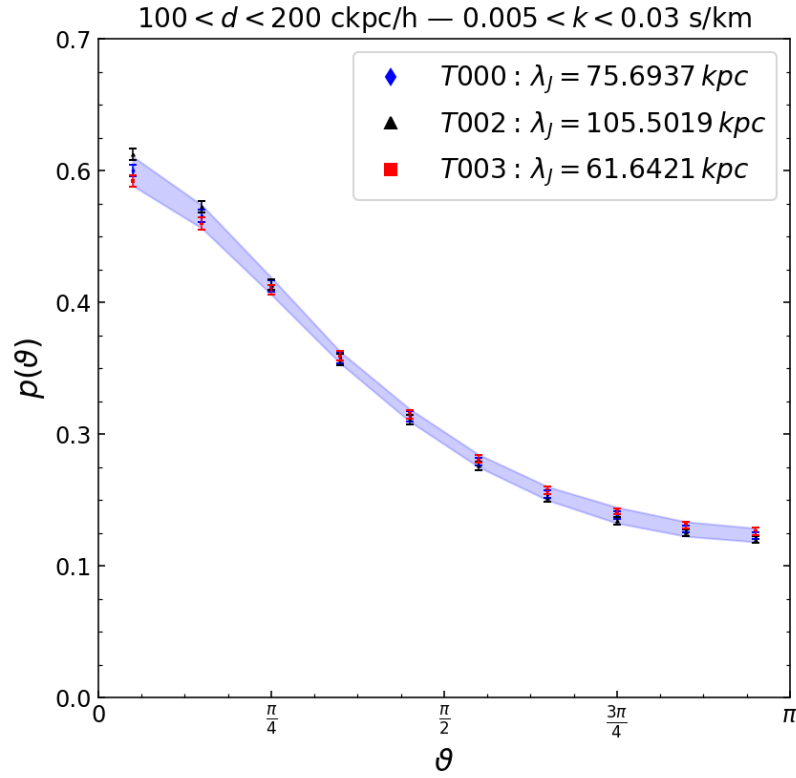
Figura 22. Resultados de la caracterización del modelo térmico T000 a distintos flujos medios con resolución espacial de 60 km/s, SNR=4 y 170 parejas de QSOs



Fuente: elaboración propia

La Figura 22 muestra el resultado de la caracterización para el modelo T000 a distintos flujos medios para las condiciones indicadas. A simple vista puede apreciarse cómo la precisión y la ausencia de errores que existía en las simulaciones la Figura 17 se ven notablemente afectadas, introduciendo un alto nivel de incertidumbre en la medida. Este deterioro en la medida no sólo afecta a las condiciones del caso más realista, sino que está presente en todos los escenarios de todos los modelos, incluidos aquellos más favorables (ver §4.2.1).

Figura 23. Resultados de la caracterización de modelos térmicos con resolución espacial de 60 km/s, SNR=4 y 170 parejas de QSOs y el rango de variación de T000 debido a la degeneración con el flujo medio.



Fuente: elaboración propia

Además, en la Figura 23 puede verse el impacto directo de la degeneración con respecto al flujo medio para las condiciones de caracterización del escenario descrito en §4.3.2. En ella aparecen los resultados obtenidos para una resolución espacial de 60 km/s, SNR=4 y 170

parejas, pero se añade una franja azul que muestra el rango de posibles valores que puede tomar la función para el modelo T000, apreciándose claramente cómo estos se superponen a los correspondientes para los modelos T002 y T003, indicando que la degeneración detectada puede inducir errores lo suficientemente importantes como para no poder diferenciar entre los tres modelos térmicos.

Este suceso supone una clara limitación del método empleado, ya que pone de manifiesto la alta susceptibilidad de los resultados a las variaciones del flujo medio. En particular, el análisis de la fase de Fourier entre líneas de visión de QSOs cercanos se ve directamente afectado por el valor medio del flujo transmitido $\langle F \rangle$, que sirve como normalización global sobre la que se construyen las simulaciones de los observables. Tal y como queda demostrado, incluso variaciones pequeñas en $\langle F \rangle$ pueden alterar de forma significativa la estadística de fase, ya que modifican la estructura relativa de los picos y valles del flujo, afectando así las propiedades del campo de fases. Esta sensibilidad introduce una degeneración sistemática en la comparación entre modelos, que puede llevar a interpretaciones erróneas si no se tiene en cuenta de manera explícita.

Por ello, es muy importante garantizar que el flujo medio de las simulaciones sea coherente con la tasa de fotoionización ultravioleta asumida. Esto puede lograrse ajustando explícitamente dicha tasa o usando un factor de normalización para que el flujo obtenido reproduzca el valor medio deseado. Estudios previos han mostrado que la calibración incorrecta de $\langle F \rangle$ no solo afecta a las medidas del espectro de potencias (Viel et al., 2004), sino que también puede inducir sesgos en enfoques que, como el análisis de fases, son sensibles a la estructura del flujo (Lidz et al., 2006). En consecuencia, cualquier estudio que emplee métodos basados en la fase del dominio de Fourier debe tratar con especial cuidado la homogeneización del flujo medio entre simulaciones y observaciones, así como entre distintos modelos simulados, con el fin de evitar que esta degeneración comprometa la interpretación física de los resultados.

5. Conclusiones y Trabajo Futuro

En este TFE se ha abordado el estudio estadístico de las propiedades del MIG a través del análisis de pares de líneas de visión de QSOs cercanos, utilizando para ello herramientas basadas en la transformada de Fourier. A lo largo de los capítulos anteriores, se han expuesto los fundamentos teóricos del enfoque empleado, el marco cosmológico en que se desarrolla, así como las características de las simulaciones hidrodinámicas utilizadas, las distintas configuraciones experimentales consideradas y la implementación en el entorno tecnológico elegido. El análisis se ha centrado en caracterizar la correlación en fase entre estas líneas de visión cercanas, identificando su dependencia con parámetros físicos como la resolución espectral o la SNR.

En este Capítulo se presentan las principales conclusiones extraídas del estudio, con especial énfasis en las limitaciones metodológicas detectadas, las implicaciones físicas de los resultados obtenidos y el potencial del enfoque propuesto para su aplicación en futuros trabajos. Asimismo, se detallan las líneas de investigación que se abren a partir de este trabajo, incluyendo posibles mejoras tanto en el tratamiento estadístico como en la consideración de aspectos teóricos en los estudios realizados. Estas reflexiones finales permiten poner en perspectiva la contribución de este TFE dentro del campo la cosmología observacional y el estudio del MIG.

5.1. Limitaciones del método y los estudios

Durante el desarrollo de estos estudios, se han identificado distintas limitaciones que conviene señalar con el objetivo de evidenciar los puntos de mejora, fortalecer la base de los futuros trabajos que se realicen en el ámbito de la Cosmología y la caracterización del MIG usando métodos de análisis espectral, y aportar transparencia a los estudios llevados a cabo. Reconocerlas contribuye a una interpretación más rigurosa de los resultados y orienta el diseño de las nuevas investigaciones.

Una de las principales limitaciones del método y los datos utilizados en este TFE está directamente relacionada con el tamaño de la muestra utilizada. Contar con un número de

parejas limitado obliga a repetir datos cuando el número de iteraciones del algoritmo es elevado y el tamaño de la muestra se aproxima al máximo establecido. Concretamente, cuando el tamaño de la muestra es de 500 parejas, el número total de parejas utilizadas es de 12000, lo que es un volumen de parejas 6 veces superior al que ofrecen los archivos de simulación. Esto implica que la estadística no sea completamente independiente, e introduce cierta incertidumbre a la hora de interpretar resultados bajo esas condiciones, ya que las muestras no están conformadas por datos exclusivamente únicos.

Además, debe considerarse la degeneración de los resultados con respecto al flujo medio transmitido, que afecta a la precisión del método implementado. La dependencia del método con el flujo medio destaca la importancia de su estimación rigurosa, con el fin de minimizar su impacto en los análisis realizados a través de posibles sesgos, y mejorar la fiabilidad de las interpretaciones que se derivan de los resultados.

5.2. Conclusiones extraídas del estudio

El desarrollo de este TFE ha permitido implementar un método que, como ha quedado reflejado a lo largo del Capítulo 4, permite la caracterización de la pequeña escala (o suavizado de Jeans) del MIG a partir del estudio del bosque de Lyman- α de parejas de QSOs cercanos, analizados en el dominio de Fourier. Este logro supone abrir una puerta a nuevas investigaciones en el campo de la Cosmología, ya que tanto el método empleado como los datos que se han procesado están en consonancia con las observaciones recientes realizadas y el modelo cosmológico estándar, Λ CDM, permitiendo que sea aplicado a los datos obtenidos de futuras observaciones y garantizando su coherencia con el marco cosmológico en que se obtendrán.

Los resultados presentados a lo largo del Capítulo 4 demuestran que, para condiciones de trabajo e instrumentación reales en las que se obtienen datos de calidad media o baja, es posible diferenciar entre tres modelos térmicos distintos del MIG. La aplicación del método detallado en el Capítulo 3 ha permitido la caracterización de la escala de suavizado del MIG para resolución espacial de 60 km/s , con SNRs 4 y 6 y tamaños de muestra de 170 y 110

parejas de QSOs, respectivamente; y para resolución espacial de 100 km/s , también con SNR de 4 y 6 y tamaños de muestra de 300 y 170 parejas de QSOs.

Estos hallazgos confirman la consecución de los objetivos iniciales de este TFE, ya que ha sido posible determinar los criterios cuantitativos de resolución mínima del MIG aplicando técnicas de análisis espectral. Además, se han desarrollado e implementado flujos de trabajo automatizados para el procesamiento eficiente de datos de simulaciones hidrodinámicas, facilitando la posterior interpretación física rigurosa de los resultados obtenidos. Durante este proceso, también se han detectado las limitaciones propias del método implementado, lo que contribuye a mejorar futuros enfoques. En conjunto, el estudio exhaustivo de las propiedades físicas del MIG ha permitido no sólo adaptar los datos al marco cosmológico actual, sino también sentar las bases para la creación de automatismos que aseguren la reproducción y extensión de los análisis en futuros estudios.

5.3. Futuras líneas de trabajo

A partir de las limitaciones detalladas en la sección anterior, se derivan distintas líneas de trabajo futuras que permitan reforzar los estudios llevados a cabo y extenderlos con una mejora de la fiabilidad y la calidad de éstos.

La primera línea de trabajo derivada de las limitaciones alude al tamaño de la muestra generada: ya que se disponen de archivos de simulación que contienen gran número de líneas de visión de QSOs, es posible generar nuevos conjuntos de parejas, para cada modelo térmico, con el objetivo de garantizar la independencia total de las muestras sucesivas tomadas a partir de estos conjuntos, incluso cuando el número de muestras y de iteraciones del algoritmo sean lo suficientemente elevadas.

Por otra parte, y aunque haya quedado demostrada la utilidad del método para una distancia e intervalo de k concretos, que se ajustan a las distancias de Jeans de los modelos térmicos, tiene especial interés la extensión de los análisis a distintas agrupaciones de k , así como a distintos intervalos de distancia transversal. Esto permitiría analizar con precisión la estadística de los subintervalos de distancia, y también de modos k , dentro del rango de distancias de Jeans cubierto por los distintos modelos térmicos. De esta manera, se pueden

generar muestras de parejas separadas por distintas distancias y poner a prueba la robustez del método y su aplicación a distintos resultados obtenidos de observaciones futuras.

Finalmente, un estudio que aporta precisión y fiabilidad a la muestra incluye, por una parte, el cálculo explícito de los intervalos de confianza de cada uno de los métodos, y por otra, la evaluación de la sensibilidad a los cambios de flujo medio con un enfoque basado en métodos bayesianos, utilizando Markov Chain Monte Carlo (MCMC) para modelar estadísticamente los datos representados en los histogramas. El ajuste de ambos parámetros de forma simultánea facilitaría la evaluación conjunta de las incertidumbres presentes en diferentes histogramas dentro de un marco probabilístico riguroso, permitiendo cuantificar de manera más robusta el grado de solapamiento entre las barras de error y mejorando así la caracterización y fiabilidad de los resultados obtenidos.

Referencias bibliográficas

- Adelberger, K. L., Steidel, C. C., Shapley, A. E., Hunt, M. P., Erb, D. K., Reddy, N. A., & Pettini, M. (2004). Optical Selection of Galaxies at Redshifts $1 < z < 3$. *The Astrophysical Journal*, 607(1), 226-240. <https://doi.org/10.1086/383221>
- Alexander, D. M., Davis, T. M., Chaussidon, E., Fawcett, V. A., X. Gonzalez-Morales, A., Lan, T.-W., Yèche, C., Ahlen, S., Aguilar, J. N., Armengaud, E., Bailey, S., Brooks, D., Cai, Z., Canning, R., Carr, A., Chabanier, S., Cousinou, M.-C., Dawson, K., De La Macorra, A., ... Zou, S. (2023). The DESI Survey Validation: Results from Visual Inspection of the Quasar Survey Spectra. *The Astronomical Journal*, 165(3), 124. <https://doi.org/10.3847/1538-3881/acacfc>
- Almgren, A. S., Bell, J. B., Lijewski, M. J., Lukić, Z., & Van Andel, E. (2013). Nyx: A MASSIVELY PARALLEL AMR CODE FOR COMPUTATIONAL COSMOLOGY. *The Astrophysical Journal*, 765(1), 39. <https://doi.org/10.1088/0004-637X/765/1/39>
- Becker, G. D., Bolton, J. S., Haehnelt, M. G., & Sargent, W. L. W. (2011). Detection of extended He II reionization in the temperature evolution of the intergalactic medium: IGM temperatures over $2 < z < 5$. *Monthly Notices of the Royal Astronomical Society*, 410(2), 1096-1112. <https://doi.org/10.1111/j.1365-2966.2010.17507.x>
- Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. (2013). A refined measurement of the mean transmitted flux in the Ly α forest over $2 < z < 5$ using composite quasar spectra. *Monthly Notices of the Royal Astronomical Society*, 430(3), 2067-2081. <https://doi.org/10.1093/mnras/stt031>
- Boera, E., Murphy, M. T., Becker, G. D., & Bolton, J. S. (2014). The thermal history of the intergalactic medium down to redshift $z = 1.5$: A new curvature measurement. *Monthly Notices of the Royal Astronomical Society*, 441(3), 1916-1933. <https://doi.org/10.1093/mnras/stu660>
- Crotts, A. P. S. (1989). Spatial structure in the Lyman-alpha forest. *The Astrophysical Journal*, 336, 550. <https://doi.org/10.1086/167034>

- Foltz, C. B., Weymann, R. J., Roser, H.-J., & Chaffee, F. H., Jr. (1984). Improved lower limits on Lyman-alpha forest cloud dimensions and additional evidence supporting the gravitational lens nature of 2345 + 007A,B. *The Astrophysical Journal*, 281, L1. <https://doi.org/10.1086/184271>
- Furlanetto, S. R., & Oh, S. P. (2009). The temperature-density relation of the intergalactic medium after hydrogen reionization. *The Astrophysical Journal*, 701(1), 94-104. <https://doi.org/10.1088/0004-637X/701/1/94>
- Gnedin, N. Y., & Hui, L. (1998). Probing the Universe with the Ly α forest – I. Hydrodynamics of the low-density intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 296(1), 44–55. <https://doi.org/10.1046/j.1365-8711.1998.01328.x>
[Sistema de Datos Astrofísicos+3](#)
- Garzilli, A., Bolton, J. S., Kim, T.-S., Leach, S., & Viel, M. (2012). The intergalactic medium thermal history at redshift $z = 1.7$ -3.2 from the Ly α forest: A comparison of measurements using wavelets and the flux distribution: The IGM thermal history at $z = 1.7$ -3.2. *Monthly Notices of the Royal Astronomical Society*, 424(3), 1723-1736. <https://doi.org/10.1111/j.1365-2966.2012.21223.x>
- Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences of the United States of America*, 15(3), 168–173. <https://doi.org/10.1073/pnas.15.3.168>
- Hui, L., & Gnedin, N. Y. (1997). Equation of state of the photoionized intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 292(1), 27-42. <https://doi.org/10.1093/mnras/292.1.27>
- Hui, L., & Haiman, Z. (2003). The Thermal Memory of Reionization History. *The Astrophysical Journal*, 596(1), 9-18. <https://doi.org/10.1086/377229>
- Jin, S., Trager, S. C., Dalton, G. B., Aguerri, J. A. L., Drew, J. E., Falcón-Barroso, J., Gäsicke, B. T., Hill, V., Iovino, A., Pieri, M. M., Poggianti, B. M., Smith, D. J. B., Vallenari, A., Abrams, D. C., Aguado, D. S., Antoja, T., Aragón-Salamanca, A., Ascasibar, Y., Babusiaux, C., ... Zurita, C.

- (2024). The wide-field, multiplexed, spectroscopic facility WEAVE: Survey design, overview, and simulated implementation. *Monthly Notices of the Royal Astronomical Society*, 530(3), 2688-2730. <https://doi.org/10.1093/mnras/stad557>
- Khan, N. K., Kulkarni, G., Bolton, J. S., Haehnelt, M. G., Iršič, V., Puchwein, E., & Asthana, S. (2024). Particle initialization effects on Lyman- α forest statistics in cosmological SPH simulations. *Monthly Notices of the Royal Astronomical Society*, 530(4), 4920-4935. <https://doi.org/10.1093/mnras/stae662>
- Kulkarni, G., Hennawi, J. F., Oñorbe, J., Rorai, A., & Springel, V. (2015). Characterizing the Pressure Smoothing Scale of the Intergalactic Medium. *The Astrophysical Journal*, 812(1), 30. <https://doi.org/10.1088/0004-637X/812/1/30>
- Lidz, A., Heitmann, K., Hui, L., Habib, S., Rauch, M., & Sargent, W. L. W. (2006). Tightening Constraints from the Ly α Forest with the Flux Probability Distribution Function. *The Astrophysical Journal*, 638(1), 27. <https://doi.org/10.1086/498699>
- Lidz, A., Faucher-Giguère, C.-A., Dall'Aglio, A., McQuinn, M., Fechner, C., Zaldarriaga, M., Hernquist, L., & Dutta, S. (2010). A measurement of small-scale structure in the $2.2 \leq z \leq 4.2$ Ly α forest. *The Astrophysical Journal*, 718(1), 199-230. <https://doi.org/10.1088/0004-637X/718/1/199>
- Lukić, Z., Stark, C., Nugent, P., White, M., Meiksin, A., & Almgren, A. (2015). The Lyman- α forest in optically-thin hydrodynamical simulations. *Monthly Notices of the Royal Astronomical Society*, 446(4), 3697-3724. <https://doi.org/10.1093/mnras/stu2377>
- Lynds, R. (1971). The Absorption-Line Spectrum of 4c 05.34. *The Astrophysical Journal*, 164, L73. <https://doi.org/10.1086/180695>
- McQuinn, M. (2016). The Evolution of the Intergalactic Medium. *Annual Review of Astronomy and Astrophysics*, 54(1), 313-362. <https://doi.org/10.1146/annurev-astro-082214-122355>
- Meiksin, A. A. (2009). The physics of the intergalactic medium. *Reviews of Modern Physics*, 81(4), 1405-1469. <https://doi.org/10.1103/RevModPhys.81.1405>

- Oñorbe, J. (s/f). *The THERMAL suite - home page*. Joseonorbe.com. Recuperado el 6 de junio de 2025, de <http://thermal.joseonorbe.com/index.html>
- Pieri, M. M., Bonoli, S., Chaves-Montero, J., Paris, I., Fumagalli, M., Bolton, J. S., Viel, M., Miralda-Escude, J., Busca, N. G., Rahmani, H., Peroux, C., Font-Ribera, A., & Trager, S. C. (2016). *WEAVE-QSO: A massive intergalactic medium survey for the william herschel telescope*.
- Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., ... Zonca, A. (2020). *Planck 2018 results: VI. Cosmological parameters*. *Astronomy & Astrophysics*, 641, A6. <https://doi.org/10.1051/0004-6361/201833910>
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing* (3rd ed.). Cambridge University Press.
- Riess, A. G., Yuan, W., Macri, L. M., Scolnic, D., Brout, D., Casertano, S., Jones, D. O., Murakami, Y., Anand, G. S., Breuval, L., Brink, T. G., Filippenko, A. V., Hoffmann, S., Jha, S. W., D'arcy Kenworthy, W., Mackenty, J., Stahl, B. E., & Zheng, W. (2022). A Comprehensive Measurement of the Local Value of the Hubble Constant with $1 \text{ km s}^{-1} \text{ Mpc}^{-1}$ Uncertainty from the Hubble Space Telescope and the SH0ES Team. *The Astrophysical Journal Letters*, 934(1), L7. <https://doi.org/10.3847/2041-8213/ac5c5b>
- Rorai, A., Hennawi, J. F., Oñorbe, J., White, M., Prochaska, J. X., Kulkarni, G., Walther, M., Lukić, Z., & Lee, K.-G. (2017). Measurement of the small-scale structure of the intergalactic medium using close quasar pairs. *Science*, 356(6336), 418-422. <https://doi.org/10.1126/science.aaf9346>
- Sargent, W. L. W., Young, P. J., Boksenberg, A., & Tytler, D. (1980). The distribution of Lyman-alpha absorption lines in the spectra of six QSOs—Evidence for an intergalactic origin. *The Astrophysical Journal Supplement Series*, 42, 41. <https://doi.org/10.1086/190644>

- Schaye, J., Theuns, T., Leonard, A., & Efstathiou, G. (1999). Measuring the equation of state of the intergalactic medium. *Monthly Notices of the Royal Astronomical Society*, 310(1), 57-70. <https://doi.org/10.1046/j.1365-8711.1999.02956.x>
- Schmidt, M. (1963). 3C 273: A Star-Like Object with Large Red-Shift. *Nature*, 197(4872), 1040-1040. <https://doi.org/10.1038/1971040a0>
- Viel, M., Haehnelt, M. G., & Springel, V. (2004). Inferring the dark matter power spectrum from the Lyman α forest in high-resolution QSO absorption spectra. *Monthly Notices of the Royal Astronomical Society*, 354(3), 684-694. <https://doi.org/10.1111/j.1365-2966.2004.08224.x>
- Viel, M., Bolton, J. S., & Haehnelt, M. G. (2009). Cosmological and astrophysical constraints from the Lyman α forest flux probability distribution function. *Monthly Notices of the Royal Astronomical Society: Letters*, 399(1), L39-L43. <https://doi.org/10.1111/j.1745-3933.2009.00720.x>
- Weinberg, S. (1973). *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. John Wiley & Sons.
- Wilson, R. W. (1980). History of the Discovery of the Cosmic Microwave Background Radiation. *Physica Scripta*, 21(5), 599-605. <https://doi.org/10.1088/0031-8949/21/5/001>
- Wolfe, A. M. (1986). New Evidence from the Lyman-Alpha Forest concerning the Formation of Galaxies. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 320(1556), 503-515. <https://doi.org/10.1098/rsta.1986.0132>
- Xu, H., Wise, J. H., & Norman, M. L. (2013). Population III stars and remnants in high-redshift galaxies. *The Astrophysical Journal*, 773(2), 83. <https://doi.org/10.1088/0004-637X/773/2/83>

Anexo A. Índice de Acrónimos

A

AGN: núcleo de galaxia activo (por sus siglas en inglés)

AMR: refinamiento de malla adaptativo (por sus siglas en inglés)

F

FWHM: anchura a media altura (por sus siglas en inglés)

H

HDF5: *Hierarchical Data Format 5*

I

IDE: entorno de desarrollo integrado (por sus siglas en inglés)

M

MCMC: Markov Chain Monte Carlo

MIG: medio intergaláctico

P

PDF: función de densidad de probabilidad (por sus siglas en inglés)

Q

QSO: cuásar, objeto cuasiestelar (por sus siglas en inglés)

S

SB: brillo superficial (por sus siglas en inglés)

SNR: relación señal-ruido (por sus siglas en inglés)

T

TFE: Trabajo de Fin de Estudios

W

WEAVE: *William Herschel Telescope Enhanced Area Velocity Explorer*