



UNIVERSIDAD NACIONAL AUTÓNOMA DE MEXICO
PROGRAMA DE POSGRADO EN ASTROFÍSICA
INSTITUTO DE ASTRONOMÍA

EL HALO DE LA GALAXIA: FENÓMENOS COLECTIVOS EN SU DINÁMICA Y
DETECCIÓN DE SUBESTRUCTURA A TRAVÉS DEL HALO ESTELAR

TESIS

QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN CIENCIAS (ASTROFÍSICA)

PRESENTA:
JOSÉ ARTURO TRELLES HERNÁNDEZ

TUTOR O TUTORES PRINCIPALES:
DR. JOSÉ OCTAVIO VALENZUELA TIJERINO
INSTITUTO DE ASTRONOMÍA
DR. SANTI ROCA-FÀBREGA
UNIVERSIDAD DE LUND, SUECIA

MIEMBROS DEL COMITÉ TUTOR:

DR. HÉCTOR MANUEL VELÁZQUEZ
DR. LUIS ALBERTO MARTÍNEZ MEDINA
DR. JOSÉ ANTONIO DE DIEGO ONSURBE
INSTITUTO DE ASTRONOMÍA

CIUDAD DE MÉXICO, CD. MX. NOVIEMBRE 2023



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*A mis padres y mi hermana,
por el tiempo que no nos vimos.*

Agradecimientos

Tengo que comenzar agradeciendo a mis padres y a mi hermana, quienes me han apoyado en todo momento y han sido un ejemplo de honestidad incondicional, esfuerzo y dedicación en la vida, sin ellos nada de esto sería posible. A Soco por iluminarme cuando la luz se hace tenue, a mis amigos por su paciencia y compañía en este camino.

Agradezco profundamente al Dr. Octavio Valenzuela quien ha sido mi tutor, mentor y maestro a lo largo de estos años y a quien le agradezco sobretodo por apoyarme durante las épocas más duras de mi aprendizaje en mis primeros pasos en el posgrado. Gracias al Dr. Santi Roca-Fàbrega por su constante e incondicional apoyo, conocimientos y sabios consejos durante todo el doctorado.

Agradezco a los miembros del comité tutor, Héctor Velázquez, Luis Martínez Medina, José Antonio de Diego Onsurbe, por sus valiosos comentarios a lo largo del doctorado.

Agradezco a los sinodales, Francesca Figueras Siñol, María Ángeles Gómez Flechoso, Pau Ramos, Alfred Castro Ginard, Luis Alberto Martínez Medina, por su esfuerzo, tiempo y dedicación para leer este trabajo de tesis y proporcionar sus comentarios detallados y sugerir correcciones que ayudaron a mejorar este trabajo considerablemente.

Agradezco a la UNAM, al Instituto de Astronomía y al posgrado por mi formación académica, que comenzó en la licenciatura, así como los recursos educativos y académicos proporcionados durante estos años.

Agradezco al CONACyT por la beca de posgrado con cvu 825451 y número de apoyo 737292, otorgada durante ocho semestres y que me permitió realizar los estudios de doctorado que derivan en este trabajo de tesis.

Agradezco el apoyo PAEP proporcionado para asistir a congresos durante el doctorado. Agradezco el apoyo de de la beca PAPIIT del proyecto AG101222 que durante cuatro meses me fue proporcionado para concluir este trabajo de tesis.

Agradezco al Instituto de Astronomía por los recursos computacionales proporcionados a lo largo del doctorado, sobre todo por permitirme utilizar el clúster Atocatl para llevar a cabo las simulaciones numérica de la tesis, en ese sentido agradezco el apoyo constante a J. Cesar Clemente en lo que se refiere a la enseñanza de las herramientas computacionales disponibles y el uso del clúster.

Acrónimos

APOGEE Apache Point Observatory Galactic Evolution Experiment

ART Adaptative Refinement Tree

CH Chandrasekhar

CMB Cosmic Microwave Background

DBSCAN Density-based spatial clustering of applications with noise

DESI Dark Energy Spectroscopic Instrument

DL Deep Learning

GANs Generative Adversarial Networks

IA Inteligencia artificial

IMF Función Inicial de Masa

K-Means K-Means

KNN K-Nearest Neighbors

ΛCDM Lambda Cold Dark Matter

LSST Large-aperture Synoptic Survey Telescope

ML Machine Learning

NFW Navarro, Frenk and White (1996)

NB N-cuerpos

RF Random Forest

RAVE The Radial Velocity Experiment

Rockstar Rockstar Phase Space Halo Finder

SA Semianalítico

SFR Tasa de Formación Estelar

SVM Support Vector Machines

SOMs Self Organized Maps

VAEs Variational Autoencoders

VL Vía Láctea

VPOS Vast POLar Structure

Índice general

Agradecimientos	4
Resumen	12
1. Introducción	15
1.1. El modelo de materia oscura fría y energía oscura	15
1.2. La Vía Láctea, su halo oscuro y su halo estelar	16
1.3. Motivación de la tesis	19
1.4. Objetivos	20
2. Revisión de la fricción dinámica: Efectos colectivos	21
2.1. Los efectos colectivos	23
2.1.1. Efecto en las orbitas con varios perturbadores	25
2.1.2. Cambios en la estructura interna y su efecto en la fricción dinámica neta	27
2.2. La vasta estructura polar	27
3. Simulaciones de N-cuerpos no-colisionales	29
3.1. Las simulaciones de N-cuerpos	29
3.1.1. Sistemas colisionales y no-colisionales	30
3.1.2. Principales ecuaciones en el problema de N-cuerpos	31
3.2. Códigos basados en árboles	32
3.2.1. Funcionamiento de gyrfalcON	33
3.2.2. Algoritmo de Cálculo de Fuerzas	33
3.2.3. Criterio de Aceptación Multipolar	34
3.3. Códigos de partícula-malla	34
3.3.1. Códigos de malla: GADGET2	35
3.4. Armado de las condiciones iniciales	36
3.4.1. Condiciones iniciales de la galaxia aislada	36
3.4.2. Fórmula de Eddington: Función de distribución para sistemas esféricos	36

3.4.3.	Fórmula de Eddington y sistemas ergódicos	37
3.5.	Algoritmo para las Condiciones iniciales	38
3.5.1.	Creando las condiciones iniciales dentro del código MKHALO	40
3.6.	El modelo de los halos de materia oscura	40
4.	Caída concurrente de satélites: Importancia de los efectos colectivos para la re- construcción orbital	41
4.1.	Un modelo semianalítico	41
4.1.1.	Variaciones en el logaritmo coulombiano en la fórmula de Chandrasekhar para la fricción dinámica	42
4.2.	Los experimentos con simulaciones de N-cuerpos	43
4.2.1.	Comparación del modelo semianalítico con simulaciones de N-cuerpos	44
4.3.	Experimentos con satélites coplanares	46
4.3.1.	Experimento con satélites coplanares a diferentes distancias inicialmente	46
4.3.2.	Satélites coplanares equidistantes	48
4.4.	Configuraciones con satélites no coplanares	51
4.4.1.	Experimento con satélites vivos en diferentes planos orbitales y equidistantes del centro	51
4.4.2.	Satélites en diferentes planos orbitales y diferentes distancias iniciales del centro	53
5.	El Aprendizaje Automático y los algoritmos de búsqueda de agrupamiento.	63
5.1.	Motivación para utilizar algoritmos del campo del Aprendizaje Automático	64
5.2.	Algoritmos de Aprendizaje Supervisado	64
5.2.1.	Algoritmos de Clasificación	65
5.2.2.	El algoritmo <i>K-Nearest-Neighbours</i>	66
5.3.	Algoritmos de aprendizaje <i>semisupervisado</i>	67
5.4.	Algoritmos de aprendizaje <i>no supervisado</i>	68
5.4.1.	Tipos de Algoritmos “ <i>no supervisados</i> ”	69
5.4.2.	Ventajas del enfoque <i>no supervisado</i>	70
5.4.3.	Desventajas del enfoque <i>no supervisado</i>	70
5.5.	El algoritmo DBSCAN para la búsqueda de grupos en conjuntos de datos	71
5.5.1.	Parámetros del algoritmo DBSCAN	71
5.5.2.	Tipos de puntos en el algoritmo DBSCAN	72
5.5.3.	Pasos del funcionamiento del algoritmo DBSCAN	73
5.5.4.	Estimación de Parámetros	74
5.5.5.	Ventajas y desventajas del algoritmo DBSCAN	75

5.5.6. Desventajas:	76
6. Metodología para buscar subestructura en las simulaciones	78
6.1. Selección de la muestra de datos de la simulación que se va a analizar . . .	79
6.2. Modelado y Generación de Muestras Coniformes para Simulaciones Astrofísicas	80
6.3. Selección y Simulación del Cono de Luz en Datos Astrofísicos	82
6.4. Calcular la densidad numérica en el conjunto de datos	84
6.5. Optimización de Parámetros para DBSCAN	85
6.6. Probando el algoritmo DBSCAN en simulaciones tipo Vía Láctea	87
6.6.1. Experimento utilizando v_x, v_y, v_z + densidad en un plano grueso . .	87
6.6.2. Experimento utilizando v_r + v_{tan} + densidad en el plano grueso . .	89
6.7. Experimento con información cinemática: v_x, v_y, v_z	92
6.7.1. Búsqueda en una zona coniforme con $v_x + v_y + v_z + \text{densidad}$	94
6.8. Comparativa: Experimento $v_x + v_y + v_z + \rho$ y experimento $v_r + v_{tan} + \rho$. .	96
6.8.1. Exploración de las partes que componen a la señal $v_{rad} + v_{tan} + \rho$.	98
6.8.2. Experimentos con solo v_{rad} y con solo v_{tan}	99
7. Simulaciones cosmológicas de formación de galaxias	101
7.1. Introducción	101
7.2. Simulaciones cosmológicas de formación de galaxias	103
7.2.1. Componente colisional en simulaciones de N-cuerpos	103
7.3. Los modelos de la simulación cosmológica de alta resolución de formación y evolución de galaxias: GARROTXA	105
7.4. Modelo de N-cuerpos+hidrodinámica	106
7.4.1. La componente Hidrodinámica	106
7.4.2. Procesos físicos	107
7.4.3. Técnica de retroalimentación dinámica	107
7.4.4. Comparación con otros modelos disponibles	107
8. Catálogos sintéticos basados en las simulaciones de formación de galaxias incluyendo modelos de errores de los censos <i>Gaia</i> y DESI	111
8.1. Introducción	111
8.1.1. Antecedentes de catálogos sintéticos	113
8.2. Las observaciones sintéticas a partir de las simulaciones	115
8.3. Método empleado en los códigos que remuestran estrellas	116
8.4. Modelo de errores de Gaia	118
8.5. Mapeo de errores en paralaje en la simulación remuestrada: $ z < 300pc$ y $ z > 300pc$	119

8.6. Modelo de errores de DESI	122
8.7. Generador de nuevas partículas con distribución Gaussiana	123
8.8. Discusión y conclusiones	125
9. Discusión y Conclusiones	126
9.1. Resumen	126
10. Trabajo a futuro	132
10.1. Análisis de proyecciones del espacio fase con técnicas de análisis de imágenes para detectar subestructura en catálogos sintético	134
10.2. Exploración de la base ortonormal de “Wavelets” como estrategia de búsqueda de estructura.	135
A. Obteniendo la fórmula de Chandrasekhar	136
A.0.1. Perturbador en un medio homogéneo	138
A.1. Casos límites de la fórmula de Chandrasekhar	139
A.1.1. Dependencia de la masa	139
A.1.2. Dependencia de la velocidad en el caso de movimiento lento	140
A.1.3. Dependencia de la velocidad en el caso de movimiento rápido	140
B. Teoría de respuesta lineal	141
C. La función de pérdida de masa en el modelo semianalítico con fricción dinámica	143
D. Software para <i>Rockstar Halo Finder</i> para identificar subhalos	148
Bibliografía	150

Resumen

Recientemente las observaciones disponibles de la componente estelar de la Vía Láctea (VL) han aumentado en órdenes de magnitud gracias a esfuerzos observacionales como Gaia (Gaia Collaboration, 2016) y APOGEE (Majewski, 2017; Blanton, 2017) o GALAH (De Silva et al., 2015), entre otros, aunque las propiedades e historia del halo estelar de la VL siguen en estudio debido a la incompletitud de los catálogos observacionales y también a la dificultad de obtener las simulaciones con la resolución suficiente y que incluyan toda la física de formación y evolución estelar así como de la materia oscura que permitan una correcta interpretación de las observaciones. En la actualidad varios autores han avanzado en este estudio mediante el uso de modelos simplificados de la VL, tanto en la física de la materia oscura y la historia de ensamblaje, así como en la dinámica orbital de los objetos estudiados, ver por ejemplo Santistevan et al. (2023) en simulaciones o Antoja et al. (2020) con datos de *Gaia*.

La mayoría de los trabajos que han abordado el estudio del ensamblaje de la VL y la historia de caída de los satélites e intentado hacer una reconstrucción orbital de los satélites (para intentar poner cotas al potencial gravitacional), suelen ser simplistas al considerar que la fricción dinámica se comporta como el modelo de Chandrasekhar (CH), (el medio es homogéneo e infinito), y no suelen considerar los modos globales de perturbación (Colpi et al., 1999) que se ven reflejados en la respuesta en densidad del medio (Garavito-Camargo et al., 2019; Aguilar-Argüello et al., 2022). Tampoco suelen considerar el hecho de que puede haber más de un satélite perturbando al medio, ni los efectos de la pérdida de masa en la historia de caída (Fujii et al., 2006; Miller et al., 2020), así como efectos debido a la forma del halo (Garavito-Camargo et al., 2021).

En este trabajo se aborda por primera vez de manera sistemática las consecuencias que tienen los efectos colectivos en la historia de caída de los satélites, considerando sistemas con varios perturbadores o satélites cayendo hacia el potencial gravitacional e interactuando con el medio y entre ellos mismos así como con los modos perturbativos que se generan, consideramos también la pérdida de masa que da lugar a la autofricción dinámica .

Mostramos que la historia de acreción de una galaxia satélite concreta, en un caso realista de acreción de múltiples satélites dentro de un halo central, difiere considerablemente del caso de estudio habitual en donde un único satélite cae hacia el centro del halo principal. Ve-

mos que cuando los satélites son compactos los tiempos de caída aumentan en comparación con el caso de un solo satélite. Algo que no ha sido reportado hasta el momento es el efecto de la autofricción dinámica en los satélites extendidos en el caso de múltiples satélites.

Abordamos el caso de los satélites presentes en la Vast POLar Structure (VPOS) (Pawlowski and Kroupa, 2014) y cuantificamos las diferencias en los apocentros y excentricidades de los satélites respecto del caso aislado, mostramos que las diferencias pueden ser mayores al 30 % para los apocentros y más del 40 % en las excentricidades en algunos casos.

Concluimos que el no considerar los efectos colectivos puede llevar a malinterpretaciones en la historia de caída de los satélites, errores en las mediciones de los tiempos de caída y reconstrucción orbital de las galaxias en la VL, así como la manera en que se intentan restringir varias propiedades del halo oscuro de la galaxia y sus satélites. La forma del halo de la VL y su historia dinámica junto con la de sus satélites también tendría que cambiar como consecuencia de estos resultados.

Los resultados más importantes en el contexto de la evolución orbital y su reconstrucción han sido publicados en la revista *Astronomy & Astrophysics*, con el título “Concurrent infall of satellites: Collective effects changing the overall picture”.

La segunda parte del trabajo de tesis se enfoca en los restos que los procesos físicos discutidos arriba dejan en el halo estelar galáctico. Se busca proporcionar predicciones en lo posible en el lenguaje de los datos observacionales de alta calidad provistos por diferentes catastros como *Gaia*, The Radial Velocity Experiment (RAVE) (Steinmetz et al., 2020), Apache Point Observatory Galactic Evolution Experiment (APOGEE), GALAH y los futuros Dark Energy Spectroscopic Instrument (DESI) (Levi and Allen, 2019) y Large-aperture Synoptic Survey Telescope (LSST) (Ivezić and Kahn, 2019), e intentar caracterizar al mayor detalle posible la subestructura del halo estelar, tarea que dista de estar completa y ser fácil.

Con este fin hemos desarrollado un código, el cual descompone partículas estelares de acuerdo a una función inicial de masa dentro de simulaciones de formación de galaxias, esto permite remuestrear y obtener estadísticas comparables con las observaciones. Se incluyen además los modelos de errores y de incompletez observacionales de *Gaia* y del censo estelar DESI, el cual complementará a *Gaia* con decenas de millones de velocidades radiales. Con ayuda de los nuevos datos de alta calidad y las nuevas simulaciones será posible una mejor caracterización de la cinemática y propiedades físicas de las estructuras del halo estelar (colas de marea y sus estructuras internas, galaxias satélites y su distribución global y cúmulos globulares), lo que a su vez nos permitirá estudiar su origen, así como las propiedades del halo de materia oscura que las contiene. Para analizar los datos simulados y observacionales planteamos el utilizar potentes herramientas de Inteligencia artificial (IA) *no supervisadas* como Density-based spatial clustering of applications with noise (DBSCAN) (Ester et al., 1996) principalmente, pero consideramos utilizar *Mapas autoorganizados* (Kohonen, 1982) en un futuro, entre otras herramientas, las cuales permiten reducir el sesgo y aprender di-

rectamente de los datos. Con este fin, en el capítulo 6 desarrollamos una metodología para búsqueda de subestructura en las simulaciones con las diferentes variables que componen el espacio 6D y en diferentes configuraciones de los datos, como cortes en un plano o en la zona de la línea de visión de las observaciones. Esta metodología se empleará sobre los catálogos sintéticos más adelante, que se detallan en el capítulo 8.

Colaboraciones en otros artículos como co-autor:

- Aguilar-Argüello et al. (2022) “Boltzmann-Poisson-like approach to simulating the galactic halo response to satellite accretion. Dependence on the halo density profile”.
- Aguilar-Argüello et al. (2020) “Assessing the hierarchical Hamiltonian Splitting integrator for collisionless N-body simulations”.

Capítulo 1

Introducción

1.1. El modelo de materia oscura fría y energía oscura

El modelo cosmológico más exitoso en explicar la formación de estructura a gran escala en el Universo y la formación y evolución de las galaxias y que incorpora las observaciones de galaxias y cúmulos es conocido como el modelo Lambda Cold Dark Matter (**ΛCDM**) (White and Rees, 1978; Davis et al., 1985; Navarro et al., 1996; Planck Collaboration, 2014).

Este modelo es también llamado modelo cosmológico estándar e implica un período de expansión exponencial llamado inflación o período inflacionario (Silk and Turner, 1987) el cual precede a la teoría conocida como *Big Bang* (Lemaître, 1927). El período inflacionario explica exitosamente el crecimiento de las inhomogeneidades en densidad que se observan en la radiación cósmica de fondo Cosmic Microwave Background (**CMB**), su homogeneidad e isotropía y su geometría plana, entre otros.

Las inhomogeneidades primordiales de las cuales observamos las huellas en el **CMB** crecen por inestabilidad gravitacional dentro de un universo en expansión, lo que da lugar a toda la estructura que observamos al día de hoy. Esta estructura a gran escala incluye cúmulos y supercúmulos de galaxias, pero también regiones vacías conocidas como “huecos”. El esqueleto principal de estas estructuras está compuesto en su mayor parte por una materia que solo interactúa gravitatoriamente con el resto del universo, la materia oscura. Las galaxias que observamos habitan dentro de estructuras compuestas en buena parte por esta materia, los halos de materia oscura (Navarro et al., 1996). La naturaleza de esta materia sigue siendo un misterio, aunque es cierto que su existencia es imprescindible para explicar varios fenómenos observados entre los cuales la dinámica interna de las galaxias, como por ejemplo, las curvas de rotación y las lentes gravitacionales (Walsh et al., 1979). Por otro lado, las galaxias contenidas en estos halos de materia oscura son en sí el resultado de procesos mucho más complejos que incluyen el enfriamiento de los bariones, la formación estelar, y la evolución de las estrellas hasta su final, en forma de supernovas o nebulosas planetarias (Turner, 1993).

En buena parte de los modelos teóricos y numéricos actuales se usa la materia oscura fría como ingrediente principal para la formación de galaxias. La materia oscura conocida como fría es la que mejor se ajusta con las observaciones, en particular, es esta materia la única que permite la formación de galaxias satélite de baja masa como las que se observan a bajo brillo superficial (White and Rees, 1978). No obstante, la falta de observaciones de objetos a muy bajo brillo superficial dificulta la distinción entre la materia oscura fría y la materia oscura tibia de baja energía (Colín et al., 2001; Weinberg et al., 2015). La prueba definitiva para el tipo de materia oscura que existe en el universo requerirá de nuevas y más profundas observaciones y mejores modelos teóricos/simulaciones, que permitan contrastar las predicciones de estas. El estudio de la caída y destrucción de galaxias satélites nos permitirá también hacer predicciones sobre la existencia, naturaleza y estructura de halos de materia oscura en torno a las galaxias y en particular en torno a la VL.

1.2. La Vía Láctea, su halo oscuro y su halo estelar

El halo oscuro de la Vía Láctea es una de las estructuras que se formaron dentro del escenario de formación Λ CDM (de Rossi et al., 2009; Strigari et al., 2010). El proceso jerárquico de la formación y crecimiento de los halos oscuros, se refleja también en la estructura de la componente estelar del halo que se conoce como halo estelar. Esto es debido a que la componente estelar se puede considerar no-colisional, a diferencia del gas que pierde energía en choques y cae al centro de los halos durante las interacciones o fusiones. Bajo el modelo Λ CDM, cuando se forman los halos oscuros, el material bariónico es atraído hacia el centro de los mismos, cuando se fusionan posteriormente los halos oscuros y las galaxias en ellos, las estrellas siguen procesos similares formando el halo estelar, mientras que el gas cae hacia regiones internas donde puede producirse nueva formación estelar y generar así un disco estelar. Dentro del escenario de formación del halo estelar, este puede dividirse en dos componentes: aquel que se formó dentro de la galaxia, la llamada componente “In-situ” (Abadi et al., 2003a; Zolotov et al., 2009; Font et al., 2011; Brook et al., 2012, e.g.) y otro que se estableció durante el proceso de formación jerárquica, la llamada componente “Ex-situ” (ver Oser et al., 2010, para más detalles).

Esta imagen está apoyada por las propiedades químicas de las estrellas (Zolotov et al., 2010; Pillepich et al., 2015), ver figura 1.1. En particular, la formación “Ex-situ” queda reflejada en la subestructura del halo estelar, tal como la presencia de galaxias satélites, las colas de marea, las corrientes estelares y otras subestructuras de bajo brillo superficial, observadas tanto en nuestra galaxia como en galaxias externas. La posición, la distribución de velocidades y el número de estas subestructuras es sensible a los procesos dinámicos que determinaron la formación del halo en su conjunto, ver por ejemplo Antoja et al. (2018),

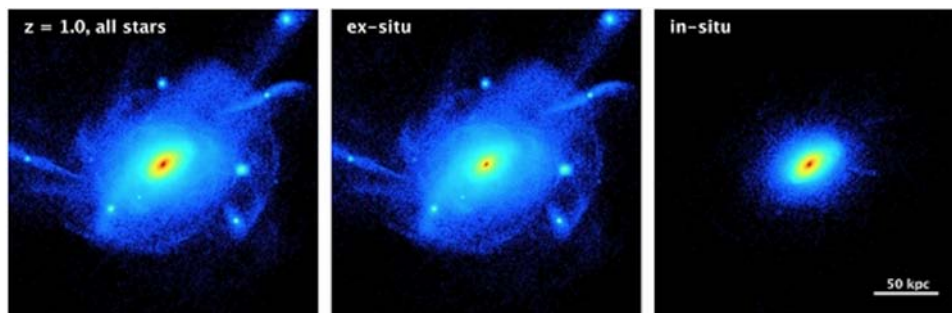


Figura 1.1: Imagen tomada de Pillepich et al. (2015). Proyección del campo de densidad estelar de la simulación ERIS. Panel izquierdo: Todas las partículas estelares. Panel medio: Partículas estelares formadas “Ex-situ”. Panel derecho: Partículas estelares formadas “In-situ”.

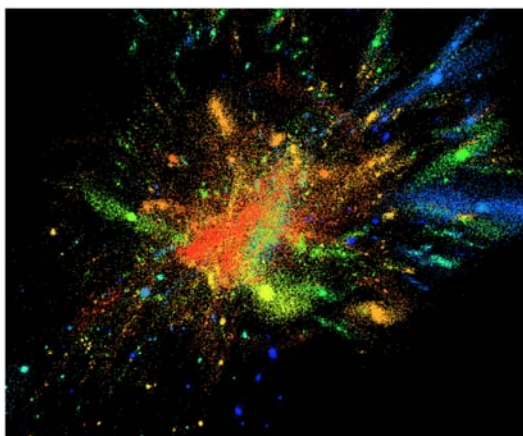


Figura 1.2: Imagen de una simulación mostrando el origen de las poblaciones estelares de distintas galaxias fusionándose (O. Valenzuela comunicación privada). En esta imagen podemos ver en distintos colores el origen de las partículas estelares contenidas en el halo estelar de una galaxia tipo Vía Láctea.

donde muestran que existe una huella dinámica de la evolución de la Galaxia, que puede ser vista, a través de las estrellas del disco en el espacio fase, o Ramos (2020) que estudia la subestructura del espacio fase en los límites de lo que ve *Gaia*, en el disco, pero también en el halo estelar a través del *Stream* de Sagitario. Es por esto que consideramos que estas observables podrían usarse para conocer mejor las propiedades de la materia oscura. Así pues, es evidente que existe mucha información de la historia de formación y evolución de la galaxia y sus satélites, contenida en la estructura y propiedades del espacio fase en el halo estelar (Helmi and White, 1999a, e.g), así como también en su química y distribución de edades (Das and Binney, 2016; Horta et al., 2022, ver, por ejemplo).

En este contexto, caracterizar al mayor detalle posible la subestructura del halo estelar es imprescindible para entender el origen y evolución de las galaxias. Esta es, no obstante, una tarea difícil, y actualmente incompleta, ya que los trabajos que se han llevado a cabo para caracterizar e interpretar esta subestructura se han obtenido mediante hipótesis sobresimpli-

ficadoras sobre el potencial galáctico o de la naturaleza de las galaxias satélite (Helmi and White, 1999b; McMillan and Binney, 2008, e.g.). Las grandes misiones espaciales planeadas para el futuro próximo van a permitir mejorar este tipo de análisis en los próximos años. No obstante, analizar estas nuevas observaciones requerirá también de nueva metodología, así como de incorporar nuevos desarrollos teóricos, como por ejemplo, la mejora de la descripción del efecto de la fricción dinámica más allá de la formulación tradicional (descrita más adelante), la descripción analítica de los efectos disruptivos que generan colas de marea, de la estructura de las galaxias satélite y la naturaleza de la materia oscura.

En este contexto también se vuelve un problema relevante el cuantificar cómo se pueden acotar las propiedades geométricas y dinámicas del halo oscuro de la galaxia y su subestructura (satélites y subhalos oscuros) mediante las observaciones de la estructura fina del halo estelar, relajando hipótesis como la esfericidad o una distribución suave. En particular, adquirirán gran importancia técnicas para caracterizar subestructuras como las que ya se han venido usando para analizar catastrófos astronómicos modernos basados en métodos tradicionales como: variables de acción-ángulo (Helmi and White, 1999b), transformadas ondículas o *wavelets* (Antoja et al., 2015) y métodos bayesianos (Mackereth et al., 2018), por mencionar algunos.

En años recientes han surgido algunos trabajos que empiezan a utilizar herramientas o algoritmos de IA en el contexto astrofísico, como por ejemplo, la búsqueda de subestructura en el halo estelar directamente sobre los datos (Myeong et al., 2018a; Hasselquist et al., 2017; Ibata et al., 2019, e.g.), también han surgido trabajos que utilizan el enfoque *no supervisado*, ver por ejemplo Castro-Ginard et al. (2018), donde utilizan un algoritmo de agrupamiento *no supervisado*, en el contexto de la búsqueda de cúmulos abiertos en la VL, utilizando el catálogo *Gaia*. Más recientemente Ostdiek et al. (2020) y Necib et al. (2020) emplean el enfoque *no supervisado* en la búsqueda de subestructura en el halo estelar, sobre datos reales, pero lo hacen sin realizar una calibración con experimentos controlados como lo son las simulaciones.

Utilizar este tipo de algoritmos en el enfoque *no supervisado* resulta altamente beneficioso. A diferencia de los métodos *supervisados*, estos algoritmos no dependen de una muestra de entrenamiento, lo que permite evitar comportamientos indeseables como el *sobreajuste*, algo que a menudo no se discute ampliamente en la literatura. Al combinar estos algoritmos con simulaciones, se obtiene un método poderoso para calibrar su desempeño y ajustar sus parámetros. Asimismo, esta combinación permite evaluar la confiabilidad de las estructuras identificadas tanto en las simulaciones como en datos reales, algo que aún no ha sido abordado hasta la fecha.

1.3. Motivación de la tesis

Aunque se han hecho muchos avances en el campo de la búsqueda de estructura en el halo estelar gracias a la gran cantidad de datos disponibles (Peñarrubia et al., 2016; Helmi et al., 2018), se desconoce aún parte del proceso de ensamblaje de las estructuras tanto estelar y oscura de la VL. Los datos proporcionados por esfuerzos observacionales como APOGEE (Majewski, 2017; Blanton, 2017), *Gaia* (Gaia Collaboration, 2016), LAMOST (Zhao et al., 2012), RAVE (Steinmetz et al., 2020) y Galah (Buder et al., 2018), entre otros, han permitido descubrir que lo que parecía ser una historia suave para el ensamblaje de la VL, no lo fué realmente, y que existen pruebas de múltiples eventos de acreción de gas y satélites recientes (Ramos et al., 2022; Vasiliev, 2023). Las características de estos eventos, por ejemplo, la masa de los objetos acretados y el momento exacto en el que ocurrieron, tiene aún muchas incertidumbres y su determinación está sujeta a hipótesis físicas o geométricas. Entre otras, algunas de las hipótesis más importantes son la de el efecto de la fricción dinámica en la caída de los satélites hacia regiones internas del halo de materia oscura, o la misma forma de dicho halo. Para entender mejor estos procesos, una buena herramienta es la caracterización de la cantidad y propiedades de las estructuras estelares remanentes (colas de marea, satélites, inhomogeneidades). Éstos observables tienen la potencialidad de permitir el discernir entre escenarios y de proveer de constricciones al ensamblaje de la galaxia y subyacentemente al escenario del modelo Λ CDM. Aunque existen un número vasto de estudios, por ejemplo Helmi and White (1999b); Bell et al. (2008), buena parte de las técnicas de búsqueda de estructura en el halo estelar asumen hipótesis geométricas u homogeneidad del halo estelar lo que potencialmente puede generar sesgos que dificulten la interpretación de los resultados. Consideramos por eso que es necesario dar pasos hacia un mayor realismo y explorar métodos de búsqueda que en lo posible relajen tales hipótesis. El uso de métodos de IA y en específico los algoritmos “no supervisados” combinados con los *supervisados* son un complemento con gran potencial para este fin. Una de las razones del interés que ha surgido en aplicar este tipo de algoritmos, es que a diferencia de los métodos comúnmente utilizados para buscar estructura, no se hace ninguna suposición sobre cuál es el potencial o estructura en densidad de fondo de la Galaxia.

Entre los métodos que aprenden de los datos automáticamente sin hacer ninguna suposición sobre el modelo se encuentra la llamada IA tanto Machine Learning (ML) (desde el enfoque *no supervisado*, es decir, que no requiere una muestra de entrenamiento para realizar su tarea, como son los algoritmos K-Means (K-Means), DBSCAN etc., así como algoritmos de Deep Learning (DL) (Redes Neuronales Artificiales que aprenden de los datos de manera análoga a neuronas en el cerebro humano).

En esta tesis doctoral se han empezado a desarrollar herramientas basadas en técnicas de IA que se pueden usar para analizar los catálogos de estrellas obtenidos de “surveys”

como *APOGEE*, *Gaia* y en el futuro *DESI* (Levi and Allen, 2019) y *LSST* (The LSST Dark Energy Science Collaboration et al., 2018). Utilizar este tipo de algoritmos en el enfoque *no supervisado*, junto con las simulaciones representará un método poderoso en cuanto a calibrar el desempeño de los algoritmos y sus parámetros, así como restringir la veracidad de la posible estructura encontrada, tanto en las simulaciones como en datos reales, lo cual es algo que no se ha realizado hasta la fecha de manera precisa. Además, los “surveys” como *DESI* (en proceso de desarrollar bases de datos y a cuya colaboración tenemos acceso) y *Gaia* (que ya cuenta con tres fases de liberación de datos), representan una oportunidad de poner a prueba diferentes herramientas, no solo de *IA* sino numéricas como las simulaciones cosmológicas (Roca-Fàbrega et al., 2016), además *DESI* proporcionará una gran cantidad de datos de velocidades radiales de estrellas en el halo estelar complementando así a *Gaia* y permitiendo ampliar la búsqueda de subestructura en otros espacios como velocidades y química, entre otros.

1.4. Objetivos

Objetivo general: Llevar a un mejor nivel de precisión los siguientes puntos a estudiar en el halo estelar de la *VL*,

- Estudiar cuantitativamente el efecto de la caída simultánea de varios perturbadores en el halo galactico sobre la respuesta física y su influencia en la caída de sistemas estelares satélites, así como sus consecuencias para la reconstrucción orbital.
- Estudiar estrategias para detectar e interpretar la estructura estelar del halo galáctico.
- Implementar una herramienta computacional para mejorar el grado de realismo de las predicciones teóricas acerca de las propiedades del halo estelar Galáctico, incluyendo un contexto cosmológico.

Objetivos Específicos:

- Revisar la eficiencia de la fricción dinámica a la luz de considerar propiedades más realistas para los halos de materia oscura.
- Buscar cuantificar la capacidad de los métodos para detectar la estructura fina del halo estelar, considerando modelos de errores y simulaciones de galaxias tipo *VL*, que incluyen bariones y propiedades observables para sus poblaciones estelares. Evaluar la posible detección de colas de marea alrededor de las galaxias enanas esferoidales a la luz de los “surveys” *Gaia*, *APOGEE* y *DESI*.

Capítulo 2

Revisión de la fricción dinámica: Efectos colectivos

En astrofísica la fricción dinámica es un proceso basado en la interacción gravitatoria que hace que los sistemas estelares como galaxias satélites o cúmulos globulares decaigan hacia el centro de sistemas mayores al intercambiar energía orbital y momento angular con el medio en el que se encuentran inmersos, ver por ejemplo el capítulo 8 de [Binney and Tremaine \(2008\)](#).

Este intercambio de energía ocurre cuando el satélite o perturbador interactúa gravitacionalmente con un medio con densidades menores que la suya, de manera que esta interacción produce una redistribución de la masa, que acaba generando una fuerza de arrastre neta sobre el perturbador. Esta fuerza de arrastre neta es la resultante de la fuerza debida a las partículas individuales del medio y la de las estructuras de densidad creadas detrás del perturbador como respuesta a su paso por el medio.

[Chandrasekhar \(1943\)](#) estudió por primera vez la fricción dinámica de forma analítica, asumiendo que existía un medio homogéneo e infinito de materia oscura, logrando así capturar los efectos físicos dominantes. No obstante, su modelo era muy simplificado y no era, en general, aplicable a sistemas galácticos reales. Ver apéndice [A](#) para una descripción analítica completa de la cuantificación de la fricción dinámica, basada en el desarrollo de [Binney and Tremaine \(2008\)](#).

Debemos tomar en cuenta que en los sistemas reales hay siempre más de un perturbador dentro del halo de la galaxia principal, además los perturbadores no son rígidos sino que pierden masa y la misma respuesta generada por los perturbadores en el medio de la galaxia principal puede interactuar con el mismo perturbador. Diversos trabajos como [Lynden-Bell and Kalnajs \(1972\)](#), [Tremaine \(1976\)](#) y [Tremaine and Weinberg \(1984\)](#) también han demostrado limitaciones de la aproximación de [CH](#). A continuación se detallan algunas de ellas:

- En la aproximación de CH la fricción dinámica solo se da sobre perturbadores que se encuentran dentro del medio de partículas. Pero trabajos como el de Tremaine and Weinberg (1984) han demostrado que la fricción dinámica se da para objetos fuera de la distribución de partículas.
- La fricción dinámica en el enfoque de CH solo estudia el caso de un perturbador que se mueve en una trayectoria recta en un medio de partículas, pero no es suficiente para explicar como es la fricción dinámica en sistemas que tienen otro tipo de órbitas.
- Los sistemas reales pueden tener más de un perturbador (la galaxia enana Fornax con cinco cúmulos globulares por ejemplo).
- La respuesta o respuestas que se generan cuando en un sistema másivo hay más de un perturbador (e.g varios cúmulos globulares) juegan un papel importante en la fricción dinámica que no puede ser estudiado con la aproximación de CH.
- No se toma en cuenta el efecto que la autofricción dinámica puede tener sobre la historia de caída de los satélites.
- No se ha estudiado a fondo el comportamiento de la fricción dinámica en sistemas triaxiales, aunque trabajos como Capuzzo-Dolcetta and Vicari (2005) si estudian el caso de sistemas triaxiales, siguen el enfoque de la fórmula de CH.

Así pues, más allá del enfoque de CH durante las décadas de 1980 y 1990 surgieron otros trabajos, por ejemplo, los trabajos basados en la “Teoría de la respuesta Lineal” (Tremaine and Weinberg, 1984; Nelson and Tremaine, 1999; Colpi et al., 1999, e.g.) y la sección B. En estos trabajos se sugiere que la fricción dinámica se puede descomponer en diversos modos de perturbación globales (subdensidades y sobredensidades), la fricción dinámica local o estela de CH, además del corrimiento del baricentro, ver por ejemplo Colpi and Pallavicini (1998) y Colpi et al. (1999), en donde se considera que la fricción dinámica consiste de varios modos perturbativos: los modos globales y los modos que son menores que la parte local, (es decir, el modo principal) así como el corrimiento del baricentro.

Otros autores han propuesto nuevas estrategias que puedan tomar en cuenta otros factores involucrados en el problema de la fricción dinámica, entre ellos, los más relevantes son aquellos que consideran un sistema anfitrión más realista, es decir, con un tamaño finito y un perfil de densidad que depende del radio (Binney, 1977; Boldrini et al., 2018, e.g.) y aquellos que consideran sistemas más realistas asemejando a los modelos de N-cuerpos, que pueden

ser descritos por un simple logaritmo Coulombiano con dependencia espacial y radial, ver Hashimoto et al. (2003). Todos estos nuevos modelos han permitido a los investigadores recuperar y estudiar la ralentización en la eficiencia de la fricción dinámica, incrementando el realismo del sistema. Más recientemente trabajos como Tamfal et al. (2021) analizan los modos perturbativos en la densidad, con simulaciones con mejor resolución con 10^8 partículas, para intentar discernir de manera visual estos modos perturbativos en densidad, pero solo considerando un satélite (vivo).

También se debe considerar que los cambios internos en el perturbador pueden afectar el proceso de caída, y esto debe tomarse en cuenta cuando se estudia la acreción de satélites por galaxias más grande. De manera general la fricción dinámica engloba varios efectos físicos más allá de la estela de CH, al conjunto de estos efectos incluyendo el modo principal o estela de CH los llamamos efectos colectivos como se explica a continuación.

2.1. Los efectos colectivos

Todos los estudios de modelos simplificados de satélites o cúmulos globulares cayendo hacia el centro de sus galaxias anfitrionas mencionados antes, pueden ser usados para entender mejor la formación y evolución, propiedades de galaxias locales incluyendo la Vía Láctea. Lejos de los modelos idealizados presentados arriba, donde un solo objeto se hunde hacia el centro de su halo anfitrión, observaciones y simulaciones muestran que los halos también sufren una continua y múltiple caída de satélites, los efectos colectivos en la caída de múltiples satélites es un proceso que requiere mayor estudio.

En este trabajo abordamos por primera vez de modo sistemático los efectos colectivos y sus consecuencias, mostrando que la órbita de cada satélite individualmente, es sensible a la configuración relativa de todos los satélites dentro del halo principal, y también de la estructura interna.

En particular, prestamos atención a una configuración específica similar a la recientemente descubierta VPOS, una configuración en la que los satélites de la VL se encuentran y es más similar a un disco plano (Lynden-Bell, 1976; Kunkel and Demers, 1976) y posteriormente estudiada por Pawlowski and Kroupa (2014). Los efectos colectivos juegan un papel importante en la historia de caída de los satélites en la Vía Láctea. Los modelos que incluyen la caída de las Nubes de Magallanes y Sagitario verán limitado su poder predictivo si no los incluimos. Recientemente otros trabajos como D'Souza and Bell (2022) refuerzan la idea de que la historia orbital de los satélites puede ser afectada por las incertidumbres relacionadas con el halo o los subhalos más masivos, el escenario donde tomamos en cuenta los efectos colectivos refuerza esta última idea. También hay que tomar en cuenta que estas incertidumbres en la eficiencia de la fricción dinámica pueden verse afectadas por la riqueza

de subestructura que existe en los halos de las galaxias. En un sistema finito, las galaxias satélites interactúan con la respuesta del medio, a su vez la respuesta del medio generada por estos sistemas pueden interactuar entre ellas y las estelas generadas por la fricción dinámica. Visualizar esta respuesta no es una tarea fácil, recientemente enfoques como el utilizado en Aguilar-Argüello et al. (2022) permiten apreciar la magnitud de la respuesta al considerar que estos sistemas se mueven en un medio continuo, ver figura 2.1, también ver 2.2, en el caso de varios perturbadores esto cambia significativamente y tiene consecuencias para trabajos con modelos que incluyen la caída de satélites (e.g Sagitario y las nubes de Magallanes LMC/SMC), ver Vasiliev (2023) y Garavito-Camargo et al. (2021).

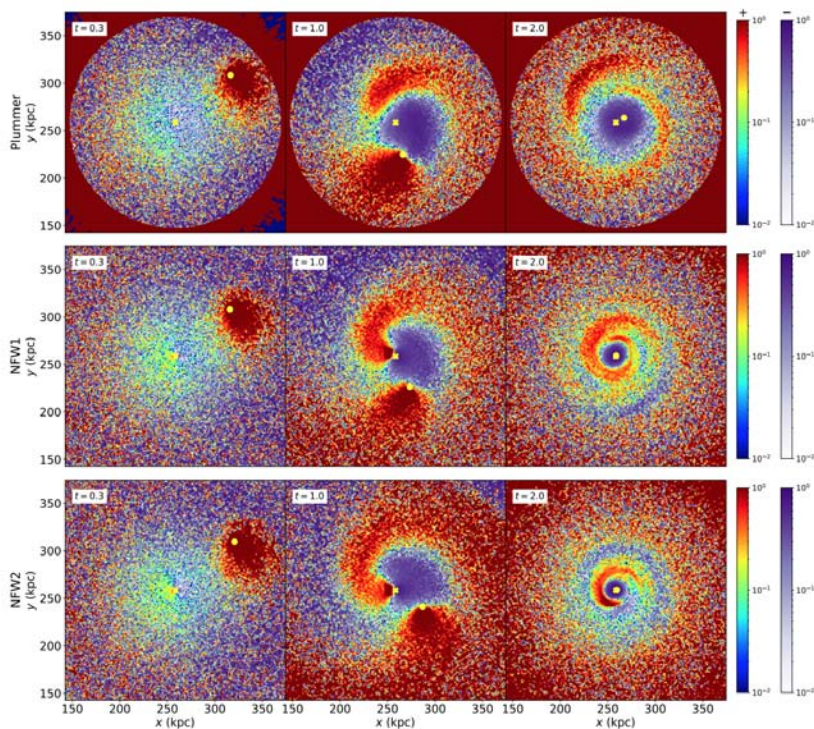


Figura 2.1: La estela de CH, y la respuesta global son claramente visibles en los tres paneles. En color morado valores para la sobredensidad negativa y rojo para la positiva. La fila superior corresponde a un modelo de Plummer (Plummer, 1911), panel de en medio e inferior corresponden a diferentes modelos Navarro, Frenk and White (1996) (NFW). Imagen: Aguilar-Argüello et al. (2022)

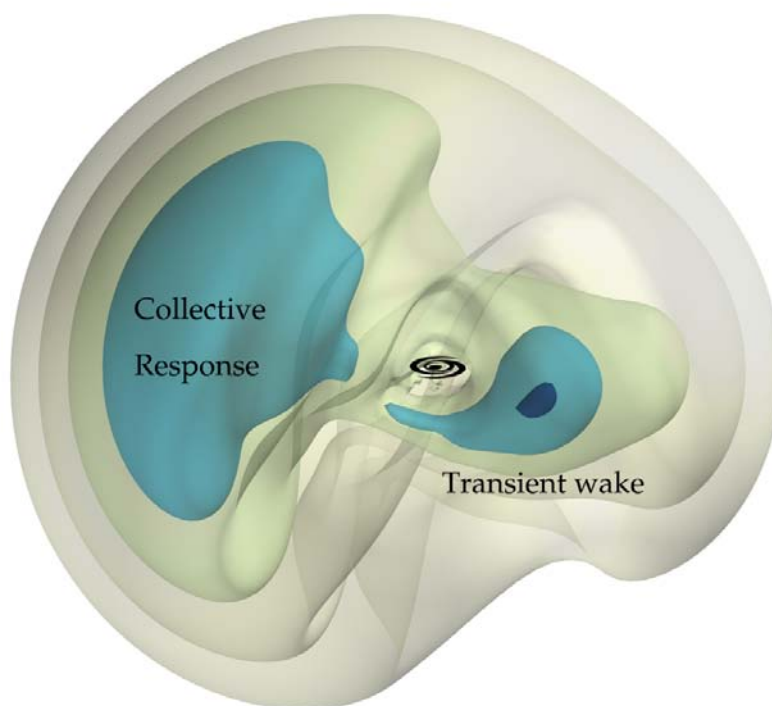


Figura 2.2: Contraste de densidad en 3D de la respuesta del halo de la Vía Láctea en presencia de la LMC. Imagen: Garavito-Camargo et al. (2021)

2.1.1. Efecto en las órbitas con varios perturbadores

Unas de las “situaciones” que van más allá del enfoque de CH para la fricción dinámica, es lo que pasa cuando hay más de un perturbador, ya que el enfoque clásico de CH no sería aplicable a esta situación, lo cual podemos ver en la dependencia en la densidad de este enfoque, si hay más de un perturbador CH no podría capturar modos perturbativos, que además, quedan correlacionados entre cada satélite.

Uno de los primeros casos en los que se reportó ante la comunidad que había consecuencias importantes no estudiadas de la presencia de múltiples perturbadores fue en Inoue (2009), en ese trabajo, como se muestra en la figura 2.3, se mostró que las órbitas pueden tener cambios abruptos cuando hay varios perturbadores, aunque su análisis estaba más enfocado en el efecto que podía tener el perfil de densidad y no en los efectos colectivos propiamente. Otro ejemplo se puede ver en Goerdt et al. (2006), ver figura 2.4, donde se puede ver que en el caso de múltiples satélites las órbitas son distintas a las del caso de un perturbador aislado, si bien es cierto que el análisis no es profundo ya que este trabajo no está enfocado en estudiar los efectos de múltiples perturbadores en las órbitas, pero cabe resaltar que las consecuencias de los efectos colectivos en el caso de múltiples satélites están a la vista en la literatura, pero no se han estudiado de manera sistemática.

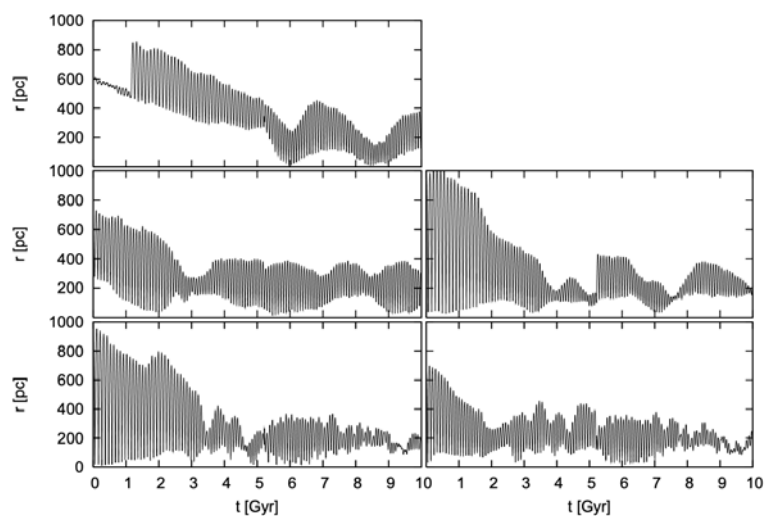


Figura 2.3: Órbita de cada uno de los 5 cúmulos globulares de Inoue (2009), cada panel muestra las variaciones del radio orbital de cada uno pero en una situación de caída simultánea. Hay que observar los cambios abruptos en las órbitas en este caso, en el que hay varios perturbadores cayendo al mismo tiempo. Imagen tomada de Inoue (2009)

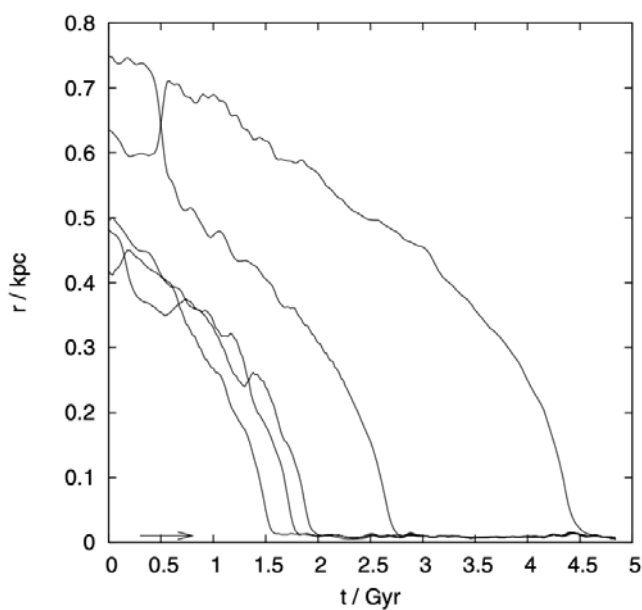


Figura 2.4: Imagen tomada de Goerdet et al. (2006): Distancia radial de cinco cúmulos globulares desde el centro de su halo como función del tiempo para una galaxia tipo fornax en un halo tipo “cusp”.

2.1.2. Cambios en la estructura interna y su efecto en la fricción dinámica neta

En situaciones realistas los perturbadores suelen ser otras galaxias (subhalos de materia oscura) y poseen una extensión, por lo tanto, se verán afectadas por las fuerzas de marea debido al halo anfitrión, perdiendo así parte de su masa. Este proceso suele dotar de estructura al satélite, como colas de marea, brazos espirales y cambios en la parte interna de los satélites (Mayer et al., 2001; Semiczuk et al., 2016). Se ha visto que cuando se toman en cuenta la distribución espacial y de velocidades del satélite, la tasa de intercambio de energía entre el satélite y el medio puede disminuir (Dominguez-Tenreiro and Gomez-Flechoso, 1998). Por ejemplo, Cowsik et al. (2009) consideran la dinámica interna en el caso de Fornax y sus cúmulos globulares, al considerar que las estrellas están inmersas en halos de materia oscura y al tomar en cuenta un modelo que considera su distribución de espacio fase, hay una reducción en el efecto de la fricción dinámica. De manera global parte de la masa perdida por el satélite, pasa a formar parte del medio. Esta masa perdida por el satélite a su vez puede tener una influencia en el efecto neto de la fricción dinámica (Fujii et al., 2006; Fellhauer and Lin, 2007; van den Bosch and Ogiya, 2018, e.g.), lo que se ha denominado auto-fricción dinámica, este efecto puede ser del orden del 15 % del efecto de la fricción dinámica como ha sido señalado por Miller et al. (2020). En este trabajo postulamos que en la realidad no es un solo sistema aislado el que está cayendo, sino que múltiples satélites caen de manera simultánea, cada uno perdiendo masa y contribuyendo al efecto de la auto-fricción dinámica, ver panel superior derecho de la figura 4.5.

2.2. La vasta estructura polar

La *Vasta Estructura Polar* o **VPOS**, fue descrita por primera vez por los trabajos de Lynden-Bell (1976) y Kunkel and Demers (1976), es una estructura en forma de plano en el que las galaxias satélites y algunos cúmulos globulares, tienen sus polos orbitales alineados alrededor de la galaxia central. Es decir, las galaxias satélites se agrupan en un plano delgado y esta estructura queda trazada por la *corriente* Magallánica, ver 2.5 (Pawlowski, 2021).

Recientemente, gracias a la mejora en las capacidades de observación de galaxias satélites en galaxias externas, muchos autores han analizado estructuras como la **VPOS**, que también se han descubierto entorno a otras galaxias como *M31* (Ibata et al., 2013) y *Centaurus A* (Müller et al., 2018). Para el caso de los satélites en la **VL**, un ejemplo de estos estudios, es el recientemente presentado por Pawlowski and Kroupa (2020). En este trabajo los autores sugieren que la **VPOS** podría ser el resultado de interacciones pasadas con la **VL**, las mareas gravitacionales y las interacciones con el halo de materia oscura de la galaxia. Estas conclusiones las obtienen estudiando la correlación de los polos orbitales de las 11 galaxias

satélites clásicas de la VL, comparando resultados de movimientos propios anteriores con los datos independientes del catálogo *Gaia* DR2.

Como parte de este trabajo decidimos enfocar nuestro análisis al caso de los satélites en la VPOS, y analizamos como cambian las historias de caída de los satélites con múltiples perturbadores, realizando un ensamble de simulaciones que toman como condiciones iniciales las posiciones y velocidades de los satélites de la VPOS, obtenidas con datos de *Gaia* DR2 y sus respectivas incertidumbres (Riley et al., 2019; Garrow et al., 2020). Cuantificamos las diferencias en los apocentros y excentricidades de las órbitas de cada experimento comparando con respecto al caso de un satélite cayendo de manera aislada para cada caso. Las diferencias pueden llegar a ser mayores al 40-50% para estos parámetros, por lo que, si uno quiere hacer una reconstrucción orbital se deben de tomar en cuenta los efectos aquí enunciados.

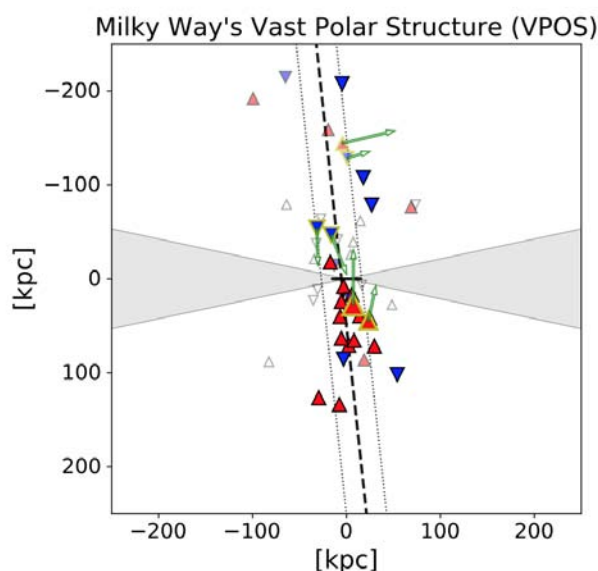


Figura 2.5: Imagen tomada de (Pawlowski, 2021). Galaxias enanas alrededor de la VL. La VL vista de canto como una línea negra en el centro, áreas grises indican una region de $\pm 12^\circ$ alrededor del plano del disco. La vista está orientada a la VPOS. Los triángulos blancos marcan galaxias que probablemente no pertenecen a la VPOS.

Capítulo 3

Simulaciones de N-cuerpos no-colisionales

3.1. Las simulaciones de N-cuerpos

Las simulaciones de N-cuerpos son herramientas computacionales utilizadas en física y astrofísica para simular la dinámica gravitacional de sistemas con un gran número de partículas, como lo son las galaxias, con N denotando el número de partículas en el sistema.

Aunque existen varias técnicas para modelar o simular sistemas de partículas como son las galaxias, por ejemplo, las simulaciones realizadas con partículas prueba que permiten explorar diversos escenarios y condiciones iniciales para sistemas galácticos, de manera simplificada (aunque no autoconsistentes y no consideran la autogravedad), las simulaciones de N-cuerpos son experimentos ideales para simular galaxias, ya que permiten modelar la evolución dinámica de las galaxias de manera autoconsistente, capturando la naturaleza no lineal de estos sistemas, ver por ejemplo la síntesis de técnicas hecha por [Sellwood \(1987\)](#)

En el contexto de la dinámica galáctica, en las simulaciones de N-cuerpos, cada partícula representa una entidad física, que podría ser una estrella, un cúmulo de estrellas, o incluso una porción de materia oscura.

El esquema más simple de modelado de galaxias como sistemas de N-cuerpos, consiste en primero definir las posiciones y velocidades iniciales de todas las partículas, (lo ideal es que estas condiciones se basen en observaciones o en modelos teóricos realistas) resolviendo simultáneamente las ecuaciones de Boltzmann y Poisson. Después, para cada partícula, se calcula la fuerza gravitacional resultante de todas las demás partículas en el sistema. Las interacciones gravitacionales entre partículas se describen usando las leyes de la gravitación de Newton. La segunda ley de Newton se utiliza para determinar cómo se mueve cada partícula bajo la influencia de las fuerzas gravitacionales ejercidas por todas las demás partículas. Las simulaciones avanzan en pasos discretos en el tiempo y en cada paso, se calculan las fuerzas

gravitacionales, se actualizan las posiciones y velocidades de las partículas y luego se avanza al siguiente paso.

Dado que calcular las interacciones entre todas las partículas puede ser computacionalmente costoso, especialmente cuando N es muy grande, se han desarrollado técnicas para hacer estas simulaciones más eficientes. Por ejemplo, algoritmos como el “código de árbol” (Barnes and Hut, 1986) y métodos de descomposición rápida como el algoritmo “Fast Multipole Method”, utilizado por Dehnen (2002), pueden acelerar los cálculos al agrupar partículas lejanas y tratarlas como una sola entidad. A medida que la simulación progresa, las partículas evolucionan y se mueven bajo sus interacciones mutuas. Estos resultados se pueden analizar para estudiar la formación de estructuras, la dinámica de cúmulos estelares, la evolución de galaxias y otros fenómenos astrofísicos.

Las simulaciones de N -cuerpos juegan un papel importante en la cosmología y dinámica modernas, ya que proporcionan información vital de la evolución de la materia oscura y también de la componente estelar de las galaxias: su agrupamiento, movimiento y las complejas interacciones entre sus componentes. El poder capturar de manera autoconsistente la evolución fuera de equilibrio o estados transitorios de los sistemas dinámicos, son algunas de las razones por las que la comunidad científica utiliza y continua desarrollando tal tipo de técnicas. En las siguientes secciones se da una descripción detallada de las ecuaciones que se resuelven, los códigos de N -cuerpos más utilizados, y la creación de las condiciones iniciales que se evolucionan con estos códigos.

3.1.1. Sistemas colisionales y *no-colisionales*

En el contexto de las simulaciones de N -cuerpos, el término *no-colisionales* no significa que las partículas no puedan acercarse entre sí o pasar muy cerca la una de la otra. Se refiere al hecho de que las partículas en la simulación no experimentan colisiones físicas como lo harían, por ejemplo, las moléculas en un gas. Las partículas en una simulación de N -cuerpos *no-colisionales* interactúan solo gravitacionalmente.

Se puede considerar como colisional por ejemplo, a un cúmulo estelar globular, donde las estrellas interactúan gravitacionalmente entre sí, con suficiente frecuencia para que el sistema alcance un estado de equilibrio en una escala de tiempo comparable a la edad del cúmulo. Las moléculas que componen un gas conforman una componente colisional en las simulaciones, donde las interacciones entre las partículas son suficientemente frecuentes para que se alcance un equilibrio térmico en un tiempo comparable al tiempo de relajación del sistema. El tiempo de relajación es el tiempo característico necesario para que una partícula individual en el sistema cambie significativamente su velocidad debido a interacciones (o colisiones) con otras partículas.

Por otro lado, en sistemas *no-colisionales*, como la mayoría de las galaxias, el tiempo de

relajación es mucho más largo que la edad del sistema, lo que significa que la mayoría de las estrellas (o partículas de materia oscura en simulaciones) no ha interactuado gravitacionalmente de manera significativa con otras estrellas más que unas pocas veces durante la vida del sistema. En estos sistemas, la estructura y dinámica están dominadas por la gravedad en lugar de por interacciones individuales entre partículas, ver el capítulo 1 de Binney and Tremaine (2008) para un análisis más detallado.

3.1.2. Principales ecuaciones en el problema de N -cuerpos

La ecuación de Boltzmann constituye la descripción más completa de un sistema de N -cuerpos a través de la función de distribución $f(\mathbf{r}, \mathbf{p}, t)$, con $\mathbf{p} = m\mathbf{v}$:

$$\frac{\partial f}{\partial t} + \mathbf{v} \frac{\partial f}{\partial \mathbf{x}} - \nabla \Phi \frac{\partial f}{\partial \mathbf{p}} = 0 \quad (3.1)$$

el potencial $\Phi(\mathbf{x})$ está determinado por la ecuación de Poisson que está dada por,

$$\nabla^2 \Phi = 4\pi G(\rho(\mathbf{x})) \quad (3.2)$$

La solución f del par de ecuaciones anteriores Boltzmann-Poisson, constituye una descripción autoconsistente de la evolución del sistema dinámico de N partículas. La posición y la velocidad se definen como vectores de 3 dimensiones, por lo que la ecuación se define en 6 dimensiones o alternativamente el problema se puede reducir a $6N$ ecuaciones. Esta descripción puede ser muy costosa particularmente en memoria, y una malla de más de 256^6 partículas, raya en el límite de las supercomputadoras modernas. La utilización de métodos de momentos o simetrías en el espacio fase, pueden permitir una descripción más eficiente, utilizando la ecuación de Boltzmann abordable con las computadoras contemporáneas. Alternativamente se puede seguir la trayectoria de cada partícula bajo la influencia gravitacional de todas las demás. Lo anterior constituye un muestreo de $f(\mathbf{r}, \mathbf{p}, t)$ o un problema de Monte-carlo para cada instante. Este es el enfoque de las simulaciones de N -cuerpos tradicionales, las cuales adoptaremos en esta tesis.

Las ecuaciones para el problema de N -cuerpos se definen dadas las coordenadas r , v , a , t y quedan relacionadas como,

$$\mathbf{v}(t) = \dot{\mathbf{r}}(t) \quad (3.3)$$

$$\mathbf{a}(t) = \dot{\mathbf{v}}(t) \quad (3.4)$$

$$\mathbf{a}_i(t) = -\sum_j^N \frac{Gm_j}{(r_i - r_j)^2} \quad (3.5)$$

Los códigos de N -cuerpos llamados de integración directa resuelven el sistema de ecuaciones mostrado arriba. Su costo computacional escala como N^2 , por lo que las simulaciones de mayor tamaño incluyen 10^6 partículas, aunque la utilización de tarjetas gráficas GPU promete ampliar estos números (Capuzzo-Dolcetta and Vicari, 2005). Como se menciona arriba, han surgido diferentes técnicas que permiten hacer eficiente la solución de estas ecuaciones, como a continuación se menciona.

3.2. Códigos basados en árboles

El código basado en árboles propuesto por Barnes and Hut (1986) organiza las partículas en una estructura jerárquica de celdas cúbicas, conocida como estructura de árbol octal, en esta estructura, cada nodo tiene ocho hijos. Inicialmente, todas las partículas quedan encerradas dentro de una celda cúbica grande. Esta celda se subdivide en ocho sub-celdas menores, y cada una de ellas aloja un subconjunto de las partículas. Este proceso de subdivisión continúa hasta que cada celda contenga únicamente una partícula. Dentro de este esquema, se determina la masa total, el centro de masa y los momentos multipolares de orden superior (hasta el término cuadrupolar) para cada celda o nodo. Esta estructura jerárquica se puede reconfigurar en cada paso temporal, optimizando así su construcción, ver fig. 3.1.

Para evaluar la fuerza en una partícula, se realiza una *Caminata* que atraviesa la estructura de árbol, comenzando desde la celda principal. En cada etapa, si una celda está lo suficientemente alejada, se utiliza para evaluar la fuerza. En caso contrario, la celda se “abre”, y sus sub-celdas se emplean para hacer el cálculo o se subdividen aún más. Esta decisión se basa en el cociente s/d , donde s es el tamaño de la celda y d es la distancia entre la partícula y el centro de masa de la celda. Si s/d es menor que un valor límite θ (ángulo de apertura), entonces se considera que la celda está lo suficientemente distante. El valor de θ se puede ajustar para variar la precisión y velocidad de la simulación.

Una vez identificadas todas las celdas y partículas que influyen en una partícula dada, se suman sus contribuciones para obtener la fuerza total. A través de este método, el número total de interacciones es mucho menor que en el enfoque directo de N -cuerpos, resultando en una complejidad temporal de $O(N \log N)$.

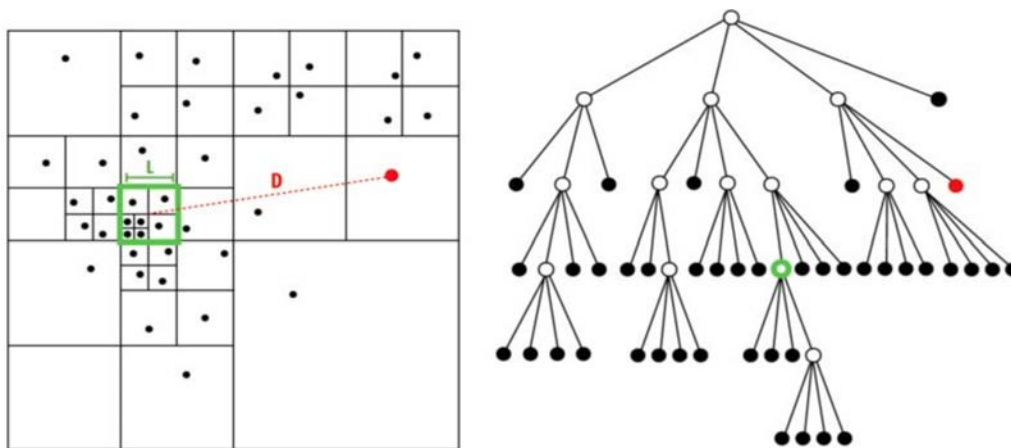


Figura 3.1: Representación gráfica del algoritmo Barnes-Hut en 2D. A la izquierda, distribución de partículas en un mallado cuadrado, mostrando cómo se agrupan en diferentes celdas según su densidad. A la derecha, se ilustra la estructura del cuad-árbol, con hojas y nodos desplegándose según la distribución de las partículas. Las zonas con más partículas generan niveles más profundos en el árbol.

3.2.1. Funcionamiento de gyrfalcON

En este estudio, utilizamos el código gyrfalcON (Dehnen, 2002) para realizar experimentos con simulaciones de N -cuerpos. El código se basa en un árbol jerárquico de celdas cúbicas y presenta interacciones mutuas entre celdas, las cuales se calculan mediante una expansión de Taylor cartesiana de manera simétrica, garantizando así la conservación del momento total. El código se beneficia de un “Criterio de Aceptación Multipolar” mejorado y sencillo que reduce el error de fuerza y el esfuerzo computacional a complejidad $O(N)$.

A continuación, se detallarán aspectos clave del funcionamiento de gyrfalcON.

3.2.2. Algoritmo de Cálculo de Fuerzas

El algoritmo de cálculo de fuerzas es fundamental para cualquier simulación de N -cuerpos. Los códigos tradicionales de árbol organizan inicialmente las partículas en una estructura jerárquica de celdas cúbicas. Luego, evalúan la fuerza en una posición arbitraria producida por el contenido de una celda a través de una expansión multipolar. El algoritmo de árbol que utiliza gyrfalcON está basado en el algoritmo de Barnes and Hut (1986), descrito en la sección anterior.

A diferencia de estos códigos tradicionales, gyrfalcON combina estructuras jerárquicas de árboles con celdas cúbicas y un innovador algoritmo que computa interacciones mutuas entre las celdas. Estas interacciones se basan en una expansión simétrica de Taylor en coordenadas cartesianas. Adicionalmente, el *Criterio de Aceptación Multipolar* utilizado en gyrfalcON optimiza el cálculo de fuerza y reduce drásticamente el costo computacional.

3.2.3. Criterio de Aceptación Multipolar

El “Criterio de Aceptación Multipolar” (en inglés, *Multipole Acceptance Criterion*, *MAC*), es una mejora significativa que hace de *gyrfalcON* una herramienta potente en el ámbito de simulaciones de N -cuerpos. Este criterio implica organizar las partículas en una jerarquía de mallas anidadas. Una vez precalculados los momentos multipolares de cada celda, las fuerzas entre las celdas se determinan mediante una expansión multipolar en armónicos esféricos.

El *Criterio de Aceptación Multipolar* es un criterio utilizado en simulaciones de N -cuerpos, especialmente en códigos basados en árboles, para decidir si una celda (o nodo) del árbol puede ser tratada como una única entidad o si se necesita dividirla y considerar sus componentes por separado.

En esencia, el “Criterio de Aceptación Multipolar” es una medida de qué tan lejos está una partícula o grupo de partículas de una celda particular, y si la distancia es suficientemente grande, la celda se trata como un único objeto con propiedades combinadas (como el centro de masa). Si la celda no está suficientemente lejos, se abre y se examinan sus componentes individuales.

Un parámetro comúnmente utilizado en el “Criterio de Aceptación Multipolar” es Θ , que es un valor umbral. Si el tamaño de la celda (por ejemplo, su diámetro) dividido por la distancia al punto de interés (por ejemplo, una partícula en la simulación) es menor que Θ , entonces se puede usar una aproximación para la celda. Si es mayor, la celda debe ser subdividida.

$$\frac{l}{d} < \theta \quad (3.6)$$

donde l es el tamaño de la celda y d es la distancia a la partícula o punto de interés.

Este criterio es fundamental para que los códigos de árbol sean eficientes, ya que evita tener que calcular interacciones entre cada par de partículas, lo que tendría un costo computacional de $O(N^2)$. Al usar el árbol con el “Criterio de Aceptación Multipolar”, el costo puede reducirse a $O(N \log N)$ o incluso $O(N)$ en algunos casos optimizados.

El uso exacto y la formulación del “Criterio de Aceptación Multipolar” pueden variar dependiendo del código y del contexto específico en el que se utilice.

3.3. Códigos de partícula-malla

El segundo método es un algoritmo de partícula-malla (P-M) donde se forma una malla cúbica uniforme sumando todas las partículas en cada celda de la malla. La ecuación de Poisson se resuelve en el espacio de Fourier y las fuerzas en cada una de las celdas de partícula-malla pueden resolverse en un algoritmo cuyo costo computacional escala con N . La limitación es que no se pueden calcular fuerzas por debajo del tamaño de la malla. Una

mallas tridimensionales pueden ser intensivas en memoria, por lo que el tamaño de la malla a menudo está limitado por la memoria.

En la práctica, las fuerzas a pequeña escala dentro de una malla se pueden calcular directamente (es decir, cálculo directo partícula-partícula, P-P) mientras que la partícula-malla funciona en escalas más grandes; a esto se le llama el método “P3M” (Couchman et al., 1995, e.g.).

En otros códigos, el algoritmo del árbol se usa para fuerzas de corto alcance y el algoritmo de partícula-malla se usa en fuerzas de largo alcance. Esto permite la mayor cantidad de rango dinámico en una caja grande por ejemplo, el código N -cuerpos Gadget-2 (Springel, 2005), usado para calcular las simulaciones del Milenio.

3.3.1. Códigos de malla: GADGET2

Gadget2 es un código del tipo partícula-malla, este código será utilizado más adelante para las simulaciones de sistemas aislados remuestreados con un número considerable de partículas.

Gadget2 es un código N -cuerpos Lagrangiano no-colisional. En él, las fuerzas gravitacionales se calculan con una expansión multipolar jerárquica, comúnmente denominada algoritmo de árbol, que opcionalmente se puede aplicar en forma de un algoritmo *TreePM* (Bagla, 2002). En el esquema del árbol, las partículas se organizan en una jerarquía de grupos, es decir, cuando se calcula la fuerza de una partícula en particular, la fuerza ejercida por grupos distantes se aproxima mediante sus momentos multipolares más bajos. De esta manera, el costo computacional para una evaluación completa de la fuerza se puede reducir al orden $O(N \log N)$, ya que en lugar de requerir $N - 1$ fuerzas parciales por partícula, como se necesita en un enfoque de suma directa, la fuerza gravitacional en una única partícula se puede calcular con solo $O(\log N)$ interacciones. Las fuerzas son más precisas si la expansión multipolar se lleva a un orden superior, pero eventualmente el creciente costo de evaluar momentos superiores hace que sea más eficiente terminar la expansión múltiple y, en cambio, usar un mayor número de nodos de árbol más pequeños para lograr una precisión de fuerza deseada. El código utiliza la construcción del árbol Barnes-Hut (Barnes and Hut, 1986) descrito anteriormente.

El método *TreePM* es un método híbrido que consiste en una síntesis del método P-M y el algoritmo de árbol. El potencial en este caso se divide explícitamente en el espacio de Fourier en una parte a largo plazo y una parte a corto plazo. El potencial a largo plazo se puede calcular de manera muy eficiente con métodos de Fourier basados en mallas. La parte a corto plazo del potencial se puede resolver fácilmente en el espacio real. Con este método, solo es necesario recorrer el árbol en una pequeña región espacial alrededor de cada partícula objetivo, y no se requieren correcciones para las condiciones de contorno

periódicas. Además, se gana precisión en la fuerza a largo plazo, que ahora es básicamente exacta y no una aproximación como en el método del árbol.

3.4. Armado de las condiciones iniciales

La creación de condiciones iniciales para sistemas galácticos es un paso fundamental en la simulación y evolución de galaxias utilizando códigos de N -cuerpos. Esto significa que hay que definir las posiciones iniciales, velocidades y masas de todas las partículas (estrellas, materia oscura, gas, etc.) que componen el sistema galáctico que se va a simular.

Cabe resaltar que todos los experimentos realizados en relación con los efectos colectivos y la fricción dinámica siguieron el enfoque de evolucionar un sistema galáctico aislado de carácter no-colisional. A continuación se describe la metodología, los modelos y códigos utilizados para la creación de las condiciones iniciales de los halos galácticos y también de los satélites cuando se modelan como sistemas vivos.

3.4.1. Condiciones iniciales de la galaxia aislada

Configurar las condiciones iniciales de una galaxia aislada es un proceso intrincado. Se inicia con la definición de los límites de la caja de simulación y culmina con la asignación de posiciones y velocidades a las partículas, asegurando la estabilidad del sistema una vez iniciada la simulación.

Es esencial comprender que, por lo general, la estructura y evolución de un sistema de N -cuerpos está regida por la ecuación diferencial de Boltzmann-Vlasov. Esta ecuación refleja la continuidad en el espacio fase y su solución proporciona una función de distribución, $f(r, v)$, que describe la densidad de partículas en un punto y velocidad específicos dentro de un rango definido. Si podemos obtener analíticamente esta función de distribución, se facilita la tarea de asignar posiciones y velocidades a las partículas, asegurando que el sistema inicial sea autoconsistente.

3.4.2. Fórmula de Eddington: Función de distribución para sistemas esféricos

Los sistemas esféricos representan una primera aproximación simplificada en el estudio de sistemas estelares. Son de especial interés porque muchas galaxias elípticas y la mayoría de los cúmulos globulares pueden ser descritos de esta manera. El desarrollo de la fórmula de Eddington presentado en esta tesis es el que se presenta en el capítulo 4 de [Binney and Tremaine \(2008\)](#). Donde, por simplicidad se supone que los sistemas estelares contienen una

única población estelar, de modo que las estrellas son idénticas y siguen una única función de distribución f .

Generalmente, si la densidad de masa que contribuye al potencial gravitacional es proporcional a $\int d^3v f$, entonces los sistemas bajo estudio pueden considerarse autoconsistentes. Esto se debe a que la distribución de densidad define el potencial mediante la ecuación de Poisson, mientras que el potencial, a su vez, define la densidad a través de la ecuación de Boltzmann no colisional.

Introduzcamos ahora dos nuevos conceptos: el potencial relativo Ψ y la energía relativa \mathcal{E} . Si definimos a Φ_0 como una constante, entonces:

$$\Psi \equiv -\Phi + \Phi_0 \quad \text{y} \quad \mathcal{E} \equiv -H + \Phi_0 = \Psi - \frac{1}{2}v^2 \quad (3.7)$$

Aseguramos que $f > 0$ para $\mathcal{E} > 0$ y $f = 0$ para $\mathcal{E} \leq 0$.

Si un sistema es aislado y se extiende hasta el infinito, entonces Φ_0 equivale a la energía de ligadura. En este contexto, el potencial relativo obedece la ecuación de Poisson:

$$\nabla^2 \Psi = -4\pi G\rho, \quad (3.8)$$

donde se cumple la condición de frontera $\Psi \rightarrow \Phi_0$ a medida que $|x| \rightarrow \infty$.

3.4.3. Fórmula de Eddington y sistemas ergódicos

Consideremos un sistema estelar esférico confinado por un potencial esférico conocido $\Phi(r)$. Para tal sistema, es posible deducir una función de distribución ergódica en función del hamiltoniano $H(\mathbf{x}, \mathbf{v})$, como se propone en el Teorema de Jeans (Binney and Tremaine (2008)).

Nuestro objetivo es determinar la función de distribución en términos de la energía relativa, es decir, $f(\mathcal{E})$. Definimos la densidad de probabilidad $\nu(r)$ al integrar f sobre todo el espacio de velocidades:

$$\nu(r) = 4\pi \int dv v^2 f(\Psi - \frac{1}{2}v^2) = 4\pi \int_0^\Psi d\mathcal{E} f(\mathcal{E}) \sqrt{2(\Psi - \mathcal{E})}, \quad (3.9)$$

Considerando que Ψ es una función monótona de r en cualquier sistema esférico, podemos expresar ν en función de Ψ en lugar de r . Así, obtenemos:

$$\frac{1}{\sqrt{8\pi}} \nu(\Psi) = 2 \int_0^\Psi d\mathcal{E} f(\mathcal{E}) \sqrt{\Psi - \mathcal{E}} \quad (3.10)$$

Diferenciando ambos lados, derivamos:

$$\frac{1}{\sqrt{8\pi}} \frac{dv}{d\Psi} = \int_0^\Psi d\mathcal{E} \frac{f(\mathcal{E})}{\sqrt{(\Psi - \mathcal{E})}} \quad (3.11)$$

Esta ecuación de Abel se soluciona como:

$$f(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \frac{d}{d\mathcal{E}} \int_0^\Psi \frac{d\Psi}{\sqrt{\mathcal{E} - \Psi}} \frac{dv}{d\Psi} \quad (3.12)$$

O de manera equivalente:

$$f(\mathcal{E}) = \frac{1}{\sqrt{8\pi^2}} \left[\int_0^\Psi \frac{d\Psi}{\sqrt{\mathcal{E} - \Psi}} \frac{d^2v}{d\Psi^2} + \frac{1}{\sqrt{\mathcal{E}}} \left(\frac{dv}{d\Psi} \right)_{\Psi=0} \right] \quad (3.13)$$

Esta última ecuación se conoce como la “Fórmula de Eddington”.

Dada una distribución de densidad para un sistema esférico, la fórmula de Eddington permite obtener una función de distribución ergódica que concuerda con la ley de densidad propuesta. No hay garantías de que la solución $f(\mathcal{E})$ cumpla siempre con la condición de ser no negativa. A partir de las ecuaciones 3.12 y 3.13, se deduce que una distribución de densidad esférica $\nu(\mathbf{r})$ en el potencial $\Phi(\mathbf{r})$ se deriva de una función de distribución ergódica si y solo si:

$$\int_0^\mathcal{E} \frac{d\Psi}{\sqrt{\mathcal{E} - \Psi}} \frac{dv}{d\Psi} \quad (3.14)$$

es una función en crecimiento respecto a \mathcal{E} .

Las condiciones iniciales que se evolucionan con los códigos de N-cuerpos, están basadas en el trabajo de [McMillan and Dehnen \(2007\)](#) y utilizan la Fórmula de Eddington, expresada en la ecuación 3.13, para derivar la función de distribución en sistemas esféricos a partir de la densidad estelar.

3.5. Algoritmo para las Condiciones iniciales

Para crear el modelo de N-cuerpos que representa al halo se busca la creación de una realización de las componentes esféricamente simétricas del sistema usando el método de [Cuddeford \(1991\)](#), el cual es una extensión del modelo de [Osipkov \(1979\)](#) y [Merritt \(1997\)](#), y busca ser una extensión de la técnica de inversión con la Fórmula de Eddington para obtener la función de distribución para un sistema esférico con distribución isotrópica de velocidad. [Cuddeford \(1991\)](#) considera funciones de distribución de la forma:

$$f(\mathcal{E}, L) = L^{2\alpha} f_0(Q).$$

Donde, $Q \equiv \mathcal{E} - L^2/2r_a^2$ como para los modelos Osipkov-Merritt ($\mathcal{E} \equiv \Psi - \frac{1}{2}v^2$, Ψ denota el

potencial gravitacional negativo) con radio de anisotropía r_a . El parámetro α está restringido a ser más grande que -1. Como mostró Cuddeford (1991) $f(Q)$ esta relacionada a la densidad a través de una integral de Abel la cual puede invertirse (asumiendo que $f(Q < 0) = 0$), dando como resultado,

$$f_0(Q) = \frac{\sin\left(n - \frac{1}{2} - \alpha\right)\pi}{\pi\lambda(\alpha)\eta(\alpha)} \frac{d}{dQ} \int_0^Q \frac{d^n \rho_{\text{red}}}{d\Psi^n} \frac{d\Psi}{(Q - \Psi)^{\alpha+3/2-n}},$$

Creamos una realización de N-cuerpos para los componentes esféricamente simétricos del sistema utilizando el método de Cuddeford (1991). Este método es una extensión del modelo propuesto por Osipkov (1979) y Merritt (1997), que a su vez ampliaron la técnica original de inversión de Eddington (1916), descrita en la sección 3.4.2, para determinar la función de distribución de un sistema esférico con una distribución de velocidad isotrópica.

Cuddeford (1991) propuso funciones de distribución de la forma:

$$f(\mathcal{E}, L) = L^{2\alpha} f_0(Q) \quad (3.15)$$

Donde $Q \equiv \mathcal{E} - \frac{L^2}{2r_a^2}$ y \mathcal{E} se define como: $\mathcal{E} \equiv -\frac{1}{2}v^2$, siendo \mathcal{E} el negativo del potencial gravitacional y r_a el radio de anisotropía.

El parámetro α está restringido a ser mayor que -1. Según lo demostrado por Cuddeford (1991), la función $f(Q)$ está relacionada con la densidad ρ mediante una ecuación integral de Abel. Bajo ciertas suposiciones, esta relación se puede expresar como:

$$f_0(Q) = \frac{\sin\left(n - \frac{1}{2} - \alpha\right)\pi}{\pi\lambda(\alpha)\eta(\alpha)} \frac{d}{dQ} \int_0^Q \frac{d^n \rho_{\text{red}}}{d\Psi^n} \frac{d\Psi}{(Q - \Psi)^{\alpha+3/2-n}}, \quad (3.16)$$

La función de distribución produce un sistema esférico simétrico con una distribución de velocidad, tal que el parámetro de anisotropía es de la forma,

$$\beta(r) \equiv 1 - \frac{\sigma_\theta^2}{\sigma_r^2} = \frac{r^2 - \alpha r_a^2}{r^2 + r_a^2} \quad (3.17)$$

El modelo del halo es un modelo tipo Navarro et al. (1996), que sigue una distribución de densidad esférica truncada y con densidad crítica ρ_c ,

$$\rho_h(r) = \rho_c \frac{\text{trunc}(r/r_t)}{(r/r_h)(1 + r/r_h)^2}, \quad (3.18)$$

donde,

$$\text{trunc}\left(\frac{r}{r_t}\right) = \begin{cases} \text{sech}\left(\frac{r}{r_t}\right) & \text{si } r_t > 0 \\ \frac{2}{\text{sech}\left(\frac{r}{r_t}\right) + \frac{1}{\text{sech}\left(\frac{r}{r_t}\right)}} & \text{si } r_t < 0 \end{cases}$$

3.5.1. Creando las condiciones iniciales dentro del código MKHALO

Para crear las condiciones iniciales, es decir, construir o crear el halo, siguiendo por ejemplo un perfil de densidad **NFW**, se hace uso de la herramienta mkhalo (McMillan and Dehnen, 2007). Esta se basa en el método de Cuddeford (1991), en una extensión del modelo de Osipkov (1979) y Merritt (1985), lo cual representa una extensión del método de inversión de Eddington (1916) para obtener la función de distribución para un sistema esférico con una distribución de velocidad isotrópica.

$$\rho(r) = C \cdot \frac{\text{trunc}(r/|r_t|)}{x^{\text{inner}} \cdot (x^\eta + 1)^{\frac{\text{outer}-\text{inner}}{\eta}}}$$

donde $\eta = 1$, $\text{outer} = 3$, $\text{inner} = 1$ para un halo con perfil de densidad tipo **NFW**.

$$\text{trunc}(x) = \begin{cases} \text{sech}(x) & \text{si } r_t > 0 \\ \frac{2}{\text{sech}(x) + \frac{1}{\text{sech}(x)}} & \text{si } r_t < 0 \end{cases}$$

donde C es una constante que depende de la masa total del sistema.

3.6. El modelo de los halos de materia oscura

El modelo **NFW** (Navarro et al., 1996) describe la densidad de materia oscura en halos galácticos utilizando la siguiente ecuación:

$$\rho(r) = \frac{\delta_c \rho_{\text{crit}}}{\left(\frac{r}{r_s}\right) \left(1 + \frac{r}{r_s}\right)^2}$$

donde: $\rho(r)$ es la densidad de materia oscura en función de la distancia r desde el centro del halo. δ_c es un parámetro relacionado con la densidad crítica del Universo en ese momento. ρ_{crit} es la densidad crítica del Universo en ese momento. r_s es la escala característica del halo que determina el tamaño del halo.

Este modelo es ampliamente utilizado en cosmología y astrofísica para describir la distribución de materia oscura en el universo y se ha demostrado que es compatible con las observaciones de la formación y evolución de estructuras cósmicas.

Capítulo 4

Caída concurrente de satélites: Importancia de los efectos colectivos para la reconstrucción orbital

En este capítulo presentamos los resultados principales en el contexto de las consecuencias que los efectos colectivos tienen sobre la historia de caída de las galaxias satélites hacia un halo masivo. Exploramos diferentes casos: el caso en el que los satélites se encuentran en una configuración coplanar y también el caso de satélites no coplanares, así como el caso en el que se encuentran a diferentes distancias inicialmente. Así como los casos en los que los satélites son partículas rígidas modeladas como un perfil tipo Plummer, satélites compactos vivos más parecidos a cúmulos globulares y satélites vivos extendidos más parecidos a galaxias enanas. El caso en el que los satélites son coplanares es interesante, ya que tiene consecuencias para los satélites en la **VPOS** de la Vía Láctea. Exploramos este caso con un ensamble de simulaciones, variando las condiciones iniciales dentro de las incertidumbres observacionales para los satélites en la **VPOS**. Nuestros resultados muestran que el caso de múltiples satélites difiere considerablemente del caso con el que se trabaja usualmente, que es el caso en el que solo se considera un perturbador masivo cayendo. También mostramos cómo la pérdida de masa en el caso de múltiples satélites puede dar lugar a un aumento de la auto-fricción dinámica mayor a lo reportado en la literatura lo cual no ha sido reportado hasta la fecha.

4.1. Un modelo semianalítico

Los trabajos que suelen trabajar en el contexto de la fricción dinámica suelen empezar comparando un modelo Semianalítico (**SA**) con las simulaciones de N-cuerpos. Los modelos **SA** tienen la ventaja de permitir explorar rápidamente una amplia gama de parámetros

y ofrecen una comprensión al menos intuitiva de los procesos físicos subyacentes. En estos modelos se usan ecuaciones simplificadas basadas en consideraciones teóricas y observacionales, los cuales pueden hacer predicciones sobre la evolución de sistemas de galaxias. Por ejemplo, un modelo SA podría basarse en la fórmula de CH para la fricción dinámica para prever cómo una galaxia satélite se acerca a su galaxia anfitriona con el tiempo. El análisis con simulaciones de N-cuerpos se ve beneficiado de los modelos SA en varios sentidos: son simplificados, permiten explorar parámetros físicos de manera rápida, sirven como pruebas de control o punto de referencia para las simulaciones de N-cuerpos, por ejemplo si la discrepancia entre un modelo SA y una simulación es muy grande, puede ser un indicador de algún problema con la simulación de N-cuerpos.

Dentro del modelo SA presentado en este trabajo, se incluye la fricción dinámica en el enfoque de CH bajo el esquema de Binney and Tremaine (1987) y descrito en la sección A. Se incluye una función de pérdida de masa de King (1962) descrito por Klypin et al. (1999) como se describe en la sección C, para incluir la variación espacial del logaritmo Coulombiano nos basamos en el trabajo de Hashimoto et al. (2003).

En estos cálculos, el modelo del satélite es el mismo que en las simulaciones de N-cuerpos, es decir, una única partícula con masa M_s y una longitud de suavizado ϵ_s . Pero en lugar de ser representado por N partículas, el potencial del halo galáctico se evalúa utilizando el potencial gravitacional de un modelo NFW, con la misma masa y escalas que las adoptadas en las simulaciones de N-cuerpos. En esta integración, la fuerza ejercida al satélite debido a la fricción dinámica del halo se evalúa usando una fórmula analítica que se resuelve en cada paso de tiempo para obtener la solución de las ecuaciones de movimiento.

4.1.1. Variaciones en el logaritmo coulombiano en la fórmula de Chandrasekhar para la fricción dinámica

Existen otras estrategias para considerar otros factores involucrados en el problema de la Fricción Dinámica, más allá de la aproximación de CH. Entre ellos, los más relevantes son aquellos que consideran un sistema anfitrión más realista, es decir, con un tamaño finito y un perfil de densidad que depende radialmente (Binney, 1977; Boldrini et al., 2018), y aquellos que consideran que sistemas anfitriones realistas, como los de modelos de N-cuerpos, pueden ser bien descritos por un simple logaritmo de Coulomb con dependencias espaciales o de masa (Hashimoto et al., 2003; Just and Peñarrubia, 2005; Colpi et al., 1999; Boylan-Kolchin et al., 2008). Por ejemplo Hashimoto et al. (2003) se basa en la fórmula de CH pero introduce una variación en la fórmula del logaritmo Coulombiano,

$$\ln \Lambda = \ln \left(R_{\text{halo}} / \epsilon_s V_s^2 \right) \tag{4.1}$$

donde R_{halo} es la longitud de escala del halo. Esta fórmula ha sido adoptada por estudios con modelos SA de la evolución orbital de galaxias satélites (Murai and Fujimoto, 1980; Johnston et al., 1995, e.g.). Hashimoto et al. (2003) también utiliza la versión propuesta por Tremaine (1976) y utilizan R_s la distancia al centro de la galaxia en vez del parámetro de escala R_{halo} , donde el factor $1,4 \epsilon_s$ ajusta mejor el modelo SA con las simulaciones de N-cuerpos.

$$\ln \Lambda = \ln (R_s / 1,4 \epsilon_s) \quad (4.2)$$

4.2. Los experimentos con simulaciones de N-cuerpos

Llevamos a cabo una serie de simulaciones autoconsistentes para examinar el comportamiento orbital de partículas masivas que se mueven dentro de un fondo de materia oscura o estelar.

Construimos halos de partículas estables utilizando las técnicas desarrolladas en la sección 3.5 descritas a detalle en McMillan and Dehnen (2007). Estos modelos tienen distribuciones de densidad que son descritas por halos de materia oscura que siguen un perfil de densidad tipo Navarro et al. (1996).

Los perturbadores se modelan tanto como partículas individuales rígidas suavizadas como un modelo de Plummer, pero también como sistemas de partículas que intercambian energía y momento angular (sistemas vivos: extendidos y compactos), en este último caso la resolución espacial y en masa debe ser consistente con las partículas del halo para evitar efectos numéricos.

Comenzamos con los perturbadores (satélites de aquí en adelante) fuera del núcleo, principalmente en órbitas circulares y los dejamos orbitar, esperando que caigan en espiral hacia el centro de sus respectivos halos anfitriones debido a la fricción dinámica. Los resultados mostrados en las siguientes secciones están en función de la distancia desde el centro del halo anfitrión como función del tiempo, $r(t)$. En el caso de los sistemas vivos, para poder monitorear $r(t)$ para cada satélite hacemos uso del software de Behroozi et al. (2013) Rockstar Phase Space Halo Finder (Rockstar) descrito a detalle en la sección D, que nos permite encontrar las coordenadas del centro del halo anfitrión y las del satélite respecto del centro de masa del sistema. Para ver los parámetros que caracterizan al halo y a los tipos de satélite ver la tabla 4.1 y el comportamiento del perfil de densidad y la curva de velocidad circular en la figura 4.1.

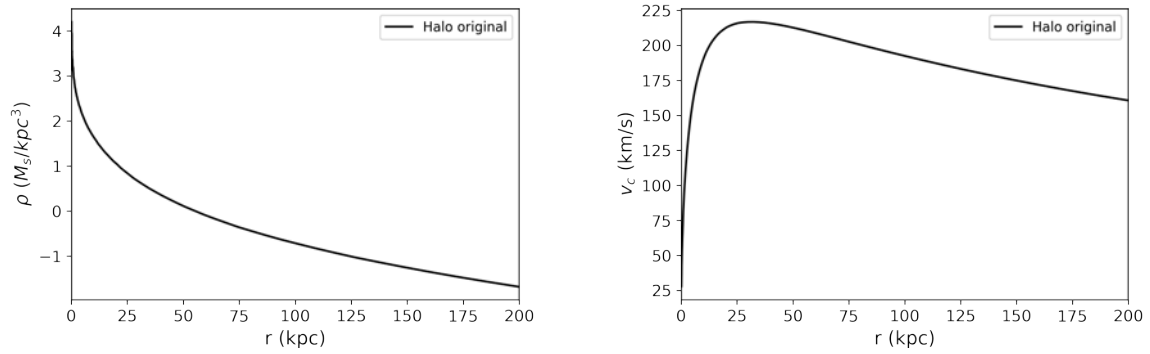


Figura 4.1: Estas figuras representan tanto el perfil de densidad como la curva de velocidad circular para el halo de materia oscura que se utiliza como condición inicial de las simulaciones de N-cuerpos.

	Halo	Sat. comp.	Sat. Intermed.	Sat. ext.
M_{200}/M_{\odot}	1.2×10^{12}	10^{10}	$10^9 - 10^{10}$	10^{10}
N_p	2×10^6	1.3×10^4	$1.3 - 13 \times 10^3$	1.3×10^3
r_s [kpc]	15	0.5	3	10
$r_{1/2}$ [kpc]	20	1.5	5.5	18
$M_{1/2}/M_{\odot}$	6×10^{11}	5×10^9	$0.5 - 5 \times 10^9$	5×10^9
R_{tr} [kpc]	200	1	5	25
ϵ [kpc]	0.3	0.3	0.2-0.4	0.6

Cuadro 4.1: Los modelos del halo y los satélites utilizados, incluyendo el caso de satélite rígido, satélite vivo compacto (cúmulo globular), el satélite de extensión intermedia (galaxia tipo enana), el satélite vivo extendido (galaxia temprana). De arriba a abajo: la masa total (M_{200}), el número de partículas (N_p), el radio de escala (r_s) del perfil de densidad tipo **NFW**, el radio a media masa ($r_{1/2}, M_{1/2}$), el radio de truncamiento (R_{tr}) y la longitud de suavizado (ϵ). Para el satélite rígido $M_{200}/M_{\odot} = 1 \times 10^{10}$, $\epsilon = 0.6$ kpc.

4.2.1. Comparación del modelo semianalítico con simulaciones de N-cuerpos

El punto de partida para comenzar el análisis de este trabajo fue comparar un modelo **SA** (con pérdida de masa, logaritmo Coulombiano variable espacialmente o constante) con modelos numéricos de simulaciones de N-cuerpos de un satélite aislado cayendo hacia un halo anfitrión en diferentes condiciones: Un satélite rígido, un satélite vivo compacto, un satélite vivo extendido.

Nuestros modelos fiduciaros **SA** asumen un perfil de halo de materia oscura tipo Navarro et al. (1996) para el halo anfitrión. Los satélites están representados por un modelo Plummer rígido (Plummer, 1911) e incluimos la fricción dinámica, tal como se describe en

las secciones anteriores. Adicionalmente, incorporamos la pérdida de masa del satélite que se hunde siguiendo a [Klypin et al. \(1999\)](#). También tomamos en cuenta un perfil de densidad no homogéneo para el anfitrión, incorporando un logaritmo Coulombiano que depende radialmente. Por simplicidad, elegimos la estrategia propuesta por [Hashimoto et al. \(2003\)](#) (ver su ecuación 2). Utilizamos estos experimentos SA como modelos de control para garantizar precisión en las simulaciones en el régimen en el cual las simulaciones del modelo SA y simulaciones de N-cuerpos son comparables, y para identificar los procesos físicos relevantes de los incluidos en los modelos SA.

Para llevar a cabo la investigación presentada aquí, generamos una serie de experimentos N-cuerpos que incluyen una distribución esférica de partículas inicialmente en equilibrio dinámico (sistema central) y uno o varios satélites. Comparamos la evolución de estos sistemas N-cuerpos con los resultados de modelos SA, el resultado se muestra en la figura 4.2

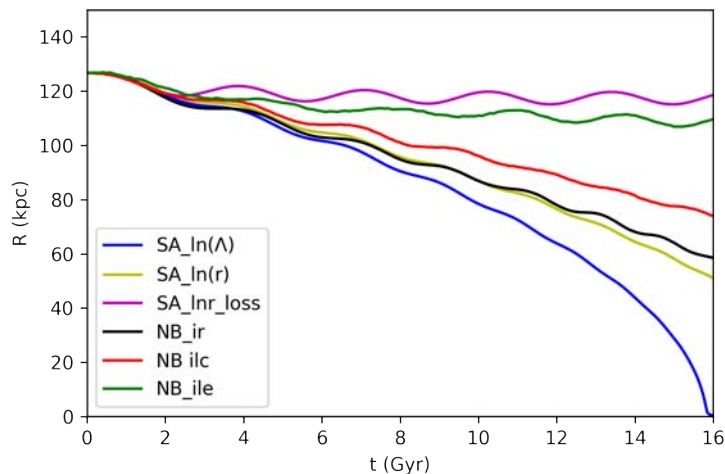


Figura 4.2: Hundimiento de satélites individuales en un halo con masa de la Vía Láctea. Línea azul: modelo fiduciario SA como un logaritmo de Coulomb sin dependencia espacial; Amarillo: SA usando un logaritmo de Coulomb con dependencia radial; Magenta: SA usando un logaritmo de Coulomb con dependencia radial y una pérdida de masa; Negro: modelo N-cuerpos (NB) con un satélite representado con un perfil de densidad Plummer suavizado rígido (sistema rígido, de ahora en adelante), $\Lambda = \frac{b_{\max}}{b_{90^\circ}} \approx 5$, siguiendo [Binney and Tremaine \(1987\)](#) ec. 8.1b; Rojo: modelo NB que incluye un satélite con un perfil de densidad NFW compacto; Verde: modelo NB que incluye un satélite con un perfil de densidad NFW extendido (para más detalles sobre los modelos, ver la tabla 4.1). La condición inicial para el satélite en los modelos es $r = 126$ kpc, $v_c = 183$ cm/s.

En la figura 4.2 podemos observar que la comparativa entre simulaciones de N-cuerpos y modelos SA es consistente en el respectivo régimen, en las simulaciones de N-cuerpos se espera que la caída sea más lenta que en los modelos SA, en el caso rígido empezamos a notar una diferencia en la caída con respecto al modelo SA sin pérdida de masa pero con un

logaritmo Coulombiano dependiente espacialmente, sin embargo, hay algunas diferencias notables. Al comparar el modelo SA con pérdida de masa con la simulación de un satélite vivo extendido vemos que hay una discrepancia, en el modelo de N-cuerpos la caída parece sufrir un retraso contrario a lo que se esperaría debido a la fricción dinámica.

4.3. Experimentos con satélites coplanares

A partir de este punto, decidimos realizar pruebas con varios satélites para poner a prueba la hipótesis de que los múltiples satélites cambian la historia de caída debido a los efectos colectivos. Nuestros primeros experimentos consisten en explorar diferentes condiciones iniciales tanto coplanares como no coplanares, satélites interactuando en diferentes planos y a diferentes distancias y satélites con diferente extensión (rígidos, compactos vivos y extendidos vivos).

4.3.1. Experimento con satélites coplanares a diferentes distancias inicialmente

En este experimento decidimos colocar 4 satélites coplanares pero a diferente distancia al inicio como se muestra en el panel inferior de la figura 4.3 y en órbita circular (ver figura 4.2). Estos satélites son extendidos, con su masa inicial y extensión como marca la columna 5 de la tabla 4.1. La gráfica del panel superior de la figura 4.3 muestra la historia de caída de los satélites cayendo simultáneamente. En la figura 4.4 podemos ver la conservación de energía durante el tiempo de la simulación en forma del error relativo en energía, el error en la energía está dentro del orden de 5×10^{-5} .

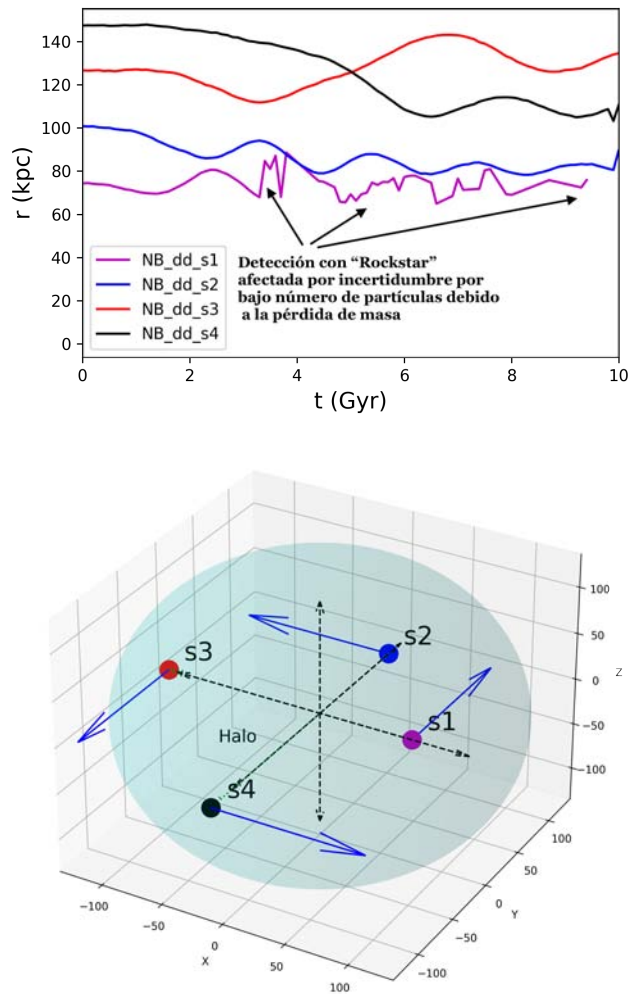


Figura 4.3: El panel superior representa la gráfica de r vs t , para cada uno de los satélites en el experimento en el que los satélites se encuentran inicialmente a diferentes distancias. La notación dd indica que los satélites están a diferente distancia. El panel inferior muestra un esquema de la posición de cada satélite con su respectivo color, las líneas azules indican la dirección del vector velocidad. La membrana azul es una representación para remarcar la presencia de las partículas del halo (no a escala).

Perturbador	Posición (kpc)			Velocidad (km/s)		
	x	y	z	V _x	V _y	V _z
NB1	74.4343	0.0314	0.0286	0.5175	203.1478	-0.1004
NB2	-0.0214	100.9429	0.0286	-192.0264	0.0701	-0.0775
NB3	-126.7771	0.0286	0.0286	0.6229	-182.7038	-0.0352
NB4	-0.0214	-147.5357	0.0286	176.5172	0.1560	-0.0217

Cuadro 4.2: Condiciones iniciales para cada satélite en un experimento con 4 satélites cayendo hacia un halo masivo, estos 4 satélites coexisten en el mismo plano pero inicialmente se encuentran a diferentes distancias.

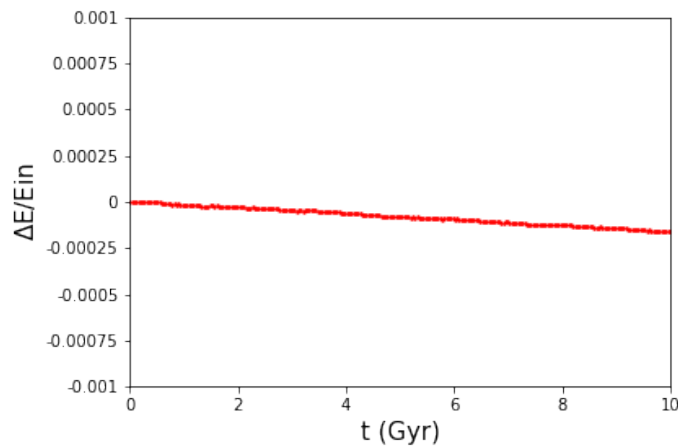


Figura 4.4: Error relativo de la energía a lo largo de la simulación en el tiempo para el experimento de cuatro satélites vivos extendidos en configuración coplanar y a diferente distancia inicial.

De la figura 4.3 podemos ver que la órbita del satélite más externo (línea negra) y el satélite más próximo a este (línea roja) tienen un cruce aparente, sin embargo la distancia en el punto de cruce es de 12 kpc. Estos cambios con cruces aparentes en la trayectoria de los dos satélites más externos se debe a la manera en que intercambian energía entre ellos a través de la respuesta en el medio. Los efectos en los satélites más internos parecen ser menores y el satélite más interno parece verse más afectado por la parte central lo que dificulta su detección con Rockstar.

4.3.2. Satélites coplanares equidistantes

En el siguiente experimento se colocaron 4 satélites en configuración coplanar y en órbita circular inicialmente, equidistantes entre ellos. Para este experimento exploramos con más detalle los casos en que los satélites se modelan como partículas rígidas, como sistemas vivos

pero compactos (más parecidos a cúmulos globulares) y como satélites vivos extendidos (más como galaxias tardías extendidas) como se muestra en la figura 4.5.

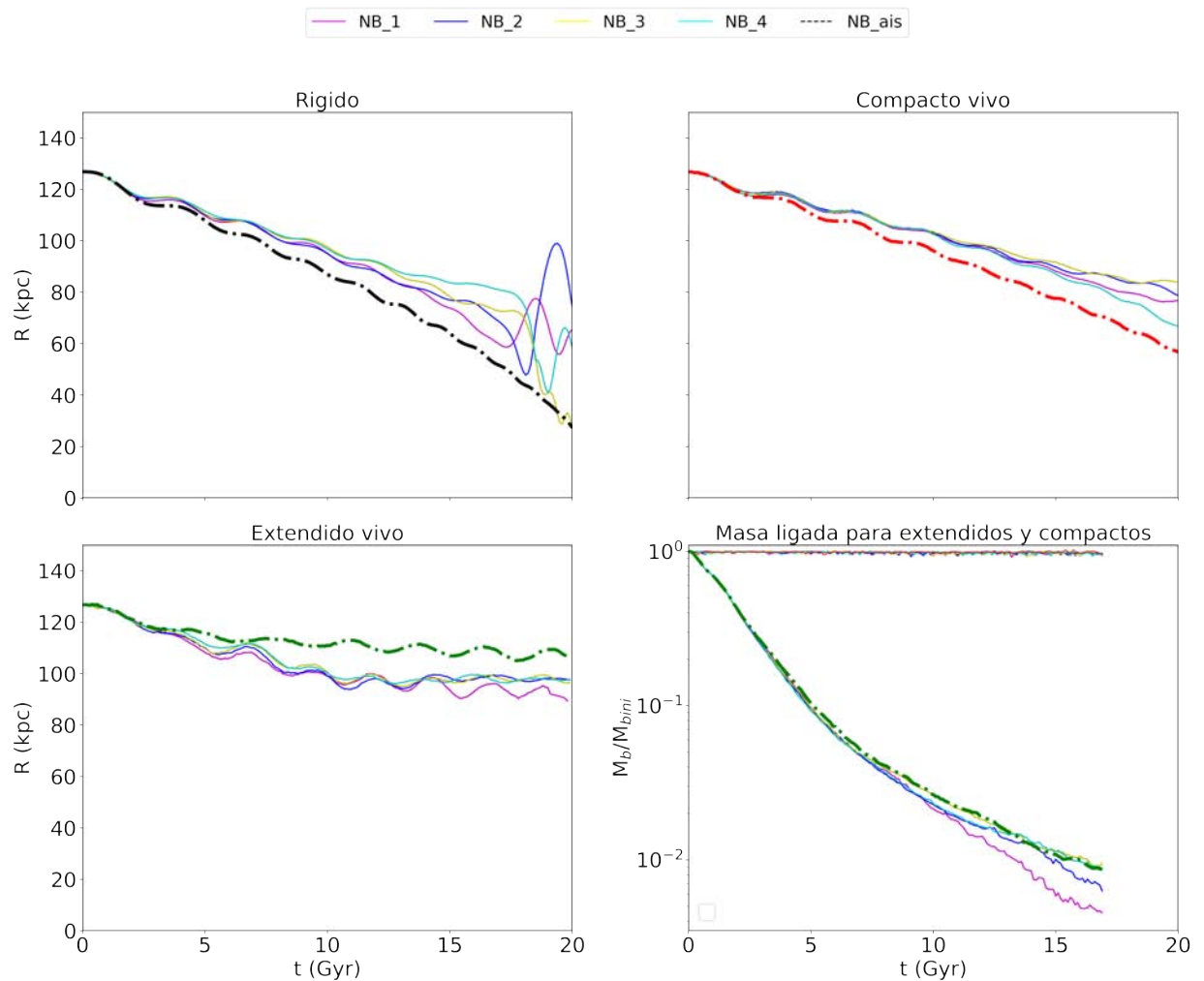


Figura 4.5: Experimento de cuatro satélites coplanares cayendo hacia el pozo de potencial en órbita circular. El panel superior izquierdo representa cuatro satélites rígidos cayendo hacia el pozo de potencial del halo masivo. El panel superior derecho representa cuatro satélites compactos. Panel inferior izquierdo: representa cuatro satélites vivos. El panel inferior derecho representa el cambio de la masa a lo largo de la simulación para los experimentos con satélites vivos.

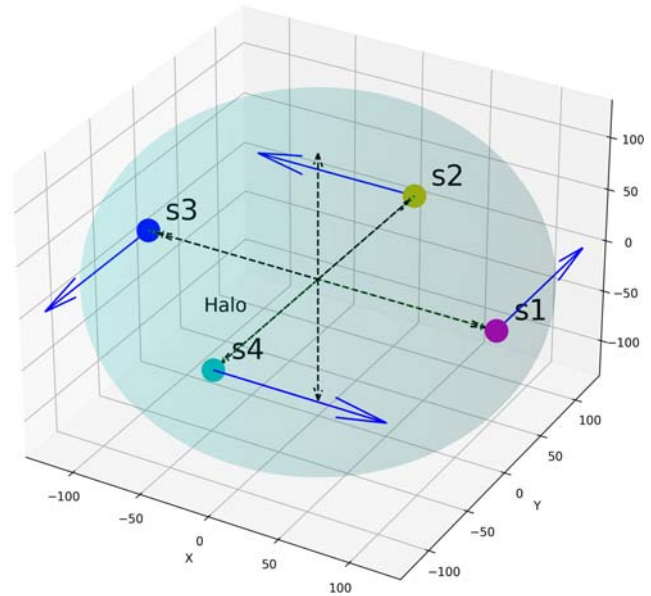


Figura 4.6: Esquema de la posición de cada satélite en el experimento con 4 satélites coplanares a la misma distancia inicialmente, los satélites se marcan con su respectivo color, las líneas azules indican la dirección del vector velocidad de cada satélite. La membrana azul es una representación para remarcar la presencia de las partículas del halo (no está a escala).

La figura 4.5 representa el experimento en el que cuatro satélites coplanares caen. El caso de satélites visto como partículas rígidas más parecidos a cúmulos globulares, el caso de satélites compactos visto como sistemas compactos más parecido a galaxias enanas, el caso de satélites extendidos más parecidos a galaxias espirales extendidas. Y su respectiva pérdida de masa.

Los experimentos con satélites vivos coplanares (caso cuatro satélites), mostraron un comportamiento diferente al caso de satélites rígidos los cuales caen más lento que el respectivo caso aislado. En el caso de satélites vivos extendidos los satélites caen simultáneamente más rápido que el caso aislado. Esto es debido a que en el caso en el que los satélites son extendidos la pérdida de masa es mayor, esto a su vez parece una contradicción con la fórmula de CH. Lo que postulamos en este trabajo es que en el caso coplanar con múltiples satélites, hay un aumento en la auto-fricción dinámica mayor al reportado en la literatura por ejemplo Miller et al. (2020), este efecto en el caso de varios satélites cayendo simultáneamente es algo que no ha sido reportado. En la figura 4.7 vemos que el error relativo en energía se mantiene dentro del orden de 5×10^{-5} .

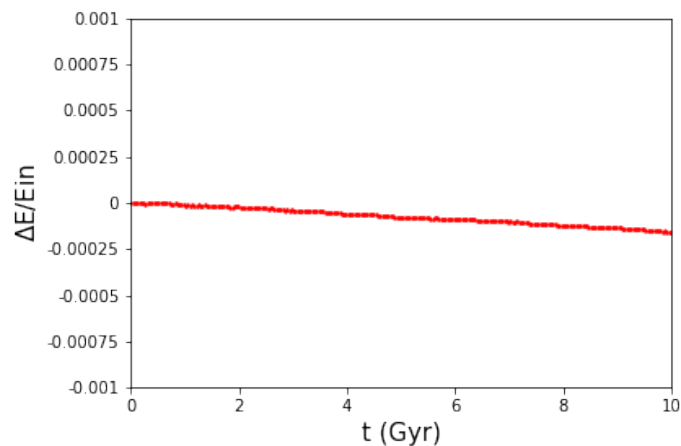


Figura 4.7: Error relativo en energía para el experimento con cuatro satélites coplanares equidistantes inicialmente.

4.4. Configuraciones con satélites no coplanares

A continuación exploramos distintos experimentos en los que los satélites coexisten en diferentes planos. Exploramos el caso en el que los satélites son equidistantes con el centro de masa del sistema y cuando están a diferentes distancias del centro. Estas pruebas servirán para reforzar la idea de que en el caso en el que los satélites son coplanares, vemos la mayor cantidad de efecto en las órbitas.

4.4.1. Experimento con satélites vivos en diferentes planos orbitales y equidistantes del centro

El siguiente experimento consiste en una configuración de cuatro satélites con sus órbitas coexistiendo en diferentes planos durante la caída e iniciando con velocidad circular, ver tabla 4.3, cabe mencionar que estos satélites son equidistantes con el centro. La línea verde sólida es el caso del satélite aislado con la órbita correspondiente. Es de notar que si bien hay algunas órbitas semejantes al caso del satélite aislado, algunas de ellas como la línea magenta queda por encima, y la línea azul cyan queda por debajo. Estas diferencias se verán acrecentadas en el caso en el que los satélites son coplanares, como se muestra en la figura del artículo más adelante.

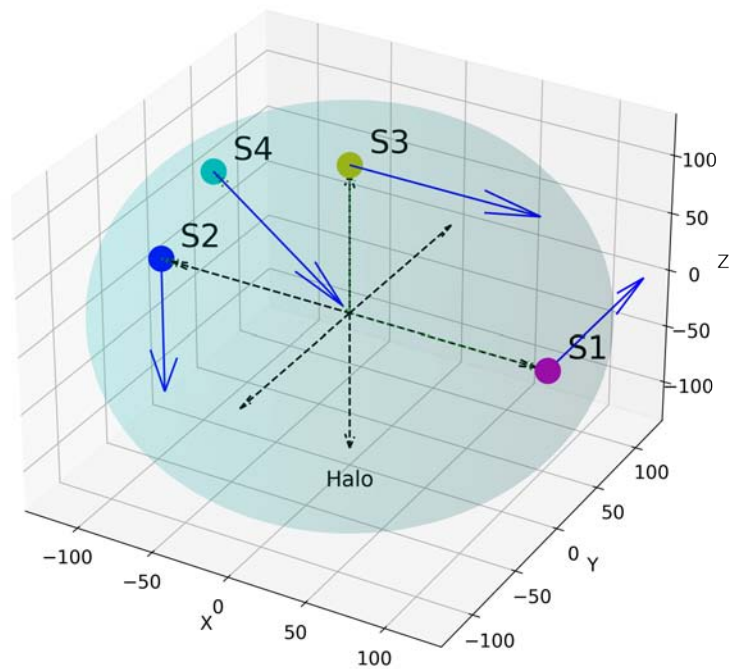
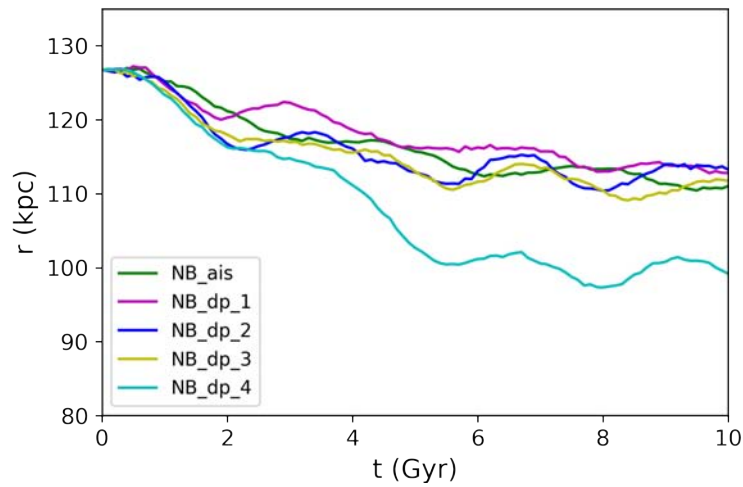


Figura 4.8: El panel superior representa la gráfica de r vs t para cada uno de los satélites en el experimento en el que el movimiento de los satélites no ocurre en el mismo plano. La notación dp indica que las órbitas de los satélites están orientadas en diferentes planos. El panel inferior muestra un esquema de la posición de cada satélite con su respectivo color, las líneas azules indican la dirección del vector velocidad. La membrana azul es una representación para remarcar la presencia de las partículas del halo (no está a escala).

Perturbador	Posición (kpc)			Velocidad (km/s)		
	x	y	z	Vx	Vy	Vz
NB1	126.756	0.0	0.0	0.0	182.77	0.0
NB2	-126.756	0.0	0.0	0.0	0.0	-182.77
NB3	0.0	0.0	126.756	182.77	0.0	0.0
NB4	-89.63003	0.0	89.63003	129.2379	0.0	-129.2379

Cuadro 4.3: Condiciones iniciales para cada satélite en un experimento con 4 satélites cayendo hacia un halo masivo, solo que estos 4 satélites tienen sus vectores velocidad orientados en diferentes direcciones o planos.

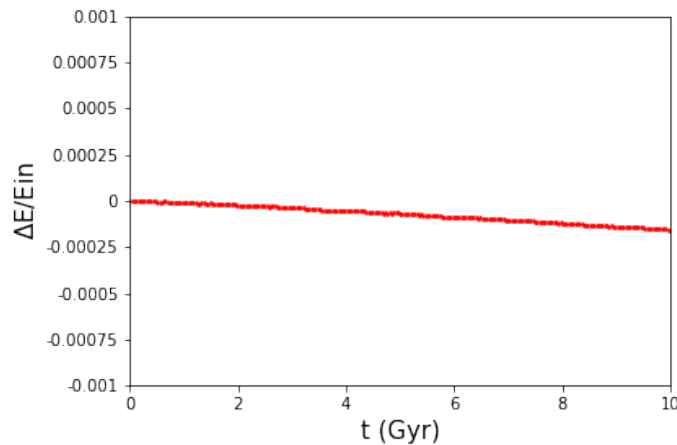


Figura 4.9: Error relativo en energía para el experimento con cuatro satélites orientados en diferentes planos inicialmente.

En el panel superior de la figura 4.8 podemos ver el experimento con 4 satélites orientados en diferentes planos inicialmente, es interesante ver que tres de los satélites mantienen su trayectoria consistente con el caso de un satélite aislado (línea verde sólida), sin embargo, uno de los satélites parece desviarse hacia la parte interna, esto puede deberse a como interactúa con la respuesta ocasionada con los otros satélites o que se ve rezagado de alguna manera debido al “shift” del baricentro, pero este caso en particular requiere un análisis más detallado.

4.4.2. Satélites en diferentes planos orbitales y diferentes distancias iniciales del centro

En este experimento exploramos el caso en el que los satélites coexisten en diferentes planos y también se encuentran a diferente distancia inicialmente.

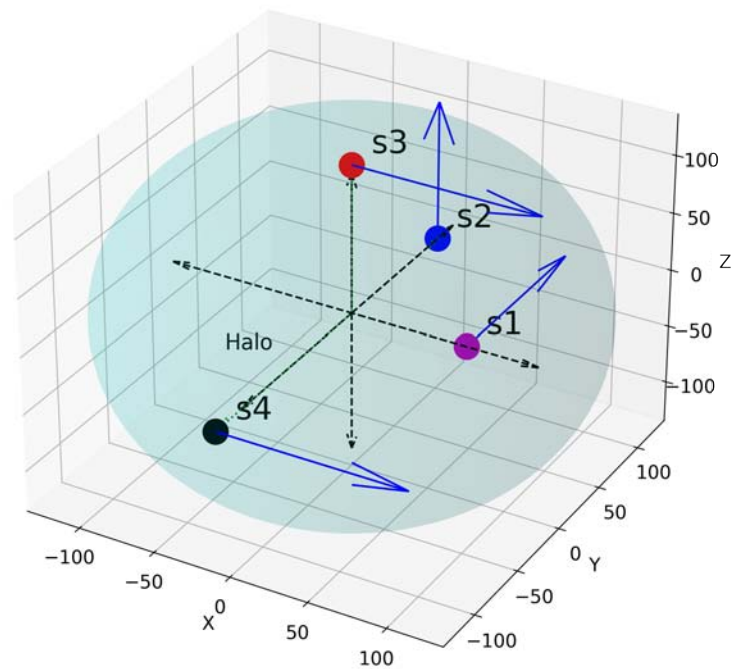
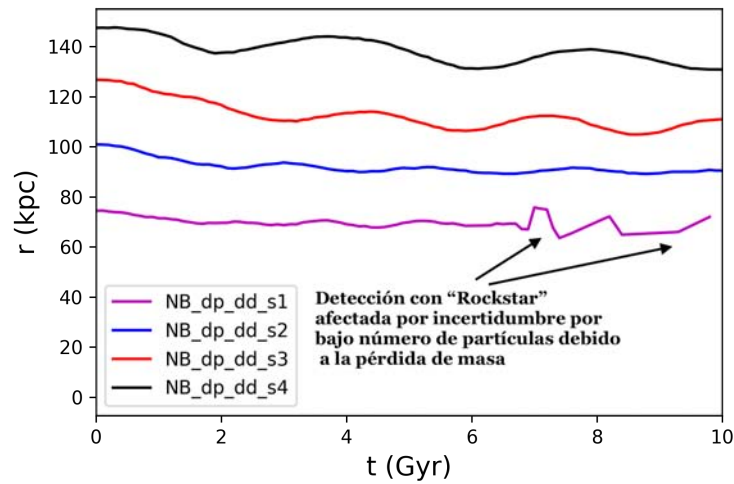


Figura 4.10: El panel superior representa la gráfica de r vs t para cada uno de los satélites en el experimento en el que el movimiento de los satélites no ocurre en el mismo plano, los satélites se encuentran a diferentes distancias desde el inicio de la simulación. La notación dp indica que las órbitas de los satélites están orientadas en diferentes planos, dd indica que los satélites se encuentran inicialmente a diferentes distancias. El panel inferior muestra un esquema de la posición de cada satélite con su respectivo color, las líneas azules indican la dirección del vector velocidad. La membrana azul es una representación para remarcar la presencia de las partículas del halo y no está a escala.

Perturbador	Posición (kpc)			Velocidad (km/s)		
	x	y	z	V _x	V _y	V _z
NB1	74.4343	0.0314	0.0286	0.5175	203.15	-0.1004
NB2	-0.0214	100.94	0.0286	0.2521	0.0377	192.18
NB3	-0.0214	0.0314	126.78	183.39	0.0662	-0.0352
NB4	0.0214	-147.54	0.0286	176.52	0.1560	-0.0217

Cuadro 4.4: Condiciones iniciales para cada satélite en un experimento con 4 satélites cayendo hacia un halo masivo, solo que estos 4 satélites tienen sus vectores velocidad orientados en diferentes direcciones y planos e inicialmente se encuentran a diferentes distancias.

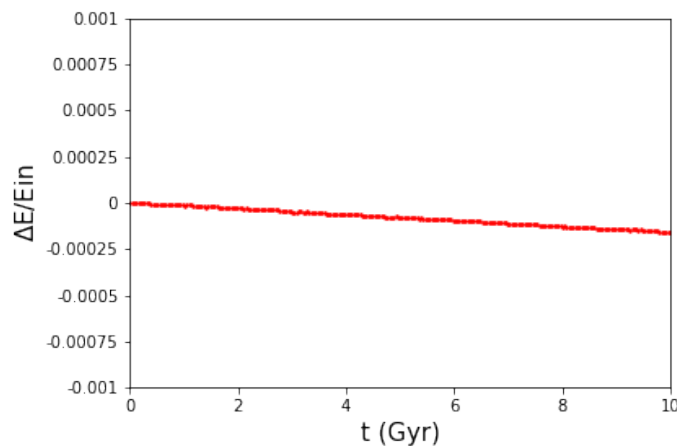


Figura 4.11: Error relativo en energía para el experimento con cuatro satélites no coplanares ni equidistantes.

En el panel superior de la figura 4.10 vemos el resultado de 4 satélites vivos extendidos cayendo simultáneamente, cuando están inicialmente orientados en diferentes planos y su distancia inicial es diferente. Lo que podemos observar de este experimento es que el efecto sobre las órbitas es pequeño, esto se debe a que los satélites no parecen interactuar con los modos principales de la respuesta en densidad debido a que las trayectorias en los planos no se cruzan entre ellos y que están a diferente distancia del centro desde el inicio. En la figura 4.11 vemos que el error relativo en energía se mantiene dentro del orden de 5×10^{-5} .

En esta tesis nos propusimos cuantificar los cambios en la historia orbital de una galaxia satélite durante su caída debido a todos los efectos aquí enunciados llamados “efectos colectivos” en la presencia de otros perturbadores o galaxias satélites. Para este propósito en este capítulo analizamos el impacto que diferentes parámetros físicos tienen sobre el proceso de caída, tales como: la estructura interna de los satélites, su distribución espacial dentro de un halo anfitrión, la pérdida de masa y el efecto combinado de la auto-fricción. En particular la auto-fricción dinámica no ha sido estudiada antes para el caso de varios satélites cayendo simultáneamente. Demostramos que los efectos colectivos pueden cambiar la historia de

caída, la comparación de las órbitas en los casos con múltiples satélites con el caso aislado muestra diferencias en todos los casos. Estas diferencias pueden llegar a ser de hasta un 40 % o más en los apocentros y excentricidades como lo muestran los experimentos inicializados con datos de los satélites en la **VPOS**. En el caso coplanar con varios satélites, similar a la **VPOS**, vimos que el efecto de la auto-fricción dinámica puede ser mayor al 15 % reportado por [Miller et al. \(2020\)](#). El no tomar en cuenta los efectos colectivos puede dar a lugar a interpretaciones erróneas en reconstrucción orbital, el perfil de densidad y la masa del satélite, la forma del halo y la estimación de los tiempos de caída de los satélites. Los resultados obtenidos para el caso de los satélites de la **VPOS** se muestran en un artículo titulado “Concurrent infall of satellites, Collective effects changing the overall picture”, publicado en la revista *Astronomy & Astrophysics*, 2210.00702.

Concurrent infall of satellites

Collective effects changing the overall picture

A. Trelles¹, O. Valenzuela¹, S. Roca-Fábrega² and H. Velázquez¹

¹ Universidad Nacional Autónoma de México, Instituto de Astronomía, AP 70-264, CDMX 04510, México
e-mail: jtrelles@astro.unam.mx

² Departamento de Física de la Tierra y Astrofísica and IPARCOS, Facultad de Ciencias Físicas, Plaza Ciencias, 1, Madrid, E-28040, Spain

October 3, 2022

ABSTRACT

Context. In recent years, high-quality observational data have allowed researchers to undertake an extensive analysis of the orbit of several Milky Way satellite galaxies, with the aim to constrain its accretion history. Although various research groups have studied the orbital decay of a satellite galaxy embedded inside a dark matter halo, a large variety of new physical processes have been proven to play an important role in this process, but its full scope not yet understood.

Aims. Our goal is to assess whether the orbital history of a satellite galaxy remains unchanged during a concurrent sinking. For this purpose, we analyzed the impact that the internal structure of the satellites and their spatial distribution inside the host halo may have on the concurrent sinking process due to both mass loss and the combined effect of self-friction – as processes that have not been studied before for the concurrent sinking of satellites.

Methods. We set up a set of N-body simulations that includes multiple satellites that are sinking simultaneously into a host halo and we compared them with models that include a single satellite.

Results. The main result of our work is that the satellite's accretion history differs from the classical isolated view when we consider the collective effects. Furthermore, the accretion history of each satellite strongly depends on the initial configuration, the number of satellites present in the halo at the time of infall, and the internal properties of each satellite. We observe that compact satellites in a flat configuration fall slower than extended satellites that have lost mass, showing a non-reported behavior of dynamical self-friction; the latter is reinforced by analytical expressions that describe the orbital decay through different approaches for the dynamical friction, including (or not) the mass loss and radial dependence of the satellite. In particular, we find that such effects are maximized when satellites are located in a flat configuration. Here, we show that in a flat configuration similar to the observed vast polar structure, deviations in the apocenters can be of about 30% with respect to the isolated case, and up to 50% on the eccentricities.

Conclusions. Overall, we conclude that ignoring the collective effects produced by the concurrent sinking of satellite galaxies may lead to large errors in the determination of the merger progenitor properties, making it considerably more challenging to trace back the accretion event. Timing constraints on host density profile may be modified by the effects discussed in this paper.

Key words. Galaxy: halo – galaxies: formation – methods: numerical

1. Introduction

The streamers of the Milky Way are true fossil records providing information about their formation and evolution history (Simon 2018; Nadler et al. 2020; Vera-Ciro & Helmi 2013; Law & Majewski 2010). Dynamical friction (DF, hereafter) is arguably the most critical physical process driving the sinking of galactic systems in a gravitational potential well. It was first studied by Chandrasekhar (1943) (CH, hereafter), based on a simple model that assumes infinite and homogeneous media, while simultaneously capturing the dominant physical effects.

Other authors have proposed new strategies to take into account other factors involved on the DF problem, beyond the Chandrasekhar approximation. Among them, the most relevant are those considering a more realistic host system, that is, with a finite size and a radially dependent density profile (Binney 1977; Boldrini et al. 2018), as well as those considering that realistic host systems such as those in N-body models can be well described by a simple Coulombian logarithm with spatial or mass dependences (Hashimoto et al. 2003; Just & Peñarrubia 2005;

Colpi et al. 1999; Boylan-Kolchin et al. 2008). All the new models have allowed researchers to recover and study the slowdown of the DF efficiency, while extending the realism of the system. Finally, internal changes in the perturber may also affect its sinking process and this should be taken into account when studying the accretion of satellite systems into central galaxy halos (Dominguez-Tenreiro & Gomez-Flechoso 1998).

All the aforementioned studies on simplified models of satellites or globular clusters sinking into central galaxies can now be used to better understand the formation, evolution, and properties of local galaxies including the Milky Way. Far from the idealized models presented above where a single object sinks into a central halo, observations and simulations show that central halos suffer a continuous and concurrent infall of satellites (e.g., Rocha et al. 2012). The collective effects of the concurrent infall of satellites is a process that is pending further study. In this work we address it, for the first time, in a systematic way, showing that the sinking orbit of each individual satellite is sensitive to the relative configuration of all satellites inside the main halo

Table 1: Initial parameters of the host and the satellite models used in the N-body simulations.

	Host	Sat. comp.	Sat. Intermed.	Sat. ext.
M_{200}/M_{\odot}	1.2×10^{12}	10^{10}	$10^9 - 10^{10}$	10^{10}
N_p	2×10^6	1.3×10^4	$1.3 - 13 \times 10^3$	1.3×10^3
r_s [kpc]	15	0.5	3	10
$r_{1/2}$ [kpc]	20	1.5	5.5	18
$M_{1/2}/M_{\odot}$	6×10^{12}	5×10^{11}	$0.5 - 5 \times 10^{11}$	5×10^{11}
R_{tr} [kpc]	200	1	5	25
ϵ [kpc]	0.3	0.3	0.2–0.4	0.6

Notes. The satellite models include rigid, compact live (globular cluster type), intermediate live (dwarf galaxy type), and extended live (early type). From top to the bottom: Total mass (M_{200}), the number of particles (N_p), the scale radius (r_s) of the NFW density profile, the half mass radius and half mass ($r_{1/2}, M_{1/2}$), the truncation radius (R_{tr}), and the smoothing length (ϵ). For rigid satellite $M_{200}/M_{\odot} = 1 \times 10^{10}$, $\epsilon = 0.6$ kpc

and also to their internal structure. In particular, we have given some attention to an specific configuration similar to the recent plane such as the configuration that has been found for several galactic systems, known as the vast polar structure (Lynden-Bell (1976); Kunkel & Demers (1976), VPOS, hereafter).

2. Simulations and methodology

2.1. Methodology

To carry out the research presented here, we generated a series of N-body experiments that include a spherical distribution of particles initially in dynamical equilibrium (central system) and one or various perturbers. We compare the evolution of these N-body systems (NB, hereafter) with the results from semianalytical models (SA, hereafter) as described below.

2.2. Semianalytical models

Our semianalytical fiducial models assume a cuspy dark matter halo profile (Navarro et al. 1996) for the host (NFW, hereafter), the satellites are represented by a rigid Plummer model and we include the CH dynamical friction. Additionally, we incorporated the mass loss of the sinking satellite following Klypin et al. (1999). We also took into account a non-homogeneous density profile for the host by incorporating a radially dependent Coulombian logarithm. For simplicity, we chose the strategy proposed by Hashimoto et al. (2003) (see their equation 2). We used these semianalytical experiments as control models to guarantee accuracy in the simulations in the regime in which SA and NB simulations are comparable and to identify the relevant physical processes from the ones included in the SA models.

2.3. N-body simulations

The set of NB simulations we used here include a central system (the host, hereafter) that is modeled as a simple dark matter halo with a NFW density profile and composed of particles that interchange energy (hereafter, live systems), plus one or several perturbers (satellites, hereafter). Our suite of models includes systems with rigid (softened Plummer), and either compact live, intermediate live, or extended live (see table 1). All N-body models are labeled as NB_x where $x=r, c, e$ and corresponds to a Plummer rigid satellite, a self-consistent compact satellite and an ex-

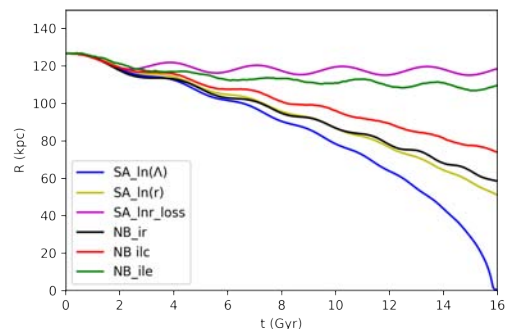


Fig. 1: Sinking of single satellites in a MW-mass halo. Blue line: SA fiducial model as a Coulombian logarithm with no spatial dependence; Yellow: SA using a Coulombian logarithm with radial dependence; Magenta: SA using a Coulombian logarithm with radial dependence and a mass loss; Black: NB model with a satellite represented with a rigid softened Plummer density profile (rigid system, hereafter), $\Lambda = \frac{b_{max}}{b_{90\%}} \approx 5$, following Binney & Tremaine (1987) eq. 8.1b ; Red: NB model including a satellite with a compact NFW density profile; Green: NB model including a satellite with an extended NFW density profile (for more details on the models see Sect. 3.1 and table1). The initial condition for the satellite in the models is $r = 126$ kpc, $v_c = 183$ cm/s.

tended NFW satellite, respectively. In all models, the host has a mass of $M_{200} = 1.2 \times 10^{12} M_{\odot}$ and a total of 2×10^6 particles. The number of particles in the satellites varies to keep the mass ratio of the host same to the satellite particles. The softening of the host and satellite particles is taken as the average particle separation within the scale radius (see Table 1). The initial conditions were obtained using a NFW model, which includes a cutoff in the form of an hyperbolic secant weight function, and depends on the truncation radius R_{tr} parameter. To create the initial conditions we used the subroutine named "mkhalo" from Dehnen & McLaughlin (2005) that can be found in the NEMO toolkit package (Barnes et al. 2010). We used the public N-body code "gyrfalcon" which is a tree code with complexity $\sim O(N)$, combined with a fast multipole method, (FMM; Dehnen (2014)) to evolve the aforementioned initial conditions for 10–20 Gyr. To track the path of live satellites in the NB simulations, we used the Rockstar halo finder (Behroozi et al. 2013).

3. Results

3.1. Dynamical Friction: Single satellite sinking using N-body and semianalytical models

In Figure 1, we present the comparison between our SA and NB simulations. In this set of six models, the complexity of the simulations is gradually increased. In the first three models, we used a SA approach where DF is first modeled the classical Chandrasekhar description using a constant Coulomb logarithm and no mass loss (blue); then we used a spatially dependent Coulomb logarithm and no mass loss (yellow); and in the last run (magenta), we used a spatially dependent Coulomb logarithm and including a mass loss equal to the one corresponding rigid satellite NB experiment (solid-black line). The last three models are NB models consisting of a live host galaxy and a single satellite that is modeled as rigid (black), compact and live (red), and extended

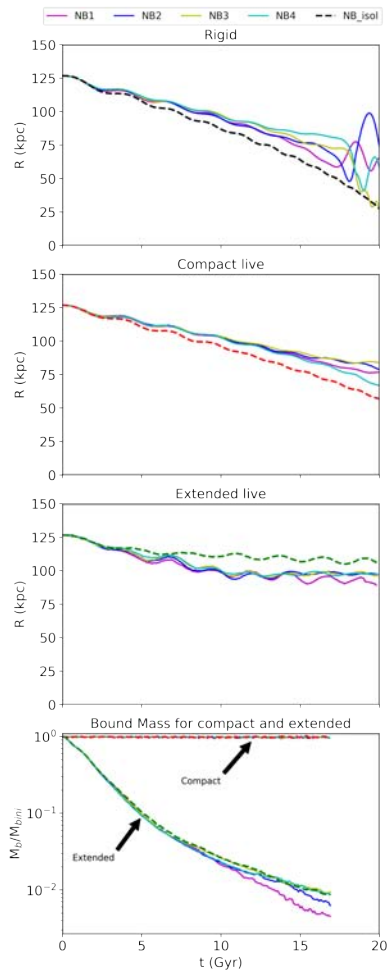


Fig. 2: NB models of a MW-mass halo host with four sinking satellites in the same orbital plane and in a fully symmetric configuration (see Figure 4), and with initial speeds corresponding to circular orbits (upper panels). From top to the bottom, satellites simulated as a rigid smoothed Plummer density distribution, as a compact NFW, and as an extended NFW. For comparison purposes, we include the sinking process of a single satellite in all panels (dashed lines). Mass evolution of the four compact and extended satellites in their collective sinking process (bottom panel). The dashed lines show the same process when only one satellite sinks. Satellites start at $r = 126$ kpc and with $v_c = 183$ km/s

and live (green). This first set of six models is relevant in our work as it illustrates agreement between the SA and NB models when the former includes the physical processes that play a dominant role in the NB simulations (e.g., yellow vs. black and pink vs. green lines in Figure 1). They serve as our reference models to compare with the multiple satellite configurations with the aim to identify (or search) for dynamical processes that are not taken into account.

3.2. Concurrent satellite sinking and internal structure

The multiple satellite case is complicated because, in addition to the standard processes discussed in the previous section, satellites interact among them as well as with the corresponding halo response and the mass lost by the satellites. These effects affect each other and their relative role in the sinking rate can not be

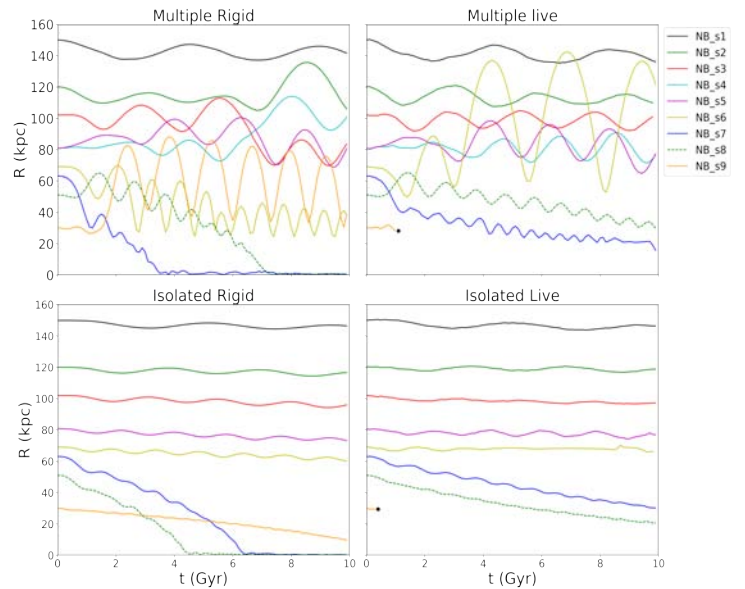


Fig. 3: NB models of a MW-mass halo host with nine sinking satellites located in the same orbital plane but falling from different angular and radial positions (see Figure 4, bottom panels). All satellites have an initial speed corresponding to the one of a circular orbit. *Top-left panel*: Simultaneous infall of the nine satellites simulated as a rigid smoothed Plummer density distribution. *Top-right panel*: Similar to the top-left panel, but satellites are now described by an "intermediate" compact NFW density distribution (see column 4 in table 1). *Bottom-left panel*: Sinking of each one of the rigid satellites in the top-left panel, individually, meaning that no other perturber is present in the halo. *Bottom-right panel*: Similar to the bottom-left panel but for the "intermediate" compact satellites. In each of the bottom panels we show eight independent experiments, where the initial mass profile of the host halo has been modified to include the mass of the remaining non-simulated satellites at its initial radial location. We excluded the cyan line case because it is the same as the magenta case. Black points indicate where the Rockstar halo finder cannot find the satellite because it lies in the central part of the host, See Table 2 for the initial conditions.

Table 2: Initial parameters for the satellites in the coplanar experiments.

NB	r (kpc)	v_t (10 km/s)	M_\odot
S1	30.16	21.68	10^9
S2	51.07	21.26	10^{10}
S3	63.18	20.95	10^{10}
S4	69.21	20.84	10^9
S5	81.07	20.31	10^9
S6	81.07	20.31	10^9
S7	102.06	19.44	10^9
S8	120.06	18.75	10^9
S9	150.05	17.71	10^9

Notes. From left to right: the radial distance, the tangential velocity to that radio, and the initial mass for the satellites in experiments from Figure 3. The corresponding live satellites in this experiments corresponds to the intermediate compact satellite that appears in column 4 in Table 1 and these satellites lie on a plane.

trivially separated. As our main objective is to study differences between isolated models and models with many perturbers, we set a spatial configuration that, after performing many experiments with random positions for the satellites, we found it maximizes the collective effects. This configuration is a fully symmetric and flat (coplanar) distribution (see Figure 4, upper panels). In Figure 2, we show the results obtained after analyzing three NB models with four rigid, compact live, and extended live satellites, respectively, sinking simultaneously through a live MW-mass halo (pink, blue, cyan, and yellow solid lines), and we compare them with the corresponding experiment of a single live satellite (black, red, and green dashed lines). In the top panel, we show that rigid Plummer satellites sink at a slower rate than the single satellite case (black-dashed). In the second panel, we show the sinking of four compact NFW live satellites and we obtain a similar result as before, satellites sink at a slower rate in the collective scenario than in the single satellite case (red-dashed). However, in agreement with the results presented in Section 3.1, we observe differences between the rigid and the compact live scenarios; as the mass loss is almost negligible in the compact live scenario (as detailed in the discussion below), the only source of these differences can be the internal structure of the live satellite that can work as a sink or source of energy and angular momentum (Dominguez-Tenreiro & Gomez-Flechoso 1998). In the third panel, we show the trajectories of satellites when these are modeled as live extended NFW; we see that extended satellites sink faster in the concurrent scenario than when they sink in isolation (green-dashed). Although it may seem unexpected the result can be explained by analyzing the mass loss suffered by satellites. In the bottom panel, we show the mass loss for the compact and extended satellites. As expected, this figure confirms that extended satellites in the collective scenario suffer higher mass loss in comparison with the compact ones. If we focus only on the extended satellites, we see that the ones sinking faster also suffer higher mass loss than the isolated ones, which is unexpected based on Chandrasekhar formula. We hypothesize that the interaction of the stripped material from these satellites among themselves and the others increases the self-friction and, thus, the sinking rate.

We assert that this contrasting behavior between the single satellite and the four extended ones shows an aspect of self-friction that has not been studied before. Therefore, the combined effect of stripped material of nearby satellites may increase their sinking speed. This may suggest that our assumption of a coplanar symmetric configuration precludes our conclusions from proving relevant in real galaxies. Although it is not infrequent to find satellite galaxies sinking in small groups, in pursuit of the answer to this question, we study a more realistic case, described in the next section.

3.3. Connection with realistic systems: Vast polar structure

In the previous section, we show that the sinking history of four satellites, initially located in a symmetric and coplanar configuration, dramatically differs from one that involves solely an individual satellite. While it may appear idealistic, such a coplanar symmetric configuration somehow is in agreement with recent studies reporting the so-called vast polar plane of satellite galaxies (VPOS) in the local group and other galaxies (Kroupa et al. 2005; Pawlowski et al. 2012). To better study the connection between our results and the VPOS problem, in this section we analyze experiments considering nine satellites in a plane configuration resembling the VPOS, using the relative orientations presented in Pawlowski & Kroupa (2020) (see also Figure 4, bot-

tom panels). It is important to clarify that we do not pretend to develop a realistic MW+satellite model; rather, we built a test system with a similar flatness as the VPOS in order to test the relevance of our former results. Additionally, we ran some experiments where the satellites were located at different distances out of the plane and we also used both the relative positions and velocities presented in Pawlowski & Kroupa (2020). The results are consistent with our general conclusions, with their statistical properties are discussed in the next section. However, a detailed discussion will be presented in a forthcoming work. In the following, we analyze two experiments, one with nine rigid satellites and another with nine "intermediate" ones (see column 4 in Table 1 and its description). In these experiments, the two satellites marked as blue and green dashed line, have an initial mass of $10^{10} M_{\odot}$, while the rest of the others have a mass of $10^9 M_{\odot}$.

The results are shown in Figure 3. As a reference, we show the isolated sinking history of each satellite in the lower panels, both for the rigid (bottom-left) and the live (bottom-right) representations. It is important to notice that in the reference models (isolated satellites) that appear in lower panels of Figure 3, we update the initial mass profile of the halo to include the mass due to the more internal satellites than the respective isolated case and that would be present in the case of multiple satellites.

Our analysis of both scenarios (rigid and live) shows that most of the trajectories of the satellites are affected when including the collective effects. We observe a wide range of variations, from subtle changes in the orbits of outermost satellites (s1, black curves) to more extreme ones such as what is seen in the s6 or s3, and also complex interactions between perturbers. However, the scope of this experiment is not to analyze the sinking of each individual satellite, but to demonstrate that collective effects change the orbits of the satellites in a way that would make it challenging to recover the infall history of satellite galaxies based on current observations. This is true for such compact systems as globular clusters (i.e., similar to rigid satellites) and for satellites such as Sagittarius (i.e., similar to the "intermediate" live satellites). Only systems in the outermost regions of the host halo suffer minor perturbations, and so their orbits can be easily tracked back in time.

3.4. Exploring the statistical robustness

To quantify the statistical stability of our results, we generated an ensemble of 20 N-Body models of a host MW-size galactic halo and satellites, as shown in Table 2 from Pawlowski & Kroupa (2020). In each model of the ensemble, we changed satellites' positions and velocities, randomly generated within the ranges defined by the observational uncertainties (Pawlowski & Kroupa 2020; Riley et al. 2019; Garrow et al. 2020); in addition, we followed a Gaussian distribution, so we could estimate the effect of changes in the initial orbital parameters. We note that our models do not include the Leo I and Leo II satellites due to their large uncertainties in GAIA DR2 proper motions.

We studied the orbital parameters of each satellite in each model of our ensemble. In particular, we analyzed the time and distance to the center of mass during apocenters and pericenters to get also the eccentricity of each orbit. For each satellite we tracked the "rms" changes of apocenters and we even computed the rms deviation of eccentricity with respect to the isolated case. In addition, we carried out an extensive study of the number and intensity of close encounters between satellites.

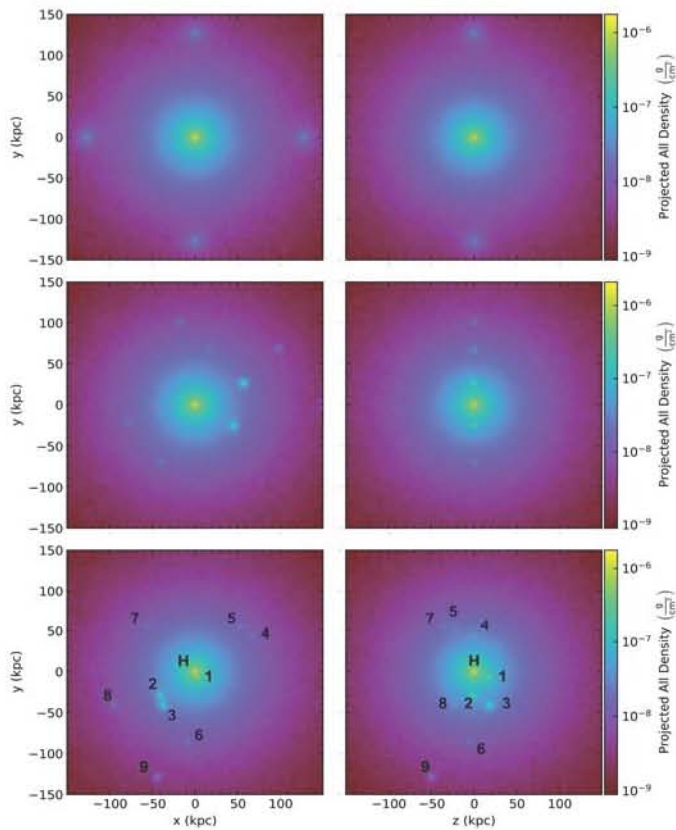


Fig. 4: Density plots showing the satellites’ initial positions in the NB models. *Top panels*: Face-on (left) and edge-on (right) views of the fully symmetric model with four satellites of the same mass that follows an extended NFW density profile. *Central panels*: Face-on (left) and edge-on (right) views of a fiducial model, where satellites lie in a plane as an approximation model to the VPS plane. *Bottom panels*: Face-on (left) and edge-on (right) views of one of the models in our N-body ensemble that follows the distribution of the satellites in the MW and the VPS plane structure; see 3.4 for details. Labels are consistent with the results showed in figure 5, where H stands for the host’s central density.

4. Discussion and conclusions

In this paper, we carry out the first systematic study of the effects of concurrent multiple sinking perturbers. More specifically, we analyze how such effects produce changes on the orbits of satellite galaxies. In conclusion, we find clear differences between the concurrent and single-satellite scenarios, which would be partially explained by the fact that the global (host + satellites) potential is time-dependent. Although close encounters between satellites may happen, we discard them as a dominant mechanism. When looking at the upper panel of Figure 6, we see that none of the ensemble orbits agree with the isolated case regardless of the presence or absence of a close encounter and the orbital drift is increasing with time. However slingshot effects still may happen, combined with the satellites interaction with the local and global halo density response, as well as interactions with the combined stripped material or dynamical self-friction (Figure 2). We enumerate our more relevant results in the following.

- The sinking process of satellities, when including collective effects of several other perturbers, is generally different from that of an isolated satellite. The difference depends on

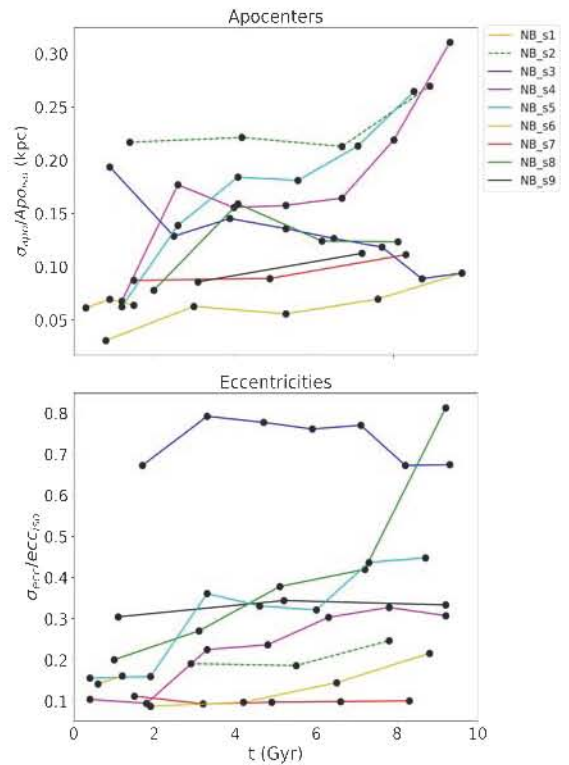


Fig. 5: Standard deviation in the apocenters (top) and eccentricities (bottom) when comparing results from our ensemble of N-body models and the isolated case (see section 3.4). Eccentricity is normalized with the initial eccentricity of the isolated case for every satellite.

the internal properties of the sinking satellites, the relative spatial and velocity configuration between them, the orbital parameters, and the mass ratio between the host and satellites, as well as the possible scatter between the different satellite masses.

- In particular, in the case with multiple compact or rigid satellites, the sinking process is slower than for the isolated ones. In this scenario, most of the effects of these interactions (satellites and host density responses) go into satellite orbital energy, delaying the sinking process (see upper panels in Figure 2). The former situation may be comparable with the compact globular clusters in dwarf spheroidal galaxies such as Fornax or Reticulus 2. However, it is necessary to carry out a future assessment of such case to test whether the rigid and compact satellite assumptions are suitable for globular clusters and also the effect of the relative globular cluster orbit orientation (see also Inoue 2011).
- For extended satellites in concurrent accretion, the mass loss is considerable. The sinking rate is faster in some situations, in comparison with the isolated case due to the combined effect of stripped material, revealing a new aspect of self-friction. Such a collective effect leads to the self-friction amplitude sometimes rising higher than the 10 – 15% quoted by Miller et al. (2020) (see lowest panel in Figure 2).
- We found that the internal density profile of the satellites makes them more or less susceptible to stripping, affecting the satellite sinking rate in combination with the already mentioned combined effect of the stripped mass from multiple satellites enhancing the DF.

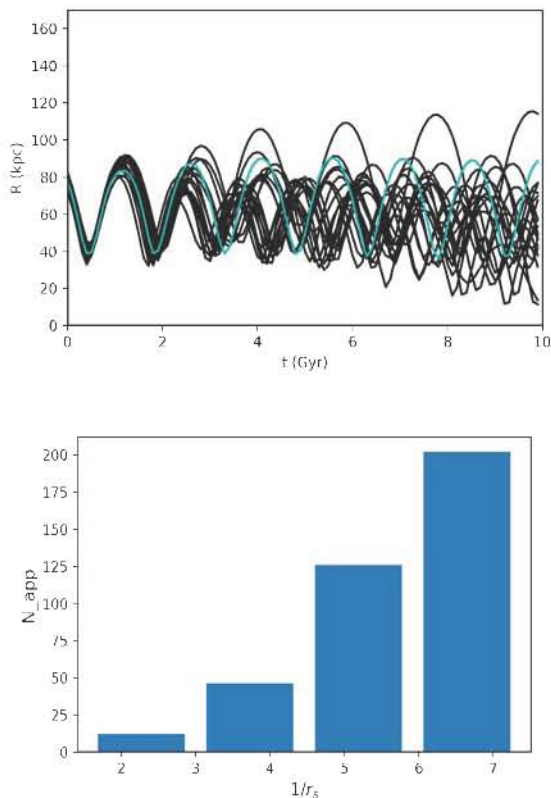


Fig. 6: Comparison of the orbit of a particular satellite (S5/6 according to table 2 nomenclature) when run in isolation (cyan line) or in each one of the 20 N-body models of our ensemble, shown in the top panel (black solid lines). This figure shows the scatter between realizations that we quantified in Figure 5, for one of the satellites. Bottom panel shows the histogram of distances of all close encounters between two satellites in the full ensemble of N-body models.

- We found that collective effects are more dramatic in such coplanar systems as the so-called Milky Way VPOS (see Figure 3), this is because the mutual perturbations are more frequent, nearby, and stronger, making the sinking satellite history quite different to the isolated case. Although we did not pretend to simulate the VPOS formation and evolution, we can point out that the sinking rate can diverge from the results of single satellite infall, it may temporally reduce or reverse the sinking process.
- We built an ensemble of twenty models for the satellite system similar to the VPOS in order to quantify the rms change regarding the isolated case for satellite orbital parameters (apocenter and eccentricity) during concurrent evolutionary trajectories (see Figures 5 and 6). In general, the orbital evolution does not correspond to the isolated case, however, in many situations the difference is mostly quantitative, although the difference is systematically growing and in some cases the orbit is completely different. We also estimate that the fraction of satellite binary encounters closer than the characteristic halo radius (3 kpc) is 5% (as we see in the lower panel of Figure 6), indicating that the observed orbital drift is triggered not only for close encounters, but also for the satellites' interaction with the local and global halo density response, as well as the tides of the combined stripped

material or dynamical self-friction. In the future, we plan to explore all of these effects with a code that efficiently captures such physical processes (Aguilar-Argüello et al. 2022).

- As a general lesson, it is not always trivial to disentangle the history of a sinking satellite accretion by only taking the instantaneous position and velocity and a simple global potential, without consider the whole galaxy system evolution or at least the most relevant perturbers. Not considering the correct complexity may result in misleading conclusions about the mass or accretion time of the perturber, or about the overall shape and evolution of the host gravitational potential.
- Thus, models that include the infall of a single satellite (e.g., Sagittarius or the LMC or SMC systems) have a limited predictive power. In this paper, we emphasize that collective effects need to be taken into account when studying the sinking of satellites – not only in the Milky Way, but in other systems as well. Recently, other works such as (D'Souza & Bell 2022), reinforced the idea that satellite orbital history is affected by large uncertainties related with the host or most massive subhalos, the collective effects that we discuss here strengthen this picture.

We emphasize again that not considering the collective effects in the orbital reconstruction may lead to misleading conclusions about the mass or accretion time of the perturber and the overall shape and evolution of the host gravitational potential.

Acknowledgements. AT and OV acknowledges support from a DGAPA-UNAM grant IN112518, IG101222 and AG101620. HV acknowledges support from PAPIIT-UNAM under grant IN101918. The authors thank for the facilities of cluster computers: Atocatl and Miztli used for this projects. SFR acknowledges support from a Comunidad de Madrid postdoctoral fellowship under grant number 2017-T2/TIC-5592. His work has been supported by the Madrid Government under the Multiannual Agreement with UCM in the line Program to Stimulate Research for Young Doctors in the context of the VPRICIT under grant number PR65/19-22462. SRF also acknowledge financial support from the MINECO under grant number AYA2017-90589-REDT, RTI2018-096188-B-I00, and S2018/NMT-429. This research was partially supported through computational and human resources provided by the LAMOD UNAM project through the clusters Atocatl and Tochtli. LAMOD is a collaborative effort between the IA, ICN, and IQ institutes at UNAM. The authors thank the anonymous referee for the comments and support.

References

- Aguilar-Argüello, G., Valenzuela, O., & Trelles, A. 2022, *A&A*, 663, A93
- Barnes, J., Hut, P., & Teuben, P. 2010, NEMO: A Stellar Dynamics Toolbox
- Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, *ApJ*, 762, 109
- Binney, J. 1977, *MNRAS*, 181, 735
- Binney, J. & Tremaine, S. 1987, *Galactic dynamics* (Addison-Wesley)
- Boldrini, P., Mohayaee, R., & Silk, J. 2018, arXiv e-prints, arXiv:1806.09591
- Boylan-Kolchin, M., Ma, C.-P., & Quataert, E. 2008, *MNRAS*, 383, 93
- Chandrasekhar, S. 1943, *ApJ*, 97, 255
- Colpi, M., Mayer, L., & Governato, F. 1999, *ApJ*, 525, 720
- Dehnen, W. 2014, *gyrfalcON: N-body code*
- Dehnen, W. & McLaughlin, D. E. 2005, *MNRAS*, 363, 1057
- Dominguez-Tenreiro, R. & Gomez-Flechoso, M. A. 1998, *MNRAS*, 294, 465
- D'Souza, R. & Bell, E. F. 2022, *MNRAS*, 512, 739
- Garrow, T., Webb, J. J., & Bovy, J. 2020, *MNRAS*, 499, 804
- Hashimoto, Y., Funato, Y., & Makino, J. 2003, *ApJ*, 582, 196
- Inoue, S. 2011, *MNRAS*, 416, 1181
- Just, A. & Peñarrubia, J. 2005, *A&A*, 431, 861
- Klypin, A., Gottlöber, S., Kravtsov, A., & Khokhlov, A. 1999, *ApJ*, 516, 530
- Kroupa, P., Theis, C., & Boily, C. M. 2005, *A&A*, 431, 517
- Kunkel, W. E. & Demers, S. 1976, in *Galaxy and the L.G.*, Vol. 182, 241
- Law, D. R. & Majewski, S. R. 2010, *ApJ*, 714, 229
- Lynden-Bell, D. 1976, *MNRAS*, 174, 695
- Miller, T. B., van den Bosch, F. C., Green, S. B., & Ogiya, G. 2020, *MNRAS*, 495, 4496
- Nadler, E. O., Wechsler, R. H., Bechtol, K., et al. 2020, *ApJ*, 893, 48
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, *ApJ*, 462, 563
- Pawlowski, M., Pflamm-Altenburg, J., & Kroupa, P. 2012, *MNRAS*, 423, 1109
- Pawlowski, M. S. & Kroupa, P. 2020, *MNRAS*, 491, 3042
- Riley, A. H., Fattahi, A., Pace, A. B., et al. 2019, *MNRAS*, 486, 2679
- Rocha, M., Peter, A. H. G., & Bullock, J. 2012, *MNRAS*, 425, 231
- Simon, J. D. 2018, *ApJ*, 863, 89
- Vera-Ciro, C. & Helmi, A. 2013, *ApJ*, 773, L4

Capítulo 5

El Aprendizaje Automático y los algoritmos de búsqueda de agrupamiento.

El aprendizaje automático, también conocido como “Machine Learning” **ML**, término acotado primero por [Samuel \(1959\)](#), implica la utilización de algoritmos computacionales diseñados para extraer y comprender patrones en los datos. Estos algoritmos se valen de funciones matemáticas para explotar las fluctuaciones estadísticas en los datos, revelando patrones de comportamiento que son difíciles de identificar mediante métodos convencionales. A través de este proceso, se obtiene información valiosa y patrones que permiten ajustar pesos o probabilidades mediante la monitorización de estas fluctuaciones. Esto se realiza al comparar los datos con muestras controladas (etiquetadas) cuyos resultados son conocidos. Este proceso, conocido como entrenamiento, sienta las bases para realizar tareas como predicciones, clasificación y agrupamiento ([Fradkov, 2020](#)).

Aunque estos algoritmos buscan automatizar tareas de clasificación, agrupamiento y predicción, suele pensarse que requieren mínima interacción con el usuario y que operan de manera autónoma al analizar los datos. Sin embargo, para lograr resultados válidos, a menudo es necesario ajustar los hiperparámetros del algoritmo. Este ajuste varía en función de los datos y el problema que se está abordando.

El algoritmo más comúnmente utilizado para lograr estos objetivos es el “Gradiente Descendente” ([Zhang, 2019](#)), un algoritmo de optimización que se utiliza en el campo del **ML**, para minimizar una función objetivo de manera iterativa. Esta función objetivo es comúnmente una función de pérdida o costo, que mide cuán bien las predicciones del modelo coinciden con los datos reales. El objetivo de este algoritmo es ajustar los parámetros del modelo (como los pesos de una red neuronal) de tal manera que minimicen esta función de pérdida. En este proceso, la función de costo dentro del “Gradiente Descendente” actúa

como un estimador, midiendo la precisión en cada iteración y actualizando los parámetros del modelo. Esta función de costo mapea eventos o valores de una o más variables a números reales que sirven para cuantificar el rendimiento del modelo.

5.1. Motivación para utilizar algoritmos del campo del Aprendizaje Automático

Si bien existen trabajos que analizan datos de galaxias para buscar agrupamientos, asumiendo un potencial rígido y fijo para la galaxia (Helmi and White, 1999b), se suelen hacer cortes bajo nociones intuitivas como cortes en el espacio de magnitud sobre los datos (Myeong et al., 2018b) o directamente analizar la distribución espacial de las estrellas con restricciones en magnitud (An and Beers, 2020). En ese sentido se busca aprovechar que los algoritmos de ML no hacen ninguna suposición sobre los datos ni trabajan con supuestos físicos o asumiendo un modelo de fondo de las galaxias.

Cabe mencionar también que algunos de estos trabajos solo consideran una variable a la vez (sobredensidades, grupos, movimientos propios) lo cual puede tener como consecuencia que la señal se diluya. Una de las ventajas que tienen los algoritmos de Aprendizaje es la posibilidad de combinar variables, esto puede reforzar que cada dato analizado pertenezca a un grupo y no al ruido. Las combinaciones pueden aumentar la señal cosa que las variables por sí solas no harían.

Los algoritmos de Aprendizaje Automático se suelen clasificar en algoritmos que funcionan en el enfoque supervisado, enfoque *no supervisado*, también existe el enfoque *semisupervisado* y el enfoque de aprendizaje reforzado. Muchos de estos algoritmos tienen su implementación a través de un módulo en python llamado llamado “Scikit-learn” (Pedregosa et al., 2011), como es el caso de los algoritmos utilizados en esta tesis. Los enfoques y sus respectivos algoritmos se describen a continuación.

5.2. Algoritmos de Aprendizaje Supervisado

En los algoritmos de Aprendizaje Supervisado, se entrena un modelo utilizando datos de entrenamiento que incluyen entradas y sus respectivas salidas o etiquetas. El objetivo de estos algoritmos es que el modelo aprenda a realizar predicciones o tomar decisiones sobre nuevos datos no vistos previamente por el algoritmo (Mitchell, 1997). Los ejemplos más comunes son:

- Regresión Lineal: La regresión lineal es un algoritmo de aprendizaje supervisado que se utiliza para predecir valores numéricos continuos. Busca establecer una relación

lineal entre una variable independiente (o predictor) y una variable dependiente (o respuesta) (Rouaud, 2013). El objetivo es encontrar la línea recta que mejor se ajuste a los datos para poder hacer predicciones precisas sobre nuevos valores. La regresión lineal se basa en el principio de minimizar la distancia entre los puntos de datos y la línea de regresión, utilizando técnicas como el método de los mínimos cuadrados.

- **Regresión Logística:** A pesar de su nombre, la regresión logística es un algoritmo utilizado principalmente para problemas de clasificación. Se emplea para predecir la probabilidad de pertenencia a una clase o categoría (Tolles and Meurer, 2016). A través de una función logística, los valores de entrada se transforman en probabilidades en el rango de 0 a 1. Luego, se establece un umbral para tomar la decisión final de clasificación. La regresión logística es particularmente útil en problemas binarios de clasificación, pero también se puede adaptar para casos multiclase utilizando técnicas como la regresión logística multinomial.
- **Árboles de Decisión:** Los árboles de decisión son algoritmos de aprendizaje supervisado que se utilizan tanto para clasificación como para regresión (von Winterfeldt, 1986). Estos algoritmos crean una estructura similar a un árbol para tomar decisiones basadas en múltiples condiciones. Cada nodo interno del árbol representa una pregunta o condición sobre los datos, mientras que las hojas representan las decisiones finales o las asignaciones de clases. Estos algoritmos pueden llegar a ser propensos al sobreajuste, lo que se puede mitigar mediante técnicas como la poda del árbol (una técnica utilizada para reducir el tamaño de un árbol de decisión y hacerlo más eficiente).
- **Bosques Aleatorios Random Forest (RF):** Los Bosques Aleatorios son un conjunto de árboles de decisión interconectados, utilizados principalmente para problemas de clasificación y regresión (Ho, 1995). Este tipo de algoritmos se basa en la idea de que la combinación de múltiples árboles de decisión puede mejorar la precisión y reducir problemas como el sobreajuste. Cada árbol se entrena en un conjunto de datos de entrenamiento ligeramente diferente, utilizando una combinación de muestras aleatorias y características aleatorias. Luego, las predicciones de los árboles individuales se combinan para obtener una predicción final más sólida. Los Bosques Aleatorios son conocidos por su robustez y capacidad para manejar datos ruidosos.

5.2.1. Algoritmos de Clasificación

Existe otra categoría dentro del aprendizaje supervisado que son los algoritmos de clasificación, los más comúnmente utilizados son:

- **Máquinas de Vectores de Soporte o Support Vector Machines (SVM):** Las máquinas de

vectores de soporte, son un conjunto de algoritmos de aprendizaje supervisados que se desarrollaron originalmente para solucionar problemas de clasificación (objetivo categórico) (Boser et al., 1996) y luego abarcaron problemas de regresión (objetivo numérico) (Drucker et al., 1996). Estos algoritmos construyen un modelo para predecir la categoría de un nuevo punto en un espacio p-dimensional, basándose en un conjunto de puntos etiquetados en dos categorías. SVM busca identificar un hiperplano que separa de manera óptima estas dos categorías, maximizando el margen, es decir, la distancia entre el hiperplano y los puntos más cercanos a él en cada categoría.

- **K-Vecinos más Cercanos o K-Nearest Neighbors (KNN):** Dentro del campo de **ML**, este algoritmo se utiliza principalmente para resolver problemas de clasificación, ver (Zhang, 2016). Su objetivo es asignar una etiqueta o categoría a un nuevo punto de datos basándose en las etiquetas de los puntos de datos cercanos en el conjunto de entrenamiento. Este algoritmo se describe con más detalle más adelante ya que es uno de los algoritmos utilizados en la tesis.

5.2.2. El algoritmo *K-Nearest-Neighbours*

El algoritmo **KNN**, fue definido en el campo de la estadística por **Fix and Hodges (1951)**, dentro del campo del **ML** es un método de aprendizaje automático que se enmarca en el enfoque supervisado y se utiliza principalmente para resolver problemas de clasificación (Zhang, 2016). Su objetivo es asignar una etiqueta o categoría a un nuevo punto de datos basándose en las etiquetas de los puntos de datos cercanos en el conjunto de entrenamiento. El proceso puede desglosarse en los siguientes pasos:

- **Conjunto de Entrenamiento Etiquetado:** Dado el enfoque supervisado, el algoritmo **KNN** requiere un conjunto de datos de entrenamiento previamente etiquetados. Cada punto de datos en este conjunto tiene características (atributos) y una etiqueta que indica a qué clase pertenece. Por ejemplo, en un conjunto de imágenes de animales, las características podrían ser los píxeles de la imagen y la etiqueta sería “gato” o “perro”.
- **Cálculo de Distancias:** Una vez que se tiene un nuevo punto de datos sin etiqueta que se desea clasificar, el algoritmo calcula la distancia entre este punto y todos los puntos del conjunto de entrenamiento. Para medir la distancia la métrica más utilizada es la distancia Euclidiana aunque existen otras métricas como “Manhattan” o “Minkowski”.
- **Selección de Vecinos:** Después de calcular las distancias, el algoritmo selecciona los “k” puntos de entrenamiento más cercanos al nuevo punto. Estos “k” vecinos serán utilizados para determinar la clase del nuevo punto.

- **Votación Mayoritaria:** En esta etapa, se realiza una votación mayoritaria entre los “k” vecinos para decidir a qué clase pertenece el nuevo punto. Cada vecino vota por su clase correspondiente, y la clase con más votos se asigna al nuevo punto. Esto significa que la clase más frecuente entre los vecinos cercanos será la predicción final.
- **Asignación de Etiqueta:** Finalmente, la etiqueta mayoritaria determinada por la votación se asigna al nuevo punto de datos, lo que completa la clasificación.

El conjunto de entrenamiento es fundamental para el funcionamiento del algoritmo **KNN**. Al utilizar ejemplos previamente etiquetados, el algoritmo aprende a reconocer patrones y relaciones entre las características y las clases. Esto le permite generalizar y clasificar nuevos puntos de datos en función de las similitudes con los ejemplos de entrenamiento.

La ventaja principal del algoritmo **KNN** es su simplicidad y flexibilidad para capturar relaciones no lineales en los datos. Sin embargo, este algoritmo se puede ver afectado por la elección del valor de “k”, la normalización de características y la sensibilidad a los valores atípicos en el conjunto de entrenamiento.

El algoritmo **KNN** es un método de **ML** del enfoque “supervisado”, que se utiliza para la clasificación de datos. Se basa en la similitud entre puntos de datos y utiliza ejemplos ya etiquetados en el conjunto de entrenamiento para asignar etiquetas a los nuevos puntos sobre los cuales se quiere hacer la clasificación. Hay que tomar en cuenta que la elección de “k” y la naturaleza del conjunto de entrenamiento son factores importantes que afectan el rendimiento y la precisión del algoritmo **KNN** en la clasificación.

5.3. Algoritmos de aprendizaje *semisupervisado*

El aprendizaje *semisupervisado* se puede categorizar entre el aprendizaje *no supervisado* (sin datos de entrenamiento que tengan etiquetas) y el aprendizaje supervisado (con datos de entrenamiento etiquetados). En la realidad puede llegar a darse que durante el entrenamiento algunos de los datos carezcan de etiquetas de entrenamiento, sin embargo, en aprendizaje automático se ha encontrado que los datos no etiquetados, cuando se utilizan en conjunto con una pequeña cantidad de datos etiquetados, pueden producir una mejora considerable en la precisión del aprendizaje (Zhu, 2007; Vapnik and Chervonenkis, 1974, e.g.).

El entrenamiento supervisado actualiza los pesos del modelo para minimizar la diferencia promedio entre las predicciones y las etiquetas. Sin embargo, con datos etiquetados limitados, el resultado puede ser válido para los puntos etiquetados pero no necesariamente se generalizará a toda la distribución. Por otro lado, el aprendizaje *no supervisado* intenta agrupar los puntos con base a similitudes en algún espacio de características, pero sin etiquetas que guíen el entrenamiento un algoritmo *no supervisado* podría encontrar grupos por debajo de la resolución óptima.

Con el aprendizaje *semisupervisado*, se entrena un modelo inicial con unas pocas muestras etiquetadas (supervisado) y luego aplicas el modelo de manera iterativa a un conjunto de datos más grande. Con el enfoque *no supervisado* se utiliza el conjunto de datos no etiquetados para explorar y entender la estructura subyacente y las distribuciones de los datos (aprendizaje *no supervisado*). Posteriormente el modelo se re-entrena utilizando tanto los datos etiquetados originales como las nuevas predicciones.

Los algoritmos dentro del enfoque *semisupervisado* pueden resumirse como:

- **Algoritmos Basados en Grafos:** Como *Label Propagation* (Zhu, 2002) y *Label Spreading* (Zhou et al., 2003), que difunden las etiquetas a través de un grafo de similitud.
- **Algoritmos de Auto-entrenamiento:** Tal como *Self-Training*, que utiliza predicciones del propio modelo para etiquetar datos no etiquetados y después reentrenar el modelo, ver Amini et al. (2022) para un resumen del funcionamiento de estos algoritmos.
- **Métodos Generativos:** Tales como Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) y Variational Autoencoders (VAEs) (Kingma and Welling, 2013), que se emplean para aprender representaciones de los datos y potenciar la clasificación.
- **Técnicas de Aprendizaje Profundo *semisupervisado*:** Como *Pseudo-Labeling* y *Consistency Training*, que usan redes neuronales para utilizar la información en los datos no etiquetados, ver el resumen hecho por Ouali et al. (2020).
- **Métodos Multi-Instancia y de Clustering:** Que utilizan relaciones o similitudes entre instancias para agrupar datos y asignar etiquetas basándose en la información disponible (Cai et al., 2023).

Mención aparte, podríamos considerar que el algoritmo de búsqueda de sub-halos; *Rockstar*, entra dentro de la categoría de algoritmos *semisupervisados*, ya que se apoya de un criterio físico como la energía aparte de su mecánica para búsqueda de vecinos (Friend-of-friends).

5.4. Algoritmos de aprendizaje *no supervisado*

El enfoque *no supervisado* en el ML se refiere a un tipo de técnica donde el algoritmo se enfrenta a un conjunto de datos sin etiquetas o categorías previas, el algoritmo interpreta automáticamente conjuntos de datos extensos y los organiza según su propia lógica. A diferencia del enfoque supervisado, donde el algoritmo se entrena con datos etiquetados para hacer predicciones o clasificaciones, el enfoque *no supervisado* se centra en descubrir patrones, relaciones o estructuras intrínsecas en los datos por sí mismo, ver Barlow (1999) para un análisis más detallado de este enfoque.

5.4.1. Tipos de Algoritmos “no supervisados”

Existen diversos tipos de algoritmos de aprendizaje *no supervisado* en ML, los cuales pueden clasificarse en varias categorías basadas en su propósito y funcionamiento. A continuación, se presentan algunas de las categorías más comunes y ejemplos de algoritmos en cada una de ellas:

■ Algoritmos de “Clustering” o agrupamiento

- **K-Means**: Agrupa datos en k clústeres basándose en la similitud de los datos, requiere que se especifique como hiperparámetro el número de grupos.
- **Jerárquico**: Crea una estructura de árbol de clústeres mediante la agrupación sucesiva de datos y clústeres.
- **DBSCAN**: Clasifica puntos en clústeres y destaca regiones de alta densidad separadas por regiones de baja densidad. Este algoritmo se describe a detalle más adelante al haber sido utilizado en este trabajo de tesis.

■ Algoritmos de Reducción de Dimensionalidad

- **Análisis de Componentes Principales (PCA)**: Reduce la dimensionalidad mientras intenta preservar la varianza de los datos.
- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**: Reduce la dimensionalidad y visualiza datos de alta dimensión.
- **Autoencoders**: Redes neuronales utilizadas para la compresión y descompresión de datos.

■ Algoritmos de Detección de Anomalías

- **Isolation Forest**: Aisla anomalías en lugar de construir perfiles de normalidad.
- **One-Class SVM**: Clasifica datos en una única clase y detecta outliers.

■ Modelos Generativos

- **Máquinas de Boltzmann**: Aprenden una distribución de probabilidad sobre sus entradas y pueden generar nuevos datos.
 - **Generative Adversarial Networks (GANs)**: Consisten en dos redes, una generadora y una discriminadora, que son entrenadas simultáneamente.
- **Mapas Autoorganizados Self-Organizing Maps (SOMs)**: Redes neuronales que realizan una proyección de datos de alta dimensión a espacios de menor dimensión, preservando la topología de los datos y este es uno de los puntos fuertes de este algoritmo.

Por ejemplo, un conjunto de datos con p variables medidas en n observaciones podría representarse como conglomerados de observaciones con valores similares para las variables. Estos conglomerados podrían visualizarse luego como un mapa bidimensional de tal manera que las observaciones en conglomerados próximos tengan valores más similares que las observaciones en conglomerados distantes. Esto puede facilitar la visualización y análisis de datos de alta dimensión.

5.4.2. **Ventajas del enfoque *no supervisado***

- Descubrimiento de patrones ocultos: Los algoritmos *no supervisados* pueden revelar patrones y relaciones complejas en los datos que podrían no ser evidentes a simple vista. Esto puede llevar a una mejor comprensión de los datos y a la identificación de características clave.
- Sin necesidad de etiquetas: No se requiere un conjunto de datos etiquetado para entrenar el modelo, lo que puede ser beneficioso cuando obtener etiquetas es costoso o difícil.
- Exploración de datos: El enfoque *no supervisado* permite explorar los datos desde diferentes ángulos y perspectivas, lo que puede llevar a descubrimientos inesperados o nuevas formas de segmentar los datos.

5.4.3. **Desventajas del enfoque *no supervisado***

- Falta de evaluación directa: Dado que no hay etiquetas para comparar las predicciones del algoritmo, evaluar la calidad del modelo puede ser más subjetivo y desafiante.
- Mayor complejidad en algunos casos: Algunos algoritmos *no supervisados* pueden ser más complejos de entender y ajustar, especialmente cuando se trata de identificar el número óptimo de clústeres o grupos en los datos.
- Interpretación de resultados: En ocasiones, los resultados de los algoritmos *no supervisados* pueden ser difíciles de interpretar o explicar de manera intuitiva, lo que puede hacer que sea complicado comunicar los hallazgos a otras personas.
- Sensibilidad a la inicialización: Algunos algoritmos *no supervisados*, como los de agrupamiento, pueden ser sensibles a la inicialización y generar resultados diferentes según cómo se inicie el proceso.

En resumen, el enfoque *no supervisado* en el **ML** es valioso para explorar datos sin etiquetas y descubrir patrones ocultos, pero también presenta desafíos en términos de evaluación y complejidad del modelo. Se utiliza en una variedad de aplicaciones, como agrupamiento de datos, reducción de dimensionalidad y detección de anomalías, entre otros. A continuación se describe el algoritmo *no supervisado* en empleado a lo largo de esta tesis, su funcionamiento y esquema de calibración. Más adelante en el siguiente capítulo se describe la preparación de los datos y la estrategia de calibración del algoritmo utilizando las simulaciones.

5.5. El algoritmo DBSCAN para la búsqueda de grupos en conjuntos de datos

El análisis de agrupamiento es un problema importante en el campo del análisis de datos. Actualmente entre los algoritmos de búsqueda de agrupamiento, el algoritmo DBSCAN es uno de los más utilizados.

DBSCAN (*Density-Based Clustering for Discovering clusters in large datasets with noise*) es un algoritmo clúster o de agrupamiento *no supervisado* basado en la densidad, que puede ser utilizado para identificar clústeres de cualquier forma en un conjunto de datos que contiene ruido y valores atípicos y fue propuesto por primera vez por Ester et al. (1996).

Los clústeres son regiones densas en el espacio de datos, separadas por regiones de menor densidad de puntos. El algoritmo **DBSCAN** se basa en la noción intuitiva de clústeres y ruido. La idea clave es que, para cada punto de un clúster, la vecindad de un radio dado tiene que contener al menos un número mínimo de puntos.

A diferencia de otros algoritmos de agrupamiento como **K-Means** (MacQueen, 1967), **DBSCAN** no necesita especificar el número de grupos. Puede detectar automáticamente el número de grupos en función de los datos de entrada y los parámetros proporcionados. Además, lo que es aún más importante, **DBSCAN** puede encontrar grupos de formas arbitrarias que **K-Means** no puede detectar. Por ejemplo, un grupo rodeado por otro grupo diferente, ver fig. 5.1.

5.5.1. Parámetros del algoritmo DBSCAN

El algoritmo **DBSCAN** requiere básicamente dos parámetros:

- Épsilon (eps): especifica lo cerca que deben estar los puntos entre sí para ser considerados parte de un clúster. Esto significa que, si la distancia entre dos puntos es menor o igual a este valor de épsilon, estos puntos se consideran vecinos.

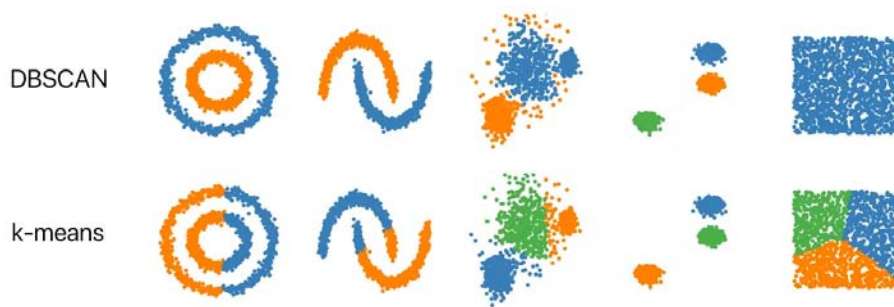


Figura 5.1: Imagen comparativa de como se desempeñan DBSCAN y K-Means en distintos conjuntos de datos con diferentes modos de agruparse. Crédito de la imagen: Pedregosa et al. (2011)

- Puntos mínimos (*minPts*): el número mínimo de puntos para formar una región densa. Por ejemplo, si establecemos *minPts* como 5, entonces necesitamos al menos 5 puntos para formar una región densa.

5.5.2. Tipos de puntos en el algoritmo DBSCAN

En este algoritmo, tenemos tres tipos de puntos de datos. Podemos categorizar cada punto de datos en punto de núcleo, punto de borde y punto de ruido.

- Punto de núcleo: Un punto es un punto de núcleo si tiene más de un número especificado de puntos mínimos *minPts* dentro de un radio de “eps” a su alrededor. El punto de núcleo siempre pertenece a una región densa. Por ejemplo, consideremos que “p” se establece como un punto central si “p” es mayor o igual de los puntos *minPts* en un radio de épsilon (eps) alrededor de ella.
- Punto de borde: Un punto es un punto de borde si tiene menos de los *minPts* dentro del valor “eps”, pero se encuentra en la vecindad de un punto de núcleo. Por ejemplo, “p” está configurado para ser un punto fronterizo si “p” no es un punto central. Ejemplo, “p” tiene menos puntos mínimos *minPts* en un radio de épsilon (“eps”). El punto p debe pertenecer al vecindario del punto “q”, donde “q” es un punto de núcleo. Donde “p” pertenece al vecindario de “q” y la distancia de “p” y “q” es menor o igual a “eps”.
- Punto de ruido: Un punto de ruido es cualquier punto que no sea un punto de núcleo o un punto de borde.

También existe el concepto de “Directamente densidad alcanzable”: Un punto “p” se considera directamente densidad alcanzable desde un punto “q” si cumple con dos condiciones:

- La distancia entre los puntos “p” y “q” es menor o igual a un valor de umbral llamado “eps”, ver fig. 5.2

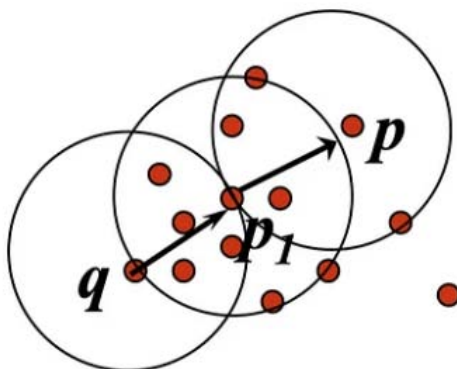


Figura 5.2: Un punto “p” es densidad alcanzable desde un punto “q” c.r.a “eps”, $minPts$, si existe una cadena de puntos P_1, \dots, P_n donde $P_1=q$ y $P_n=p$ de modo que P_{i+1} es densidad alcanzable directamente desde el punto P_i .

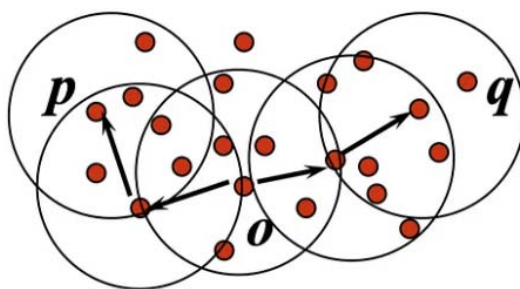


Figura 5.3: Un punto “p” es conectado en densidad a un punto “q” c.r.a “eps”, $minPts$, si existe un punto “o” de modo que “p” y “q” son densidad alcanzables desde “o”.

- El número de puntos en la vecindad “eps” de “q” (denominada $NEps(q)$) es igual o mayor que un valor de umbral denominado $minPts$, ver fig. 5.3

5.5.3. Pasos del funcionamiento del algoritmo DBSCAN

- El algoritmo comienza con un punto arbitrario que no ha sido visitado y la información de su vecindario se recupera desde el parámetro de ϵ .
- Si el punto contiene puntos mínimos dentro del vecindario “eps”, se inicia la formación de clústeres. De lo contrario, el punto se etiqueta como ruido. Este punto se puede encontrar más tarde dentro de la vecindad de “eps” de un punto diferente y por lo tanto, se puede hacer parte del clúster. El concepto de densidad alcanzable y los puntos de conexión de densidad son importantes aquí.

- Si se encuentra que un punto es un punto de núcleo, entonces los puntos dentro del vecindario de ϵ también son parte del grupo. Así que todos los puntos que se encuentran dentro del vecindario de ϵ se agregan, junto con su propio vecindario ϵ , si también son puntos de núcleo.
- El proceso anterior continúa hasta que se encuentra completamente el clúster conectado a la densidad.
- El proceso se reinicia con un nuevo punto que puede ser parte de un nuevo clúster o etiquetado como ruido.

5.5.4. Estimación de Parámetros

Cada tarea de minería de datos se enfrenta al problema de los parámetros. Cada parámetro influye en el algoritmo de maneras específicas. En el caso de **DBSCAN**, se requieren los parámetros “ ϵ ” y $minPts$.

- $minPts$: Como regla general, se puede derivar un valor mínimo para $minPts$ a partir del número de dimensiones D en el conjunto de datos, donde $minPts \geq D + 1$. El valor mínimo muy bajo, como $minPts = 1$, no tiene sentido, ya que cada punto por sí solo sería un clúster. Con $minPts \leq 2$, el resultado será similar al del agrupamiento jerárquico con la métrica de enlace único, con el dendrograma cortado en la altura “ ϵ ”. Por lo tanto, $minPts$ debe ser al menos 3. Sin embargo, en general, valores mayores suelen ser mejores para conjuntos de datos con ruido y producirán clústeres más significativos. Como guía aproximada, se puede utilizar $minPts = 2 \times dim$, pero en casos de conjuntos de datos muy grandes, datos ruidosos o datos con muchas duplicaciones, puede ser necesario elegir valores mayores.
- “ ϵ ”: El valor de “ ϵ ” se puede elegir utilizando un gráfico de distancia k , donde se traza la distancia al k -ésimo vecino más cercano ($k = minPts - 1$) ordenado de mayor a menor valor. Los valores más óptimos de “ ϵ ” son aquellos en los que este gráfico muestra una señal o pico: si “ ϵ ” se elige demasiado pequeño, gran parte de los datos no se agruparán; mientras que si “ ϵ ” es demasiado grande, los clústeres se fusionarán y la mayoría de los objetos estarán en el mismo clúster. En general, valores pequeños de “ ϵ ” son preferibles, y como regla general, solo una pequeña fracción de puntos debe estar dentro de esta distancia entre sí.
- Función de Distancia: La elección de la función de distancia está estrechamente relacionada con la elección de “ ϵ ” y tiene un impacto significativo en los resultados del algoritmo. Por lo general, será necesario identificar primero una medida razonable

de similitud para el conjunto de datos antes de elegir el parámetro “eps”. No existe una estimación para este parámetro, pero las funciones de distancia deben elegirse de manera apropiada según las características del conjunto de datos.

En resumen, la elección de *minPts*, “eps” y la función de distancia es fundamental para el éxito de **DBSCAN**. Estos parámetros deben seleccionarse cuidadosamente, mediante prueba y error, teniendo en cuenta las características específicas del conjunto de datos y los resultados de agrupamiento deseados.

5.5.5. Ventajas y desventajas del algoritmo DBSCAN

Es importante mencionar que al comparar el algoritmo de *clustering* **DBSCAN** con un *halo finder* como **Rockstar** para identificar halos en simulaciones cosmológicas o un algoritmo de búsqueda de agrupación como **K-Means**, hay varias ventajas y desventajas a considerar. A continuación se enlistan las ventajas y desventajas del algoritmo **DBSCAN**.

Ventajas:

- **Formas de Clusters Flexibles:** **DBSCAN** puede identificar clusters de formas irregulares o arbitrarias en lugar de asumir que tienen formas esféricas o elipsoidales por ejemplo [referencia].
- **Número de clusters:** Al basarse en la densidad, el algoritmo decide autónomamente el número de clusters, liberando al usuario de esta tarea. A diferencia de algoritmos como **K-Means**, **DBSCAN** no requiere que el número de clusters sea especificado *a priori*.
- **Tratamiento de Ruido:** Una característica distintiva de **DBSCAN** es su capacidad para identificar y separar datos atípicos o “ruido”. Aunque esto puede ser una ventaja al tratar con las simulaciones donde pueden existir partículas que no pertenecen a ningún halo.
- **Pocos Parámetros:** Aunque requiere de parámetros, estos son relativamente pocos en comparación con otros algoritmos, solo depende de un valor “eps” y un valor *minPts*.
- **DBSCAN no realiza suposiciones ni asume ningún modelo sobre los datos.** Aunque “**Rockstar**” no asume un modelo específico para los halos (como podría ser un perfil tipo **NFW**), si opera bajo supuestos generales sobre la naturaleza de las simulaciones cosmológicas. Por ejemplo, se espera que los halos tengan una cierta jerarquía y estructura.

- El límite de las variables que componen el espacio en el que DBSCAN puede trabajar no se limita al espacio fase. Otras variables como la metalicidad pueden ser incluidas por ejemplo, sin embargo, hay que tener precauciones y tomar en cuenta el problema de “La maldición de la dimensionalidad” (A medida que aumenta el número de dimensiones o características, la cantidad de datos necesaria para generalizar con precisión el modelo de ML aumenta de manera exponencial).

5.5.6. Desventajas:

- Elección de Parámetros: Elegir los valores adecuados para “eps” y “minPts” puede representar un desafío y se requiere mayor análisis a la hora de analizar conjuntos de datos específicos, como es el caso de detectar rastros de la estructura de galaxias satélites en el halo estelar.
- Variabilidad en la Densidad: DBSCAN puede ver afectado su rendimiento en conjuntos de datos con grandes variaciones en densidad de sus grupos, ya que solo se maneja en un solo valor del parámetro “eps” y puede no adecuarse a todas las regiones.
- Dimensión Alta: Con conjuntos de datos multidimensionales, DBSCAN puede enfrentar dificultades debido a problemas inherentes de esos espacios, como la “maldición de la dimensionalidad”.
- Escalabilidad: Si bien es eficiente en ciertos contextos, con grandes volúmenes de datos puede presentar retos de eficiencia y el costo computacional puede ser significativo. *Rockstar*, por su diseño, tiene una metodología para identificar subestructuras dentro de halos. DBSCAN, en su forma estándar, no tiene una manera directa de hacer esto.
- Eficiencia computacional: En grandes conjuntos de datos, como las simulaciones de N-cuerpos, la eficiencia de DBSCAN puede ser un problema. *Rockstar*, por otro lado, está optimizado específicamente para este tipo de análisis (las simulaciones suelen generar conjuntos de datos grandes).

El algoritmo DBSCAN ha mostrado ser eficiente en la búsqueda de grupos en datos que tienen grupos de diferentes densidades, por lo que parece viable utilizarlo para la búsqueda de subestructura, como colas de marea alrededor de galaxias satélites o “streams” en el halo estelar. En los siguientes capítulos desarrollamos un esquema para calibrar los hiperparámetros del algoritmo DBSCAN, con simulaciones de N-cuerpos aisladas por ahora, pero en el futuro planeamos utilizar este esquema con simulaciones cosmológicas de alta resolución para la búsqueda de subestructura en observaciones sintéticas realistas, basadas en simulaciones más precisas, esto se describe en los siguientes capítulos. En un futuro planeamos

utilizar otros algoritmos de búsqueda de agrupamiento aparte de **DBSCAN**, como los “Mapas Auto Organizados” Self Organized Maps (**SOMs**).

Capítulo 6

Métodología para buscar subestructura en las simulaciones

Como se mencionó en el capítulo 5, **DBSCAN** es un algoritmo no supervisado que no requiere una muestra de entrenamiento. Sin embargo, para garantizar la fidelidad de los grupos o subestructuras identificados, es necesario realizar una calibración previa de sus hiperparámetros. En este capítulo, se detalla la metodología que se utilizará en esta tesis para calibrar el algoritmo de búsqueda de subestructura **DBSCAN** mediante simulaciones de N-cuerpos de sistemas aislados. Esta metodología implica la combinación de variables como posición, velocidades radiales y tangenciales, densidad y metalicidades, y el análisis de sus correlaciones (ver sección 6.6). También exploramos la búsqueda en situaciones realistas, utilizando el cono de luz, similar a las observaciones reales, o realizando la búsqueda en un plano de las simulaciones que contiene las partículas del satélite (ver sección 6.7.1).

Es fundamental destacar que, si bien estamos aplicando esta calibración a simulaciones de N-cuerpos aisladas en este momento, tenemos la intención de extender esta técnica a observaciones sintéticas basadas en simulaciones cosmológicas de formación de galaxias en el futuro. Alguien podría preguntarse cómo podemos estar seguros de que esta calibración será efectiva en las observaciones sintéticas, dado que las simulaciones son diferentes. La respuesta radica en la naturaleza de los sistemas que estamos estudiando, que incluyen desde satélites, cúmulos globulares y estrellas del halo de fondo, entre otros.

Independientemente del tipo de simulación, tenemos la certeza de que habrá grupos con diferentes densidades en estos sistemas. Incluso las mismas galaxias satélites en las simulaciones presentan diferentes densidades, por mencionar un ejemplo, las colas de marea y la parte central, así como una parte de las partículas del halo que se alinea con el movimiento de las partículas del satélite. Por lo tanto, la técnica que describimos en este capítulo se puede aplicar a las simulaciones cosmológicas, pero reconocemos que será necesario realizar una calibración adicional de los hiperparámetros específicamente para las simulaciones

cosmológicas. Lo que extraemos de este capítulo es la metodología para identificar grupos con distintas densidades, lo que nos servirá como punto de partida para futuros trabajos en simulaciones cosmológicas.

Es importante señalar que **DBSCAN** trabaja con una sola densidad una vez calibrada, lo que puede llevar a situaciones en las que cierta información de interés se destaque en los datos, como las colas de marea de las galaxias satélites, mientras que parte de la información quede oculta debido a la similitud con las partículas del halo. En la sección 6.8, hemos abordado este problema al reducir variables en la muestra, resaltando la información de las partículas del satélite en comparación con las del halo. Esto nos permite comprender mejor la naturaleza de los grupos identificados y las limitaciones de la búsqueda calibrada. En la sección 6.8.1, aprovechamos esto e intentamos discernir las diferentes componentes que pueden tener los grupos encontrados, tanto lo que se puede identificar como lo que no se puede, permitiendo una clasificación de la estructura encontrada y la no encontrada.

6.1. Selección de la muestra de datos de la simulación que se va a analizar

La primera parte del proceso consiste en la preparación de datos para la aplicación del algoritmo **DBSCAN** en los datos de la simulación. El proceso de preparación de datos garantiza que solo las partículas relevantes sean consideradas en el análisis, lo que resultará en una detección más precisa de los clústeres y estructuras en la simulación del satélite.

Este paso es necesario para equilibrar la muestra, si se intentarían tomar las partículas del halo, probablemente sería imposible extraer la señal de los datos del satélite ya que esta quedaría oculta por el ruido de fondo debido a la presencia de partículas del halo.

- Importar y leer el archivo de datos de la simulación, los modelos del halo y el satélite en la simulación correspondiente a la muestra de datos analizada en este capítulo corresponden a la columna 2 y a la columna 5 de la tabla 4.1 respectivamente. El archivo que se analiza corresponde a una simulación con un satélite vivo extendido y cayendo solo a 3 Gyr de la simulación.
- Después de la lectura, se ajustan las coordenadas de las partículas en función de la posición del satélite, ver figura 6.1.
- Se realiza un recorte en los datos para seleccionar la región específica del espacio donde se encuentra el satélite, ver los tres paneles de la fig. 6.1 y la tabla 6.1. Este corte puede ser sobre un plano grueso que contiene a las partículas del satélite, sin embargo, en las pruebas que se muestran más adelante se realizaron cortes en una

región conoidal que contiene a las partículas del satélite para simular el modo en que un catastro puede visualizar al satélite, ver más detalles de esto en la sección 6.2.

- A continuación, se añade una nueva columna que corresponde a las etiquetas de las partículas en función de si son pertenecientes al satélite o al halo, según corresponda.
- Este proceso permite diferenciar y categorizar las partículas en función de su asociación con el satélite.

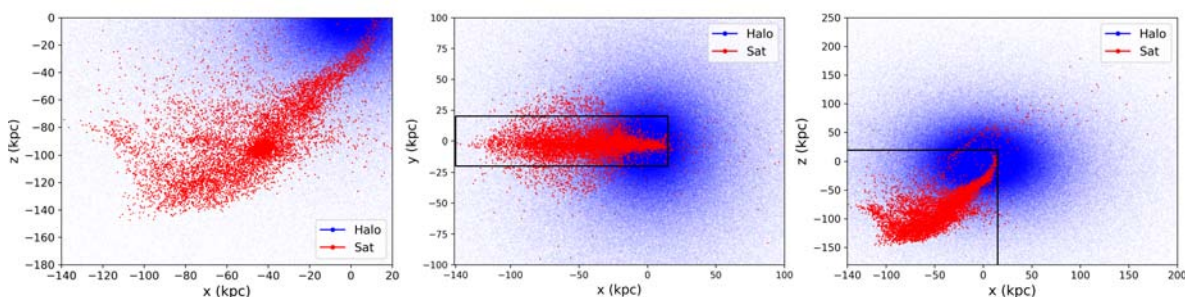


Figura 6.1: Selección de las partículas que conforman la muestra a analizar con DBSCAN, las regiones seleccionadas están acotadas por líneas negras sólidas. Izq: Partículas etiquetadas del satélite y el halo en la región de la partículas del satélite. Centro: Selección o corte del espacio en el plano xy que contiene las partículas del satélite. Der: Selección o corte en el plano xz que contiene las partículas del satélite. Las partículas del satélite son puntos rojos, las partículas del halo puntos azules atenuados en la imagen para efectos de visualización.

Categoría de Partículas	Cantidad
Partículas del satélite en la muestra	10,290
Partículas del halo en la muestra	165,036
Partículas totales en la muestra	175,326

Cuadro 6.1: Número de partículas del halo y del satélite presentes en el conjunto de datos final en el plano que contiene a las partículas del satélite.

6.2. Modelado y Generación de Muestras Conformes para Simulaciones Astrofísicas

La precisión y fidelidad de las simulaciones astrofísicas dependen en gran medida de su capacidad para capturar la naturaleza y las características de los objetos observados en el universo. Al abordar la representación de galaxias satélite u otros cuerpos celestes, es fundamental considerar cómo se llevan a cabo las observaciones reales.

- **Función de selección:** La función de selección se refiere a la probabilidad de que un objeto en el espacio sea seleccionado para ser observado en un estudio específico. En el caso de observaciones reales, esta función tiene en cuenta las limitaciones instrumentales, las condiciones de observación y los criterios de selección del objeto. Para entender correctamente la distribución y propiedades de los objetos en el estudio, es esencial tener en cuenta esta función de selección, ver por ejemplo [Chen et al. \(2018\)](#) para LAMOST ([Zhao et al., 2012](#)).
- **Cono de luz:** Para representar la región del espacio-tiempo que sería observable desde un punto específico en el universo. El cono de luz se forma trazando líneas de visión desde ese punto en todas las direcciones posibles. Los objetos que caen dentro de este cono de luz son considerados observables desde ese punto en el espacio, ver por ejemplo [Merson et al. \(2013\)](#) donde crean catálogos sintéticos con “conos de luz”.

La relación entre la función de selección y el “cono de luz” se encuentra en el hecho de que ambos conceptos abordan la idea de cuales objetos son observables en un estudio determinado. En el caso de la función de selección, se trata de la probabilidad de selección de objetos reales, mientras que en el caso del “cono de luz”, se trata de la región en el espacio-tiempo que puede ser simulada como observable.

En este contexto, el concepto de “cono de luz” y también el de “campo de visión” desempeñan un papel crucial en la configuración y diseño de las simulaciones para lograr una correspondencia más acorde con la realidad observacional.

El cono de luz se refiere a la región en el espacio que está iluminada por la luz emitida desde una fuente astronómica hacia el observador. Cuando miramos al cielo nocturno, la luz de las estrellas, galaxias y otros objetos viaja a través del espacio formando conos de luz que se extienden desde cada fuente hacia nosotros. En consecuencia, este cono de luz define la zona en la que podemos captar la luz de esos objetos. Sin embargo, el proceso de observación va más allá de la mera captación de luz, ya que también está influenciado por el “campo de visión”.

El “campo de visión” es el área del cielo que un instrumento de observación, como un telescopio, puede abarcar en un instante dado. La forma y el tamaño del campo de visión varían en función del diseño y características del instrumento utilizado para la observación. En la práctica, el campo de visión limita la cantidad de objetos que pueden ser vistos simultáneamente y afecta la capacidad de captar detalles específicos en el espacio.

Las galaxias satélite, que orbitan alrededor de galaxias más grandes, pueden ser menos brillantes y más pequeñas, lo que las hace más desafiantes de detectar. La observación exitosa de una galaxia satélite requiere que el telescopio esté apuntando hacia la dirección correcta para que el cono de luz de la galaxia alcance el instrumento.

A continuación se describe una implementación de como se puede modelar un “cono de luz” en el análisis de los datos de las simulaciones (utilizando las simulaciones de N-cuerpos aislados), estas pruebas tienen como fin poner a prueba los algoritmos de búsqueda de estructura en una situación parecida al “cono de luz”, pero de ningún modo representa el cono de luz de catastros observacionales.

6.3. Selección y Simulación del Cono de Luz en Datos Astrofísicos

En este procedimiento, se aplican ajustes y restricciones a la muestra de datos para simular la región con forma cónica que sería iluminada y observada por un telescopio. A continuación se describen las componentes de este proceso:

- **Preparación Inicial de los Datos:** Se inicia cargando un archivo de datos que contiene información sobre las partículas en el espacio. Estos datos son procesados para eliminar columnas no necesarias, y las coordenadas (x, y, z) se ajustan para centrarlas en la posición del observador.
- **Definición del Cono de Luz:** Para simular la región observada en el espacio, se establecen parámetros como el radio del cono de luz (representado por la variable rs) y el factor de atenuación (representado por la variable g). Estos parámetros definen la forma y el alcance del cono de luz que se simulará.

Dado un conjunto de partículas X_N con coordenadas (x, y, z) , y utilizando las constantes (cxs, cys, czs) que representan las coordenadas del centro de masa del satélite, r_s y g , el proceso de filtrado para simular el cono de luz y las restricciones geométricas en el plano xy se describe de la siguiente manera:

1. Se calculan las coordenadas $NewX$ y $NewY$ restando las coordenadas x y y de cada partícula por las coordenadas del centro de masa del satélite:

$$NewX = x - cxs$$

$$NewY = y - cys$$

2. Se realiza el filtrado en función de las siguientes condiciones:

- La coordenada z de la partícula debe ser negativa (para abarcar la región donde está el satélite): $z < 0$
- El valor absoluto de la coordenada z de la partícula debe ser menor que el valor absoluto de g multiplicado por cz_s :

$$|z| < g \cdot cz_s$$

- La distancia al cuadrado desde el punto $(NewX, NewY)$ al origen cumple la restricción de estar dentro del círculo de radio r_s en el plano xy , escalado por la relación entre z y cz_s :

$$(NewX)^2 + (NewY)^2 < (rs^2) \cdot \left(\frac{z}{g \cdot cz_s}\right)^2$$

3. Las partículas que cumplen con estas condiciones son seleccionadas y forman el nuevo conjunto de datos X_N .

Estas ecuaciones describen cómo las coordenadas x , y , z de las partículas se transforman y cómo se aplican restricciones geométricas y de cono de luz en el plano xy para filtrar las partículas que serán consideradas en el análisis posterior.

- **Filtrado de Datos:** Se realiza un filtrado de los datos para seleccionar las partículas que estarían dentro del cono de luz simulado. Se aplican restricciones tanto en el eje z como en el plano xy para garantizar que las partículas seleccionadas estén en la región iluminada por el cono de luz y dentro del campo de visión del observador.
- **Etiquetado de Partículas:** Una vez que se han seleccionado las partículas que se encuentran en el cono de luz simulado, se asignan etiquetas a ellas para indicar si son parte del objeto observado (por ejemplo, el satélite) o no.
- **Resultados y Significado:** Al seguir las restricciones y consideraciones geométricas del cono de luz, se logra modelar con mayor precisión la distribución de partículas que estarían dentro del campo de visión y la región iluminada por el cono de luz. Ver los tres paneles de la figura 6.5. Hay que tomar en cuenta que la ventaja de tener un experimento controlado como las simulaciones nos permite etiquetar las partículas como pertenecientes al satélite o al halo, esto será relevante más adelante para contrastar lo obtenido con **DBSCAN**.

Este proceso no solo es relevante para la simulación, sino que también tiene aplicaciones en la generación de muestras de datos que se asemejen a las observaciones reales. Al incorporar las limitaciones observacionales y geométricas, se puede obtener una muestra de datos más realista que refleje cómo las observaciones astronómicas se comportarían en el mundo real. Esta técnica brinda una valiosa herramienta para validar y mejorar los métodos de análisis utilizados en el campo de las simulaciones de galaxias.

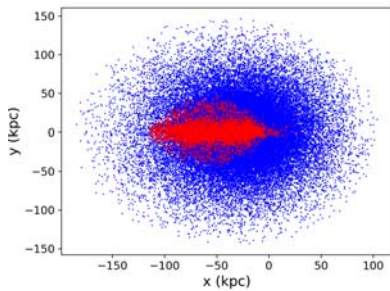


Figura 6.2

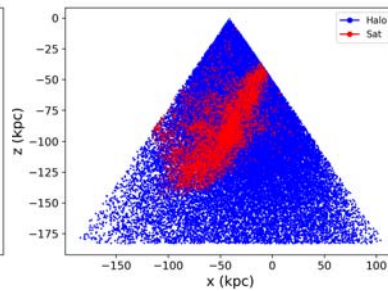


Figura 6.3

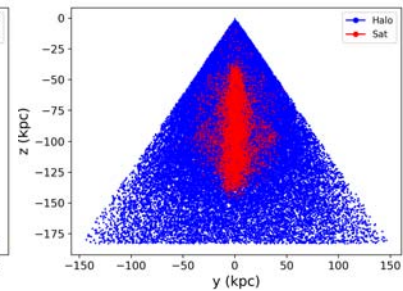


Figura 6.4

Figura 6.5: Selección de las partículas que conforman la muestra a analizar con DBSCAN, tomando una región conforme simulando el cono de luz. Izq: Selección o corte en la región del cono que contiene las partículas del satélite visto desde el plano xy. Centro: Selección o corte del espacio en el cono visto desde el plano xz . Der: Selección o corte en el espacio del cono visto desde el plano xz. Las partículas está etiquetadas por su pertenencia al satélite con puntos rojos y las partículas del halo con puntos azules.

Categoría de Partículas	Cantidad
Partículas del satélite en la muestra	10,749
Partículas del halo en la muestra	67250
Partículas totales en la muestra	77999

Cuadro 6.2: Número de partículas del halo y del satélite presentes en el conjunto de datos final en el cono que contiene a las partículas del satélite.

6.4. Calcular la densidad numérica en el conjunto de datos

Este procedimiento se enfoca en calcular la densidad numérica de partículas en la simulación mediante el uso del algoritmo KNN.

- Primero, se utiliza el algoritmo Nearest Neighbors para estimar la densidad alrededor de cada partícula. Para esto, se calculan las distancias y los índices de las partículas vecinas en función de las coordenadas (x, y, z) de las partículas.

- Luego, se define un radio de búsqueda y se procede a calcular la densidad numérica alrededor de cada partícula en ese radio. Este radio de búsqueda se corresponde con la región de mayor densidad en la simulación.
- Después, para cada partícula, se cuenta la cantidad de partículas vecinas que se encuentran dentro del radio establecido. Estos conteos son utilizados para estimar la densidad de partículas en esa región específica.
- La muestra de datos consistirá de las diferentes combinaciones de variables; posiciones x, y, z velocidades v_x, v_y, v_z , densidad y aquí se pueden agregar otras variables como la metalicidad.
- Estos datos se reorganizan y normalizan mediante la estandarización para tener la seguridad de que estén en una escala similar y adecuada para el análisis.

En resumen, este procedimiento emplea el método de *k-nearest neighbors* para calcular la densidad numérica alrededor de cada partícula en función de un radio de búsqueda. Luego, se preparan los datos para su posterior análisis mediante transformaciones y normalización, es decir, se estandariza la muestra de datos. Este proceso de cálculo de densidad y preparación de datos es crucial para obtener información sobre la distribución y densidad de partículas en diferentes regiones de la simulación.

6.5. Optimización de Parámetros para DBSCAN

Antes de utilizar el algoritmo **DBSCAN**, es esencial determinar los valores óptimos de los hiperparámetros para garantizar su eficacia en la tarea de segmentación de los datos. En este proceso, se emplea el algoritmo **KNN** para llevar a cabo una búsqueda sistemática de los mejores valores de hiperparámetros. Además, se crea un histograma que visualiza la distribución del número de vecinos en función de la distancia, como a continuación se describe:

- Con el objetivo de identificar los valores óptimos del hiperparámetro $n_neighbors$, se implementa el algoritmo **KNN** en una fase de búsqueda. Se establece un conjunto de valores $n_neighbors_values$ que se probarán en el algoritmo. Mediante un bucle de iteración, se ajusta el modelo **KNN** con cada valor de $n_neighbors$. Luego, se calcula una métrica de separación que refleja la diferencia entre las distancias de los vecinos correspondientes a las partículas del satélite y al halo. La métrica de separación se obtiene al calcular la diferencia entre la media de las distancias de los vecinos satélite y la media de las distancias de los vecinos halo. El objetivo es encontrar el valor de

$n_neighbors$ que maximice esta métrica, lo que indicaría la mejor separación entre las dos clases de partículas.

- Una vez completada la búsqueda, se identifica el valor óptimo de $n_neighbors$ que resulta en la mayor métrica de separación. Este valor se almacena como $best_n_neighbors$ para su posterior uso.
- Utilizando el valor óptimo de $n_neighbors$ ($best_n_neighbors$), se procede a entrenar nuevamente el modelo de **KNN**. Se calculan las distancias de los $best_n_neighbors$ vecinos más cercanos para cada muestra. Estas distancias se recopilan en un conjunto de datos $dist$.
- Con el conjunto $dist$ se crea un histograma que ilustra la distribución de las distancias. El eje horizontal del histograma representa las distancias (bines de distancia) y el eje vertical muestra el número de vecinos en bins de distancia. El rango de distancias se extiende desde 1 hasta $best_n_neighbors$, y el objetivo es visualizar cómo se distribuyen los vecinos en función de la distancia.
- La visualización del histograma facilita la elección del parámetro eps para el algoritmo **DBSCAN**. Al observar el histograma, se busca el valor de distancia que proporciona la separación más clara entre las distribuciones de las partículas del satélite y del halo, ver figura 6.6 como ejemplo.

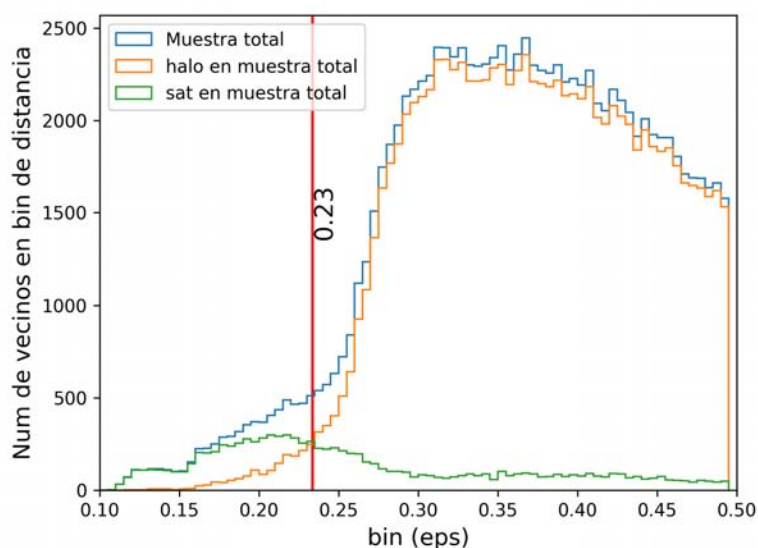


Figura 6.6: Histograma de distribución de vecinos en función de la distancia. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo **DBSCAN**.

Este proceso de optimización de hiperparámetros garantiza que el algoritmo **DBSCAN** se aplique con valores adecuados de $n_neighbors$ y eps , permitiendo una segmentación más precisa de las partículas en el espacio astrofísico. La combinación de la búsqueda basada en **KNN** y la visualización del histograma proporciona una metodología sólida para determinar los parámetros que mejor se ajustan a la naturaleza de los datos y a la tarea de segmentación.

6.6. Probando el algoritmo DBSCAN en simulaciones tipo Vía Láctea

A continuación, se exponen resultados preliminares en relación a la detección de subestructuras mediante el uso del algoritmo **DBSCAN**. Estas evaluaciones involucran la aplicación del algoritmo **DBSCAN** en una instantánea (“snap”) de una simulación que representa un halo similar a la Vía Láctea, con un satélite vivo correspondiente al progenitor de la galaxia dSph Sagitario. Se llevaron a cabo experimentos con diversas configuraciones de variables, incluyendo información cinemática (v_x , v_y , v_z) e información espacial (densidad local alrededor de cada partícula).

Estas primeras exploraciones consisten en aplicar el algoritmo **DBSCAN** a una muestra que incorpora tanto la información espacial en términos de densidad como la información cinemática completa en sus componentes V_x , V_y y V_z , en una región cercana al satélite. Las imágenes presentadas a continuación representan proyecciones de los resultados tanto en el espacio de velocidades, donde se lleva a cabo la búsqueda, como en el espacio de configuración, donde la estructura del satélite es más distinguible.

6.6.1. Experimento utilizando v_x, v_y, v_z + densidad en un plano grueso

En esta sección se muestra el resultado de aplicar **DBSCAN** a la simulación con partículas del halo y del satélite, pasándole solo las velocidades más un parámetro de densidad, también se calcula el nivel de contaminación de los grupos encontrados.

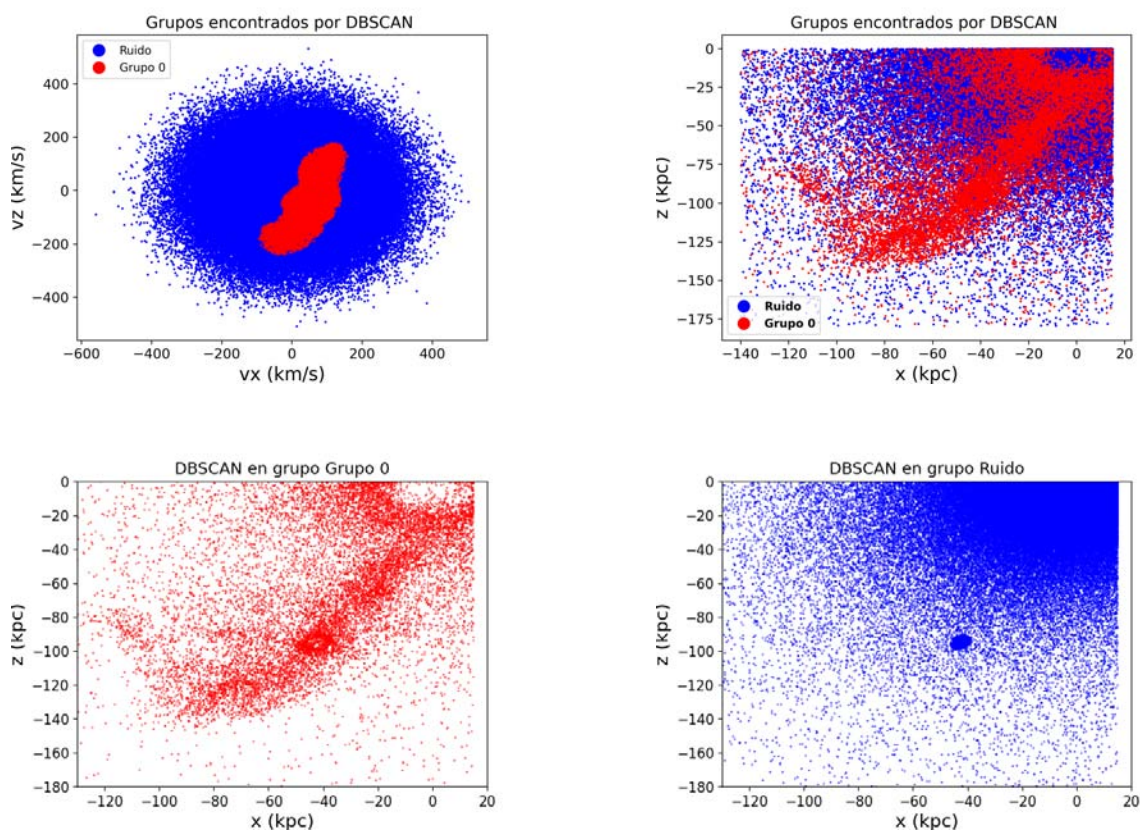


Figura 6.7: Sup izq: Grupos encontrados por **DBSCAN** en un corte del espacio de velocidades vx vs vz. Sup. der: Grupos encontrados por **DBSCAN** en un corte del espacio x vs z. Inf izq: Partículas del grupo 0 conformado en mayor parte por partículas del satélite. Inf derecha: Grupo encontrado por **DBSCAN**, compuesto en su mayor parte por partículas del halo, se considera mayormente como ruido.

El nivel de contaminación en un grupo se puede estimar calculando la proporción de partículas que se consideran como contaminantes en relación con el total de partículas en ese grupo. En este caso, tenemos un solo grupo, el grupo 0, discernible del ruido etiquetado con -1, el grupo 0 en su mayoría contiene partículas del satélite. La contaminación estaría compuesta por las partículas del halo en ese grupo.

Para medir la contaminación nos basamos en el grupo 0 que contiene la mayor parte de las partículas del satélite, la suma de las partículas del halo en los puntos core y frontera con etiqueta 0, es decir, $1648 + 5352 = 7000$ dividido con el número total de partículas en ese grupo, 13810.

La proporción de contaminación se calcula como:

Proporción de contaminación:

$$P = \frac{\text{Contaminacion}}{\text{Totaldeparticulasenelgrupo}} = \frac{7000}{13810} = 0.5065$$

Grupo	H en Core (Halo)	Sat en Core (Sat)	H en Frontera (Halo)	Sate Front (Sat)	Sate tot Sat tot
-1 (Ruido)	0	0	158036	3480	3480
0	1648	4845	5352	2965	6810

Cuadro 6.3: Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera para el experimento vx+vy+vz+rho.

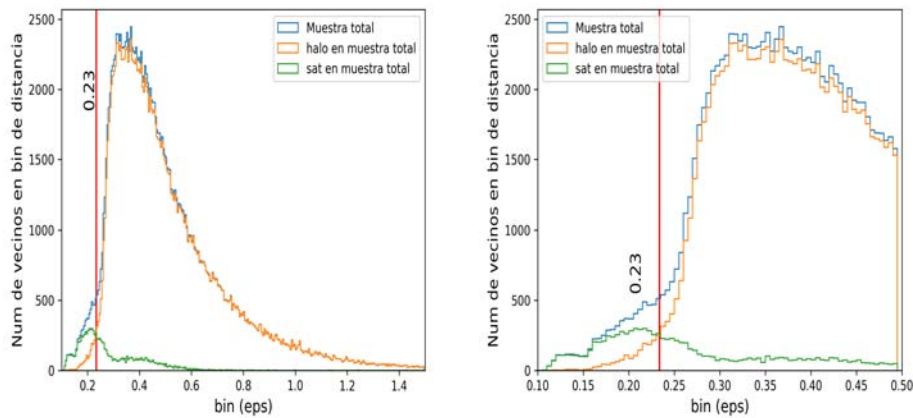


Figura 6.8: Izq: Histograma de distribución de vecinos en función de la distancia para el experimento con vx+vy+vz+densidad. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN. Der: Amplificación del histograma.

6.6.2. Experimento utilizando vr + vtan + densidad en el plano grueso

Se pueden utilizar las componentes de velocidad v_x , v_y y v_z para calcular la velocidad radial y tangencial en el contexto de una galaxia satélite cayendo hacia una galaxia más grande.

- **Calcular la Magnitud de la Velocidad Total (v):** Primero, calcula la magnitud de la velocidad total utilizando las componentes de velocidad v_x , v_y y v_z :

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

- **Calcular la Velocidad Radial (V_r):** La velocidad radial se refiere a la componente de velocidad a lo largo de la línea de visión. Si tienes la posición x , y y z de la partícula, puedes calcular la distancia r desde el centro de la galaxia grande a la partícula:

$$r = \sqrt{x^2 + y^2 + z^2}$$

- Luego, se calcula la velocidad radial V_r utilizando el producto escalar entre el vector de posición (x, y, z) y el vector de velocidad (v_x, v_y, v_z) :

$$V_r = \frac{v_x \cdot x + v_y \cdot y + v_z \cdot z}{r}$$

- **Calcular la Velocidad Tangencial (V_t):** La velocidad tangencial se refiere a la componente de velocidad perpendicular a la línea de visión. Se calcula utilizando la magnitud de la velocidad total v y la velocidad radial V_r :

$$V_t = \sqrt{v^2 - V_r^2}$$

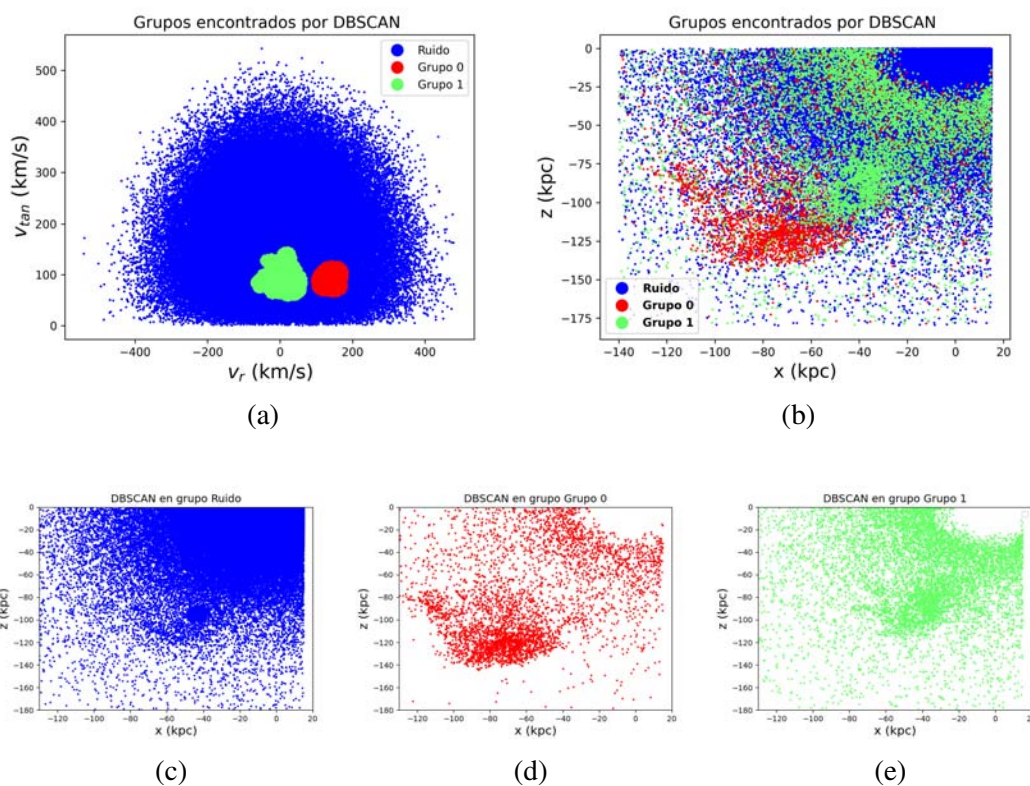


Figura 6.9: Sup izq: Grupos encontrados por DBSCAN en un corte del espacio de velocidades v_r vs v_{tan} . Sup. der: Grupos encontrados por DBSCAN en un corte del espacio x vs z . Paneles inferiores: a) Ruido b) grupo 0 c) grupo 1.

Grupo	H en Core (Halo)	Sat en Core (Sat)	H en Frontera (Halo)	Sate Front (Sat)	Sate tot Sat tot
-1	0	0	156456	6174	6174
0	669	1857	1286	585	2242
1	1549	733	5076	941	1674

Cuadro 6.4: Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera para el experimento con velocidad+densidad.

- Para el Grupo 0:

Total de Partículas: $669 + 1857 + 1286 + 585 = 4397$ Nivel de Contaminación: $\frac{669+1286}{4397} \times 100 \approx 57,77 \%$

- Para el Grupo 1:

Total de Partículas: $1549 + 733 + 5076 + 941 = 8309$ Nivel de Contaminación: $\frac{1549+5076}{8309} \times 100 \approx 80,04 \%$

El nivel de contaminación promedio para los grupos 0 y 1 se calcula sumando los niveles de contaminación de cada grupo y luego dividiendo por el número total de grupos considerados. En este caso, tenemos dos grupos (0 y 1). Entonces:

$$\text{Contaminación Promedio} = \frac{\text{Contaminación Grupo 0} + \text{Contaminación Grupo 1}}{2}$$

Usando los valores calculados previamente:

$$\text{Nivel de Contaminación Promedio} = \frac{57,77 + 80,04}{2} \approx 68,91 \%$$

Por lo tanto, el nivel de contaminación promedio para los grupos 0 y 1 sería aproximadamente 68.91 %.

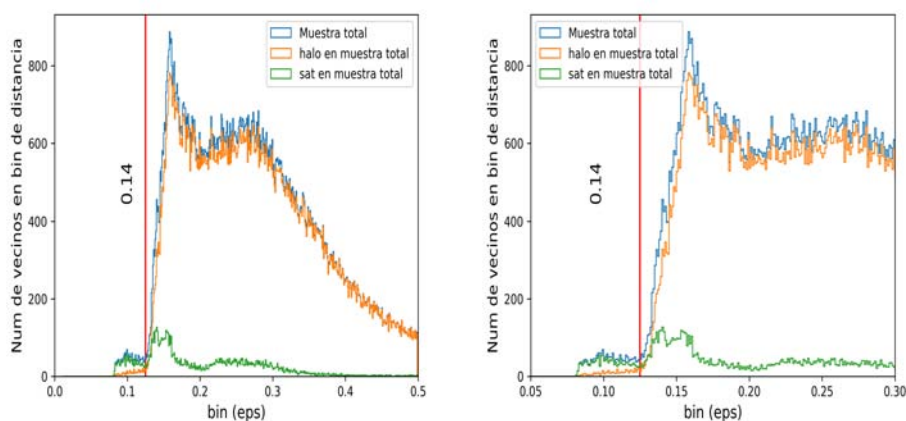


Figura 6.10: Izq: Histograma de distribución de vecinos en función de la distancia para el experimento con $v_{rad}+v_{tan}+\rho$, la roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo **DBSCAN**. Der: Amplificación del histograma.

6.7. Experimento con información cinemática: v_x, v_y, v_z

El siguiente experimento es una prueba que incluye sólo la información cinemática. Los resultados de esta prueba se muestran en la tabla 6.5. La figura 6.11 muestra los grupos encontrados con este conjunto de datos, tanto en el espacio de velocidades como en el espacio de configuración.

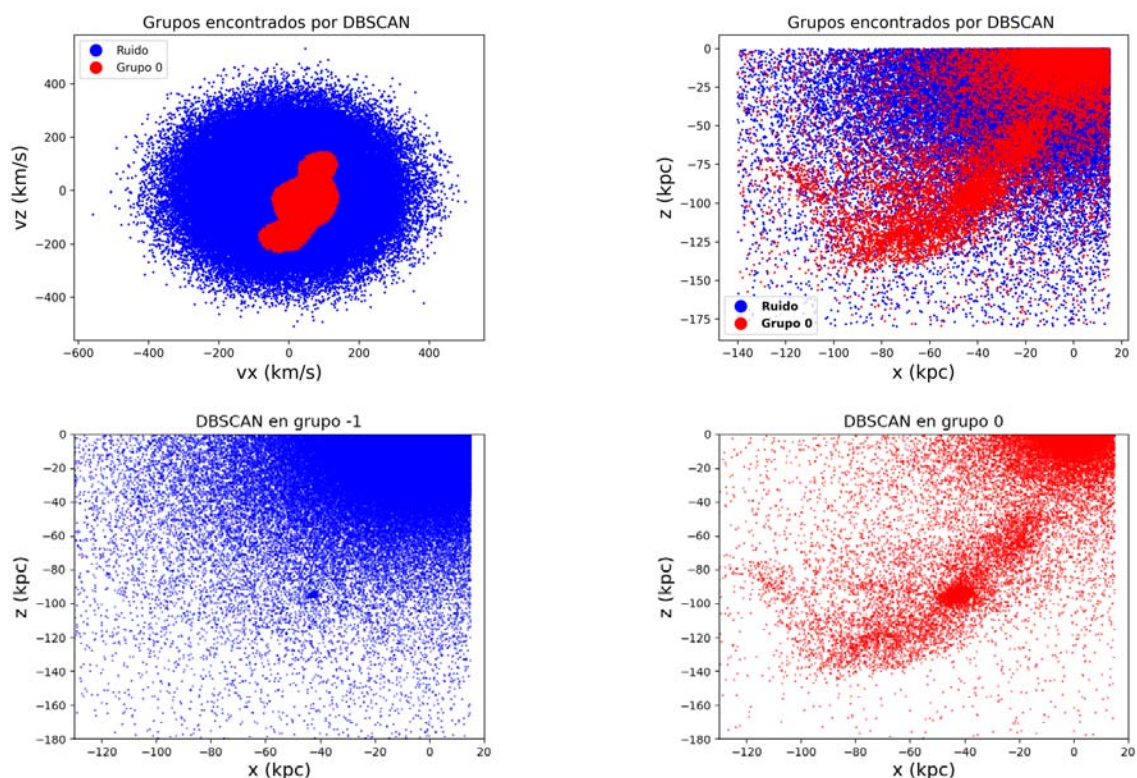


Figura 6.11: Sup izq: Grupos encontrados por **DBSCAN** en un corte del espacio de velocidades vx vs vz. Sup. der: Grupos encontrados por **DBSCAN** en un corte del espacio x vs z. Inf izq: Partículas del grupo 0 conformado en mayor parte por partículas del satélite. Inf derecha: Grupo encontrado por **DBSCAN**, compuesto en su mayor parte por partículas del halo, se considera mayormente como ruido.

Grupo	H en Core (Halo)	Sat en Core (Sat)	H en Frontera (Halo)	Sate Front (Sat)	Sate tot Sat tot
-1 (Ruido)	0	0	150349	1506	1506
0	4273	5674	10414	3110	8784

Cuadro 6.5: Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera para el experimento con vx+vy+vz.

- Para el Grupo 0:

Total de Partículas: 32255 Nivel de Contaminación: $\frac{14687}{32255} \times 100 \approx 45,5 \%$

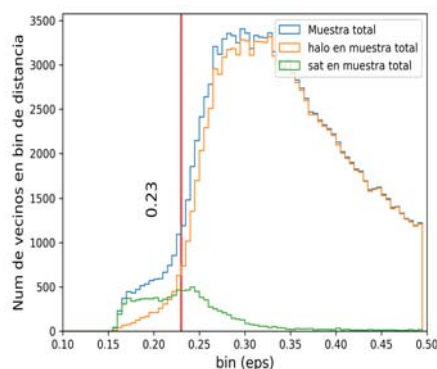


Figura 6.12: Izq: Histograma de distribución de vecinos en función de la distancia para el experimento con solo velocidades. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo **DBSCAN**. Der: Amplificación del histograma.

6.7.1. Búsqueda en una zona conforme con $v_x+v_y+v_z$ +densidad

En esta sección realizamos una búsqueda sobre las simulaciones utilizando **DBSCAN** con la metodología de la sección 6.2 utilizando la información en velocidades y el parámetro de densidad, se presenta también el análisis de la fidelidad de cada grupo encontrado en términos de la contaminación con partículas del halo.

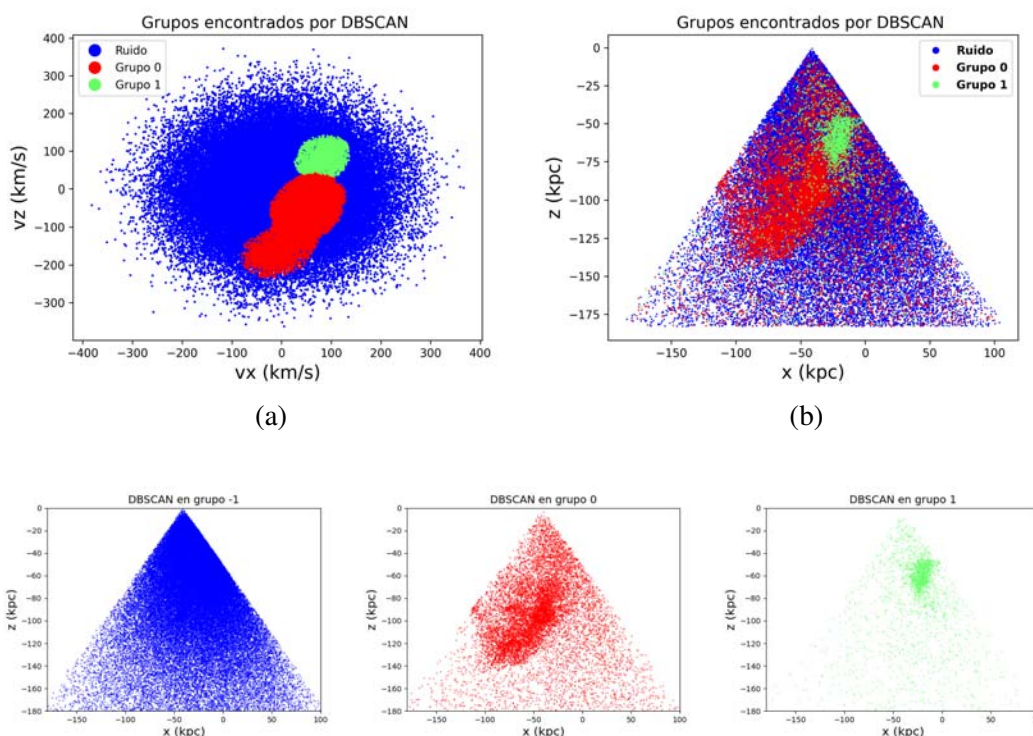


Figura 6.13: Superior: Grupos encontrados por DBSCAN búsqueda en el cono, visto en un corte del espacio de velocidades vx vs vz. Inf: Grupos individuales encontrados por DBSCAN.

Grupo	H en Core (Halo)	Sat en Core (Sat)	H en Frontera (Halo)	Sate Front (Sat)	Sat tot (Sat tot)
-1	0	0	60340	3891	3891
0	1365	3448	4416	2332	5780
1	112	248	1017	830	878

Cuadro 6.6: Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera.

- Para el Grupo 0:
 Total de Partículas: $1365 + 3448 + 4416 + 2332 = 11561$
 Nivel de Contaminación: $\frac{1365+4416}{11561} \times 100 \approx 50,05 \%$
- Para el Grupo 1:
 Total de Partículas: $112 + 248 + 1017 + 830 = 2207$
 Nivel de Contaminación: $\frac{112+1017}{2207} \times 100 \approx 51,16 \%$

El nivel de contaminación promedio para los grupos 0 y 1 se calcula sumando los niveles de contaminación de cada grupo y luego dividiendo por el número total de grupos considerados.

En este caso, tenemos dos grupos (0 y 1). Entonces:

$$\text{Nivel de Contaminación Promedio} = \frac{50,05 + 51,16}{2} \approx 50,61 \%$$

Por lo tanto, el nivel de contaminación promedio para los grupos 0 y 1 sería aproximadamente 50.61 %.

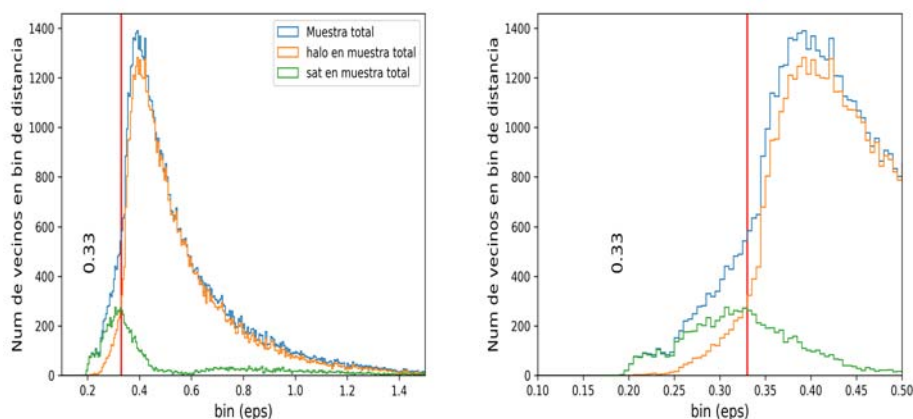


Figura 6.14: Izq: Histograma de distribución de vecinos en función de la distancia para el experimento en una zona conforme utilizando las variables vxvyvz+densidad. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN. Der: Amplificación del histograma.

6.8. Comparativa: Experimento vx+vy+vz+rho y experimento vr+vtan+rho

Cuando observamos los histogramas correspondientes a los experimentos con velocidades y densidad y el experimento con velocidad radial, velocidad tangencial y densidad notamos algunas diferencias en la forma en que se representa la señal (número de vecinos en cada bin de espaciamento), ver fig. 6.15. En esta sección nos proponemos dar una explicación a este fenómeno en las señales.

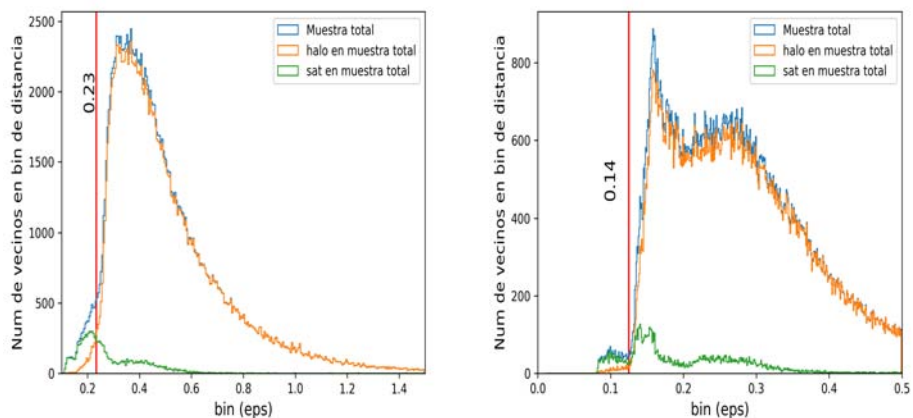


Figura 6.15: Izq: Histograma de distribución de vecinos para el experimento con $v_x+v_y+v_z+densidad$. Der: Histograma de distribución de vecinos para el experimento con $v_r+v_{tan}+densidad$. Observar que la señal tiene una forma distinta para cada caso.

La figura 6.16 ilustra lo que sucede con los histogramas cuando reducimos el número de variables. Cuando reducimos el número de variables a dos dimensiones de velocidad y densidad recuperamos los picos en la señal, muy parecido a lo que sucede en el caso del experimento con velocidad radial y tangencial de la figura 6.15.

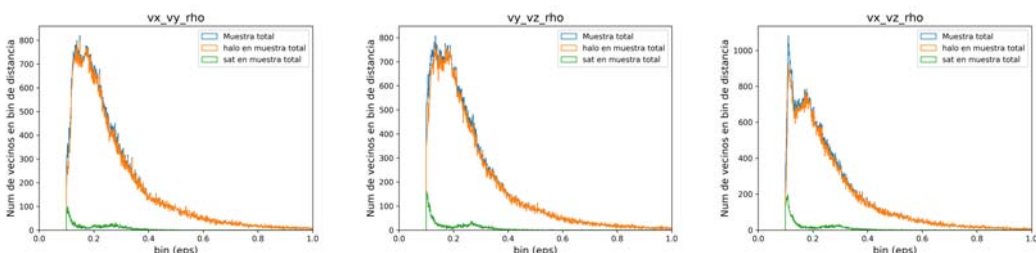


Figura 6.16: Inspección de los tres casos de combinación de variables de velocidad para el experimento con velocidades+densidad.

6.8.1. Exploración de las partes que componen a la señal $v_{rad}+v_{tan}+\rho$.

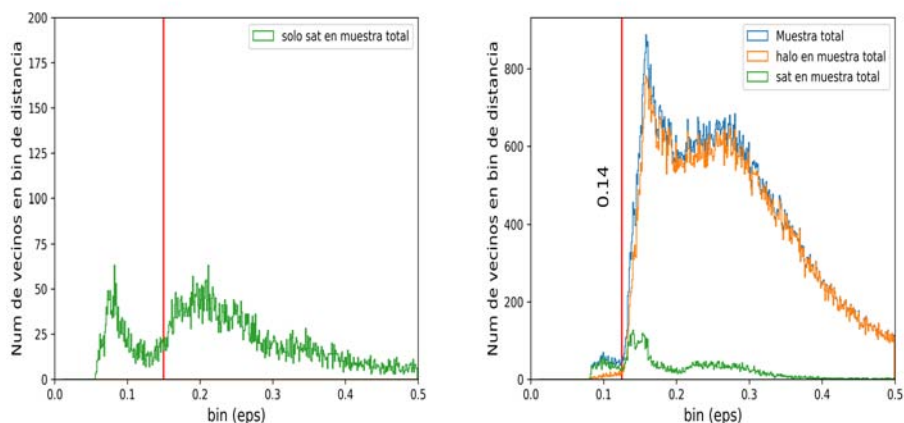


Figura 6.17: Izq: Histograma de distribución de vecinos para solo el satélite del experimento con $v_{tan}+densidad$. Der: Histograma de distribución de vecinos para el experimento con $v_r+v_{tan}+densidad$ con todas las partículas (Halo+satélite).

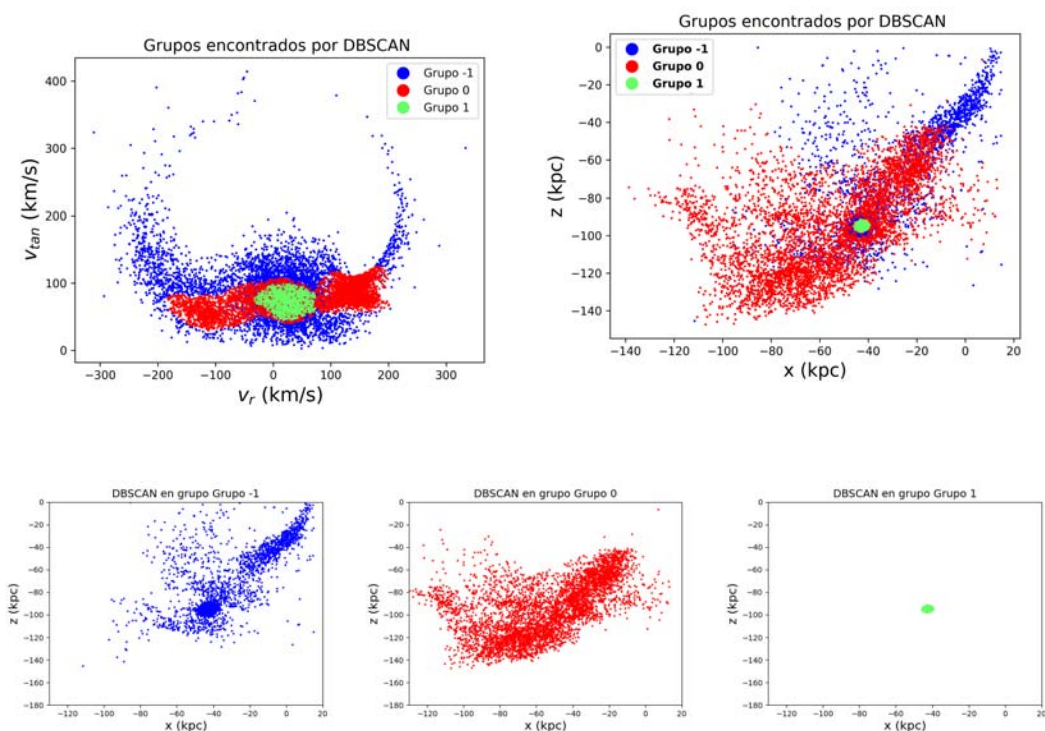


Figura 6.18: Esta figura representa lo que **DBSCAN** clasifica como grupos al pasarle solo las partículas del satélite, sin las del halo, en el espacio de velocidad radial y velocidad tangencial. Los paneles superiores representan los grupos encontrados en el espacio de $v_{rad}+v_{tan}$ lado izquierdo y en el espacio de configuración x vs z lado derecho. Los paneles inferiores representan como se ve cada grupo en el espacio de configuración.

6.8.2. Experimentos con solo vrad y con solo vtan

A continuación, decidimos hacer un experimento en el que solo tuvieramos disponible la velocidad radial y uno donde solo tuvieramos la velocidad tangencial, para medir el efecto que tiene el no disponer de la información 6D completa.

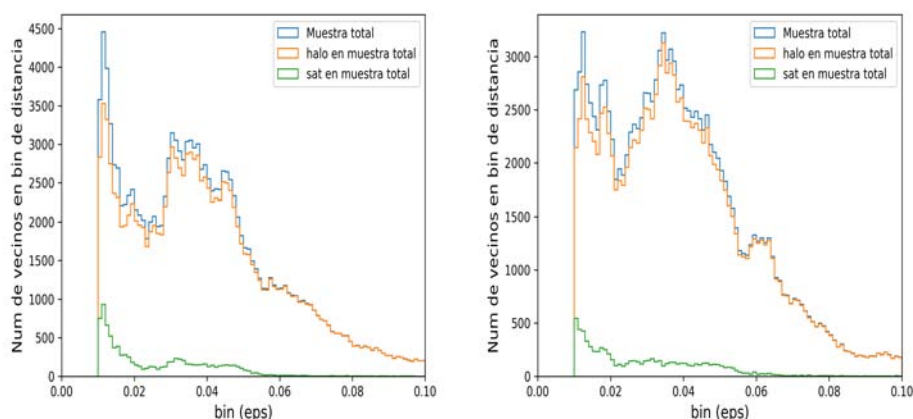


Figura 6.19: Izq: Histograma de distribución de vecinos en función de la distancia para el experimento solo utilizando las variables vrad+densidad. Der: Histograma de distribución de vecinos en función de la distancia para el experimento solo utilizando las variables vtan+densidad

Podemos ver en la figura 6.19 el histograma para un conjunto de datos que sólo utiliza vrad (izq), el histograma para un conjunto de datos que sólo utiliza vtan (der). Es claro que no hay una señal del satélite que sobresalga sobre los datos del halo, por lo que no hay manera de poder encontrar algún tipo de agrupamiento asociado al satélite, sólo utilizando vrad o vtan.

En esta sección implementamos una metodología para realizar la búsqueda de subestructura en las simulaciones con el algoritmo **DBSCAN**. Probamos diferentes conjuntos de datos, una muestra de datos en un plano que contiene al satélite y en una zona coniforme más parecido a las observaciones. Probamos diferentes combinaciones de variables entre velocidades (vrad, vtan), posiciones y densidad.

En cada caso analizamos la fidelidad de los grupos encontrados en términos de la contaminación del halo. Concluimos que incluir la información del espacio 6D para nuestro caso permite encontrar los grupos con una fidelidad de hasta 70 %, esto mejorará al utilizar otras variables que permitan resaltar la señal de la subestructura como la metalicidad y con la mejora en la resolución de los satélites lo cual se logrará con las simulaciones cosmológicas, así como tener una mejor representación de los datos del satélite con respecto de los del halo en una zona coniforme. En el caso de utilizar solo vr+rho o solo vtan+rho, el resultado es prácticamente nulo y no es posible resaltar la señal de la subestructura del satélite sobre el

halo estelar, lo que le da más peso al hecho de que tanto *Gaia* como DESI proporcionarán velocidades radiales en grandes cantidades en los próximos años.

Capítulo 7

Simulaciones cosmológicas de formación de galaxias

7.1. Introducción

El estudio de galaxias y la evolución del universo se suele abordar desde distintos enfoques: las simulaciones de N-cuerpos de sistemas aislados y las simulaciones de N-cuerpos cosmológicas de formación de galaxias.

Las simulaciones de N-cuerpos de sistemas aislados tal como se discutieron en la sección 3, se concentran en estudiar un sistema específico, como una galaxia. Su propósito es poner a prueba hipótesis que permitan entender su dinámica y evolución de manera detallada. Para ello, se basan en condiciones iniciales que representan estados concretos o idealizados, como una galaxia con parámetros determinados. Estas simulaciones se enfocan en escalas galácticas y abarcan regiones pequeñas, permitiendo incluir más detalles y resolver estructuras finas. Están diseñadas para no considerar interacciones con estructuras externas. Son ideales para estudiar la evolución de galaxias satélites y la dinámica de discos galácticos.

Por otro lado, las Simulaciones cosmológicas de formación de galaxias que incluyen hidrodinámica y N-cuerpos tienen como objetivo comprender la evolución de estructuras en el universo a gran escala. Sus condiciones iniciales provienen del espectro de potencia primordiales del universo temprano, fundamentado en observaciones del CMB (Turner, 1993). Estas simulaciones abarcan desde escalas galácticas hasta supercúmulos. Dada su magnitud, es necesario hacer compromisos en detalles y resolución. Aquí, las galaxias y estructuras interactúan entre sí y con su entorno. Son esenciales para entender la evolución del universo desde el Big Bang en el contexto de un modelo cosmológico, ver fig 7.1.

En contraste con las simulaciones de N-cuerpos de sistemas aislados, las simulaciones cosmológicas de formación de galaxias consideran un panorama más extenso, modelando vastas regiones del universo y la evolución de estructuras dentro del marco de un modelo

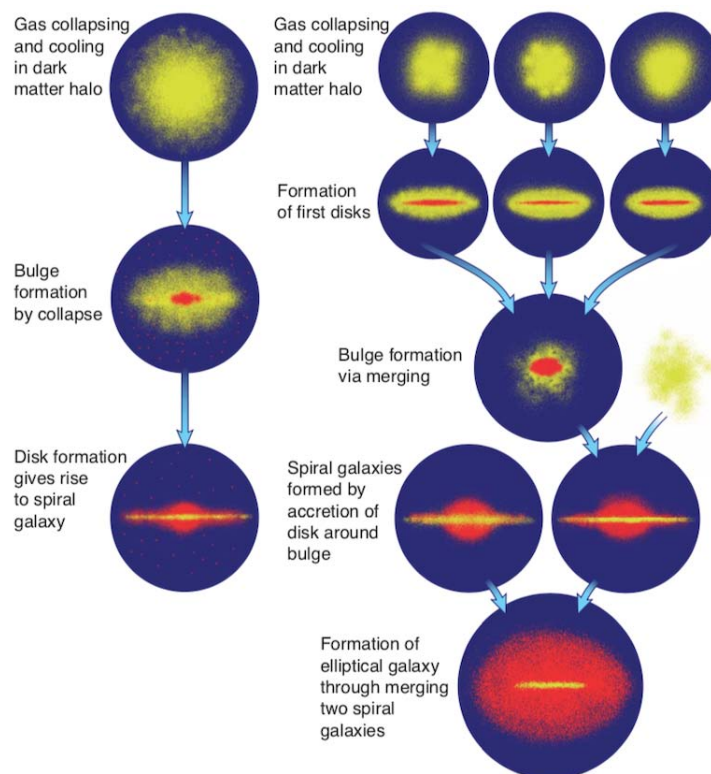


Figura 7.1: Modelos de evolución galáctica. Izq. Colapso Monolítico: El gas se enfría primero y colapsa en el halo oscuro. El gas se acomoda en el disco. Una galaxia espiral se forma con estrellas, gas y polvo en el disco. Si el gas primordial se convierte en estrellas prontamente, se forma una galaxia elíptica en lugar de una galaxia espiral. Der: Modelo jerárquico: El gas enfriándose y colapsando es seguido por la formación de los primeros discos. Varias de estas protogalaxias surgen a través de fusiones de dos galaxias espirales. El gas residual puede acomodarse en un disco delgado. Halo oscuro (Azul), gas (Amarillo), estrellas (rojo). Crédito de imagen y texto: Bertola (2002)

cosmológico dado. Estas buscan comprender el universo en su totalidad: la formación de estructuras, la evolución temporal de las galaxias y la influencia de la materia y energía oscura en dichos procesos. A través de estas simulaciones, es posible hacer predicciones sobre la distribución a gran escala de la materia en el universo (Vogelsberger et al., 2020; Han et al., 2022).

En este capítulo se describe una de las primeras simulaciones cosmológicas realistas de alta resolución de formación y evolución galáctica GARROTXA (Roca-Fàbrega et al., 2016), así como sus modelos parecidos en masa a la Vía Láctea. El objetivo principal de esta simulación es establecer un puente entre los procesos dinámicos y la parte cosmológica.

7.2. Simulaciones cosmológicas de formación de galaxias

Las simulaciones cosmológicas son herramientas que permiten estudiar el origen y la evolución de estructuras cósmicas a gran escala, y que sirven para interpretar a las observaciones.

En el caso de la formación de galaxias tales simulaciones se utilizan para interpretar las observaciones, estudiar en un contexto cosmológico los procesos físicos que dan lugar a la formación de galaxias: formación estelar, dinámica del gas, estrellas y materia oscura, efectos del medio ambiente, evolución química.

La primera simulación gravitacional de N-cuerpos de galaxias interactivas se realizó usando un ordenador óptico analógico (Holmberg, 1941): En este modelo, la gravedad se representaba por el flujo de bombillas, con fotocélulas y galvanómetros usados para medir y mostrar la fuerza de la ley del cuadrado inverso que sigue el comportamiento de la luz. Los primeros cálculos astronómicos de N-cuerpos usando ordenadores digitales fueron realizados a comienzos de la década de 1960 (von Hoerner, 1960; Aarseth, 1963). Pero estos primeros intentos de simulaciones estaban limitadas a un máximo de cien partículas aproximadamente. Las simulaciones de la dinámica de gases en la formación de galaxias comenzaron con los cálculos esféricamente simétricos de Larson (1969) y durante la década de los años 1970 se realizaron simulaciones cada vez más grandes del colapso y evolución de los cúmulos (ver por ejemplo, Peebles (1970), White (1976)).

7.2.1. Componente colisional en simulaciones de N-cuerpos

Para estudiar procesos altamente afectados por la presencia de componentes hidrodinámicas se vuelve necesario agregar una componente que es colisional. Las simulaciones que incluyen la componente colisional permiten estudiar estas propiedades a gran escala, como la distribución de edades en galaxias con diferentes morfologías, su evolución química y dinámica y la distribución de gas interestelar y su interacción con sistemas galácticos, un ingrediente que tiene impacto en el proceso de formación y evolución de galaxias, ver figura 7.2 para una comparativa directa de considerar la parte colisional.

Durante las últimas tres décadas surgieron las primeras simulaciones cosmológicas de formación de galaxias que empezaron a tomar en cuenta métodos para tomar en cuenta la formación estelar en las galaxias, uno los primeros intentos se puede ver en el trabajo de Yepes et al. (1997) donde introducen el enfriamiento del gas en varias etapas con tiempos de enfriamiento cortos y con formación de estrellas. También Abadi et al. (2003b) introducen el “feedback” estelar en sus simulaciones, pero la formación estelar no adecuada da lugar a un esferoide (transferir momento angular hacia el halo) que afecta las curvas de rotación y la relación de Tully-fisher para galaxias espirales. Por ejemplo Robertson et al. (2004) logran

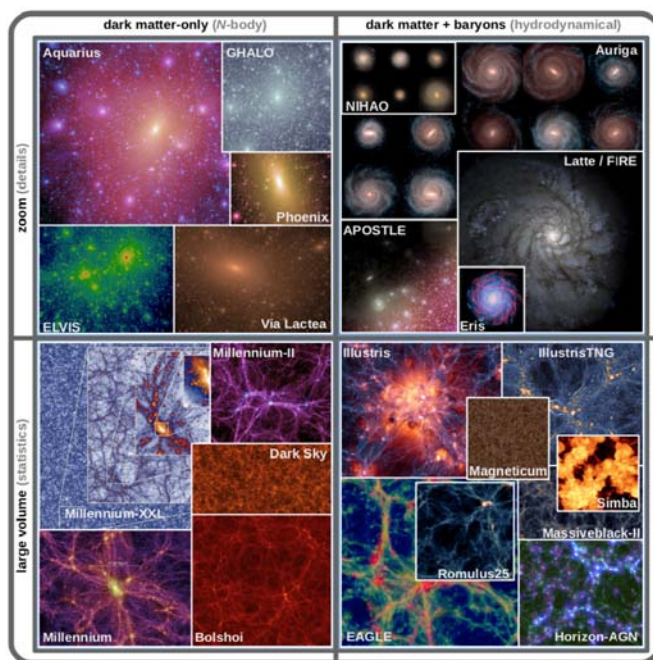


Figura 7.2: Crédito de la imagen: (Vogelsberger et al., 2020). Representaciones visuales de simulaciones recientes de formación de estructuras y galaxias seleccionadas. Algunas son de gran volumen, lo que proporciona una amplia visión estadística de las galaxias, mientras que otras son de zoom-in y resuelven detalles más pequeños. También hay dos tipos de simulaciones: las de materia oscura (N-cuerpos) y las hidrodinámicas, que incluyen la materia oscura y los bariones. Las simulaciones de materia oscura han convergido en predicciones sobre la distribución a gran escala y la estructura de los halos de materia oscura. Las simulaciones hidrodinámicas coinciden sorprendentemente bien con observaciones reales en cuanto a la formación de galaxias, pero algunos detalles aún dependen de cómo se implementa la física de los bariones en el modelo.

reproducir la relación de Tully-fisher pero disponen de baja resolución en sus simulaciones.

Posteriormente otros trabajos como Governato et al. (2007) reproducen la relación Tully-fisher para galaxias espirales, por ejemplo Governato et al. (2010) logran relajar la tensión entre las observaciones y las predicciones del modelo Λ CDM. Okamoto et al. (2005) también logran reproducir observables como la relación de Kennicutt–Schmidt.

Estos trabajos demostraron que era importante introducir el “feedback” de supernovas para lograr que la formación estelar en las simulaciones fuera eficiente para dar lugar a la formación de estructuras como el disco y reproducir así las observaciones.

Las simulaciones GARROTXA de Roca-Fàbrega et al. (2016) logran reproducir las observables, tanto la curva de rotación, como la fracción de masa estelar y la relación de Tully-fisher para galaxias espirales.

La generación más reciente de simulaciones incluyendo las simulaciones GARROTXA de Roca-Fàbrega et al. (2016), Horizon de Lee et al. (2021), Illustris de Vogelsberger et al. (2014), TNG de Nelson et al. (2019), IllustrisTNG de Nelson et al. (2019), LATTE (GIZMO) de Wetzel et al. (2016) y AURIGA de Grand et al. (2018) son simulaciones cosmológicas de formación de galaxias en el estado del arte que logran reproducir observables y tienen mejoras considerables en cuanto a la resolución ver tabla 7.2 para una comparación directa con GARROTXA.

7.3. Los modelos de la simulación cosmológica de alta resolución de formación y evolución de galaxias: GARROTXA

Las simulaciones GARROTXA son simulaciones cosmológicas de alta resolución de formación y evolución de galaxias, realizadas por Roca-Fàbrega et al. (2016) utilizando el código Adaptive Refinement Tree (ART) (Hidrodinámica Euleriana + malla adaptiva + N-cuerpos) (Kravtsov et al., 1997; Kravtsov, 2003), en la versión de Colín et al. (2010) para incluir galaxias de baja masa, GARROTXA consiste de un conjunto de cuatro simulaciones cosmológicas de alta resolución espacial (~ 100 pc) con hidrodinámica, de la formación de galaxias tipo Vía Láctea, en una cosmogonía tipo Λ CDM. La masa virial final del sistema es de aproximadamente $7,4 \times 10^{11} M_{\odot}$. Cada una de las simulaciones se realizó con un valor distinto para la resolución en masa de la componente bariónica, lo que, a su vez, también condujo a un número mayor de celdas de alta resolución en la región del disco galáctico y de las corrientes estelares. Todos los modelos incluyen procesos bariónicos como formación y evolución estelar, retroalimentación de supernovas, enfriamiento y calentamiento de gas y evolución química (elementos alfa). La receta de retroalimentación consiste en depositar

toda la energía proveniente de los vientos estelares y las supernovas en el gas inmediatamente después del nacimiento de la partícula estelar.

7.4. Modelo de N-cuerpos+hidrodinámica

En estas simulaciones se inicia con un espectro impuesto de fluctuaciones en densidad derivado de CMB (Spergel et al., 2007). La simulación corre bajo una cosmología inicial adoptada definida por la constante de Hubble, el espectro de la materia oscura fría y las componentes de la densidad del universo.

Al comienzo las sobredensidades iniciales colapsan debido a la auto-gravedad y formando paredes cosmológicas, filamentos y huecos y a pequeñas escalas halos de materia oscura.

Después cuando el universo se enfría, los bariones comienzan a seguir los procesos físicos implementados dando lugar a los halos de materia oscura. El surgimiento de los bariones ocurre a diferentes tasas, dependiendo de la masa del halo de materia oscura.

Dependiendo del ambiente (i.e interacciones con otros sistemas), los bariones pueden dar lugar a discos por la conservación de momento angular o una distribución esferoidal si los sistemas sufren interacciones fuertes (Read et al., 2009; Halle and Combes, 2013)

Entonces las estrellas comienzan a formarse, cuando esto pasa los procesos de retroalimentación y evolución química comienzan a seguir la física implementada para SNeIa, SNeII y vientos estelares.

Como consecuencia la evolución químico-dinámica comienza. Este proceso es fundamental para obtener sistemas galácticos realistas.

Algunos investigadores han replicado sistemas parecidos a la Vía Láctea en propiedades a gran escala, pero no en aspectos como galaxias satélites o distribución de gas. Las discrepancias con las observaciones de la Vía Láctea surgen al analizar detalles como formación estelar y distribuciones de edad. Para mejorar la concordancia, es crucial entender y aplicar correctamente los procesos físicos internos, considerando que algunos requieren aproximaciones debido a limitaciones de resolución. Abordar este desafío es la motivación del surgimiento de mejores modelos cosmológicos de N-cuerpos, como los que incluyen la parte hidrodinámica.

7.4.1. La componente Hidrodinámica

En la malla sin colisiones solo actúa la gravedad, mientras que en la malla hidrodinámica se considera también el modelado de una componente colisional que sigue las leyes de la hidrodinámica.

A continuación se mencionan los procesos físicos que regulan las propiedades globales como la formación estelar o la fracción de bariones en halos de materia oscura.

7.4.2. Procesos físicos

En esta sección se abordan los procesos físicos del código.

- **Enfriamiento y calentamiento:** Estos procesos regulan propiedades como la tasa de formación estelar. Se incluyen mecanismos como el calentamiento/compton, enfriamiento atómico y molecular, y el calentamiento por radiación UV del fondo cósmico.
- **Formación estelar:** Basada en observaciones, se necesita imitar físicas a escalas más pequeñas que la resolución actual (físicas “subgrid”). Se observa que la tasa de formación estelar depende de la masa de gas denso en complejos nubosos. El proceso es lento, y suele modelarse en regiones frías y densas.
- **Evolución estelar, retroalimentación de supernovas y advección de metales:** Las partículas estelares, mucho más masivas que una estrella individual, se consideran poblaciones estelares únicas. Eyectan metales y energía a través de vientos estelares y explosiones de supernovas.

7.4.3. Técnica de retroalimentación dinámica

Para evitar el conocido como el problema de exceso de enfriamiento, el cual se produce cuando la energía liberada en explosiones de supernovas se pierde eficientemente por procesos de radiación resultando en una pérdida continua de energía y momentum angular que causa picos centrales no realistas en las curvas de rotación, en GARROTXA se usa la técnica de retroalimentación dinámica. A diferencia de otras soluciones propuestas que se basaban en retrasar artificialmente el enfriamiento en regiones donde se forman estrellas o implementar nuevas físicas “subgrid” como la presión de radiación, en GARROTXA se deposita toda la energía proveniente de los vientos estelares y las supernovas en el gas inmediatamente después del nacimiento de la partícula estelar de modo que esta inyección repentina de energía es capaz de elevar la temperatura del gas en la celda en el orden de 10^7 K, un valor suficientemente alto para hacer que el tiempo de enfriamiento sea mayor que el tiempo de cruce y así se evita la mayor parte del exceso de enfriamiento.

7.4.4. Comparación con otros modelos disponibles

El modelo de GARROTXA de más alta resolución, aunque se limita a un único sistema galáctico, compite con respecto a esfuerzos más recientes, como los modelos del proyecto AURIGA (Grand et al., 2017), el cual fué uno de los más ambiciosos en su tipo, u otros esfuerzos contemporáneos como NIHAO (Wang et al., 2015), incluyendo también una física bariónica similar a la de Wadsley et al. (2004).

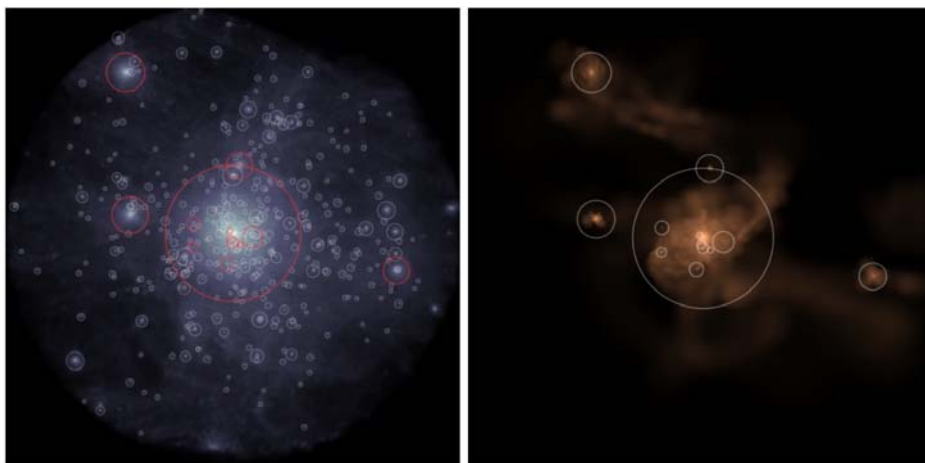


Figura 7.3: Densidad superficial de materia oscura (izquierda) y estelar de la simulación GARROTXA G.321, at $z = 0$. (derecha). Los círculos indican la posición y radio del virial de los halos que contienen estrellas, es decir, de la galaxia central y de sus galaxias satélites, obtenidos usando el buscador de halos Rockstar” (Behroozi et al., 2013).

Aunque la resolución espacial no es tan buena como en modelos mucho más recientes, ver por ejemplo los modelos LATTE del grupo de FIRE (Wetzell et al., 2016), sí que permite capturar la formación de estructuras no axisimétricas como brazos espirales o la caída de gas frío a través de filamentos (ver la figura 7.4).

A continuación describimos las propiedades generales del modelo GARROTXA, en su versión de mayor resolución. Los valores se han calculado dentro del radio del virial de la galaxia principal definido como $R_{vir}=R_{200}=175.6$ pc.

- Número de partículas de materia oscura: 7×10^6 .
- Número de partículas estelares: 2.34×10^6 .
- Número de celdas de gas: 6.8×10^6 .
- Masa mínima de partículas bariónicas y de materia oscura: $1.2 \times 10^3 M_{\odot}$ y $9.3 \times 10^4 M_{\odot}$, respectivamente.
- Resolución espacial: 109 pc por celda en el disco de la galaxia principal, 1 kpc en las regiones de más baja resolución del halo.
- Resolución temporal (paso de tiempo mínimo): 10^3 años
- Masa virial de la galaxia central: $\sim 7 \times 10^{11} M_{\odot}$.

El tener simulaciones con semejanzas cuantitativas como GARROTXA (esta tesis), LATTE, AURIGA, TNG, Illustris, etc., permite que los estudios observacionales se

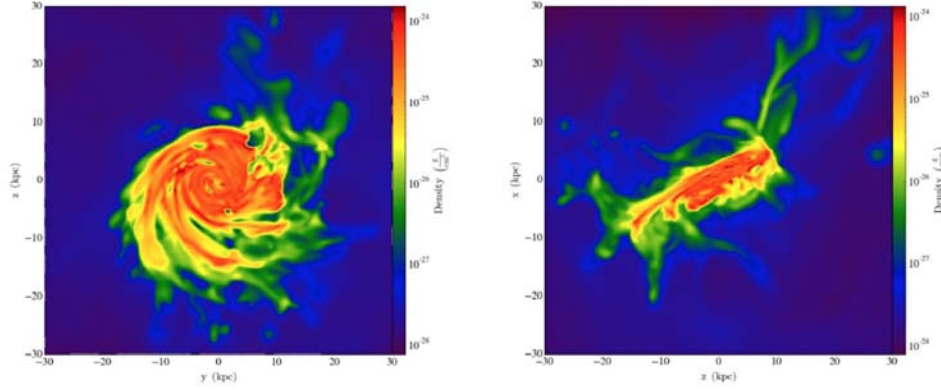


Figura 7.4: Proyecciones de la densidad superficial de gas en la galaxia principal del modelo GARROTXA. La resolución espacial de ~ 100 pc en el disco galáctico permite capturar la formación de estructuras espirales y las corrientes frías de gas que alimentan la formación estelar en la galaxia.

	r_{vir} [kpc]	M_{vir} [M_{\odot}]	M_* [M_{\odot}]	M_{gas} [M_{\odot}]	M_{hotgas} [M_{\odot}]	$M_{warmgas}$ [M_{\odot}]	$M_{coldgas}$ [M_{\odot}]
G.321	230.1	$7,33 \times 10^{11}$	$6,1 \times 10^{10}$	$2,70 \times 10^{10}$	$1,22 \times 10^{10}$	$5,66 \times 10^9$	$9,34 \times 10^9$
MW obs.	-	$1,0 \pm 0,30 \times 10^{12}$	$4,9 - 5,5 \times 10^{10}$	-	-	-	$7,3 - 9,5 \times 10^9$
	r_{200} [kpc]	M_{200} [M_{\odot}]	$F_{b,U}$	m_{DM1sp} [M_{\odot}]	m_* mín [M_{\odot}]	m_* máx [M_{\odot}]	m_{SPH} [M_{\odot}]
G.321	175.6	$6,84 \times 10^{11}$	0.120	$0,93 \times 10^5$	$0,24 \times 10^4$	$120,0 \times 10^4$	-
MW obs.	-	-	-	-	-	-	-
	N_{part} [10^6]	N_{DM1sp} [10^6]	N_* [10^6]	N_{gas} [10^6]	Resolution [pc]	CPU time [10^4 h]	CODE
G.321	7.52	7.12	0.39	2.0	109(1 cell)	2.5	ART + hydro
MW obs.	-	-	$\sim 100000,0$	-	-	-	-
	$c = r_{vir}/r_s$	R_d [kpc]	$h_{z,young}$ [pc]	$h_{z,old}$ [pc]	M_{hotgas}/M_{vir}	α_{χ}	SFR ($z=0$) [$M_{\odot}yr^{-1}$]
G.321	28.5	2,56(4,89/2,21)	277(exp)	1356	0.016	-0.62	0.27
MW obs.	$21,1^{+14,8}_{-8,3}$	2,0 - 4,5	300 ± 60	$600 - 1100 \pm 60$	-	-	$0,68 - 1,45$
	$V_{c\odot}(R=8kpc)$ [kms^{-1}]	R_{peak} [kpc]	$V_c(R_{peak})$ [kms^{-1}]	$V_{2,2}/V_{200}$ 2,27 = 200	\bar{n}_{SF} [cm^{-3}]	\bar{T}_{SF} [10^3 K]	ϵ_{SF}
G.321	239.8	5.69	243.8	1.90	1.0	9	0.65
MW obs.	221 ± 18	-	-	$1,11^{+0,22}_{-0,20}$	-	-	-
	Box [$Mpch^{-1}$]	z_{ini}	Ω_0	Ω_{Λ}	$\bar{\Omega}_b$	H_0 [$km\ s^{-1}Mpc^{-1}$]	
G.321	20	60	0.30	0.70	0.045	70	0.80
MW obs.	-	-	-	-	-	-	-

Cuadro 7.1: Parametros de las simulaciones GARROTXA G.321, a $z=0$. Para más detalles ver Tabla 1 de (Roca-Fàbrega et al., 2016)

Proyecto-Código ódigo	Mt[M \odot]	M _{vir} [M \odot]	# Nt Partículas	mgas[M \odot]	mDM[M \odot]	mstar [M \odot]	ϵ_{gas} [pc]	ϵ_D M[pc]	ϵ_{star} [pc]	Ref
Latte [Proyecto FIRE2- Código GIZMO+hydro]	$1,3 \times 10^{12}$	$1,1 \times 10^{12}$	$3,1 \times 10^7$	$7,07 \times 10^3$	$3,5 \times 10^4$	$7,07 \times 10^3$	1	40	40	Philip F. Hopkins et al. 2018
Proyecto Auriga- Código AREPO+hydro Level 3 Resolution	$1 \sim 2 \times 10^{12}$	$1,5 \times 10^{12}$	$3,7 \times 10^7$	$4 \sim 5 \times 10^4$	$3 \sim 4 \times 10^4$	6×10^3	369	184	184	Robert J. Grand et al. 2017
GARROTXA-ART+ hydro HART MW model G.323	$\sim 1 \times 10^{12}$	$6,90 \times 10^{11}$	$9,44 \times 10^6$	$2,88 \times 10^3$	$1,07 \times 10^5$	$2,64 \times 10^4$	109	109	109	Roca-F ábrega, Valenzuela, et al 2016
GARROTXA- RAMSES	$\sim 2,3 \times 10^{12}$	$2,2 \times 10^{12}$	$3,76 \times 10^8$	—	5×10^4	1×10^3	80	80	80	Roca-F ábrega, Valenzuela, en proceso

Cuadro 7.2: Resolución mínima dentro del conjunto de simulaciones Latte del proyecto FIRE2 (Wetzel et al., 2023), de las simulaciones dentro del proyecto Auriga (Grand et al., 2017) y del proyecto Garrotxa-HART y GARROTXA-RAMSES (Roca-Fàbrega et al., 2016)

interpreten como pruebas de la formación y evolución de las galaxias incluida la Vía Láctea así como de la cosmología, desde la naturaleza de la materia oscura hasta las condiciones iniciales.

Capítulo 8

Cátálogos sintéticos basados en las simulaciones de formación de galaxias incluyendo modelos de errores de los censos *Gaia* y DESI

8.1. Introducción

Tanto en la última como dentro de la próxima década, catastró del cielo como *Gaia*, *Euclid*, *DESI*, *LSST* buscan cambiar el paradigma o visión que tenemos acerca del universo e incluso de la propia Vía Láctea. Se espera que estos catastró lleguen a proporcionar posiciones y velocidades para miles de millones de estrellas en la Galaxia, datos astrométricos y fotométricos por parte de catastró como *Gaia* (1.4 mil millones de fuentes en tres bandas) y SDSS/APOGEE (Majewski, 2017), sin embargo, también se proporcionarán velocidades radiales, extinción y parámetros estelares como la temperatura efectiva con mejor significancia debido a las mejoras tecnológicas en estos catastró. Por ejemplo, *DESI* en su fase I proporcionará velocidades radiales para aproximadamente seis millones de estrellas en los próximos años y aproximadamente 30 millones en su segunda etapa, complementando así a *Gaia* DR3 (Gaia Collaboration et al., 2021) que cuenta con números semejantes de velocidades radiales.

El disponer de una gran cantidad de datos observacionales requiere como paso adicional para su interpretación la creación de modelos teóricos y computacionales que guarden cierta consistencia con las observaciones de la Vía Láctea aunque no necesariamente un ajuste a las mismas, pero que ayuden a la interpretación de las mismas referentes a degeneraciones geométricas y de la evolución de la Vía Láctea y así se pueda generar una visión de la estructura y evolución de la Galaxia (Hunt and Kawata, 2014; Nitschai et al., 2021; Khrapov

et al., 2021, e.g.).

Estas técnicas de modelado abarcan desde simulaciones de N-cuerpos que incluyan las componentes de la Galaxia (bulbo, disco, halo), que incluyan los efectos de los procesos evolutivos, y que tienen como fin el poder ofrecer predicciones verificables.

Los datos simulados permiten a los investigadores comprender mejor las limitaciones, errores y sesgos presentes en los datos reales. Para llevar a cabo esta tarea, es importante incluir varios factores, como son la función de selección, errores y extinción por polvo, así como el tamaño de la muestra, el volumen de estrellas del catastro, la precisión en la medición de propiedades del espacio fase y la información obtenida mediante espectroscopía en los datos simulados.

En ese sentido, llevar a cabo esa tarea es la principal motivación para querer crear catálogos sintéticos de las simulaciones, con el fin de estar preparado para poder explotar y comprender adecuadamente los datos de los catálogos generados por catastros como *Gaiay* futuros catastros como **DESI** y **LSST**.

A continuación se enlistan algunas razones específicas y los beneficios que se obtienen al crear catálogos sintéticos con simulaciones basados en el proyecto/catastro:

- **Etapas de validación y prueba:** Los catálogos sintéticos permiten verificar y poner a prueba las herramientas y métodos de análisis en un entorno controlado como lo son las simulaciones, como una fase previa a la generación de los datos reales generados del catastro. Permiten estudiar la pregunta de si se alcanzará la precisión o completitud estadística en un censo observacional.
- **Modelaje Predictivo:** Se busca hacer comparaciones entre las predicciones teóricas y los datos reales de las observaciones, esto da una idea de lo que esperaríamos ver en los datos reales. Esto puede ser útil para identificar características o patrones específicos que podrían estar presentes en los datos observacionales, por ejemplo la abundancia y geometría de la subestructura en el halo de la Galaxia. Dimensiona estrategias para la búsqueda de estructura en el halo estelar es uno de los propósitos de esta tesis.
- **Analizar Limitaciones y Sesgos:** Los datos simulados permiten comprender las limitaciones, errores y sesgos que podrían estar presentes en los datos reales.
- **Extrapolación y complementación:** Algunos fenómenos o características pueden no ser fácilmente observables en los datos reales debido a diversas limitaciones. Los catálogos sintéticos pueden llenar esos vacíos al simular lo que sería observable bajo ciertas condiciones ideales o debido al uso de estrategias o algoritmos de búsqueda o de procesamiento de datos.

En este trabajo desarrollamos una herramienta para crear observaciones sintéticas basadas en las simulaciones cosmológicas de alta resolución de formación y evolución de galaxias de Roca-Fàbrega et al. (2016). Esta herramienta incluye el modelo de errores (paralaje y fotométricos) de Gaia y el modelo de errores en velocidad radial de DESI. La herramienta y las observaciones sintéticas serán publicadas más adelante con el nombre de “MC Balrogs” (Mock Catalogs Based on Large Resolution Galform Simulations).

8.1.1. Antecedentes de catálogos sintéticos

Uno de los primeros intentos de crear catálogos sintéticos fue el modelo de Besançon (Robin and Crézé, 1986), este modelo de la Galaxia es un modelo de síntesis de población estelar. Su principal objetivo es simular y comprender la estructura, cinemática y evolución de la Vía Láctea, aunque no construía un modelo dinámicamente autoconsistente. Las características principales del modelo original incluyen: El proceso de producción de estrellas en el modelo se basaba en lo que se conoce como diagramas de Hess. Estos diagramas se calculaban para cada población galáctica en función de una serie de parámetros, como la Función Inicial de Masa (IMF), que corresponde a la distribución de masas iniciales de las estrellas en una población estelar recién formada en una galaxia. En otras palabras, describe la probabilidad de que una estrella tenga una masa específica al comienzo de su vida, también la Tasa de Formación Estelar (SFR), trayectorias evolutivas y la relación edad-metalicidad. Una vez establecidos estos parámetros y otros relacionados con la evolución estelar, estos permanecían fijos en el modelo. Esto significaba que el modelo no podía ajustarse para probar diferentes escenarios de evolución galáctica. El modelo original se creó con el propósito de comparar sus simulaciones con datos reales, ayudando así a entender mejor la estructura y evolución de la Vía Láctea. Posteriormente se propusieron el modelo de Allen and Santillan (1991) y posteriormente Pichardo et al. (2004), enfocados principalmente a la parte dinámica y orbital.

También están las técnicas tales como “Make to measure” (M2M, por sus siglas en inglés), que buscan encontrar un modelo de partículas que represente fielmente las observaciones de la galaxia, cuando su distribución de densidad superficial y cinemática proyectada coincide con las de la galaxia (Syer and Tremaine, 1996; De Lorenzi et al., 2007; Hunt and Kawata, 2013, e.g.). En resumen, en el método “M2M”, se cambian iterativamente los pesos de las partículas individuales para que el conjunto de partículas coincida con las observaciones. Una vez logrado esto, el modelo se puede utilizar para aprender sobre la estructura intrínseca de la galaxia. Estos modelos pueden utilizarse para modelar el disco galáctico (Hunt and Kawata, 2014), modelar la dinámica de galaxias con datos observados, (Lowing et al., 2014). Cabe mencionar que la autoconsistencia de tales modelos se compromete cuando se varían los pesos de las partículas tales como densidad o masas. Una generalización

un poco más complicada que utiliza como función de variación la ecuación de Boltzmann a través de integradores simplécticos pero que no pierde la autoconsistencia, fue introducida por Rodionov et al. (2009), aunque es poco utilizada.

Existen catálogos simulados completos de estrellas de *Gaia*, por ejemplo, el modelo “Gaia Universe Model Snapshot” (gums) ofrece una visión del modelo de la Galaxia de Besançon como se ve desde *Gaia* (Robin et al., 2012). Este modelo toma en cuenta la extinción por polvo asumiendo que no hay errores observacionales. Esta predicción detallada de las observaciones de *Gaia* proporciona una excelente indicación del volumen y calidad de datos que estarán disponibles a través de *Gaia*.

El modelo de Besançon, a pesar de sus contribuciones, tenía limitaciones, en particular la incapacidad de probar diferentes escenarios de evolución galáctica debido a la fijación de ciertos parámetros y el hecho de que estos modelos no son dinámicamente consistentes y simplifican demasiado la estructura de la Galaxia.

Sharma et al. (2011) presentan GALAXIA un código que permite generar catálogos sintéticos de estrellas, basado en un modelo predefinido de la Galaxia, que puede ser determinado a través de distribuciones de densidad específicas o usando simulaciones de N-cuerpos. A diferencia de los esquemas previos que se centraban en generar estrellas en puntos específicos, este código puede crear una distribución continua de estrellas en cualquier volumen definido. Esta flexibilidad lo hace adaptable a diferentes escenarios y modelos de la Vía Láctea.

El código acepta diversas entradas, como tasas de formación estelar, relaciones de edad-metalicidad, relaciones de edad-dispersión de velocidad y funciones de distribución de densidad. Luego, con la ayuda de las isothermas teóricas, se genera un catálogo de estrellas que puede adaptarse a múltiples bandas fotométricas.

Más recientemente Hunt et al. (2015); Kawata et al. (2014a) han utilizado el enfoque de Sharma et al. (2011) para generar un catálogo estelar sintético a partir de un modelo galáctico N-cuerpos o analítico (Kawata et al., 2014b), para amplias secciones del cielo como las vería *Gaia*, con un esquema de muestreo que genera una muestra de estrellas distribuida de manera uniforme. Estos catálogos sintéticos son generados a partir de modelos dinámicos de la Galaxia a diferencia del modelo de Besançon.

En años recientes Grand et al. (2018) analizan catálogos estelares/sintéticos que incluyen criterios de selección y observables, incluidas las incertidumbres de la segunda liberación de *Gaia*, *Gaia*-DR2, estos catálogos se basan en las simulaciones cosmológicas de alta resolución magneto-hidrodinámicas del proyecto AURIGA (Grand et al., 2017). Un problema típico de las simulaciones es el bajo número de partículas que pueden tener con respecto a los sistemas que quieren representar, formación estelar en las galaxias y su evolución, debido a limitaciones computacionales, exceder la memoria RAM o utilizar muchos procesadores. Los códigos de remuestreo, permiten crear catálogos sintéticos de galaxias. Utilizando la

información estadística del sistema, estos códigos pueden crear nuevos conjuntos de datos que estén mejor representados por el número de partículas pero manteniendo las propiedades estadísticas del sistema original. Tener una mejor representación en el número de partículas ayuda a mejorar la resolución y la precisión de las simulaciones, evitando posibles errores numéricos derivados del ruido “Poissoniano” entre otros efectos.

En este trabajo de tesis nos basamos en las simulaciones cosmológicas de alta resolución de formación y evolución de galaxias (Roca-Fàbrega et al., 2016), descritas en el capítulo 7, para la creación de catálogos sintéticos. Estos catálogos sintéticos servirán para explorar las próximas liberaciones de datos de catastró como DESI y LSST y permitirán comparar las simulaciones con las observaciones reales. Serán útiles para la búsqueda de subestructura en el halo estelar, explorar la posibilidad de colas de marea alrededor de galaxias satélites, etc. A continuación se describe el procedimiento para llevar a cabo el remuestreo hecho con simulaciones.

8.2. Las observaciones sintéticas a partir de las simulaciones

El código de remuestreo calcula la edad, la masa estelar, el flujo y la magnitud en diferentes bandas con el fin de aplicar la función de selección del catastro correspondiente, se incluye en el remuestreo un modelo errores del catastro, con esto se busca crear “Mock Catalogs”, es decir, catálogos sintéticos del catastro, a partir de las simulaciones.

Los catálogos sintéticos que se presentarán en este trabajo, aprovechan la complementariedad que hay entre el catastro Gaia y el catastro DESI, sobre todo de la parte astrométrica de Gaia y a su vez se incluyen los errores en velocidad radial de DESI en el remuestreo, ver secciones 8.4 y 8.6.

Este código se emplea sobre las simulaciones de formación de galaxias tipo Vía Láctea como Garrotxa, tomando las simulaciones y haciendo una proyección en el cielo de la misma incluyendo los errores, lo que permite “observar” las estrellas de la simulación permitiendo aproximarse a lo que observarían Gaia o DESI.

Una parte importante consiste en integrar un generador de nuevas partículas al código de remuestreo, esto se describe mejor en la sección 8.7, sin embargo, cabe mencionar que esta parte aun está en progreso. Esto permitirá tener catálogos sintéticos de simulaciones cosmológicas de alta resolución que estén bien representados con el número de partículas.

8.3. Método empleado en los códigos que remuestrean estrellas

A continuación se describe el método que samplea modelos de simulaciones de N-cuerpos en catálogos sintéticos de observaciones estelares, tomando en cuenta modelo de poblaciones estelares, leyes de extinción y modelo de errores de Gaia y DESI, la ley de extinción o mapas de extinción por polvo son tomados de (Schlegel et al., 1998).

Este código permite generar catálogos sintéticos de un modelo de N-cuerpos o modelo analítico de galaxias, mapeando o remuestreando secciones del cielo con un esquema que genera muestras de estrellas con un suavizamiento sobre la distribución de espacio fase que puede ser un kernel Gaussiano.

Este código divide cada partícula de N-cuerpos en un número apropiado de partículas estelares, permite crear un catálogo sintético de estrellas observables de un modelo de N-cuerpos.

Es necesario escoger una IMF, por ejemplo, la utilizada en este trabajo de tesis es Salpeter (1955). También es necesario escoger un conjunto de isocronas. La IMF se define en cada intervalo de masa. Se hace uso de las isocronas de Padova (Bertelli et al., 1994; Marigo et al., 2008). A cada partícula estelar remuestreada se le asigna una edad y metalicidad. Cada partícula se acomoda a su isocrona más cercana en metalicidad y edad de la malla de isocronas. Una vez seleccionada la isocrona se identifica la masa máxima inicial de la isocrona, se determina una masa límite mínima requerida para que la estrella sea mas brillante que la magnitud aparente en el límite de selección, tomando en cuenta el valor de extinción en la posición de la partícula madre. Se determina cuantas estrellas será remuestreadas de la partícula de N-cuerpos, se integra la IMF sobre el rango de masa deseado. De manera aleatoria se remuestrean masas estelares de la sección de la isocrona, “Ns” veces (Ns: Numero de particulas a remuestrear). Se pesa la selección aleatoria por la IMF. Las isocronas se forman de datos estelares y se interpolan dentro de los valores cercanos de la isocrona a magnitudes M_V y color $V-I_c$, las particulas generadas tienen las mismas posiciones y velocidad de las particulas madre. A continuación se describen los pasos generales de manera resumida, también ver figura 8.1.

- Extinción: Utilizando el mapa de extinción de la Vía Láctea tomado del trabajo de Sharma et al. (2011), se aplica la extinción por polvo a los datos de la simulación de la galaxia.
- Síntesis de poblaciones estelares: Se elige una IMF y un conjunto de isocronas. Se asigna la IMF (Salpeter, 1955) a cada intervalo de masa, y se normaliza para la masa deseada. Se seleccionan las isocronas de Padova (Girardi et al., 2002) y se elige un

límite inferior de masa para la síntesis de población. Se asignan edades y metalicidades a cada partícula estelar de la simulación. Para cada partícula estelar, se selecciona una isocrona cercana en metalicidad y edad. Se determina el número de estrellas a remuestrear a partir de cada partícula estelar, integrando la **IMF** sobre el rango de masas deseado y teniendo en cuenta el límite de magnitud aparente. Se seleccionan masas estelares aleatorias de acuerdo con la **IMF** y se interpolan los valores de magnitud absoluta y color en las isocronas para las masas generadas. Se asume que las estrellas generadas tienen la misma posición y velocidad que sus partículas progenitoras.

- Errores observacionales: Se agregan errores observacionales basados en las estimaciones de rendimiento científico de *Gaia*. Se utiliza la estimación de errores posterior al lanzamiento proporcionada por Mercè Romero-Gómez vía comunicación privada. Se asume que la posición y velocidad Solares son conocidas y se ubica al observador en una posición específica. Se establece un límite de magnitud para las estrellas más brillantes, que se incluirán en el catálogo simulado. Se agregan errores a la paralaje, el movimiento propio y las velocidades de línea de visión utilizando los coeficientes proporcionados en el artículo de *Gaia*.

- Errores de velocidad radial de **DESI**: A cada partícula remuestreada se le asigna un valor en error de velocidad radial de acuerdo a su posición en un arreglo de magnitud “r” y color “g-r” del modelo de errores original, proporcionado por S. Koposov vía comunicación privada.

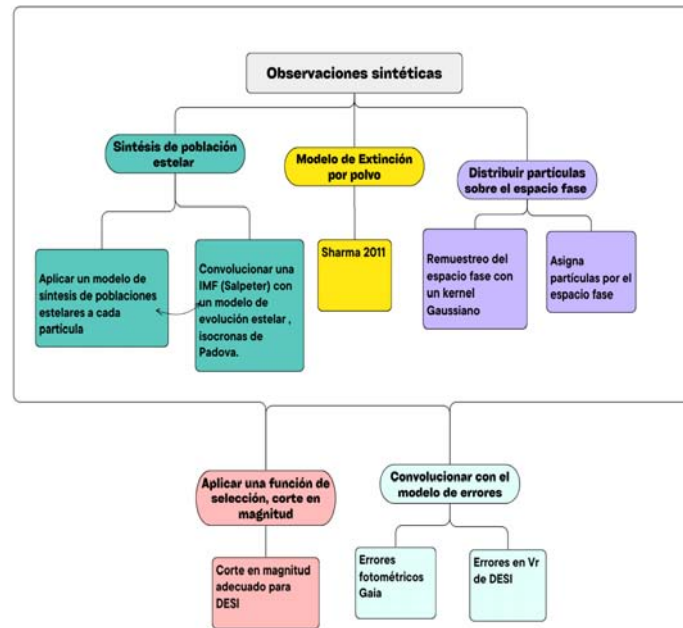


Figura 8.1: Diagrama de flujo que explica de manera resumida el esquema para crear observaciones sintéticas a partir de las simulaciones

8.4. Modelo de errores de Gaia

Los errores de *Gaia* en las observaciones sintéticas, se agregan de acuerdo a GAIA-C9-TN-UB-RMC-001-1 como se describe a continuación. Los errores se agregan dependiendo de los catálogos/versiones con los que se comparará el catálogo simulado. Proporcionando columnas adicionales para los parámetros astrométricos y fotométricos observados que resultan al considerar tres casos diferentes dependiendo de si se quiere comparar datos modelo con: datos reales provenientes de Tycho-2, del experimento TGAS y de los datos de *Gaia*.

Para simular los datos astrométricos de Tycho-2, se asume que los errores en los parámetros (posición y movimientos propios) dependen de la magnitud aparente V_T de la estrella. Los errores se asignan siguiendo la tabla 2 de Høg et al. (2000). Los errores se asignan al azar asumiendo una distribución gaussiana con una desviación estándar igual a las incertidumbres en astrometría indicadas en la mencionada tabla. Se ha asumido la misma desviación estándar para el movimiento propio tanto en ascensión recta como en declinación. Para simular los datos astrométricos de TGAS, se utilizan los valores proporcionados por Michalik et al. (2015) (tabla 1). En este último caso se pueden asignar errores a la paralaje, por lo que las paralajes observadas se proporcionan en el catálogo simulado de TGAS. Para la fotometría de Tycho-2, se utiliza la estrategia descrita en Czekaj et al. (2014). Se asume que no se proporciona nueva fotometría con el experimento TGAS.

Para simular los datos astrométricos y fotométricos de *Gaia*, se sigue la página “web” de rendimiento científico de *Gaia*. El código para agregar errores de *Gaia* es una actualización del código de M. Romero-Gomez y JM. Carrasco. Los códigos proporcionados por M. Romero-Gomez (vía comunicación privada) calculan los errores de *Gaia* estimados después de la puesta en marcha para los datos de fin de misión (se asume una duración total de 5 años de la misión). Para simular los datos de la misión *Gaia* de 6 meses, se agregan los factores correspondientes:

- Para la paralaje el factor es $\sqrt{\frac{5}{L}}$, L es la longitud de la misión en años.

- Para los movimientos propios: $\sigma_{\mu}^{(L)} = \frac{5}{L} g_{ratio} \sigma_{\pi}^{(L)} = \left(\frac{5}{L}\right)^{\frac{3}{2}} g_{ratio} \sigma_{\pi}^{(5)}$

8.5. Mapeo de errores en paralaje en la simulación remuestreada: $|z| < 300pc$ y $|z| > 300pc$

A continuación se muestran pruebas del funcionamiento del código de remuestreo a través de un mapeo del error relativo en paralaje (modelo de errores de *Gaia*). Estas pruebas consisten en reproducir el modelo de errores de *Gaia* a partir de la simulación remuestreada a través del diagrama en el espacio de configuración pesado por el error relativo de paralaje. El mapeo de las estrellas de la simulación en el rango de $|z| < 300pc$ y $|z| > 300pc$ debe ser consistente con la figura 4 del trabajo de Romero-Gómez et al. (2015).

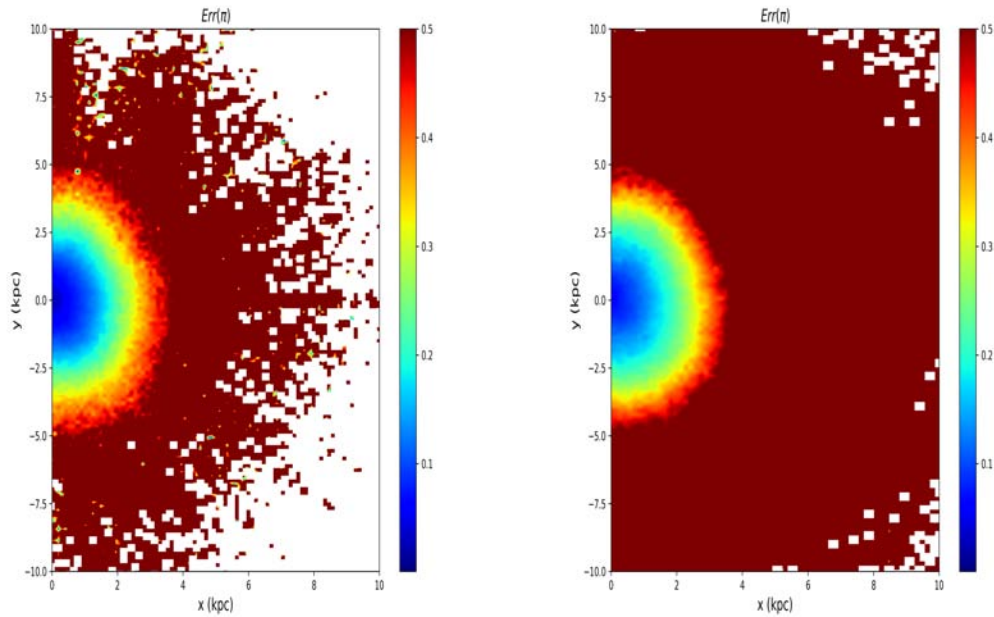


Figura 8.2: El error relativo en la paralaje para las estrellas de la simulación remuestreada (izquierda) para estrellas más cercanas al plano galáctico, $|z| < 300$ pc y más allá de $|z| > 300$ pc (derecha). El sol se encuentra en la posición central debido a que durante el remuestreo la simulación esta centrada en la posición solar a diferencia de los mapas mostrados en los paneles superiores de la figura 8.3 Romero-Gómez et al. (2015).

Podemos ver en ambos paneles de la figura 8.2 que la zona mejor representada por los errores en paralaje es la zona central del disco, vemos que hacia la parte externa hay un aumento gradual del error en paralaje.

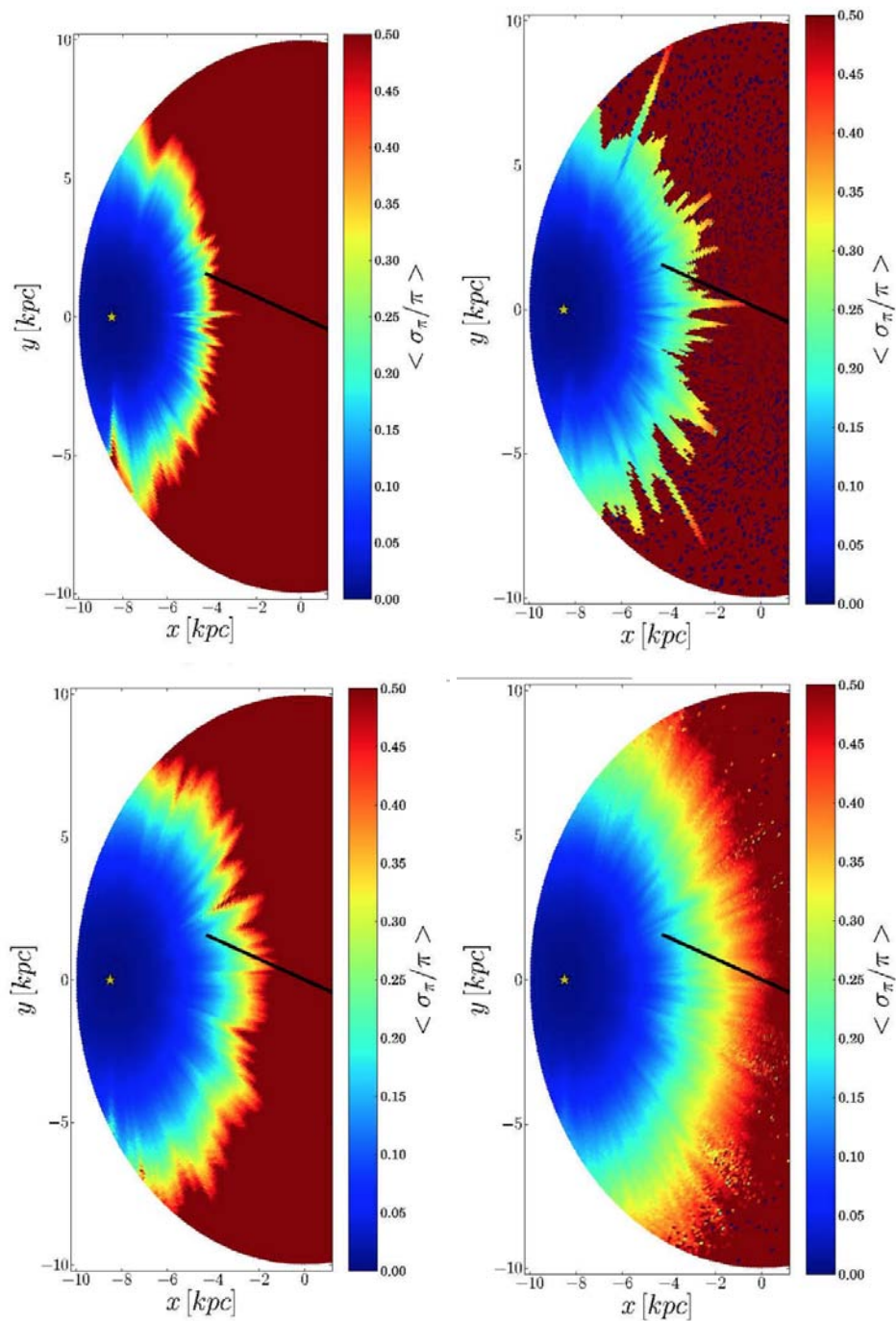


Figura 8.3: Imagen tomada directamente de Romero-Gómez et al. (2015) para efectos de comparación del modelo de errores de Gaia. El error relativo promedio en paralaje para las muestras RC-G20 (izquierda) y RC-RVS (derecha) para estrellas más cercanas al plano galáctico, $|z| < 300 pc$ (arriba) y más allá de $|z| > 300 pc$ (abajo). La estrella amarilla en todos los paneles muestra la posición del Sol, mientras que la línea sólida negra marca la posición y longitud de la barra galáctica.

8.6. Modelo de errores de DESI

A continuación se muestran pruebas del funcionamiento del código que permite crear las observaciones sintéticas. Estas pruebas consisten en reproducir el modelo de errores de velocidad radial, a partir de la simulación ya remuestreada, esto se muestra a través del diagrama color-magnitud construido con la magnitud “r” de la banda Johnson y el color “g-r”. El mapa de densidad numérica en el diagrama color magnitud muestra las zonas en las que se espera encontrar más puntos, ver figura 8.5, lo que explica las zonas con menor densidad de puntos en la figura 8.4 del modelo de errores.

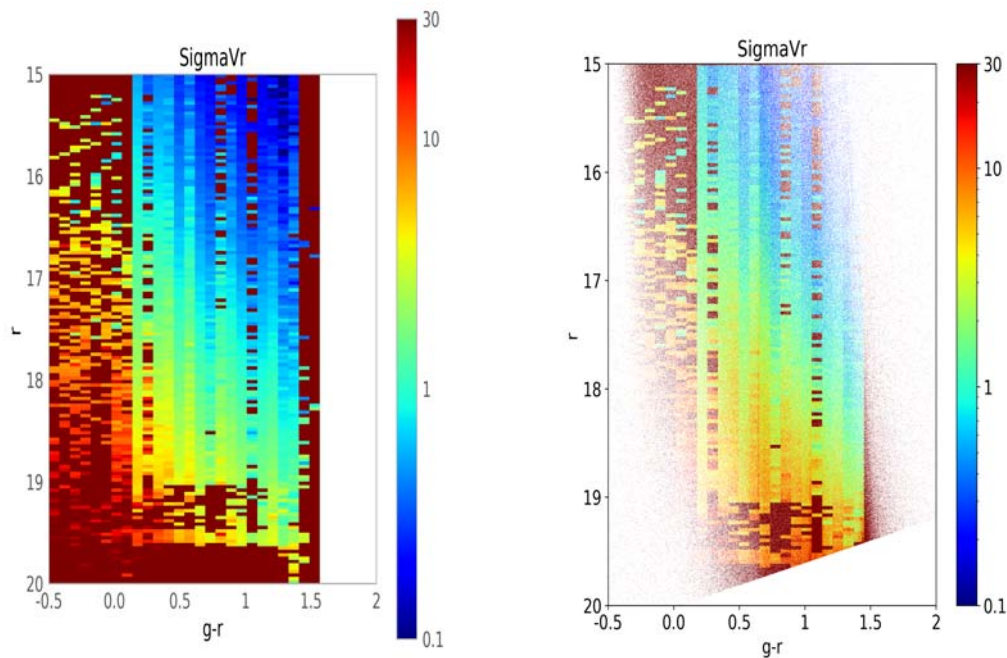


Figura 8.4: Izq: Diagrama color magnitud pesado por error en velocidad radial de acuerdo al modelo de errores de DESI. Der: Mismo diagrama construido directamente de las estrellas remuestreadas de la simulación. Las partes más tenues son consistentes con la densidad de estrellas en cada región del mapa que siguen la función de selección, como se muestra en la figura 8.5

De acuerdo con lo que vemos en la figura 8.4 en el modelo de errores original, esperaríamos que las estrellas más azules con magnitudes de “r” entre 18 y 20 estén más afectadas por los errores en velocidad radial (inferior izquierda), mientras que las estrellas rojas a “r” cercanas a 15 (rojas y brillantes) tienen menores errores en velocidad radial (superior derecha).

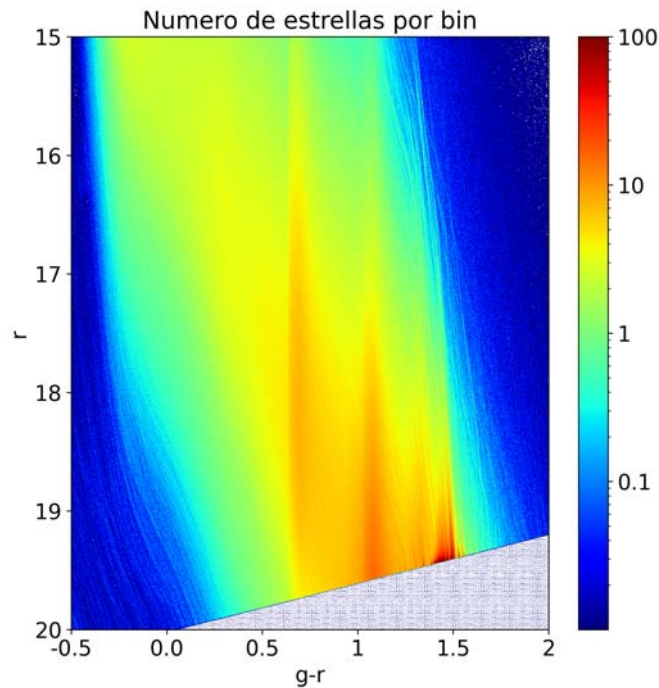


Figura 8.5: Mapa de densidad numérica de estrellas en el diagrama color magnitud construido a partir de las estrellas remuestreadas de la simulación.

La figura 8.5 muestra el diagrama color-magnitud como un mapa de densidad de estrellas por píxel. Como podemos ver en la zona inferior izquierda y en la zona superior derecha la densidad de estrellas es baja y hay una zona entre $g-r=0.5$ y $g-r=2$ donde prácticamente no hay información, por lo que en el modelo de errores extraído de las observaciones sintéticas esperaríamos tener pocos datos. Como podemos ver la zona media del diagrama color magnitud es la que presenta mayor densidad de datos.

8.7. Generador de nuevas partículas con distribución Gaussiana

El suavizamiento durante la creación de catálogos sintéticos con funciones kernel, como el kernel gaussiano se utiliza en el procesamiento de los datos de las simulaciones, especialmente en la creación de catálogos sintéticos basados en simulaciones cosmológicas. Esta técnica se utiliza para suavizar o difuminar datos discretos, lo que puede ayudar a la reducción de ruido en y por lo tanto tener una mejor representación de datos observacionales a través de las simulaciones, mejorando la resolución del espacio fase que podría verse afectado por el ruido.

El kernel gaussiano es una de las funciones kernel más utilizadas en astrofísica y cosmología debido a sus propiedades suaves y bien definidas. Se aplica un kernel gaussiano a cada punto de datos en el catálogo sintético para promediar su valor con los valores de puntos cercanos, por ejemplo Lowing et al. (2014) utilizan este enfoque

Un código que lee en un archivo externo posiciones x,y,z velocidades v_x,v_y,v_z , divide el sistema en una malla de celdas cubicas con las coordenadas espaciales. Guarda las celdas en arreglo matriciales y calcula el valor promedio de diferentes propiedades del sistema en cada celda, para obtener la dispersion en posiciones y velocidades del sistema original. El esquema que utiliza para calcular los promedios en cada celda es el esquema de interpolacion “Cloud in cell” (Vasileska et al., 2011), el cual pesa los promedios en cada celda con la distancia de cada particula al centro de la celda y viene normalizado por el volumen de la celda. Esta informacion estadistica es pasada posteriormente a un generador de numeros gaussianos donde la anchura de la gaussiana es la dispersion obtenida anteriormente en cada celda. Esto es el remuestreo que permite generar un sistema con mas particulas. El nuevo sistema generado debe mantener las propiedades del sistema original como se muestra en la figura 8.6 y 8.7

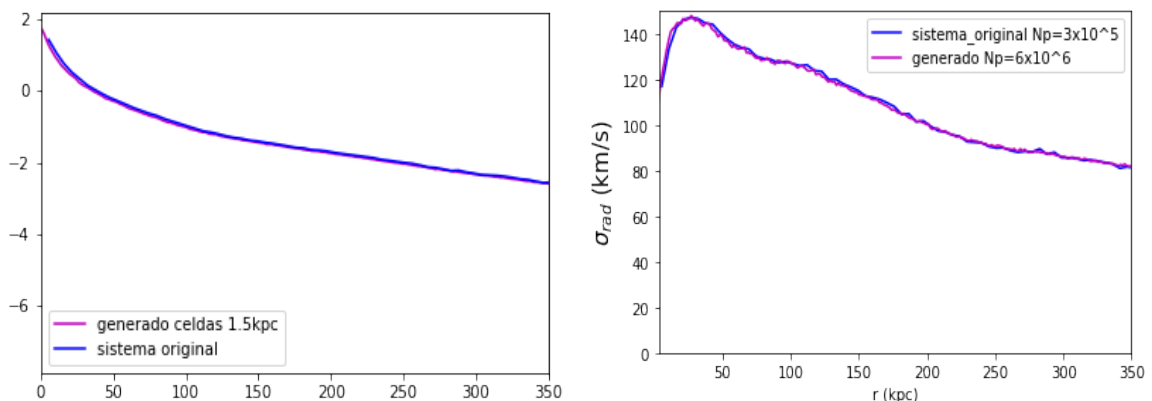


Figura 8.6: Comparación de los perfiles de densidad (en escala logarítmica) y también de la dispersión de velocidad total (radial+tangencial) tanto del sistema original como del sistema generado a través del remuestreo. Izquierda: Perfil de densidad. Derecha: Dispersión de velocidad total (radial+tangencial)

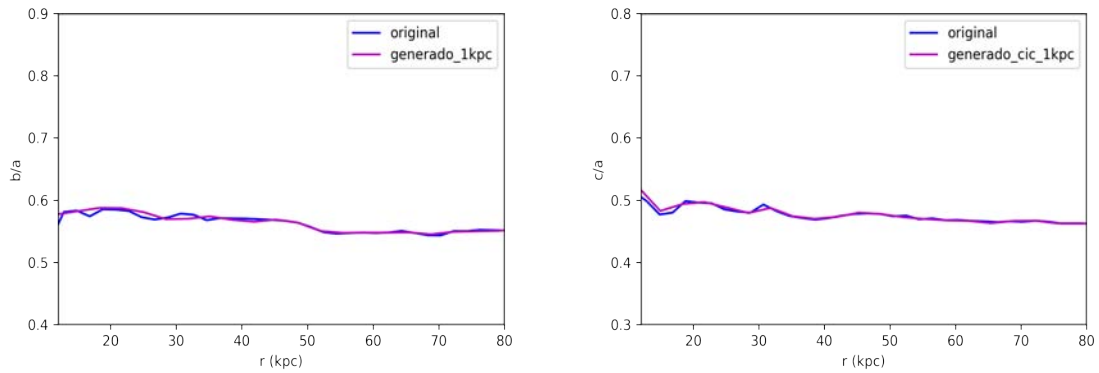


Figura 8.7: Razones de ejes: semi-eje mayor c/a y semi-eje menor b/a entre el sistema original y el nuevo sistema generado, que muestran la triaxialidad del sistema.

8.8. Discusión y conclusiones

En este capítulo mostramos el avance que hemos hecho en la creación de las observaciones sintéticas, usando como base las simulaciones cosmológicas de formación y evolución de galaxias de GARROTXA, específicamente los modelos tipo Vía Láctea, las partículas del halo y estelares. Las pruebas que hicimos para verificar que los modelos de errores tanto de *Gaia* como *DESI* fueron correctamente implementados consistieron en poder reproducir ambos modelos de errores una vez creadas las observaciones sintéticas. El panel derecho de la figura 8.5 consiste en un diagrama color-magnitud con cada píxel pesado por el valor en el error en velocidad radial, el modelo de errores original proporcionado por S. Koposov (por comunicación privada) del grupo de la colaboración que estudia la Vía Láctea de *DESI*, es el panel izquierdo de la figura 8.4, como podemos ver son comparables, la baja densidad de puntos en algunas zonas de mapa se debe a la función de selección de *Gaia*, en ese sentido faltaría incluir la función de selección adecuada para el “survey” *DESI*. También reproducimos el modelo de los errores relativos de paralaje para *Gaia* con las estrellas del disco 8.2 comparando con el trabajo de Romero-Gómez et al. (2015), hay que tomar en cuenta que las diferencias se deben a que las observaciones sintéticas no corresponden a la galaxia real y esta comparación la estamos haciendo entre simulaciones con observaciones, sin embargo, lo que se quiere mostrar es la consistencia entre las observaciones sintéticas y los datos reales. (ver figura 8.3).

Capítulo 9

Discusión y Conclusiones

9.1. Resumen

En la presente tesis, hemos explorado un mayor realismo sobre los procesos dinámicos que definen la historia de evolución dinámica de los satélites cayendo simultáneamente a una galaxia como podría ser el caso de la Vía Láctea, empleando metodologías que van desde modelos analíticos hasta simulaciones de N-cuerpos, tanto aisladas como en un contexto cosmológico. También hemos analizado la estrategia de aplicar algoritmos no supervisados de **ML**, como **DBSCAN** para detectar estructura en el halo estelar, la interpretación de tales estructuras está relacionada con los procesos dinámicos discutidos y con la historia evolutiva de la Vía Láctea. Presentamos además una nueva herramienta que en base a simulaciones cosmológicas de formación de galaxias puede generar catálogos de estrellas con cierto grado de realismo, y con estadística de estrellas variable, pudiendo alcanzar números comparables con las observaciones, incluyendo los modelos de errores de los censos Gaia y **DESI**. Esta herramienta es de gran valor tanto para probar estrategias de estudio de la Vía Láctea como para cuantificar la capacidad de los censos Gaia y **DESI** y otros en desarrollo como **LSST** en pruebas combinadas.

Inicialmente, utilizamos simulaciones de N-cuerpos aislados, obtenidos mediante códigos de N-cuerpos de sistemas no-colisionales, soportadas por modelos semianalíticos de diversas configuraciones de satélites cayendo simultáneamente y las consecuencias que esto tiene en los efectos colectivos y por lo tanto en la historia de caída de satélites. También exploramos la importancia de nuestros resultados en el caso más realista de los satélites en la Vasta Estructura Polar **VPOS**. Realizamos también pruebas de convergencia (Errores en energía) para cada experimento, para asegurar que los resultados de estos modelos no estaban afectados por efectos numéricos.

Basados en los modelos cosmológicos de GARROTXA de Roca-Fàbrega et al. (2016), desarrollamos un código de remuestreo de estrellas de esta simulación. El código de remues-

treo tiene tanto los errores de Gaia fotométricos, como los errores de **DESI** para las velocidades radiales. Generamos una serie de experimentos para comprobar el funcionamiento de este generador que tendrá como resultado la generación de catálogos sintéticos.

Estos últimos modelos son mucho más realistas que los que utilizamos anteriormente, por lo tanto su análisis es mucho más complejo y requiere más tiempo. Hemos comenzado el primer análisis de estos nuevos modelos, obteniendo resultados del modelo de errores de **DESI** y de los errores en paralaje de Gaia. En un futuro estos catálogos servirán para la búsqueda y análisis de subestructura en catastrófos como Gaia, **DESI** y **LSST**.

Primera parte: Efectos Colectivos

En la primera parte de este trabajo de tesis nos hemos propuesto catalogar y cuantificar los cambios en la historia orbital de una galaxia satélite durante su historia de caída, ante los efectos colectivos, por la presencia de otros perturbadores o galaxias satélites. Se analizó el impacto que diferentes parámetros físicos y configuraciones entre satélites tienen sobre los procesos de caída, tales como: la estructura interna de los satélites, su distribución espacial dentro de un halo anfitrión, la pérdida de masa y el efecto combinado de la autofricción. En particular la auto-fricción dinámica no ha sido estudiada antes para el caso de varios satélites cayendo simultáneamente.

A continuación, se enumeran los principales hallazgos derivados de la primera parte de este trabajo:

- Se observaron claras diferencias entre la caída concurrente de galaxias y su correspondiente caso aislado. El proceso de inmersión de los satélites, al incluir efectos colectivos de varios perturbadores, varía con respecto al de un satélite aislado en todas las situaciones que estudiamos.
- En el caso en el que los satélites son coplanares la velocidad de caída puede llegar a ser más rápida, en comparación con el caso aislado, debido al efecto combinado del material perdido del satélite, lo que proponemos que es consecuencia de la autofricción dinámica.
- Encontramos claras diferencias entre la caída concurrente de galaxias y su respectivo caso aislado (que ha sido la manera tradicional de abordar la historia de caída, al considerar un solo perturbador), estas diferencias pueden explicarse al menos parcialmente por qué el potencial global galáctico (Galaxia + satélites) es dependiente del tiempo. Aunque los encuentros cercanos entre satélites pueden suceder, no representan el mecanismo principal o dominante (ver figura 6 en el artículo). Cabe resaltar que la historia orbital de los satélites cayendo simultáneamente nunca es igual que su respectivo caso

aislado a pesar de la presencia o ausencia de encuentros cercanos y que el corrimiento orbital incrementa con el tiempo. Efectos de “boomerang” pueden suceder, pero esto queda combinado con la interacción de los satélites con la respuesta en densidad global y local, así como con las interacciones con el material desprendido de los satélites, así como la autofricción dinámica.

- El proceso de hundimiento de los satélites, al incluir efectos colectivos de varios perturbadores, en general difiere del de un satélite aislado. La diferencia depende de las propiedades internas de los satélites hundiéndose, la configuración espacial y de velocidad relativa entre ellos, los parámetros orbitales y la relación de masa entre el anfitrión y los satélites o la dispersión entre las diferentes masas de los satélites.
- En particular, en el caso de múltiples satélites compactos o rígidos, el proceso de hundimiento es más lento que para los satélites aislados. En este escenario, la mayor parte del efecto de las interacciones (respuestas de densidad de los satélites y del anfitrión) va a la energía orbital del satélite, retrasando el proceso de hundimiento (ver paneles izquierdos de la figura 4.5). La situación anterior puede ser comparable con los cúmulos globulares compactos en galaxias esferoidales enanas como Fornax o Reticulus 2. Sin embargo, la evaluación de tal caso requiere comprobar si las suposiciones de satélites rígidos y compactos son adecuadas para los cúmulos globulares y también el efecto de la orientación orbital relativa del cúmulo globular (Inoue, 2011, ver también).
- Para satélites extendidos en acreción concurrente, la pérdida de masa es considerable. La velocidad de hundimiento es más rápida en algunas situaciones, en comparación con el caso aislado debido al efecto combinado del material arrancado, revelando un nuevo aspecto de la fricción dinámica. Tal efecto colectivo hace que la amplitud de la fricción dinámica a veces sea mayor que el 10 – 15 % citado por Miller et al. (2020) (ver la pérdida de masa en el panel inferior derecho en la figura 4.5).
- Concluimos que el perfil de densidad interno de los satélites los hace más/menos susceptibles al arrancamiento, afectando la velocidad de hundimiento del satélite en combinación con el efecto combinado ya mencionado de la masa arrancada de múltiples satélites, mejorando la función de distribución (PDF).
- Descubrimos que los efectos colectivos son más dramáticos en sistemas coplanares como la llamada “Vasta Estructura Polar”, VPOS (ver figura 3 del artículo), esto se debe a que las perturbaciones mutuas son más frecuentes y cercanas y más fuertes, haciendo que la historia del satélite hundiéndose sea bastante diferente al caso aislado. Aunque no pretendimos simular la formación y evolución de la VPOS. Podemos

señalar que la velocidad de hundimiento puede divergir de los resultados de una sola caída de satélite, puede reducir temporalmente o revertir el proceso de hundimiento.

- Construimos un conjunto de veinte modelos para el sistema de satélites similar al **VPOS** para cuantificar el cambio “rms” con respecto al caso aislado para los parámetros orbitales de la caída (apocentros y excentricidades) de cada satélite. Obteniendo que para los apocentros los cambios puede llegar a ser del 30 % y hasta 50 % en las excentricidades.
- Si uno quiere ir a alta precision hay que mejorar el realismo y los efectos colectivos son parte importante de eso, ver por ejemplo (D’Souza and Bell, 2022) y (Pagnini et al., 2023) donde hay efectos en la caída de cúmulos globulares. El nuestro es uno de los primeros que señala esto, pero hay otros y continúa el grado de realismo. Esto es importante para alta precision. Esto sera clave a la luz de los datos de una nueva generacion de censos **LSST**, **DESI**.
- La promesa de la nueva generación de censos estelares de la Vía Láctea abre la posibilidad de la detección de una gran riqueza en subestructura fósil impresa en el halo estelar. Sin embargo para planear e interpretar cuantitativamente las estrategias de búsqueda, son indispensables catálogos de observaciones sintéticas, incluyendo los modelos de errores y la función de selección de tales censos. En el capítulo 8 mostramos los resultados preeliminares de construir tales catálogos sintéticos utilizando simulaciones cosmológicas de formación de galaxias asemejadas a la Vía Láctea.
- Las nuevas metodologías permiten extraer el maximo contenido de informacion, no sin el riesgo de no entender las incertidumbres de estos metodos, llámese **DBSCAN**, **SOMs**, **K-Means**,
- El proceso tradicional de fricción dinamica, de manera gruesa en primera aproximacion está correcto como lo ha capturado Chandrasekhar, pero para la alta precision hay una mayor riqueza, cuando hay varios perturbadores la auto-friccion puede cambiar la historia de caída, los fenómenos a gran escala son más importantes que los encuentros binarios y es algo que la comunidad sigue discutiendo.
- Si bien, no incluimos las otras componentes como el disco o el bulbo, se puede pensar que el disco puede ser importante, en este trabajo vemos los efectos incluso en las partes mas externas de la galaxia, ocurriendo independientemente del disco, por supuesto si uno quiere ajustar la evolución de un satélite en particular, si sería adecuado introducir el disco, sin embargo, no hay pérdida de generalidad, ya que el mecanismo físico general es un potencial gravitacional variable en el tiempo y eso va a ocurrir con o sin disco. La parte del disco será importante hacia la parte central, sobre todo si se quieren

establecer conclusiones sobre un satélite en específico, pero lo que se busca mostrar es el efecto que tiene la caída concurrente, considerar el disco cambiará los detalles del cambio, aunque no la amplitud, ya que el disco cambia a escalas de tiempo más largas.

Segunda parte: Estrategias para la búsqueda de subestructura

La segunda parte de este trabajo consistió en elaborar una estrategia de búsqueda automatizada en datos realistas de las simulaciones. Para esta tarea nos basamos en la eficiencia que ha demostrado el algoritmo **DBSCAN** antes (Castro-Ginard et al., 2018), aunque para búsqueda de cúmulos abiertos. Decidimos aplicar el algoritmo **DBSCAN** a datos de las simulaciones en las que tenemos control de las etiquetas de las partículas del satélite. Nos propusimos hacer la búsqueda del satélite y su estructura con diferentes combinaciones de variables entre posiciones, velocidades y densidad.

A continuación se enlistan los principales resultados de esta parte:

- El combinar la información espacial con la información cinemática completa da lugar a un conjunto de datos en los que se puede discernir la estructura como galaxias satélites y sus componentes con grado de confiabilidad aceptable, entre el 60 % y el 70 %.
- El hecho de buscar la subestructura en una zona conoidal a semejanza de una observación real hace más robusta la búsqueda en el sentido de que permite tener una muestra más equilibrada con respecto a la señal que puede haber de la presencia de estructura con respecto al ruido de fondo de las partículas del halo.
- Las pruebas que se realizaron utilizando la velocidad radial, tangencial y la densidad dieron mejores resultados en la detectabilidad con respecto al caso en el que no se dispone de la componente radial. Esto sugiere que posiblemente estructuras como las colas de marea alrededor de galaxias dSpH no han sido detectadas debido a que los estudios en este sentido carecen de dicha variable. Esto será diferente con el advenimiento de nuevas bases de datos como Gaia y **DESI** que tendrán velocidades radiales.
- El análisis con velocidad radial y tangencial puede permitir discernir las diferentes partes de la estructura que podemos detectar y clasificar, así como aquellas que no podríamos detectar, como demostraron los resultados de calibración con histogramas de número de vecinos, aplicados solo a las partículas del satélite.

La nueva generación de censos estelares de la Vía Láctea tales como Gaia (Gaia Collaboration, 2016), **DESI** (Levi and Allen, 2019), **LSST** (Ivezić and Kahn, 2019), prometen una

mejora dramática en la estadística de sus datos así como la integración de muestras espectroscópicas, edad o química o con movimientos propios con completez sin precedente. Esto permitirá buscar por restos fósiles del ensamblaje de la Galaxia y sus satélites en la tenue distribución de poblaciones de estrellas en el halo galactico.

Conclusión General

Es imperativo mejorar el realismo en la forma en la que se modela la formación y evolución de las galaxias. Para alcanzar una alta precisión, considerar los efectos colectivos de la presencia de varias galaxias satélites es una parte crucial. Concluimos de manera general que es importante tomar en cuenta los efectos colectivos en los modelos que restringen la historia orbital de las galaxias, ya que tienen consecuencias en la historia de evolución de las galaxias, en la medición de los tiempos de caída de las galaxias, el perfil de densidad e incluso en la forma del halo anfitrión, el no tomar estos efectos en cuenta puede conducir a interpretaciones erróneas.

La nueva generación de censos estelares de la Vía Láctea como Gaia actualmente, **DESI** y **LSST** proporcionarán observaciones exquisitas de prácticamente todo el halo estelar. El aprovechamiento completo de tales cuerpos de observaciones requiere el auxilio de modelos razonablemente realistas en términos de la distribución y la estadística de trazadores, incluyendo incertidumbres realistas. Lo anterior permitirá la detección de una rica subestructura fósil en el halo estelar. Lo anteriormente establecido enfatiza el valor de una herramienta como *Balrogs*, que hace posible el tener catálogos sintéticos realistas que puedan guiar la interpretación de las observaciones. Un derivado de *Balrogs* es la posibilidad de caracterizar y remuestrear el espacio fase, lo cual permite aumentar la resolución de simulaciones de N-cuerpos sin introducir o borrar información dinámica. Todas las estrategias discutidas en la tesis contribuirán a nuevos estudios de la historia de la Vía Láctea, como los discutidos en la tesis.

Capítulo 10

Trabajo a futuro

En esta tesis hemos cuantificado y establecido cómo los efectos colectivos tienen consecuencias importantes para la reconstrucción de la historia orbital de las galaxias satélites en la Vía Láctea. El proceso de destrucción de galaxias satélites deja como resultado subestructura (colas de marea, cúmulos globulares, “streams”) que estas galaxias siembran en el halo estelar en su historia de caída y a su vez los satélites también se verán afectados. Por lo que un estudio cada vez más preciso, cubriendo diferentes casos y lo más generales posibles se convertirá en una necesidad.

Esta búsqueda de subestructura también puede ayudar a restringir varias propiedades del halo oscuro de la galaxia y sus satélites: La forma del halo de la Vía Láctea y su historia dinámica junto con la de sus satélites, en particular el origen de las galaxias más pequeñas conocidas como dSph (por sus siglas en inglés), incluyendo las galaxias del tipo “ultra-faint” difíciles de detectar con las herramientas actuales.

Esta búsqueda en si no es una tarea fácil de realizar directamente sobre los catálogos o catastros astronómicos, incluso aquellos más actuales como **DESI** o **LSST**. El caso de la detección de colas de marea en las galaxias satélites por ejemplo, es una tarea que ha representando un reto debido al bajo brillo superficial de estos objetos. En ese sentido haremos uso de simulaciones cosmológicas de alta resolución que incluyen información de los catastros como un modelo de los errores, es decir, se busca simular las condiciones con las que observan estos catastros, pero con la ventaja del control sobre las condiciones de las galaxias que ofrecen las simulaciones.

Estado actual: En cuanto al contexto de los efectos colectivos y sus consecuencias para la historia de caída, a la luz de nuestros resultados estos efectos deben de ser tomados en cuenta en los trabajos que buscan ofrecer restricciones a la historia orbital de las galaxias satélites en la Vía Láctea, si bien existen trabajos que reconocen la complejidad de la evolución orbital de las galaxias en este contexto, el caso de considerar los efectos colectivos no se ha abordado de manera sistemática. Es muy probable que estos trabajos no puedan dar una predicción

exacta de la órbita de los satélites debido a la complejidad de los efectos colectivos, sino que tendrán que ofrecer más bien una zona de probabilidad de las condiciones iniciales, es decir, una función de distribución de probabilidad (PDF). También hemos realizado pruebas de los efectos colectivos con múltiples satélites con los halos triaxiales, estudiando el fenómeno de la circularización en las órbitas, pero este es un trabajo en proceso que planeamos continuar.

En cuanto a la búsqueda de subestructura en el halo estelar, las herramientas de inteligencia artificial han tenido una explosividad en la última década, algoritmos de “Aprendizaje Automático” o **ML**, han demostrado su potencial para problemas de búsqueda de grupos en datos de manera autónoma. Se han empezado a implementar en tareas de búsqueda de subestructura en las galaxias, si, pero esto dista de ser una tarea completa. Los algoritmos no supervisados como los que se utilizan en este trabajo requieren un trabajo previo de calibración y obtener catálogos sintéticos basados en “surveys” astronómicos que tengan la resolución adecuada para este tipo de problemas es una tarea compleja en si misma.

En ese sentido este trabajo ofrece una línea de trabajo que permitirá explorar la capacidad de búsqueda de grupos o subestructura de herramientas de **ML**, *no supervisados* principalmente, pero también *supervisados*, en catálogos astronómicos sintéticos realistas. También ofrece una puerta hacia nuevas restricciones sobre el potencial del halo de la Galaxia. Robustez en el análisis de propiedades de la Galaxia como es la forma de su halo, el cual se ha sugerido que es triaxial, este trabajo puede reforzar ese panorama.

Propuestas y recomendaciones: Agregar una función de suavizamiento sobre el espacio fase al código de remuestreo, llamado *kernel* gaussiano que permitan obtener una mayor resolución y tener una comparación con las observaciones reales, lo cual ya hemos empezado a realizar, ver 8.7. Esta tarea requiere hacer pruebas, el introducir un *kernel* de suavizamiento gaussiano puede introducir información espuria en las simulaciones por lo que se requieren hacer pruebas para constatar que los halos generados con este *kernel* son consistentes con el halo original. Si el suavizado gaussiano es el adecuado, las características generales de estadísticas como la dispersión en los datos del espacio fase y la forma del perfil deben mantenerse y no debe haber cambios drásticos o incoherencias significativas. Lo que se busca es conciliar las simulaciones con un número adecuado de partículas que permitan capturar los procesos físicos en las galaxias. El peligro de los remuestreos es añadir información que no es real.

Si bien, tenemos una técnica de calibración para los algoritmos de búsqueda de grupos como **DBSCAN**, esta debe llevarse a cabo sobre los halos que se van a analizar, debe calibrarse el algoritmo con estos halos generados con el *kernel* gaussiano. Esta calibración dependerá de la información de los datos, por lo que se requieran pruebas para asegurarse que el algoritmo esté bien calibrado, es decir, tener una certidumbre de si se encontrará o no dicha subestructura.

Posibles colaboraciones: Directamente con el grupo de “Milky Way Survey” de **DESI**

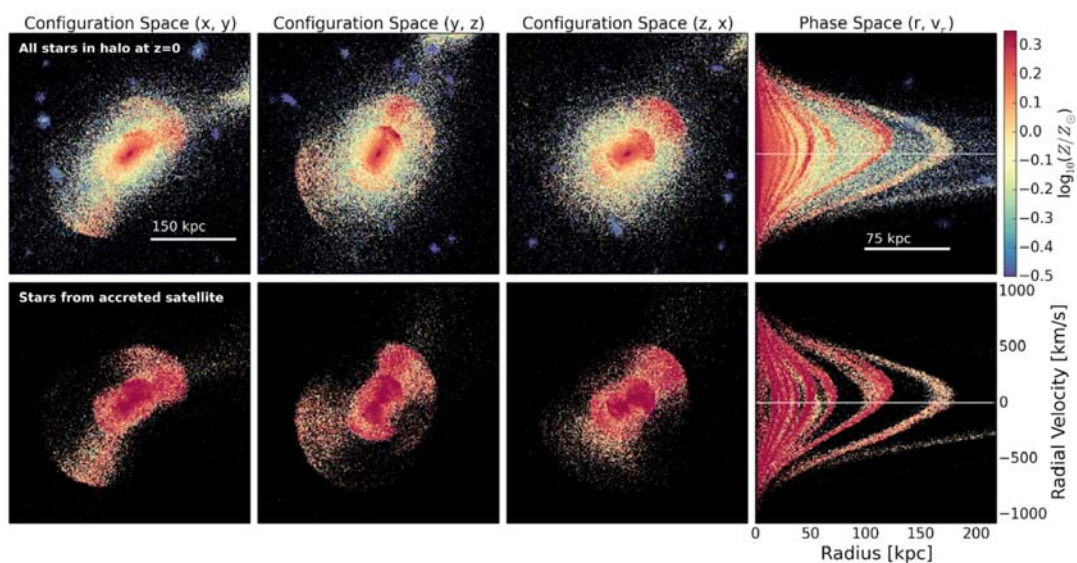


Figura 10.1: Ejemplo de como se ve la subestructura en el espacio fase a través de las simulaciones.

complementando el trabajo que ellos realizan con estas herramientas de búsqueda de subestructura.

10.1. Análisis de proyecciones del espacio fase con técnicas de análisis de imágenes para detectar subestructura en catálogos sintético

Como trabajo a futuro se propone explorar otras herramientas de búsqueda de grupos como algoritmos que trabajan sobre imágenes, uno de ellos es Min-Hash, pero también proponemos utilizar *Mapas Auto Organizados*, SOMs, estos mapas tienen la capacidad de organizar y visualizar datos multidimensionales de manera estructurada y comprensible, pero requieren un entrenamiento (enfoque supervisado). La propuesta es utilizar estos esquemas directamente sobre las imágenes del espacio fase de las galaxias, extraídos de los catálogos sintéticos, ver por ejemplo 10.1.

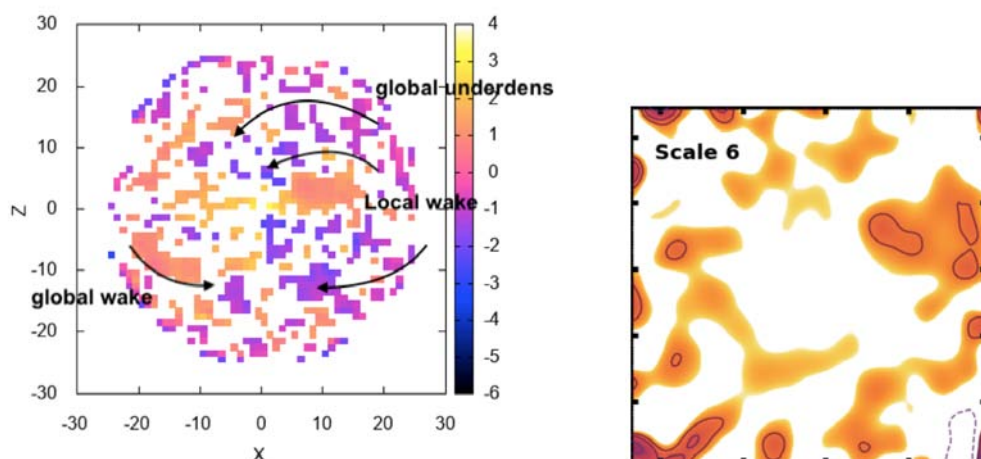


Figura 10.2: Para una de las simulaciones aplicamos el esquema “Cloud-in-cell” (Birdsall and Fuss, 1997) sobre una de las simulaciones con un perturbador o satélite vivo para detectar las diferentes partes de la respuesta que ocasiona. Izquierda: En la imagen se aprecian la respuesta tanto local y global. Derecha: Aplicación del esquema “Wavelets” implementado por P. Ramos para el mismo experimento.

10.2. Exploración de la base ortonormal de “Wavelets” como estrategia de búsqueda de estructura.

En colaboración con Pau Ramos y Teresa Antoja (U. de Barcelona), comenzamos a explorar la utilización de los “wavelets” para buscar a diferentes escalas la presencia de estructura en el espacio de configuraciones de simulaciones de N-cuerpos. Ver figura 10.2. Continuaremos en la medida de lo posible con este proyecto.

Apéndice A

Obteniendo la fórmula de Chandrasekhar

A continuación, se reproduce el desarrollo de [Arcavi \(2011\)](#) y [Binney and Tremaine \(1987\)](#) para obtener la fórmula de Chandrasekhar para la fricción dinámica.

Consideremos el caso de un cuerpo de masa M que se mueve a través de una población de estrellas, cada una de masa m .

Para facilitar el problema, consideremos primero el encuentro de la partícula de masa M con una partícula de masa m perteneciente al medio de partículas.

Sea $\mathbf{X}_M, \mathbf{V}_M$ y $\mathbf{X}_m, \mathbf{V}_m$ las posiciones y velocidades de M y m respectivamente. Además, sea $\mathbf{r} = \mathbf{X}_m - \mathbf{X}_M$ el vector de separación relativo.

La ecuación de movimiento para una partícula cuya masa es la masa reducida es:

$$\frac{mM}{m+M} \ddot{\mathbf{r}} = -G \frac{Mm}{r^2} \hat{\mathbf{e}}_r \quad (\text{A.1})$$

Esta ecuación representa el movimiento de una partícula ficticia con masa $m = 1$ en un potencial Kepleriano generado por un cuerpo con masa $M + m$.

Definiendo,

$$\Delta V = \Delta V_m - \Delta V_M \quad (\text{A.2})$$

Y por conservación del momento, tenemos:

$$m \Delta V_m + M \Delta V_M = 0 \quad (\text{A.3})$$

Al sustituir ΔV_m de la ecuación anterior en esta última, el cambio en la velocidad de M debido al arrastre es:

$$\Delta V_M = -\frac{m}{m+M} \Delta V \quad (\text{A.4})$$

Donde ΔV es el cambio en la velocidad de la partícula reducida.

Cuando $t \rightarrow -\infty$, $V = V_0$ y b es el parámetro de impacto. Así, $L = bV_0$ es el tamaño del momento angular por unidad de masa conservado.

Considerando un potencial Kepleriano, la solución a la ecuación de movimiento es:

$$\frac{1}{r} = C \cos(\Psi - \Psi_0) + G \frac{M+m}{b^2 V_0^2} \quad (\text{A.5})$$

donde C y Ψ_0 son constantes que dependen de las condiciones iniciales.

Diferenciando dos veces con respecto al tiempo:

$$\dot{r} = C r^2 \dot{\Psi} \sin(\Psi - \Psi_0) = C b V_0 \sin(\Psi - \Psi_0) \quad (\text{A.6})$$

por conservación del momento angular, $r^2 \dot{\Psi} = bV_0$.

Por lo tanto, cuando $t \rightarrow -\infty$, $r \rightarrow \infty$ y $\Psi = 0$, obtenemos que:

$$-V_0 = C b V_0 \sin(-\Psi_0) \quad (\text{A.7})$$

$$C \cos(-\Psi_0) + G \frac{M+m}{b^2 V_0^2} \quad (\text{A.8})$$

Eliminando C :

$$\tan(\Psi_0) = -\frac{b V_0}{G(M+m)} \quad (\text{A.9})$$

Donde Ψ_0 es el ángulo de máxima aproximación al centro de fuerza. El ángulo de deflexión está dado por:

$$\theta_{\text{defl}} = 2\Psi_0 - \pi \quad (\text{A.10})$$

Descomponiendo el problema en dos direcciones

Componente perpendicular de la velocidad.

$$|\Delta V_{\perp}| = V_0 \sin(\theta_{\text{defl}}) = \frac{2bV_0^3}{G(M+m)} \left[1 + \frac{b^2 V_0^2}{G^2 (M+m)^2} \right]^{-1} \quad (\text{A.11})$$

Componente paralela de la velocidad:

$$|\Delta V_{\parallel}| = V_0 - V_0 \cos(\theta_{\text{defl}}) = 2V_0 \left[1 + \frac{b^2 V_0^2}{G^2 (M+m)^2} \right]^{-1} \quad (\text{A.12})$$

ΔV_{\parallel} apunta en la dirección contraria de V_0 .

Las componentes del cambio en velocidad del perturbador de masa M son,

$$|\Delta V_{M\perp}| = \frac{2bmV_0^3}{G(M+m)^2} \left[1 + \frac{b^2 V_0^2}{G^2 (M+m)^2} \right]^{-1} \quad (\text{A.13})$$

$$|\Delta V_{M\parallel}| = \frac{2mV_0}{(M+m)} \left[1 + \frac{b^2 V_0^2}{G^2 (M+m)^2} \right]^{-1} \quad (\text{A.14})$$

A.0.1. Perturbador en un medio homogéneo

Si el perturbador de masa M viaja a través de un medio homogéneo de estrellas, por simetría, tenemos:

$$\sum \Delta V_{M\perp} = 0 \quad \text{pero} \quad \sum \Delta V_{M\parallel} \neq 0$$

ya que todos los vectores apuntan en la misma dirección. Esto se debe a que M desvía las partículas de masa menor m , creando una sobredensidad detrás de ella.

La función de distribución $f(V)$ nos da el número de estrellas encontradas con velocidades y parámetros de impacto en el intervalo $v_m \pm d^3 v_m$ y $b \pm db$.

Entonces la tasa de encuentros de estas estrellas es:

$$2\pi b db V_0 f(v_m) d^3 v_m \quad (\text{A.15})$$

El cambio neto de V_M debido a estos encuentros es:

$$\begin{aligned} \frac{dV_M}{dt} \Big|_{v_m} &= V_0 f(V_m) d^3 V_m \int_0^{b_{\max}} \frac{2m V_0}{m+M} \left[1 + \frac{b^2 V_0^2}{G^2 (M+m)} \right]^{-1} 2\pi b db \\ &= 2\pi \ln(1 + \Lambda^2) G^2 m (m+M) f(V_m) d^3 V_m \frac{V_m - V_M}{|V_m - V_M|^3} \end{aligned} \quad (\text{A.16})$$

Donde:

$$\Lambda = \frac{b_{\max} V_0^2}{G (M+m)} \quad (\text{A.17})$$

Y dado que Λ es muy grande, podemos hacer la siguiente aproximación:

$$\frac{\ln(1 + \Lambda^2)}{2} = \ln(\Lambda) \quad (\text{A.18})$$

La ecuación (A.16) establece que las estrellas que tienen velocidad V_m ejercen una fuerza sobre M que actúa paralela a $V_m - V_M$ y es inversamente proporcional al cuadrado de este vector.

Al integrar la aceleración $\frac{dV_M}{dt} \Big|_{v_m}$ sobre todos los v_m 's, es equivalente a encontrar el campo gravitacional en el punto con vector de velocidad V_m generado por la distribución de densidad de masa:

$$\rho(V_m) = 4\pi \ln(\Lambda) G m (m+M) f(v_m) \quad (\text{A.19})$$

Si las estrellas se mueven de manera isotrópica, la distribución de densidad es esférica y aplicando los teoremas de Newton, la aceleración total de M es igual a G veces la masa total que yace a velocidades $V_m < v_M$. Por lo tanto, para una distribución isotrópica de velocidades:

$$\frac{d\mathbf{V}_M}{dt} = -16\pi^2 \ln(\Lambda) G^2 m (M+m) \int_0^{v_M} f(v_m) v_m^2 dv_m \frac{\mathbf{V}_M}{v_M^3} \quad (\text{A.20})$$

A.1. Casos límites de la fórmula de Chandrasekhar

A.1.1. Dependencia de la masa

$$\frac{d\mathbf{V}_M}{dt} \propto M \quad (\text{A.21})$$

La fricción dinámica es proporcional a la densidad de masa de las partículas del medio y proporcional a la suma de las masas individuales de las partículas de campo y a la masa de

la partícula de prueba de masa M . Por lo tanto, entre más masivo sea el perturbador, se verá afectado en mayor medida por la fricción dinámica.

A.1.2. Dependencia de la velocidad en el caso de movimiento lento

En el límite de movimiento lento, es decir, $v_M \ll v_m$:

$$\frac{d\mathbf{V}_M}{dt} \propto - \int_0^{v_M} f(v_m) v_m^2 dv_m \frac{\mathbf{V}_M}{v_M^3} \quad (\text{A.22})$$

Podemos decir que $f(v_m) = f(0)$, y la fricción dinámica es:

$$\frac{d\mathbf{V}_M}{dt} = -16 \pi^2 \ln(\Lambda) G^2 m (M + m) f(0) \mathbf{v}_M \quad (\text{A.23})$$

Entre más lento se mueva el perturbador, la fricción dinámica le afectará en mayor medida.

A.1.3. Dependencia de la velocidad en el caso de movimiento rápido

En el límite de movimiento rápido, es decir, $v_M \gg v_m$:

$$\frac{d\mathbf{V}_M}{dt} \propto - \int_0^{v_M} f(v_m) v_m^2 dv_m \quad (\text{A.24})$$

Y también:

$$\frac{d\mathbf{V}_M}{dt} \propto \frac{1}{v_M^2} \quad (\text{A.25})$$

Si v_M es grande, el efecto de la fricción dinámica disminuye; el efecto disminuye para velocidades suficientemente grandes de la partícula prueba.

Apéndice B

Teoría de respuesta lineal

La teoría de respuesta lineal que se desarrolla a continuación es necesaria para entender las diferentes extensiones que han surgido a la fórmula de Chandrasekhar debido a sus restricciones. Entre los trabajos que utilizan este enfoque podemos encontrar los de [Lynden-Bell and Kalnajs \(1972\)](#) y [Tremaine and Weinberg \(1984\)](#) que estudian la fricción dinámica en sistemas esféricos obteniendo conclusiones que se contradicen con las suposiciones de la fórmula de Chandrasekhar como veremos más adelante.

La teoría de respuesta lineal suele utilizarse para estudiar la estabilidad y respuesta en sistemas de fluidos y sistemas estelares. Los sistemas estelares más simples a estudiar son los sistemas esféricos homogéneos y autogravitantes. El enfoque presentado a continuación fue tomado de [Binney and Tremaine \(1987\)](#). Empezaremos examinando un sistema en equilibrio con una densidad ρ_s que es forzada por un potencial gravitacional externo $-\epsilon\nabla\Phi_e$, donde $|\nabla\Phi_e|$ es del mismo orden que el campo gravitacional en el sistema en equilibrio y $\epsilon \ll 1$. La distribución de densidad que generaría este campo es $\epsilon\rho_e(\mathbf{x}, t)$ donde,

$$\nabla^2\Phi_e = 4\pi G\rho_e \quad (\text{B.1})$$

Ya que la perturbación es débil, la respuesta es lineal y por lo tanto también proporcional a ϵ ; de este modo podemos escribir la perturbación de densidad inducida como $\epsilon\rho_e(\mathbf{x}, t)$. La *función de respuesta* $R(x, x', \tau)$ se define como,

$$\rho_{s1}(x, t) = \int d^3x dt' R(x, x', t - t')\rho_e(x', t') \quad (\text{B.2})$$

y relaciona la densidad respuesta $\rho_{s1}(x, t)$ a la densidad forzada $\rho_e(x', t')$. La respuesta depende solamente de las diferencias entre t y t' ya que el sistema en equilibrio es independiente del tiempo, así que la respuesta al tiempo t desde un impulso instantáneo a t' pueden depender solamente del retraso $t-t'$. La causalidad requiere que $R(x, x', \tau) = 0$ para $\tau < 0$ (El

efecto no puede preceder a la causa).

La densidad forzada o forzante ρ_e y la densidad respuesta contribuyen al potencial gravitacional y el potencial total perturbante $\Phi_1 \equiv \Phi_e + \Phi_{s1}$ el que determina la dinámica del sistema. La densidad correspondiente es $\rho_1 = \rho_e + \rho_{s1}$. La función de polarización $P(x, x', \tau)$ relaciona la densidad respuesta ρ_{s1} a la densidad total ρ_1 :

$$\rho_{s1}(x, t) = \int d^3x dt' P(x, x', t - t') \rho_1(x', t) \quad (\text{B.3})$$

donde, $P(x, x', \tau) = 0$ para $\tau < 0$.

Hay una diferencia física entre la función de respuesta y la función de polarización. La función de respuesta describe la densidad respuesta a un potencial externo perturbativo, mientras que la función de polarización describe la densidad respuesta a la fuerza total perturbativa. Si la autogravedad de la densidad respuesta es despreciable, las funciones de polarización y respuesta son idénticas.

Toda función de respuesta se puede descomponer en Fourier y estas perturbaciones en Fourier pueden a su vez conformar un sistema de modos perturbativos.

Apéndice C

La función de pérdida de masa en el modelo semianalítico con fricción dinámica

Dado que el modelo semianalítico utilizado en este trabajo representa un perfil de densidad tipo NFW, nos basamos en el esquema de construcción de la función de pérdida de masa establecido en el trabajo de Klypin et al. (1999), reproducido más adelante.

Los perfiles de densidad de los halos de materia oscura en los modelos CDM ahora se conocen razonablemente bien para un amplio rango de masas y para una variedad de modelos cosmológicos. A continuación, proporcionamos estimaciones analíticas para varias propiedades del halo en el modelo estándar CDM con $\Omega = 1$.

Hay que tomar en cuenta que los procesos físicos que dan lugar al dispersamiento de subestructura incluyen a la fricción dinámica, disrupciones por efectos de marea y calentamiento de las partículas de los halos que interactúan gravitacionalmente.

La fricción dinámica hace que un halo de materia oscura se dirija hacia las partes centrales donde hay mayor densidad, ocasionando así que el material sea disruptado y se fusionen con el objeto masivo central. Que tan importante es la fricción dinámica dependerá de la masa del halo y detalles específicos de la órbita del perturbador.

El perfil de NFW está dado por:

$$\begin{aligned} \rho(r) &= \frac{\rho_0 r_s^3}{r(r+r_s)^2}, & M(r) &= M_{\text{vir}} \times \frac{f(x)}{f(C)}, \\ f(x) &\equiv \ln(1+x) - \frac{x}{1+x}, & x &\equiv \frac{r}{r_s}. \end{aligned} \quad (\text{C.1})$$

donde r_s y ρ_0 son el radio característico y la densidad del halo, M_{vir} es la masa virial, r_{vir} es el radio de virial.

C es la concentración del halo definida como:

$$C \equiv \frac{r_{\text{vir}}}{r_s},$$

$$r_{\text{vir}}(M_{\text{vir}}) = 443 h^{-1} \text{kpc} \left(\frac{M_{\text{vir}} / 10^{11} h^{-1} M_{\odot}}{\Omega_0 \delta_{\text{th}}} \right)^{1/3}, \quad (\text{C.2})$$

$$M_{\text{vir}} \equiv \frac{4\pi}{3} \rho_{\text{cr}} \Omega_0 \delta_{\text{th}} r_{\text{vir}}^3.$$

Donde, ρ_{cr} es la densidad crítica del universo y δ_{th} es la sobredensidad ($\delta\rho/\rho_{\text{matter}}$) de un objeto colapsado de acuerdo al modelo top-hat del colapso esférico. Para el modelo CDM $\delta_{\text{th}} \approx 200$.

La definición de r_{vir} difiere de la usada en NFW, quienes usan $\delta_{\text{th}} = 200$ para todos los modelos cosmológicos. Se usan valores predichos por el modelo top-hat, que da por ejemplo $\delta_{\text{th}} \approx 340$ para el modelo Λ CDM con $\Omega_0 = 0,3$. La concentración C es una función de la masa M_{vir} .

La concentración C es una función de la masa M_{vir} . En el caso del modelo CDM:

$$C \approx 124 \left(M_{\text{vir}} / 1 h^{-1} M_{\odot} \right)^{-0,084} \quad (\text{C.3})$$

Valores típicos de la concentración C van desde el rango de $C \approx 12$ para $M_{\text{vir}} = 10^{12} h^{-1} M_{\odot}$ hasta $C \approx 7$ para $M_{\text{vir}} = 10^{15} h^{-1} M_{\odot}$ (NFW). Usando estas definiciones, se puede escribir la masa $M(r)$, frecuencia orbital $\Omega(r)$, y el potencial gravitacional $\phi(r)$:

$$\Omega^2(r) = \frac{GM}{r^3} = \frac{GM_{\text{vir}}}{r_s^3 f(C)} \times \frac{f(x)}{x^3},$$

$$\phi(r) = -\frac{GM_{\text{vir}}}{r_s f(C)} \times \frac{f(x) + x/(1+x)}{x}. \quad (\text{C.4})$$

En algunos casos es más conveniente definir propiedades de los halos usando la velocidad rotacional máxima $V_{\text{máx}} = (GM/r)^{1/2}|_{\text{máx}}$, el parámetro del halo que es más fácilmente relacionable a una cantidad observable, en vez de la concentración C o la masa del virial M_{vir} .

Para el perfil de la ecuación (numero), la velocidad rotacional máxima ocurre en $r_{\text{máx}} \approx 2r_s$. Esto da

$$\begin{aligned}
 V_{\text{máx}}^2 &= \frac{G M_{\text{vir}}}{r_s} \frac{f(2)}{2f(C)}, \quad f(2) \approx 0,432, \\
 M(r) &= \frac{r_s V_{\text{máx}}^2}{G} \frac{2f(x)}{f(2)}, \quad \Omega^2(r) = \frac{V_{\text{máx}}^2}{r_s^2} \times \frac{2f(x)}{x^3 f(2)}, \\
 V_{\text{esc}}^2 &= -2\phi(r) = 4 V_{\text{máx}}^2 \times \frac{\ln(1+x)}{x f(2)},
 \end{aligned} \tag{C.5}$$

donde V_{esc} es la velocidad de escape a una distancia del centro de la galaxia.

El radio de marea r_t de un halo pequeño con masa m y velocidad rotacional máxima $vR_{\text{máx}}$ moviéndose a un radio R desde el centro de un gran halo con masa $M(R)$ es $V_{\text{máx}}$, es el mínimo de dos radios:

1. Un radio en el cual la fuerza gravitacional del halo pequeño es igual a la fuerza de marea del halo grande, $F_{\text{grav}} = F_{\text{tide}}$.
2. Un radio definido por las resonancias tren la fuerza que el halo pequeño ejerce sobre las partículas y las fuerzas de marea debidas al halo grande.

Si r es la distancia de una partícula desde el centro del pequeño halo, entonces la condición es

$F_{\text{grav}}(r) = F_{\text{tide}}(r; R)$ da una ecuación para el radio de marea r_t :

$$\begin{aligned}
 \left(\frac{R}{r_t}\right)^3 \frac{m(r_t)}{M(R)} &= 2 - \frac{R}{M} \frac{\partial M}{\partial R}, \\
 \frac{f(x_r)}{f(x_R)} &= \left(\frac{x_r}{x_R}\right)^3 \left(\frac{r_s V_{\text{máx}}}{R_s v_{\text{máx}}}\right)^2 \left[2 - \frac{x_R^2}{(1+x_R)^2 f(x_R)} \right], \\
 x_r &\equiv \frac{r_t}{r_s}, \quad x_R \equiv \frac{R}{R_s}
 \end{aligned} \tag{C.6}$$

La última ecuación se puede resolver numéricamente. Se asume que la disrupción ocurre principalmente en resonancias primarias. $\Omega(r)|_{\text{small}} = \Omega(R)|_{\text{large}}$. Esto da lugar a la ecuación para el radio de marea:

$$\frac{f(x_r)}{f(x_R)} = \left(\frac{x_r}{x_R}\right)^3 \left(\frac{r_s V_{\text{máx}}}{R_s v_{\text{máx}}}\right)^2 \tag{C.7}$$

Se toma el valor más pequeño de los dos estimados para r_t . Para $x_R < 2,2$ el radio de marea queda definido por la condición de fuerza. A distancias pequeñas, la resonancia orbital-interna define el radio de marea.

La fricción dinámica es otro efecto que contribuye al borrado de la subestructura. Este no es un efecto numérico, pero al llevar galaxias al centro de la galaxia, donde serán destruidas por marea, puede potenciar los efectos numéricos. El tiempo de fricción dinámica para un pequeño halo con masa m moviéndose en una órbita circular de radio R alrededor del gran halo puede estimarse utilizando la fórmula de Chandrasekhar (Binney and Tremaine, 2008) con suposiciones de equilibrio y una distribución isotrópica maxwelliana de velocidades de las partículas de materia oscura. La tasa de decadencia del radio orbital debido a la fricción dinámica está dada por

$$\begin{aligned} \frac{dR}{dt} &= -\frac{2R}{t_{\text{fric}}} \left[\frac{\partial \ln M(R)}{\partial \ln R} + 1 \right], \\ t_{\text{fric}} &= \frac{V_{\text{circ}}^3}{4\pi G^2 (\ln \Lambda) m(r_i) \rho(R) [\text{erf}(X) - 2X e^{-X^2/\sqrt{\pi}}]}, \\ V_{\text{circ}}^2 &= \frac{GM(R)}{R}, \quad X = \frac{V_{\text{circ}}}{\sqrt{2}\sigma_r}, \\ \ln \Lambda &= \ln \left[\frac{R_{\text{vir}} M(R)}{R m(r_i)} \right], \\ M(R) &= -\frac{R\sigma_r^2}{G} \left[\frac{d \ln \rho}{d \ln R} + \frac{d \ln \sigma_r^2}{d \ln R} \right]. \end{aligned} \quad (\text{C.8})$$

Para la ecuación de perfil de densidad, tenemos $d \ln \rho / d \ln R = -(1 + 3x)/(1 + x)$, donde $x = R/R_s$. La última de las ecuaciones para X se puede reescribir como una ecuación para la dispersión de velocidad:

$$\frac{d\sigma_r^2}{dx} - \frac{1 + 3x}{x(1 + x)} \sigma_r^2 = -\frac{GM_{\text{vir}}}{R_s f(C)} \times \frac{f(x)}{x^2}. \quad (\text{C.9})$$

La solución de la ecuación es

$$\sigma_r^2 = \frac{x(1 + x)^2}{4} \left[\sigma_0^2 - 4 \frac{GM_{\text{vir}}}{R_s f(C)} \int_1^x \frac{f(x)}{x^3 (1 + x)^2} dx \right] \quad (\text{C.10})$$

$$\begin{aligned} &= V_{\text{máx}}^2 \frac{2x(1 + x)^2}{f(2)} \int_x^\infty \frac{f(x)}{x^3 (1 + x)^2} dx \quad (\text{C.11}) \\ \sigma_0^2 &= 4 \frac{GM_{\text{vir}}}{R_s f(C)} \int_1^\infty \frac{f(x)}{x^3 (1 + x)^2} dx \int dx \approx 0,432 V_{\text{máx}}^2 \end{aligned}$$

Aquí, σ_0 es la dispersión de velocidad unidimensional en $R = R_s$. La dispersión de velocidad σ_r tiene un máximo $\sigma_r \approx \sigma_0$ en $R \approx 0,8R_s$. Declina en radios más pequeños y más grandes, pero el máximo es plano: $\sigma_r \approx 0,78\sigma_0$ en $R = 0,1R_s$ y $\sigma_r \approx 0,69\sigma_0$ en $R = r_{\text{vir}}$. Las

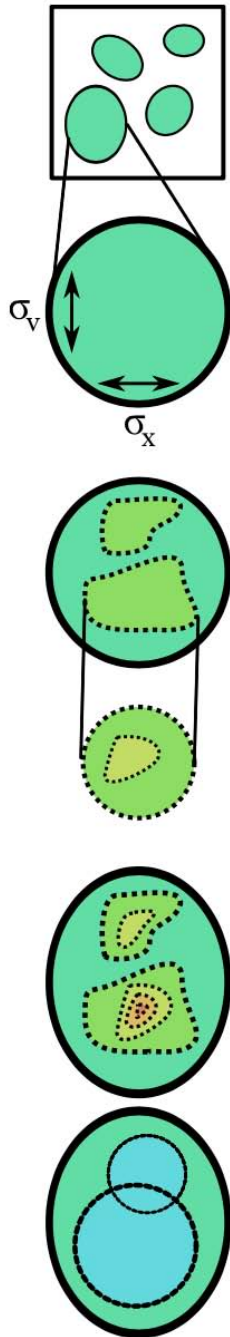
ecuaciones para X y σ_r definen la fricción dinámica.

Apéndice D

Software para *Rockstar Halo Finder* para identificar subhalos

Rockstar (*Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement*) (Behroozi, Wechsler, and Wu, 2013) es un software que sirve para identificar halos de materia oscura y subestructura como la asociada a efectos de marea. Su enfoque está basado en un refinamiento jerárquico adaptivo de grupos *friend-of-friends* (Tempel et al. (2016)), es decir, agrupamiento de partículas que se encuentran cercanas entre ellas dentro de una vecindad, en el espacio de seis dimensiones y en una dimensión temporal lo que permite un rastreo robusto de subestructura, *independiente-de-malla*, *independiente-de-forma* y resistente al ruido.

Rockstar utiliza información tanto del espacio de configuración, como del espacio de fase (posición y velocidad) para identificar sub-halos. Esto le permite identificar con precisión estructuras que otros métodos podrían pasar por alto o malinterpretar, ver fig D.1. Comienza identificando halos potenciales en el espacio de configuración utilizando un método de vinculación de amigos-de-amigos. Una vez identificados los halos iniciales, ROCKSTAR refina estas detecciones utilizando el espacio de fase. Esto permite identificar subhalos con más precisión que si sólo se usara el espacio de configuración. Se identifican y eliminan Sub-halos artificiales que surgieran debido a efectos numericos derivados de las limitaciones presentes en las simulaciones. Utilizando densidades locales y un enfoque adaptativo, se identifican estructuras dentro de halos más grandes, esto es crucial para estudios de la estructura interna de halos y la formación de galaxias satélites, por ejemplo. Una vez identificados los halos y los subhalos, ROCKSTAR procede a calcular una serie de propiedades para cada uno de los subhalos, como su masa, velocidad, forma, etc.



- El volumen de la simulación se divide en grupos *friend-of-friends* para una fácil paralelización.
- Para cada grupo, las posiciones de las partículas y sus velocidades se dividen (normalizadas) por la posición del grupo y las dispersiones de velocidad, lo que da una métrica natural del espacio fase.
- Una longitud del espacio fase se escoge de manera adaptiva de modo que el 70 % de las partículas del grupo se unan en grupos.
- Se repite el proceso para cada subgrupo: renormalización de una nueva *longitud-enlace* y un nuevo nivel de subestructura se calculan.
- Una vez que todos los niveles de subestructura son encontrados, los halos semilla son colocados al nivel más bajo de subestructura y las partículas son asignadas jerárquicamente al halo semilla más cercano en el espacio fase.
- Una vez que las partículas han sido asignadas a halos, las partículas no ligadas son removidas y las propiedades del halo (posiciones y velocidades) son calculadas.

Figura D.1: Resumen visual del algoritmo de asignación partícula-halo. Crédito: (Behroozi, Wechsler, and Wu, 2013)

Bibliografía

- S. J. Aarseth. Dynamical evolution of clusters of galaxies, I. *MNRAS*, 126:223, January 1963. doi: 10.1093/mnras/126.3.223.
- Mario G. Abadi, Julio F. Navarro, Matthias Steinmetz, and Vincent R. Eke. Simulations of Galaxy Formation in a Λ Cold Dark Matter Universe. I. Dynamical and Photometric Properties of a Simulated Disk Galaxy. *ApJ*, 591(2):499–514, July 2003a. doi: 10.1086/375512.
- Mario G. Abadi, Julio F. Navarro, Matthias Steinmetz, and Vincent R. Eke. Simulations of Galaxy Formation in a Λ Cold Dark Matter Universe. II. The Fine Structure of Simulated Galactic Disks. *ApJ*, 597(1):21–34, November 2003b. doi: 10.1086/378316.
- G.üello@, O. Valenzuela, J. C. Clemente, H.ázquez@, and J. A. Trelles. Assessing the Hierarchical Hamiltonian Splitting Integrator for Collisionless N-body Simulations. *arXiv e-prints*, art. arXiv:2009.06133, September 2020. doi: 10.48550/arXiv.2009.06133.
- G.üello@, O. Valenzuela, and A. Trelles. Boltzmann-Poisson-like approach to simulating the galactic halo response to satellite accretion. Dependence on the halo density profile. *aap*, 663:A93, July 2022. doi: 10.1051/0004-6361/202141988.
- Christine Allen and Alfredo Santillan. An improved model of the galactic mass distribution for orbit computations. , 22:255, October 1991.
- Massih-Reza Amini, Vasilii Feofanov, Loic Pauletto, Lies Hadjadj, Emilie Devijver, and Yury Maximov. Self-Training: A Survey. *arXiv e-prints*, art. arXiv:2202.12040, February 2022. doi: 10.48550/arXiv.2202.12040.
- Deokkeun An and Timothy C. Beers. A blueprint for the milky way’s stellar populations: The power of large photometric and astrometric surveys. *The Astrophysical Journal*, 897(1):39, jun 2020. doi: 10.3847/1538-4357/ab8d39. URL <https://dx.doi.org/10.3847/1538-4357/ab8d39>.
- T. Antoja, C. Mateu, L. Aguilar, F. Figueras, E. Antiche, F.ández-Pérez@, A. G. A. Brown, O. Valenzuela, A. Aparicio, S. Hidalgo, and H.ázquez@. Detection of satellite remnants in the Galactic halo with Gaia- III. Detection limits for ultrafaint dwarf galaxies. *MNRAS*, 453(1):541–560, October 2015. doi: 10.1093/mnras/stv1622.
- T. Antoja, A. Helmi, M.ómez@, D. Katz, C. Babusiaux, R. Drimmel, D. W. Evans, F. Figueras, E. Poggio, C.é@, A. C. Robin, G. Seabroke, and C. Soubiran. A dynamically young and perturbed Milky Way

- disk. *nat*, 561(7723):360–362, September 2018. doi: 10.1038/s41586-018-0510-7.
- T. Antoja, P. Ramos, C. Mateu, A. Helmi, F. Anders, C. Jordi, and J. A. Carballo-Bello. An all-sky proper-motion map of the Sagittarius stream using Gaia DR2. *aap*, 635:L3, March 2020. doi: 10.1051/0004-6361/201937145.
- I. Arcavi. Introduction to astrophysics, Dynamical Friction. *astrocourse*, 8, May 2011.
- J. S. Bagla. TreePM: A Code for Cosmological N-Body Simulations. *Journal of Astrophysics and Astronomy*, 23(3-4): 185–196, December 2002. doi: 10.1007/BF02702282.
- H B Barlow. Unsupervised learning: introduction. In Geoffrey E. Hinton and Terrence Joseph Sejnowski, editors, *Unsupervised Learning: Foundations of Neural Computation*, pages 1–17. Bradford Company Scituate, MA, USA, 1999. URL <http://books.google.com/books?id=yj04Y01je4cC>.
- J. Barnes and P. Hut. A hierarchical $O(N \log N)$ force-calculation algorithm. *nat*, 324: 446–449, December 1986. doi: 10.1038/324446a0.
- Peter S. Behroozi, Risa H. Wechsler, and Hao-Yi Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *ApJ*, 762(2):109, January 2013. doi: 10.1088/0004-637X/762/2/109.
- Eric F. Bell, Daniel B. Zucker, Vasily Belokurov, Sanjib Sharma, Kathryn V. Johnston, James S. Bullock, David W. Hogg, Knud Jahnke, Jelte T. A. de Jong, Timothy C. Beers, N. W. Evans, Eva K. Grebel, Željko Ivezić, Sergey E. Koposov, Hans-Walter Rix, Donald P. Schneider, Matthias Steinmetz, and Adi Zolotov. The Accretion Origin of the Milky Way’s Stellar Halo. *ApJ*, 680(1):295–311, June 2008. doi: 10.1086/588032.
- G. Bertelli, A. Bressan, C. Chiosi, F. Fagotto, and E. Nasi. Theoretical isochrones from models with new radiative opacities. *ApJ*, 106:275–302, August 1994.
- Francesco Bertola. Putting galaxies on the scale. *Science*, 295(5553):283–284, 2002. doi: 10.1126/science.1067918. URL <https://www.science.org/doi/abs/10.1126/science.1067918>.
- J. Binney. Dynamical friction in aspherical clusters. *MNRAS*, 181:735–746, December 1977. doi: 10.1093/mnras/181.4.735.
- J. Binney and S. Tremaine. *Galactic dynamics*. Addison-Wesley, 1987.
- James Binney and Scott Tremaine. *Galactic Dynamics: Second Edition*. 2008.
- Charles K. Birdsall and Dieter Fuss. Clouds-in-clouds, clouds-in-cells physics for many-body plasma simulation. *Journal of Computational Physics*, 135(2): 141–148, 1997. ISSN 0021-9991. doi: <https://doi.org/10.1006/jcph.1997.5723>. URL <https://www.sciencedirect.com>.

- [com/science/article/pii/S0021999197957235](https://doi.org/10.1093/mnras/sty1281).
- Michael R. et al. Blanton. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. *aj*, 154(1):28, July 2017. doi: 10.3847/1538-3881/aa7567.
- P. Boldrini, R. Mohayaee, and J. Silk. Does Fornax have a cored halo? Implications for the nature of dark matter. *arXiv e-prints*, art. arXiv:1806.09591, June 2018.
- Bernhard Boser, Isabelle Guyon, and Vladimir Vapnik. A training algorithm for optimal margin classifier. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 5, 08 1996. doi: 10.1145/130385.130401.
- Michael Boylan-Kolchin, Chung-Pei Ma, and Eliot Quataert. Dynamical friction and galaxy merging time-scales. *MNRAS*, 383(1):93–101, January 2008. doi: 10.1111/j.1365-2966.2007.12530.x.
- C. B. Brook, G. S. Stinson, B. K. Gibson, D. Kawata, E. L. House, M. S. Miranda, A. V. Macciò, K. Pilkington, R.škar@, J. Wadsley, and T. R. Quinn. Thin disc, thick disc and halo in a simulated galaxy. *MNRAS*, 426(1):690–700, October 2012. doi: 10.1111/j.1365-2966.2012.21738.x.
- Sven Buder, Martin Asplund, Ly Duong, Janez Kos, Karin Lind, Melissa K. Ness, Sanjib Sharma, Joss Bland-Hawthorn, Andrew R. Casey, Gayandhi M. de Silva, Valentina D’Orazi, Ken C. Freeman, Geraint F. Lewis, Jane Lin, Sarah L. Martell, Katharine J. Schlesinger, Jeffrey D. Simpson, Daniel B. Zucker, Tomaz Zwitter, Anish M. Amarsi, Borja Anguiano, Daniela Carollo, Luca Casagrande, Klemen Čotar, Peter L. Cottrell, Gary da Costa, Xudong D. Gao, Michael R. Hayden, Jonathan Horner, Michael J. Ireland, Prajwal R. Kafle, Ulisse Munari, David M. Nataf, Thomas Nordlander, Dennis Stello, Yuan-Sen Ting, Gregor Traven, Fred Watson, Robert A. Wittenmyer, Rosemary F. G. Wyse, David Yong, Joel C. Zinn, Maruša Žerjal, and Galah Collaboration. The GALAH Survey: second data release. *MNRAS*, 478(4):4513–4552, August 2018. doi: 10.1093/mnras/sty1281.
- Jianghui Cai, Jing Hao, Haifeng Yang, Xujun Zhao, and Yuqing Yang. A review on semi-supervised clustering. *Information Sciences*, 632:164–200, 2023. ISSN 0020-0255. doi: <https://doi.org/10.1016/j.ins.2023.02.088>. URL <https://www.sciencedirect.com/science/article/pii/S0020025523002840>.
- R. Capuzzo-Dolcetta and A. Vicari. Dynamical friction on globular clusters in core-triaxial galaxies: is it a cause of massive black hole accretion? *MNRAS*, 356(3): 899–912, January 2005. doi: 10.1111/j.1365-2966.2004.08433.x.
- A. Castro-Ginard, C. Jordi, X. Luri, F. Julbe, M. Morvan, L.úñez@, and T. Cantat-Gaudin. A new method for unveiling open clusters in Gaia. New nearby open clusters confirmed by DR2. *aap*, 618:

- A59, Oct 2018. doi: 10.1051/0004-6361/201833390.
- S. Chandrasekhar. Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. *ApJ*, 97:255, March 1943. doi: 10.1086/144517.
- B. Q. Chen, X. W. Liu, H. B. Yuan, M. S. Xiang, Y. Huang, C. Wang, H. W. Zhang, and Z. J. Tian. The selection function of the LAMOST Spectroscopic Survey of the Galactic Anti-centre. *MNRAS*, 476(3):3278–3289, May 2018. doi: 10.1093/mnras/sty454.
- P.ín@, V. Avila-Reese, and O. Valenzuela. Substructure and Halo Density Profiles in a Warm Dark Matter Cosmology. In José G. Funes and Enrico Maria Corsini, editors, *Galaxy Disks and Disk Galaxies*, volume 230 of *Astronomical Society of the Pacific Conference Series*, pages 651–652, January 2001. doi: 10.48550/arXiv.astro-ph/0009317.
- Pedro Colín, Vladimir Avila-Reese, Enrique Vázquez-Semadeni, Octavio Valenzuela, and Daniel Ceverino. Low-mass Galaxy Formation in Cosmological Adaptive Mesh Refinement Simulations: The Effects of Varying the Sub-grid Physics Parameters. *ApJ*, 713(1):535–551, April 2010. doi: 10.1088/0004-637X/713/1/535.
- M. Colpi and A. Pallavicini. Drag on a Satellite Moving across a Spherical Galaxy: Tidal and Frictional Forces in Short-lived Encounters. *ApJ*, 502:150–166, July 1998. doi: 10.1086/305877.
- Monica Colpi, Lucio Mayer, and Fabio Governato. Dynamical Friction and the Evolution of Satellites in Virialized Halos: The Theory of Linear Response. *ApJ*, 525(2):720–733, November 1999. doi: 10.1086/307952.
- H. M. P. Couchman, P. A. Thomas, and F. R. Pearce. Hydra: an Adaptive-Mesh Implementation of P 3M-SPH. *ApJ*, 452:797, October 1995. doi: 10.1086/176348.
- Ramanath Cowsik, Kasey Wagoner, Emanuele Berti, and Amit Sircar. Internal dynamics and dynamical friction effects in the dwarf spheroidal galaxy in fornax. *The Astrophysical Journal*, 699(2):1389, jun 2009. doi: 10.1088/0004-637X/699/2/1389. URL <https://dx.doi.org/10.1088/0004-637X/699/2/1389>.
- Phil Cuddeford. An analytic inversion for anisotropic spherical galaxies. *Monthly Notices of the Royal Astronomical Society*, 253(3):414–426, 12 1991. ISSN 0035-8711. doi: 10.1093/mnras/253.3.414. URL <https://doi.org/10.1093/mnras/253.3.414>.
- M. A. Czekaj, A. C. Robin, F. Figueras, X. Luri, and M. Haywood. The Besançon Galaxy model renewed. I. Constraints on the local star formation history from Tycho data. *aap*, 564:A102, April 2014. doi: 10.1051/0004-6361/201322139.
- Payel Das and James Binney. Characterizing stellar halo populations - I. An extended distribution function for halo K giants.

- MNRAS*, 460(2):1725–1738, August 2016. doi: 10.1093/mnras/stw744.
- M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292:371–394, May 1985. doi: 10.1086/163168.
- Flavio De Lorenzi, Victor P. Debattista, Ortwin Gerhard, and Niranjan Sambhus. nmagic: a fast parallel implementation of a xi2-made-to-measure algorithm for modeling observational data. *Monthly Notices of the Royal Astronomical Society*, 376(1):71–88, 02 2007. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2007.11434.x. URL <https://doi.org/10.1111/j.1365-2966.2007.11434.x>.
- Maria E. de Rossi, Patricia B. Tissera, Gabriella De Lucia, and Guinevere Kauffmann. Milky Way type galaxies in a Λ CDM cosmology. *MNRAS*, 395(1):210–217, May 2009. doi: 10.1111/j.1365-2966.2009.14560.x.
- G. M. De Silva, K. C. Freeman, J. Bland-Hawthorn, S. Martell, E. Wylie de Boer, M. Asplund, S. Keller, S. Sharma, D. B. Zucker, T. Zwitter, B. Anguiano, C. Baciagalupo, D. Bayliss, M. A. Beavis, M. Bergemann, S. Campbell, R. Cannon, D. Carollo, L. Casagrande, A. R. Casey, G. Da Costa, V. D’Orazi, A. Dotter, L. Duong, A. Heger, M. J. Ireland, P. R. Kafle, J. Kos, J. Lattanzio, G. F. Lewis, J. Lin, K. Lind, U. Munari, D. M. Nataf, S. O’Toole, Q. Parker, W. Reid, K. J. Schlesinger, A. Sheinis, J. D. Simpson, D. Stello, Y. S. Ting, G. Traven, F. Watson, R. Wittenmeyer, D. Yong, and M. Žerjal. The GALAH survey: scientific motivation. *MNRAS*, 449(3):2604–2617, May 2015. doi: 10.1093/mnras/stv327.
- Walter Dehnen. A Hierarchical $E10_0$ Force Calculation Algorithm. *Journal of Computational Physics*, 179(1):27–42, June 2002. doi: 10.1006/jcph.2002.7026.
- R. Dominguez-Tenreiro and M. A. Gomez-Flechoso. Dynamical friction for compound bodies. *MNRAS*, 294(3):465–477, March 1998. doi: 10.1046/j.1365-8711.1998.01224.x.
- Harris Drucker, Chris J. C. Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. Support vector regression machines. In *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS’96, page 155–161, Cambridge, MA, USA, 1996. MIT Press.
- Richard D’Souza and Eric F. Bell. Uncertainties associated with the backward integration of dwarf satellites using simple parametric potentials. *arXiv e-prints*, art. arXiv:2202.05707, February 2022.
- A. S. Eddington. The distribution of stars in globular clusters. *MNRAS*, 76:572–585, May 1916. doi: 10.1093/mnras/76.7.572.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Confe-*

- rence on Knowledge Discovery and Data Mining, KDD'96, page 226–231. AAAI Press, 1996.
- M. Fellhauer and D. N. C. Lin. The influence of mass-loss from a star cluster on its dynamical friction - I. Clusters without internal evolution. *MNRAS*, 375(2):604–614, February 2007. doi: 10.1111/j.1365-2966.2006.11308.x.
- Evelyn Fix and Joseph L. Hodges. *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*. USAF School of Aviation Medicine, Randolph Field, Texas, 1951. URL <http://original-url-if-available.com>. Archived (PDF) from the original on September 26, 2020.
- A. S. Font, I. G. McCarthy, R. A. Crain, T. Theuns, J. Schaye, R. P. C. Wiersma, and C. Dalla Vecchia. Cosmological simulations of the formation of the stellar haloes around disc galaxies. *MNRAS*, 416(4): 2802–2820, October 2011. doi: 10.1111/j.1365-2966.2011.19227.x.
- Alexander L. Fradkov. Early history of machine learning. *IFAC-PapersOnLine*, 53(2):1385–1390, 2020. ISSN 2405-8963. doi: <https://doi.org/10.1016/j.ifacol.2020.12.1888>. URL <https://www.sciencedirect.com/science/article/pii/S2405896320325027>. 21st IFAC World Congress.
- Michiko Fujii, Yoko Funato, and Junichiro Makino. Dynamical Friction on Satellite Galaxies. *Publications of the Astronomical Society of Japan*, 58(4):743–752, 08 2006. ISSN 0004-6264. doi: 10.1093/pasj/58.4.743. URL <https://doi.org/10.1093/pasj/58.4.743>.
- Gaia Collaboration. The Gaia mission. *aap*, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272.
- Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, M. Biermann, O. L. Creevey, D. W. Evans, L. Eyer, A. Hutton, F. Jansen, C. Jordi, S. A. Klioner, U. Lammers, L. Lindegren, X. Luri, F. Migard, C. Panem, D. Pourbaix, S. Randich, P. Sartoretti, C. Soubiran, N. A. Walton, F. Arenou, C. A. L. Bailer-Jones, U. Bastian, M. Cropper, R. Drimmel, D. Katz, M. G. Lattanzi, F. van Leeuwen, J. Bakker, C. Cacciari, J. ñeda@, F. De Angeli, C. Ducourant, C. Fabricius, M. Fouesneau, Y. émat@, R. Guerra, A. Guerrier, J. Guiraud, A. Jean-Antoine Piccolo, E. Masana, R. Messineo, N. Mowlavi, C. Nicolas, K. Nienartowicz, F. Pailler, P. Panuzo, F. Rielet, W. Roux, G. M. Seabroke, R. Sordo, P. Tanga, F. évenin@, G. Gracia-Abril, J. Portell, D. Teyssier, M. Altmann, R. Andrae, I. Bellas-Velidis, K. Benson, J. Berthier, R. Blomme, E. Brugaletta, P. W. Burgess, G. Busso, B. Carry, A. Cellino, N. Cheek, G. Clementini, Y. Damerджи, M. Davidson, L. Delchambre, A. Dell'Oro, J. ández-Hernández@, L. Galluccio, P. ía-Lario@, M. Garcia-Reinaldos, J. ález-Núñez@, E. Gosset, R. Haigron, J. L. Halbwachs, N. C. Hambly, D. L. Harrison, D. Hatzidimi-

- triou, U. Heiter, J.ández@, D. Hestroffer, S. T. Hodgkin, B. Holl, K.ßen@, G. Jevardat de Fombelle, S. Jordan, A. Krone-Martins, A. C. Lanzafame, W. Loffler, A. Lorca, M. Manteiga, O. Marchal, P. M. Marrese, A. Moitinho, A. Mora, K. Muinonen, P. Osborne, E. Pancino, T. Pauwels, J. M. Petit, A. Recio-Blanco, P. J. Richards, M. Riello, L. Rimoldini, A. C. Robin, T. Roegiers, J. Rybizki, L. M. Sarro, C. Siopis, M. Smith, A. Sozzetti, A. Ulla, E. Utrilla, M. van Leeuwen, W. van Reeve, U. Abbas, A. Abreu Aramburu, S. Accart, C. Aerts, J. J. Aguado, M. Ajaj, G. Altavilla, M. A. Álvarez, J.Álvarez Cid-Fuentes@, J. Alves, R. I. Anderson, E. Anglada Varela, T. Antoja, M. Audard, D. Baines, S. G. Baker, L.úñez@, E. Balbinot, Z. Balog, C. Barache, D. Barbato, M. Barros, M. A. Barstow, S.é@, J. L. Bassilana, N. Bauchet, A. Baudesson-Stella, U. Becciani, M. Bellazzini, M. Bernet, S. Bertone, L. Bianchi, S. Blanco-Cuaresma, T. Boch, A. Bombrun, D. Bossini, S. Bouquillon, A. Bragaglia, L. Bramante, E. Breedt, A. Bressan, N. Brouillet, B. Bucciarelli, A. Burlacu, D. Busonero, A. G. Butkevich, R. Buzzi, E. Caffau, R. Cancelliere, H.ánovas@, T. Cantat-Gaudin, R. Carballo, T. Carlucci, M. I. Carnerero, J. M. Carrasco, L. Casamiquela, M. Castellani, A. Castro-Ginard, P. Castro Sampol, L. Chaoul, P. Charlot, L. Chemin, A. Chiavassa, M. R. L. Cioni, G. Comoretto, W. J. Cooper, T. Cornez, S. Cowell, F. Crifo, M. Crosta, C. Crowley, C. Dafonte, A. Dapergolas, M. David, P. David, P. de Laverny, F. De Luise, R. De March, J. De Ridder, R. de Souza, P. de Teodoro, A. de Torres, E. F. del Peloso, E. del Pozo, M. Delbo, A. Delgado, H. E. Delgado, J. B. Delisle, P. Di Matteo, S. Diakite, C. Diener, E. Distanico, C. Dolding, D. Eappachen, B. Edvardsson, H. Enke, P. Esquej, C. Fabre, M. Fabrizio, S. Faigler, G. Fedorets, P. Fernique, A. Fienga, F. Figueras, C. Fournon, F. Fragkoudi, E. Fraile, F. Franke, M. Gai, D. Garabato, A. Garcia-Gutierrez, M.ía-Torres@, A. Garofalo, P. Gavras, E. Gerlach, R. Geyer, P. Giacobbe, G. Gilmore, S. Girona, G. Giuffrida, R. Gommel, A. Gomez, I. Gonzalez-Santamaria, J. J. González-Vidal, M. Granvik, R.érrez-Sánchez@, L. P. Guy, M. Hauser, M. Hayward, A. Helmi, S. L. Hidalgo, T. Hilger, N. Hladczuk, D. Hobbs, G. Holland, H. E. Huckle, G. Jasiewicz, P. G. Jonker, J. Juaristi Campillo, F. Julbe, L. Karbevská, P. Kervella, S. Khanna, A. Kochoska, M. Kontizas, G. Kordopatis, A. J. Korn, Z. Kostrzewa-Rutkowska, K.ńska@, S. Lambert, A. F. Lanza, Y. Lagneau, J. F. Le Campion, Y. Le Fustec, Y. Lebreton, T. Lebzelter, S. Leccia, N. Leclerc, I. Lecoœur-Taïbi, S. Liao, E. Licata, E. P. Lindstrom, T. A. Lister, E. Livanou, A. Lobel, P. Madrero Pardo, S. Managau, R. G. Mann, J. M. Marchant, M. Marconi, M. M. S. Marcos Santos, S. Marinoni, F. Marocco, D. J. Marshall, L. Martin Polo, J. M. Martín-Fleitas, A. Masip, D. Massari, A. Mastrobuono-Battisti, T. Mazeh, P. J. McMillan, S. Messina, D. Michalik, N. R. Millar, A. Mints, D. Molina, R. Molinaro, L.ár@, P. Montegriffo,

- R. Mor, R. Morbidelli, T. Morel, D. Morris, A. F. Mulone, D. Munoz, T. Muraveva, C. P. Murphy, I. Musella, L. Noval, C.énovic@, G.ù@, J. Osinde, C. Pagani, I. Pagano, L. Palaversa, P. A. Palićio, A. Panahi, M. Pawlak, X.ñalosa Esteller@, A. Penttila, A. M. Piersimoni, F. X. Pineau, E. Plachy, G. Plum, E. Poggio, E. Poretti, E. Poujoulet, A. Prsa, L. Pulone, E. Racero, S. Ragaini, M. Rainer, C. M. Raiteri, N. Rambaux, P. Ramos, M. Ramos-Lerate, P. Re Fiorentin, S. Regibo, C.é@, V. Ripepi, A. Riva, G. Rixon, N. Robichon, C. Robin, M. Roelens, L. Rohrbasser, M.ómez@, N. Rowell, F. Royer, K. A. Rybicki, G. Sadowski, A.és@, J. Sahlmann, J. Salgado, E. Salguero, N. Samaras, V. Sanchez Gimenez, N. Sanna, R.ña@, M. Sarasso, M. Schultheis, E. Sciacca, M. Segol, J. C. Segovia, D.égransan@, D. Semeux, S. Shahaf, H. I. Siddiqui, A. Siebert, L. Siltala, E. Slezak, R. L. Smart, E. Solano, F. Solitro, D. Souami, J. Souchay, A. Spagna, F. Spoto, I. A. Steele, H. Steidelmüller, C. A. Stephenson, M. Suveges, L. Szabados, E. Szegedi-Elek, F. Taris, G. Tauran, M. B. Taylor, R. Teixeira, W. Thuillot, N. Tonello, F. Torra, J. Torra, C. Turon, N. Unger, M. Vaillant, E. van Dillen, O. Vanel, A. Vecchiato, Y. Viala, D. Vicente, S. Voutsinas, M. Weiler, T. Wevers, L. Wyrzykowski, A. Yoldas, P. Yvard, H. Zhao, J. Zorec, S. Zucker, C. Zurbach, and T. Zwitter. Gaia Early Data Release 3. Summary of the contents and survey properties. *aap*, 649:A1, May 2021. doi: 10.1051/0004-6361/202039657.
- Nicolas Garavito-Camargo, Gurtina Besla, Chervin F. P. Laporte, Kathryn V. Johnston, Facundo A. Gómez, and Laura L. Watkins. Hunting for the Dark Matter Wake Induced by the Large Magellanic Cloud. *ApJ*, 884(1):51, October 2019. doi: 10.3847/1538-4357/ab32eb.
- Nicolás Garavito-Camargo, Gurtina Besla, Chervin F. P. Laporte, Adrian M. Price-Whelan, Emily C. Cunningham, Kathryn V. Johnston, Martin Weinberg, and Facundo A. Gómez. Quantifying the Impact of the Large Magellanic Cloud on the Structure of the Milky Way’s Dark Matter Halo Using Basis Function Expansions. *ApJ*, 919(2):109, October 2021. doi: 10.3847/1538-4357/ac0b44.
- Nicolás Garavito-Camargo, Gurtina Besla, Chervin F. P. Laporte, Adrian M. Price-Whelan, Emily C. Cunningham, Kathryn V. Johnston, Martin Weinberg, and Facundo A. Gómez. Quantifying the impact of the large magellanic cloud on the structure of the milky way’s dark matter halo using basis function expansions. *The Astrophysical Journal*, 919(2): 109, sep 2021. doi: 10.3847/1538-4357/ac0b44. URL <https://dx.doi.org/10.3847/1538-4357/ac0b44>.
- Turner Garrow, Jeremy J. Webb, and Jo Bovy. The effects of dwarf galaxies on the orbital evolution of galactic globular clusters. *MNRAS*, 499(1):804–813, November 2020. doi: 10.1093/mnras/staa2773.
- L. Girardi, G. Bertelli, A. Bressan, C. Chio-

- si, M. A. T. Groenewegen, P. Marigo, B. Salasnich, and A. Weiss. Theoretical isochrones in several photometric systems. I. Johnson-Cousins-Glass, HST/WFPC2, HST/NICMOS, Washington, and ESO Imaging Survey filter sets. *aap*, 391:195–212, August 2002. doi: 10.1051/0004-6361:20020612.
- Tobias Goerdt, Ben Moore, J. I. Read, Joachim Stadel, and Marcel Zemp. Does the fornax dwarf spheroidal have a central cusp or core? *Monthly Notices of the Royal Astronomical Society*, 368(3):1073–1077, 2006. doi: 10.1111/j.1365-2966.2006.10182.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2006.10182.x>.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv e-prints*, art. arXiv:1406.2661, June 2014. doi: 10.48550/arXiv.1406.2661.
- F. Governato, B. Willman, L. Mayer, A. Brooks, G. Stinson, O. Valenzuela, J. Wadsley, and T. Quinn. Forming disc galaxies in Λ CDM simulations. *MNRAS*, 374(4):1479–1494, February 2007. doi: 10.1111/j.1365-2966.2006.11266.x.
- F. Governato, C. Brook, L. Mayer, A. Brooks, G. Rhee, J. Wadsley, P. Jonsson, B. Willman, G. Stinson, T. Quinn, and P. Madau. Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows. *nat*, 463(7278):203–206, January 2010. doi: 10.1038/nature08640.
- Robert J. J. Grand, Facundo A. Gómez, Federico Marinacci, Rüdiger Pakmor, Volker Springel, David J. R. Campbell, Carlos S. Frenk, Adrian Jenkins, and Simon D. M. White. The Auriga Project: the properties and formation mechanisms of disc galaxies across cosmic time. *MNRAS*, 467(1): 179–207, May 2017. doi: 10.1093/mnras/stx071.
- Robert J. J. Grand, John Helly, Azadeh Fattahi, Marius Cautun, Shaun Cole, Andrew P. Cooper, Alis J. Deason, Carlos Frenk, Facundo A. Gómez, Jason A. S. Hunt, Federico Marinacci, Rüdiger Pakmor, Christine M. Simpson, Volker Springel, and Dandan Xu. Aurigaia: mock Gaia DR2 stellar catalogues from the AURIGA cosmological simulations. *MNRAS*, 481(2):1726–1743, December 2018. doi: 10.1093/mnras/sty2403.
- A. Halle and F. Combes. Influence of baryonic physics in galaxy simulations: A semi-analytic treatment of the molecular component. *aap*, 559:A55, November 2013. doi: 10.1051/0004-6361/201220952.
- Shichao Han, Weicheng Sun, and Binsong Xiong. Cosmological simulations and models of galaxy formation. *Journal of Physics: Conference Series*, 2200(1):012001, feb 2022. doi: 10.1088/1742-6596/2200/1/012001. URL <https://dx.doi.org/10.1088/1742-6596/2200/1/012001>.
- Y. Hashimoto, Y. Funato, and J. Makino. To Circularize or Not To Circularize?—Orbital Evolution of Satellite Galaxies. *ApJ*, 582:

- 196–201, January 2003. doi: 10.1086/344260.
- Sten Hasselquist, Matthew Shetrone, Verne Smith, Jon Holtzman, Andrew McWilliam, J. G. Fernández-Trincado, Timothy C. Beers, Steven R. Majewski, David L. Nidever, Baitian Tang, Patricia B. Tissera, Emma Fernández Alvar, Carlos Allende Prieto, Andres Almeida, Borja Anguiano, Giuseppina Battaglia, Leticia Carigi, Gloria Delgado Inglada, Peter Frinchaboy, D. A. García-Hernández, Doug Geisler, Dante Minniti, Vinicius M. Placco, Mathias Schultheis, Jennifer Sobeck, and Sandro Villanova. Apogee chemical abundances of the sagittarius dwarf galaxy. *The Astrophysical Journal*, 845(2): 162, aug 2017. doi: 10.3847/1538-4357/aa7ddc. URL <https://dx.doi.org/10.3847/1538-4357/aa7ddc>.
- Amina Helmi and Simon D. M. White. Building up the Stellar Halo of the Milky Way. In Brad K. Gibson, Rim S. Axelrod, and Mary E. Putman, editors, *The Third Stromlo Symposium: The Galactic Halo*, volume 165 of *Astronomical Society of the Pacific Conference Series*, page 89, January 1999a. doi: 10.48550/arXiv.astro-ph/9811108.
- Amina Helmi and Simon D. M. White. Building up the stellar halo of the Galaxy. *MNRAS*, 307(3):495–517, Aug 1999b. doi: 10.1046/j.1365-8711.1999.02616.x.
- Amina Helmi, Carine Babusiaux, Helmer H. Koppelman, Davide Massari, Jovan Veljanoski, and Anthony G. A. Brown. The merger that led to the formation of the Milky Way’s inner stellar halo and thick disk. *nat*, 563(7729):85–88, October 2018. doi: 10.1038/s41586-018-0625-x.
- Tin Kam Ho. Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1, ICDAR ’95*, pages 278–282, M, 1995. IEEE Computer Society. ISBN 0-8186-7128-9.
- E.øg@, C. Fabricius, V. V. Makarov, S. Urban, T. Corbin, G. Wycoff, U. Bastian, P. Schwekendiek, and A. Wicenec. The Tycho-2 catalogue of the 2.5 million brightest stars. *aap*, 355:L27–L30, March 2000.
- Erik Holmberg. On the Clustering Tendencies among the Nebulae. II. a Study of Encounters Between Laboratory Models of Stellar Systems by a New Integration Procedure. *ApJ*, 94:385, November 1941. doi: 10.1086/144344.
- Danny Horta, Ricardo P Schiavon, J Ted Mackereth, David H Weinberg, Sten Hasselquist, Diane Feuillet, Robert W O’Connell, Borja Anguiano, Carlos Allende-Prieto, Rachael L Beaton, Dmitry Bizyaev, Katia Cunha, Doug Geisler, D A García-Hernández, Jon Holtzman, Henrik Jönsson, Richard R Lane, Steve R Majewski, Szabolcs Mészáros, Dante Minniti, Christian Nitschelm, Matthew Shetrone, Verne V Smith, and Gail Zasowski. The chemical characterization of halo substructure in the Milky Way based on APOGEE. *Monthly Notices of the*

- Royal Astronomical Society*, 520(4):5671–5711, 11 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac3179. URL <https://doi.org/10.1093/mnras/stac3179>.
- J. A. S. Hunt and D. Kawata. A PRIMAL view of the Milky Way, made possible by Gaia and M2M modelling. In *EAS Publications Series*, volume 67-68 of *EAS Publications Series*, pages 83–86, July 2014. doi: 10.1051/eas/1567014.
- Jason A. S. Hunt and Daisuke Kawata. Disc galaxy modelling with a particle-by-particle made-to-measure method. *Monthly Notices of the Royal Astronomical Society*, 430(3):1928–1939, 02 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt021. URL <https://doi.org/10.1093/mnras/stt021>.
- Jason A. S. Hunt and Daisuke Kawata. M2M modelling of the Galactic disc via primal: fitting to Gaia error added data. *Monthly Notices of the Royal Astronomical Society*, 443(3):2112–2125, 07 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu1306. URL <https://doi.org/10.1093/mnras/stu1306>.
- Jason A. S. Hunt, Daisuke Kawata, Robert J. J. Grand, Ivan Minchev, Stefano Pasetto, and Mark Cropper. The stellar kinematics of corotating spiral arms in Gaia mock observations. *MNRAS*, 450(2):2132–2142, June 2015. doi: 10.1093/mnras/stv765.
- Rodrigo A. Ibata, Geraint F. Lewis, Anthony R. Conn, Michael J. Irwin, Alan W. McConnachie, Scott C. Chapman, Michelle L. Collins, Mark Fardal, Annette M. N. Ferguson, Neil G. Ibata, A. Dougal Mackey, Nicolas F. Martin, Julio Navarro, R. Michael Rich, David Valls-Gabaud, and Lawrence M. Widrow. A vast, thin plane of corotating dwarf galaxies orbiting the Andromeda galaxy. *nat*, 493(7430):62–65, January 2013. doi: 10.1038/nature11717.
- Rodrigo A. Ibata, Michele Bellazzini, Khyati Malhan, Nicolas Martin, and Paolo Bianchini. Identification of the long stellar stream of the prototypical massive globular cluster ω Centauri. *Nature Astronomy*, 3:667–672, April 2019. doi: 10.1038/s41550-019-0751-x.
- S. Inoue. The test for suppressed dynamical friction in a constant density core of dwarf galaxies. *MNRAS*, 397:709–716, August 2009. doi: 10.1111/j.1365-2966.2009.15066.x.
- Shigeki Inoue. Corrective effect of many-body interactions in dynamical friction. *MNRAS*, 416(2):1181–1190, September 2011. doi: 10.1111/j.1365-2966.2011.19122.x.
- Željko Ivezić and Kahn. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2):111, March 2019. doi: 10.3847/1538-4357/ab042c.
- Kathryn V. Johnston, David N. Spergel, and Lars Hernquist. The Disruption of the Sagittarius Dwarf Galaxy. *ApJ*, 451:598, October 1995. doi: 10.1086/176247.

- A. Just and J.ñarrubia@. Large scale inhomogeneity and local dynamical friction. *aap*, 431(3):861–877, March 2005. doi: 10.1051/0004-6361:20041175.
- D. Kawata, J. A. S. Hunt, R. J. J. Grand, A. Siebert, S. Pasetto, and M. Cropper. Stellar Motion around Spiral Arms: Gaia Mock Data. In *EAS Publications Series*, volume 67-68 of *EAS Publications Series*, pages 247–250, July 2014a. doi: 10.1051/eas/1567044.
- Daisuke Kawata, Jason A. S. Hunt, Robert J. J. Grand, Stefano Pasetto, and Mark Cropper. Gas and stellar motions and observational signatures of corotating spiral arms. *MNRAS*, 443(3):2757–2765, September 2014b. doi: 10.1093/mnras/stu1292.
- Sergey Khrapov, Alexander Khoperskov, and Vladimir Korchagin. Modeling of Spiral Structure in a Multi-Component Milky Way-Like Galaxy. *Galaxies*, 9(2):29, April 2021. doi: 10.3390/galaxies9020029.
- I. King. The structure of star clusters. I. an empirical density law. *aj*, 67:471, October 1962. doi: 10.1086/108756.
- Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv e-prints*, art. arXiv:1312.6114, December 2013. doi: 10.48550/arXiv.1312.6114.
- Anatoly Klypin, Stefan Gottlöber, Andrey V. Kravtsov, and Alexei M. Khokhlov. Galaxies in N-Body Simulations: Overcoming the Overmerging Problem. *ApJ*, 516(2):530–551, May 1999. doi: 10.1086/307122.
- Anatoly Klypin, Stefan Gottlöber, Andrey V. Kravtsov, and Alexei M. Khokhlov. Galaxies in n-body simulations: Overcoming the overmerging problem. *The Astrophysical Journal*, 516(2):530, may 1999. doi: 10.1086/307122. URL <https://dx.doi.org/10.1086/307122>.
- Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, 43:59–69, 1982. doi: 10.1007/BF00337288.
- Andrey V. Kravtsov. On the origin of the global schmidt law of star formation. *The Astrophysical Journal*, 590(1):L1, may 2003. doi: 10.1086/376674. URL <https://dx.doi.org/10.1086/376674>.
- Andrey V. Kravtsov, Anatoly A. Klypin, and Alexei M. Khokhlov. Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations. *apjs*, 111(1):73–94, July 1997. doi: 10.1086/313015.
- William E. Kunkel and Serge Demers. The Magellanic Plane. In *Galaxy and the L.G.*, volume 182, page 241, January 1976.
- Richard B. Larson. Numerical calculations of the dynamics of collapsing proto-star. *MNRAS*, 145:271, January 1969. doi: 10.1093/mnras/145.3.271.
- Jaehyun Lee, Jihye Shin, Owain N. Snaith, Yonghwi Kim, C. Gareth Few, Julien Devriendt, Yohan Dubois, Leah M. Cox, Sungwook E. Hong, Oh-Kyoung Kwon, Chan Park, Christophe Pichon, Juhan Kim, Brad K. Gibson, and Changbom Park. The

- Horizon Run 5 Cosmological Hydrodynamical Simulation: Probing Galaxy Formation from Kilo- to Gigaparsec Scales. *ApJ*, 908(1):11, February 2021. doi: 10.3847/1538-4357/abd08b.
- G.ître@. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59, January 1927.
- Michael Levi and Allen. The Dark Energy Spectroscopic Instrument (DESI). In *Bulletin of the American Astronomical Society*, volume 51, page 57, September 2019. doi: 10.48550/arXiv.1907.10688.
- Ben Lowing, Wenting Wang, Andrew Cooper, Rachel Kennedy, John Helly, Shaun Cole, and Carlos Frenk. Creating mock catalogues of stellar haloes from cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 446(3):2274–2290, 11 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu2257. URL <https://doi.org/10.1093/mnras/stu2257>.
- D. Lynden-Bell. Dwarf galaxies and globular clusters in high velocity hydrogen streams. *MNRAS*, 174:695–710, March 1976. doi: 10.1093/mnras/174.3.695.
- D. Lynden-Bell and A. J. Kalnajs. On the generating mechanism of spiral structure. *MNRAS*, 157:1, 1972. doi: 10.1093/mnras/157.1.1.
- J Ted Mackereth, Ricardo P Schiavon, Joel Pfeffer, Christian R Hayes, Jo Bovy, Borja Anguiano, Carlos Allende Prieto, Sten Hasselquist, Jon Holtzman, Jennifer A Johnson, Steven R Majewski, Robert O’Connell, Matthew Shetrone, Patricia B Tissera, and J G Fernández-Trincado. The origin of accreted stellar halo populations in the Milky Way using APOGEE, Gaia, and the EAGLE simulations. *Monthly Notices of the Royal Astronomical Society*, 482(3):3426–3442, 11 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty2955. URL <https://doi.org/10.1093/mnras/sty2955>.
- J. B. MacQueen. Some methods for classification and analysis of multivariate observations. 1:281–297, 1967. Consultado el 7 de abril de 2009.
- Steven R. et al. Majewski. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *aj*, 154(3):94, September 2017. doi: 10.3847/1538-3881/aa784d.
- P. Marigo, L. Girardi, A. Bressan, M. A. T. Groenewegen, L. Silva, and G. L. Granato. Evolution of asymptotic giant branch stars. II. Optical to far-infrared isochrones with improved TP-AGB models. *aap*, 482(3):883–905, May 2008. doi: 10.1051/0004-6361:20078467.
- Lucio Mayer, Fabio Governato, Monica Colpi, Ben Moore, Thomas Quinn, James Wadsley, Joachim Stadel, and George Lake. Tidal Stirring and the Origin of Dwarf Spheroidals in the Local Group. *apjl*,

- 547(2):L123–L127, February 2001. doi: 10.1086/318898.
- P. J. McMillan and W. Dehnen. Initial conditions for disc galaxies. *MNRAS*, 378:541–550, June 2007. doi: 10.1111/j.1365-2966.2007.11753.x.
- Paul J. McMillan and James J. Binney. Disassembling the Galaxy with angle-action coordinates. *Monthly Notices of the Royal Astronomical Society*, 390(1):429–437, 10 2008. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2008.13767.x. URL <https://doi.org/10.1111/j.1365-2966.2008.13767.x>.
- D. Merritt. Distribution functions for spherical galaxies. *MNRAS*, 214:25P–28P, June 1985. doi: 10.1093/mnras/214.1.25P.
- D. Merritt. Dynamics of Triaxial Stellar Systems. In M. Arnaboldi, G. S. Da Costa, and P. Saha, editors, *The Nature of Elliptical Galaxies; 2nd Stromlo Symposium*, volume 116 of *Astronomical Society of the Pacific Conference Series*, page 32, 1997.
- Alexander I. Merson, Carlton M. Baugh, John C. Helly, Violeta Gonzalez-Perez, Shaun Cole, Richard Bielby, Peder Norberg, Carlos S. Frenk, Andrew J. Benson, Richard G. Bower, Cedric G. Lacey, and Claudia del P. Lagos. Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection. *MNRAS*, 429(1):556–578, February 2013. doi: 10.1093/mnras/sts355.
- Daniel Michalik, Lennart Lindegren, and David Hobbs. The Tycho-Gaia astrometric solution . How to get 2.5 million parallaxes with less than one year of Gaia data. *aap*, 574:A115, February 2015. doi: 10.1051/0004-6361/201425310.
- Tim B. Miller, Frank C. van den Bosch, Sheridan B. Green, and Go Ogiya. Dynamical self-friction: how mass loss slows you down. *MNRAS*, 495(4):4496–4507, July 2020. doi: 10.1093/mnras/staa1450.
- Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997. ISBN 978-0-07-042807-2.
- Oliver Müller, Marcel S. Pawlowski, Helmut Jerjen, and Federico Lelli. A whirling plane of satellite galaxies around Centaurus A challenges cold dark matter cosmology. *Science*, 359(6375):534–537, February 2018. doi: 10.1126/science.aao1858.
- T. Murai and M. Fujimoto. The Magellanic stream and the galaxy with a massive halo. *pasj*, 32:581–603, January 1980.
- G. C. Myeong, N. W. Evans, V. Belokurov, N. C. Amorisco, and S. E. Koposov. Halo substructure in the SDSS-Gaia catalogue: streams and clumps. *MNRAS*, 475(2):1537–1548, April 2018a. doi: 10.1093/mnras/stx3262.
- G. C. Myeong, N. W. Evans, V. Belokurov, J. L. Sanders, and S. E. Koposov. The Milky Way Halo in Action Space. *apjl*, 856(2):L26, April 2018b. doi: 10.3847/2041-8213/aab613.

- J. F. Navarro, C. S. Frenk, and S. D. M. White. The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, May 1996. doi: 10.1086/177173.
- Lina Necib, Bryan Ostdiek, Mariangela Lisanti, Timothy Cohen, Marat Freytsis, and Shea Garrison-Kimmel. Chasing accreted structures within gaia dr2 using deep learning. *The Astrophysical Journal*, 903(1):25, oct 2020. doi: 10.3847/1538-4357/abb814. URL <https://dx.doi.org/10.3847/1538-4357/abb814>.
- Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.
- Robert W. Nelson and Scott Tremaine. Linear response, dynamical friction and the fluctuation dissipation theorem in stellar dynamics. *Monthly Notices of the Royal Astronomical Society*, 306(1):1–21, 06 1999. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1999.02101.x. URL <https://doi.org/10.1046/j.1365-8711.1999.02101.x>.
- Maria Selina Nitschai, Anna-Christina Eilers, Nadine Neumayer, Michele Cappellari, and Hans-Walter Rix. Dynamical model of the milky way using apogee and gaia data. *The Astrophysical Journal*, 916(2): 112, aug 2021. doi: 10.3847/1538-4357/ac04b5. URL <https://dx.doi.org/10.3847/1538-4357/ac04b5>.
- Takashi Okamoto, Vincent R. Eke, Carlos S. Frenk, and Adrian Jenkins. Effects of feedback on the morphology of galaxy discs. *MNRAS*, 363(4):1299–1314, November 2005. doi: 10.1111/j.1365-2966.2005.09525.x.
- Ludwig Oser, Jeremiah P. Ostriker, Thorsten Naab, Peter H. Johansson, and Andreas Burkert. The Two Phases of Galaxy Formation. *ApJ*, 725(2):2312–2323, December 2010. doi: 10.1088/0004-637X/725/2/2312.
- L. P. Osipkov. Spherical systems of gravitating bodies with ellipsoidal velocity distribution. *Pisma v Astronomicheskii Zhurnal*, 5:77–80, February 1979.
- B. Ostdiek, L. Necib, T. Cohen, M. Freytsis, M. Lisanti, S. Garrison-Kimmel, A. Wetzel, R. E. Sanderson, and P. F. Hopkins. Cataloging accreted stars within Gaia DR2 using deep learning. *aap*, 636: A75, April 2020. doi: 10.1051/0004-6361/201936866.
- Yassine Ouali, Céline Hudelot, and Myriam Tami. An Overview of Deep Semi-Supervised Learning. *arXiv e-prints*, art. arXiv:2006.05278, June 2020. doi: 10.48550/arXiv.2006.05278.
- G. Pagnini, P. Di Matteo, S. Khoperskov, A. Mastrobuono-Battisti, M. Haywood, F. Renaud, and F. Combes. The distribution of globular clusters in kinematic spa-

- ces does not trace the accretion history of the host galaxy. *aap*, 673:A86, May 2023. doi: 10.1051/0004-6361/202245128.
- Marcel S. Pawlowski. Phase-Space Correlations among Systems of Satellite Galaxies. *Galaxies*, 9(3):66, September 2021. doi: 10.3390/galaxies9030066.
- Marcel S. Pawlowski and Pavel Kroupa. The Vast Polar Structure of the Milky Way Attains New Members. *ApJ*, 790(1):74, July 2014. doi: 10.1088/0004-637X/790/1/74.
- Marcel S. Pawlowski and Pavel Kroupa. The Milky Way’s disc of classical satellite galaxies in light of Gaia DR2. *MNRAS*, 491(2):3042–3059, January 2020. doi: 10.1093/mnras/stz3163.
- Jorge Peñarrubia, Facundo A. Gómez, Gurtina Besla, Denis Erkal, and Yin-Zhe Ma. A timing constraint on the (total) mass of the Large Magellanic Cloud. *MNRAS*, 456(1):L54–L58, February 2016. doi: 10.1093/mnrasl/slv160.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, October 2011. doi: 10.48550/arXiv.1201.0490.
- P. J. E. Peebles. Structure of the Coma Cluster of Galaxies. *aj*, 75:13, February 1970. doi: 10.1086/110933.
- Bárbara Pichardo, Marco Martos, and Edmundo Moreno. Models for the gravitational field of the galactic bar. an application to stellar orbits in the galactic plane and orbits of some globular clusters. *The Astrophysical Journal*, 609, 02 2004. doi: 10.1086/421008.
- Annalisa Pillepich, Piero Madau, and Lucio Mayer. Building late-type spiral galaxies by in-situ and ex-situ star formation. *The Astrophysical Journal*, 799(2):184, jan 2015. doi: 10.1088/0004-637X/799/2/184. URL <https://dx.doi.org/10.1088/0004-637X/799/2/184>.
- Planck Collaboration. Planck 2013 results. XVI. Cosmological parameters. *aap*, 571:A16, Nov 2014. doi: 10.1051/0004-6361/201321591.
- H. C. Plummer. On the problem of distribution in globular star clusters. *MNRAS*, 71:460–470, March 1911. doi: 10.1093/mnras/71.5.460.
- P. Ramos, T. Antoja, Z. Yuan, A. Arentsen, P. A. Oria, B. Famaey, R. Ibata, C. Ma-

- teu, and J. A. Carballo-Bello. The Sagittarius stream in Gaia Early Data Release 3 and the origin of the bifurcations. *aap*, 666:A64, October 2022. doi: 10.1051/0004-6361/202142830.
- Pau Ramos. Substructure in the phase-space of the galaxy with gaia, 09 2020. <http://hdl.handle.net/10803/670916>.
- J. I. Read, L. Mayer, A. M. Brooks, F. Governato, and G. Lake. A dark matter disc in three cosmological simulations of Milky Way mass galaxies. *Monthly Notices of the Royal Astronomical Society*, 397(1):44–51, 07 2009. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2009.14757.x. URL <https://doi.org/10.1111/j.1365-2966.2009.14757.x>.
- Alexander H. Riley, Azadeh Fattahi, Andrew B. Pace, Louis E. Strigari, Carlos S. Frenk, Facundo A. Gómez, Robert J. J. Grand, Federico Marinacci, Julio F. Navarro, Rüdiger Pakmor, Christine M. Simpson, and Simon D. M. White. The velocity anisotropy of the Milky Way satellite system. *MNRAS*, 486(2):2679–2694, June 2019. doi: 10.1093/mnras/stz973.
- Brant Robertson, Naoki Yoshida, Volker Springel, and Lars Hernquist. Disk galaxy formation in a lambda cold dark matter universe. *The Astrophysical Journal*, 606(1):32, may 2004. doi: 10.1086/382871. URL <https://dx.doi.org/10.1086/382871>.
- A. C. Robin, X. Luri, C.é@, Y. Isasi, E. Grux, S. Blanco-Cuaresma, F. Arenou, C. Babusiaux, M. Belcheva, R. Drimmel, C. Jordi, A. Krone-Martins, E. Masana, J. C. Mauduit, F. Mignard, N. Mowlavi, B. Rocca-Volmerange, P. Sartoretti, E. Slezak, and A. Sozzetti. Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue. *aap*, 543:A100, July 2012. doi: 10.1051/0004-6361/201118646.
- Annie Robin and Michel Crézé. Stellar populations in the milky way—a synthetic model. *Astronomy and Astrophysics (ISSN 0004-6361)*, vol. 157, no. 1, March 1986, p. 71–90., 157:71–90, 1986.
- Santi Roca-Fàbrega, Octavio Valenzuela, Pedro Colín, Francesca Figueras, Yair Krongold, Héctor Velázquez, Vladimir Avila-Reese, and Hector Ibarra-Medel. GA-RROTXA Cosmological Simulations of Milky Way-sized Galaxies: General Properties, Hot-gas Distribution, and Missing Baryons. *ApJ*, 824(2):94, Jun 2016. doi: 10.3847/0004-637X/824/2/94.
- S. A. Rodionov, E. Athanassoula, and N. Ya. Sotnikova. An iterative method for constructing equilibrium phase models of stellar systems. *MNRAS*, 392(2):904–916, January 2009. doi: 10.1111/j.1365-2966.2008.14110.x.
- M.ómez@, F. Figueras, T. Antoja, H. Abedi, and L. Aguilar. The analysis of realistic stellar Gaia mock catalogues - I. Red clump stars as tracers of the central bar. *MNRAS*, 447(1):218–233, February 2015. doi: 10.1093/mnras/stu2457.

- Mathieu Rouaud. *Probability, Statistics and Estimation Chapter 2: Linear Regression, Linear Regression with Error Bars and Nonlinear Regression*. 2013.
- Edwin E. Salpeter. The Luminosity Function and Stellar Evolution. *ApJ*, 121:161, January 1955. doi: 10.1086/145971.
- A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959. doi: 10.1147/rd.33.0210.
- Isaiah B. Santistevan, Andrew Wetzel, Erik Tollerud, Robyn E. Sanderson, and Jenna Samuel. Orbital dynamics and histories of satellite galaxies around Milky Way - mass galaxies in the FIRE simulations. *MNRAS*, 518(1):1427–1447, January 2023. doi: 10.1093/mnras/stac3100.
- David J. Schlegel, Douglas P. Finkbeiner, and Marc Davis. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *ApJ*, 500(2):525–553, June 1998. doi: 10.1086/305772.
- J. A. Sellwood. The art of N-body building. *ApJ*, 25:151–186, January 1987. doi: 10.1146/annurev.aa.25.090187.001055.
- Marcin Semczuk, Ewa L. Łokas, and Andrés del Pino. Tidal origin of spiral arms in galaxies orbiting a cluster. *The Astrophysical Journal*, 834(1):7, dec 2016. doi: 10.3847/1538-4357/834/1/7. URL <https://dx.doi.org/10.3847/1538-4357/834/1/7>.
- Sanjib Sharma, Joss Bland-Hawthorn, Kathryn V. Johnston, and James Binney. Galaxia: A Code to Generate a Synthetic Survey of the Milky Way. *ApJ*, 730(1):3, March 2011. doi: 10.1088/0004-637X/730/1/3.
- Joseph Silk and Michael S. Turner. Double inflation. *Phys. Rev. D*, 35:419–428, Jan 1987. doi: 10.1103/PhysRevD.35.419. URL <https://link.aps.org/doi/10.1103/PhysRevD.35.419>.
- D. N. Spergel, R. Bean, O.é@, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *apjs*, 170(2):377–408, June 2007. doi: 10.1086/513700.
- Volker Springel. The cosmological simulation code GADGET-2. *MNRAS*, 364(4):1105–1134, December 2005. doi: 10.1111/j.1365-2966.2005.09655.x.
- Matthias Steinmetz, Gal Matijevič, Harry Enke, Tomaz Zwitter, Guillaume Guiglion, Paul J. McMillan, Georges Kordopatis, Marica Valentini, Cristina Chiappini, Luca Casagrande, Jennifer Wojno, Borja Anguiano, Olivier Bienaymé, Albert Bijaoui, James Binney, Donna Burton, Paul Cass, Patrick de Laverny, Kristin Fiegert, Kenneth Freeman, Jon P. Fulbright, Brad K.

- Gibson, Gerard Gilmore, Eva K. Grebel, Amina Helmi, Andrea Kunder, Ulisse Munari, Julio F. Navarro, Quentin Parker, Gregory R. Ruchti, Alejandra Recio-Blanco, Warren Reid, George M. Seabroke, Alessandro Siviero, Arnaud Siebert, Milorad Stupar, Fred Watson, Mary E. K. Williams, Rosemary F. G. Wyse, Friedrich Anders, Teresa Antoja, Danijela Birkko, Joss Bland-Hawthorn, Diego Bossini, Rafael A. García, Ismael Carrillo, William J. Chaplin, Yvonne Elsworth, Benoit Famaey, Ortwin Gerhard, Paula Jofre, Andreas Just, Savita Mathur, Andrea Miglio, Ivan Minchev, Giacomo Monari, Benoit Mosser, Andreas Ritter, Thaise S. Rodrigues, Ralf-Dieter Scholz, Sanjib Sharma, Kseniia Sysoliatina, and RAVE Collaboration. The Sixth Data Release of the Radial Velocity Experiment (RAVE). I. Survey Description, Spectra, and Radial Velocities. *aj*, 160(2):82, August 2020. doi: 10.3847/1538-3881/ab9ab9.
- Louis E. Strigari, Carlos S. Frenk, and Simon D. M. White. Kinematics of Milky Way satellites in a Lambda cold dark matter universe. *MNRAS*, 408(4):2364–2372, November 2010. doi: 10.1111/j.1365-2966.2010.17287.x.
- D. Syer and S. Tremaine. Made-to-measure N-body systems. *MNRAS*, 282(1):223–233, September 1996. doi: 10.1093/mnras/282.1.223.
- Tomas Tamfal, Lucio Mayer, Thomas R. Quinn, Pedro R. Capelo, Stelios Kazantzidis, Arif Babul, and Douglas Potter. Revisiting dynamical friction: The role of global modes and local wakes. *The Astrophysical Journal*, 916(1):55, jul 2021. doi: 10.3847/1538-4357/ac0627. URL <https://dx.doi.org/10.3847/1538-4357/ac0627>.
- E. Tempel, R. Kipper, A. Tamm, M. Gramann, M. Einasto, T. Sepp, and T. Tuvikene. Friends-of-friends galaxy group finder with membership refinement. Application to the local Universe. *aap*, 588:A14, April 2016. doi: 10.1051/0004-6361/201527755.
- The LSST Dark Energy Science Collaboration, Rachel Mandelbaum, Tim Eifler, Renée Hložek, Thomas Collett, Eric Gawiser, Daniel Scolnic, David Alonso, Humna Awan, Rahul Biswas, Jonathan Blazek, Patricia Burchat, Nora Elisa Chisari, Ian Dell’Antonio, Seth Digel, Josh Friedman, Daniel A. Goldstein, Isobel Hook, Željko Ivezić, Steven M. Kahn, Sowmya Kamath, David Kirkby, Thomas Kitching, Elisabeth Krause, Pierre-François Leget, Philip J. Marshall, Joshua Meyers, Hiro-ao Miyatake, Jeffrey A. Newman, Robert Nichol, Eli Rykoff, F. Javier Sanchez, Anže Slosar, Mark Sullivan, and M. A. Troxel. The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document. *arXiv e-prints*, art. arXiv:1809.01669, September 2018. doi: 10.48550/arXiv.1809.01669.
- J. Tolles and W. J. Meurer. Logistic regression: Relating patient characteristics to outcomes. *JAMA*, 316(5):533–

- 534, 2016. doi: 10.1001/jama.2016.7653. URL <https://doi.org/10.1001/jama.2016.7653>.
- S. Tremaine and M. D. Weinberg. Dynamical friction in spherical systems. *MNRAS*, 209:729–757, August 1984. doi: 10.1093/mnras/209.4.729.
- S. D. Tremaine. The effect of dynamical friction on the orbit of the Magellanic Clouds. *ApJ*, 203:72–74, January 1976. doi: 10.1086/154049.
- Michael S. Turner. Why is the temperature of the universe 2.726 kelvin? *Science*, 262(5135):861–867, 1993. doi: 10.1126/science.262.5135.861. URL <https://www.science.org/doi/abs/10.1126/science.262.5135.861>.
- Frank C van den Bosch and Go Ogiya. Dark matter substructure in numerical simulations: a tale of discreteness noise, runaway instabilities, and artificial disruption. *Monthly Notices of the Royal Astronomical Society*, 475(3):4066–4087, 01 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty084. URL <https://doi.org/10.1093/mnras/sty084>.
- V. Vapnik and A. Chervonenkis. *Theory of Pattern Recognition [in Russian]*. Nauka, Moscow, 1974. (German Translation: W. Wapnik & A. Tscherwonenkis, *Theorie der Zeichenerkennung*, Akademie-Verlag, Berlin, 1979).
- Dragica Vasileska, Katerina Raleva, and Stephen Goodnick. *Monte Carlo Device Simulations*. 02 2011. ISBN 978-953-307-691-1. doi: 10.5772/16190.
- Eugene Vasiliev. The Effect of the LMC on the Milky Way System. *Galaxies*, 11(2):59, April 2023. doi: 10.3390/galaxies11020059.
- Mark Vogelsberger, Shy Genel, Volker Springel, Paul Torrey, Debora Sijacki, Dandan Xu, Greg Snyder, Dylan Nelson, and Lars Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS*, 444(2):1518–1547, October 2014. doi: 10.1093/mnras/stu1536.
- Mark Vogelsberger, Federico Marinacci, Paul Torrey, and Ewald Puchwein. Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66, January 2020. doi: 10.1038/s42254-019-0127-2.
- S. von Hoerner. Die numerische Integration des n-Körper-Problemes für Sternhaufen. I. *zap*, 50:184–214, January 1960.
- Ward von Winterfeldt, Detlof; Edwards. *Decision Analysis and Behavioral Research*. Cambridge University Press, New York, 1986. ISBN ISBN 0-521-27304-8.
- J. W. Wadsley, J. Stadel, and T. Quinn. Gasoline: a flexible, parallel implementation of TreeSPH. *na*, 9(2):137–158, February 2004. doi: 10.1016/j.newast.2003.08.004.
- D. Walsh, R. F. Carswell, and R. J. Weymann. 0957+561 A, B: twin quasistellar objects or gravitational lens? *nat*, 279:381–384, May 1979. doi: 10.1038/279381a0.

- Liang Wang, Aaron A. Dutton, Gregory S. Stinson, Andrea V. Macciò, Camilla Penzo, Xi Kang, Ben W. Keller, and James Wadsley. NIHAO project - I. Reproducing the inefficiency of galaxy formation across cosmic time with a large sample of cosmological hydrodynamical simulations. *MNRAS*, 454(1):83–94, November 2015. doi: 10.1093/mnras/stv1937.
- David H. Weinberg, James S. Bullock, Fabio Governato, Rachel Kuzio de Naray, and Annika H. G. Peter. Cold dark matter: Controversies on small scales. *Proceedings of the National Academy of Sciences*, 112(40):12249–12255, 2015. doi: 10.1073/pnas.1308716112. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1308716112>.
- Andrew Wetzel, Christopher C. Hayward, Robyn E. Sanderson, Xiangcheng Ma, Daniel Anglés-Alcázar, Robert Feldmann, T. K Chan, Kareem El-Badry, Coral Wheeler, Shea Garrison-Kimmel, Farnik Nikakhtar, Nondh Panithanpaisal, Arpit Aroora, Alexander B. Gurvich, Jenna Samuel, Omid Sameie, Viraj Pandya, Zachary Hafen, Cameron Hummels, Sarah Loebman, Michael Boylan-Kolchin, James S. Bullock, Claude-André Faucher-Giguère, Dušan Kereš, Eliot Quataert, and Philip F. Hopkins. Public data release of the fire-2 cosmological zoom-in simulations of galaxy formation. *The Astrophysical Journal Supplement Series*, 265(2):44, mar 2023. doi: 10.3847/1538-4365/acb99a. URL <https://dx.doi.org/10.3847/1538-4365/acb99a>.
- Andrew R. Wetzel, Philip F. Hopkins, Ji-hoon Kim, Claude-André Faucher-Giguère, Dušan Kereš, and Eliot Quataert. Reconciling Dwarf Galaxies with Λ CDM Cosmology: Simulating a Realistic Population of Satellites around a Milky Way-mass Galaxy. *apjl*, 827(2):L23, August 2016. doi: 10.3847/2041-8205/827/2/L23.
- S. D. M. White. The dynamics of rich clusters of galaxies. *MNRAS*, 177:717–733, December 1976. doi: 10.1093/mnras/177.3.717.
- S. D. M. White and M. J. Rees. Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, May 1978. doi: 10.1093/mnras/183.3.341.
- G. Yepes, R. Kates, A. Khokhlov, and A. Klypin. Hydrodynamical simulations of galaxy formation: effects of supernova feedback. *MNRAS*, 284(1):235–256, January 1997. doi: 10.1093/mnras/284.1.235.
- Jiawei Zhang. Gradient Descent based Optimization Algorithms for Deep Learning Models Training. *arXiv e-prints*, art. arXiv:1903.03614, March 2019. doi: 10.48550/arXiv.1903.03614.
- Zhongheng Zhang. Introduction to machine learning: k-nearest neighbors. *Annals of Translational Medicine*, 4(11), 2016. ISSN 2305-5847. URL <https://atm.amegroups.org/article/view/10170>.

- Gang Zhao, Yong-Heng Zhao, Yao-Quan Chu, Yi-Peng Jing, and Li-Cai Deng. LA-MOST spectral survey — An overview. *Research in Astronomy and Astrophysics*, 12(7):723–734, July 2012. doi: 10.1088/1674-4527/12/7/002.
- Dengyong Zhou, Olivier Bousquet, Thomas Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003. URL https://proceedings.neurips.cc/paper_files/paper/2003/file/87682805257e619d49b8e0dfdc14affa-Paper.pdf.
- Xiaojin Zhu. Learning from labeled and unlabeled data with label propagation, 2002. CiteSeerX 10.1.1.14.3864.
- Xiaojin Zhu. Semi-supervised learning literature survey. *University of Wisconsin-Madison Department of Computer Sciences*, 2007. Page 5.
- Adi Zolotov, Beth Willman, Alyson M. Brooks, Fabio Governato, Chris B. Brook, David W. Hogg, Tom Quinn, and Greg Stinson. The Dual Origin of Stellar Halos. *ApJ*, 702(2):1058–1067, September 2009. doi: 10.1088/0004-637X/702/2/1058.
- Adi Zolotov, Beth Willman, Alyson M. Brooks, Fabio Governato, David W. Hogg, Sijing Shen, and James Wadsley. The dual origin of stellar halos. ii. chemical abundances as tracers of formation history. *The Astrophysical Journal*, 721(1): 738, sep 2010. doi: 10.1088/0004-637X/721/1/738. URL <https://dx.doi.org/10.1088/0004-637X/721/1/738>.

Índice de figuras

1.1.	Imagen tomada de Pillepich et al. (2015) . Proyección del campo de densidad estelar de la simulación ERIS. Panel izquierdo: Todas las partículas estelares. Panel medio: Partículas estelares formadas “Ex-situ”. Panel derecho: Partículas estelares formadas “In-situ”.	17
1.2.	Imagen de una simulación mostrando el origen de las poblaciones estelares de distintas galaxias fusionándose (O. Valenzuela comunicación privada). En esta imagen podemos ver en distintos colores el origen de las partículas estelares contenidas en el halo estelar de una galaxia tipo Vía Láctea.	17
2.1.	La estela de CH, y la respuesta global son claramente visibles en los tres paneles. En color morado valores para la sobredensidad negativa y rojo para la positiva. La fila superior corresponde a un modelo de Plummer (Plummer, 1911), panel de en medio e inferior corresponden a diferentes modelos NFW. Imagen: Aguilar-Argüello et al. (2022)	24
2.2.	Contraste de densidad en 3D de la respuesta del halo de la Via Lactea en presencia de la LMC. Imagen: Garavito-Camargo et al. (2021)	25
2.3.	Órbita de cada uno de los 5 cúmulos globulares de Inoue (2009) , cada panel muestra las variaciones del radio orbital de cada uno pero en una situación de caída simultánea. Hay que observar los cambios abruptos en las órbitas en este caso, en el que hay varios perturbadores cayendo al mismo tiempo. Imagen tomada de Inoue (2009)	26
2.4.	Imagen tomada de Goerd et al. (2006) : Distancia radial de cinco cúmulos globulares desde el centro de su halo como función del tiempo para una galaxia tipo fornax en un halo tipo “cusp”.	26
2.5.	Imagen tomada de (Pawlowski, 2021). Galaxias enanas alrededor de la VL. La VL vista de canto como una línea negra en el centro, áreas grises indican una region de +/- 12° alrededor del plano del disco. La vista está orientada a la VPOS. Los triángulos blancos marcan galaxias que probablemente no pertenecen a la VPOS.	28

- 3.1. Representación gráfica del algoritmo Barnes-Hut en 2D. A la izquierda, distribución de partículas en un mallado cuadrado, mostrando cómo se agrupan en diferentes celdas según su densidad. A la derecha, se ilustra la estructura del cuad-árbol, con hojas y nodos desplegándose según la distribución de las partículas. Las zonas con más partículas generan niveles más profundos en el árbol. 33
- 4.1. Estas figuras representan tanto el perfil de densidad como la curva de velocidad circular para el halo de materia oscura que se utiliza como condición inicial de las simulaciones de N-cuerpos. 44
- 4.2. Hundimiento de satélites individuales en un halo con masa de la Vía Láctea. Línea azul: modelo fiduciario SA como un logaritmo de Coulomb sin dependencia espacial; Amarillo: SA usando un logaritmo de Coulomb con dependencia radial; Magenta: SA usando un logaritmo de Coulomb con dependencia radial y una pérdida de masa; Negro: modelo NB con un satélite representado con un perfil de densidad Plummer suavizado rígido (sistema rígido, de ahora en adelante), $\Lambda = \frac{b_{\max}}{b_{90^\circ}} \approx 5$, siguiendo Binney and Tremaine (1987) ec. 8.1b; Rojo: modelo NB que incluye un satélite con un perfil de densidad NFW compacto; Verde: modelo NB que incluye un satélite con un perfil de densidad NFW extendido (para más detalles sobre los modelos, ver la tabla 4.1). La condición inicial para el satélite en los modelos es $r = 126$ kpc, $v_c = 183$ cm/s. 45
- 4.3. El panel superior representa la gráfica de r vs t , para cada uno de los satélites en el experimento en el que los satélites se encuentran inicialmente a diferentes distancias. La notación dd indica que los satélites están a diferente distancia. El panel inferior muestra un esquema de la posición de cada satélite con su respectivo color, las líneas azules indican la dirección del vector velocidad. La membrana azul es una representación para remarcar la presencia de las partículas del halo (no a escala). 47
- 4.4. Error relativo de la energía a lo largo de la simulación en el tiempo para el experimento de cuatro satélites vivos extendidos en configuración coplanar y a diferente distancia inicial. 48

4.5. Experimento de cuatro satélites coplanares cayendo hacia el pozo de potencial en órbita circular. El panel superior izquierdo representa cuatro satélites rígidos cayendo hacia el pozo de potencial del halo masivo. El panel superior derecho representa cuatro satélites compactos. Panel inferior izquierdo: representa cuatro satélites vivos. El panel inferior derecho representa el cambio de la masa a lo largo de la simulación para los experimentos con satélites vivos.	49
4.6. Esquema de la posición de cada satélite en el experimento con 4 satélites coplanares a la misma distancia inicialmente, los satélites se marcan con su respectivo color, las líneas azules indican la dirección del vector velocidad de cada satélite. La membrana azul es una representación para remarcar la presencia de las partículas del halo (no está a escala).	50
4.7. Error relativo en energía para el experimento con cuatro satélites coplanares equidistantes inicialmente.	51
4.8. El panel superior representa la gráfica de r vs t para cada uno de los satélites en el experimento en el que el movimiento de los satélites no ocurre en el mismo plano. La notación dp indica que las órbitas de los satélites están orientadas en diferentes planos. El panel inferior muestra un esquema de la posición de cada satélite con su respectivo color, las líneas azules indican la dirección del vector velocidad. La membrana azul es una representación para remarcar la presencia de las partículas del halo (no está a escala).	52
4.9. Error relativo en energía para el experimento con cuatro satélites orientados en diferentes planos inicialmente.	53
4.10. El panel superior representa la gráfica de r vs t para cada uno de los satélites en el experimento en el que el movimiento de los satélites no ocurre en el mismo plano, los satélites se encuentran a diferentes distancias desde el inicio de la simulación. La notación dp indica que las órbitas de los satélites están orientadas en diferentes planos, dd indica que los satélites se encuentran inicialmente a diferentes distancias. El panel inferior muestra un esquema de la posición de cada satélite con su respectivo color, las líneas azules indican la dirección del vector velocidad. La membrana azul es una representación para remarcar la presencia de las partículas del halo y no está a escala.	54
4.11. Error relativo en energía para el experimento con cuatro satélites no coplanares ni equidistantes.	55

5.1.	Imagen comparativa de como se desempeñan DBSCAN y K-Means en distintos conjuntos de datos con diferentes modos de agruparse. Crédito de la imagen: Pedregosa et al. (2011)	72
5.2.	Un punto “p” es densidad alcanzable desde un punto “q” c.r.a “eps”, $minPts$, si existe una cadena de puntos P_1, \dots, P_n donde $P_1=q$ y $P_n=p$ de modo que P_{i+1} es densidad alcanzable directamente desde el punto P_i	73
5.3.	Un punto “p” es conectado en densidad a un punto “q” c.r.a “eps”, $minPts$, si existe un punto “O” de modo que “p” y “q” son densidad alcanzables desde “O”.	73
6.1.	Selección de las partículas que conforman la muestra a analizar con DBSCAN, las regiones seleccionadas están acotadas por líneas negras sólidas. Izq: Partículas etiquetadas del satélite y el halo en la región de la partículas del satélite. Centro: Selección o corte del espacio en el plano xy que contiene las partículas del satélite. Der: Selección o corte en el plano xz que contiene las partículas del satélite. Las partículas del satélite son puntos rojos, las partículas del halo puntos azules atenuados en la imagen para efectos de visualización. . .	80
6.2.	84
6.3.	84
6.4.	84
6.5.	Selección de las partículas que conforman la muestra a analizar con DBSCAN, tomando una región coniforme simulando el cono de luz. Izq: Selección o corte en la región del cono que contiene las partículas del satélite visto desde el plano xy. Centro: Selección o corte del espacio en el cono visto desde el plano xz . Der: Selección o corte en el espacio del cono visto desde el plano xz. Las partículas está etiquetadas por su pertenencia al satélite con puntos rojos y las partículas del halo con puntos azules.	84
6.6.	Histograma de distribución de vecinos en función de la distancia. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN.	86
6.7.	Sup izq: Grupos encontrados por DBSCAN en un corte del espacio de velocidades vx vs vz. Sup. der: Grupos encontrados por DBSCAN en un corte del espacio x vs z. Inf izq: Partículas del grupo 0 conformado en mayor parte por partículas del satélite. Inf derecha: Grupo encontrado por DBSCAN, compuesto en su mayor parte por partículas del halo, se considera mayormente como ruido.	88

- 6.8. Izq: Histograma de distribución de vecinos en función de la distancia para el experimento con $v_x+v_y+v_z$ +densidad. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN. Der: Amplificación del histograma. 89
- 6.9. Sup izq: Grupos encontrados por DBSCAN en un corte del espacio de velocidades v_r vs v_{tan} . Sup. der: Grupos encontrados por DBSCAN en un corte del espacio x vs z . Paneles inferiores: a) Ruido b) grupo 0 c) grupo 1. . . . 90
- 6.10. Izq: Histograma de distribución de vecinos en función de la distancia para el experimento con $v_{rad}+v_{tan}+\rho$, la roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN. Der: Amplificación del histograma. 92
- 6.11. Sup izq: Grupos encontrados por DBSCAN en un corte del espacio de velocidades v_x vs v_z . Sup. der: Grupos encontrados por DBSCAN en un corte del espacio x vs z . Inf izq: Partículas del grupo 0 conformado en mayor parte por partículas del satélite. Inf derecha: Grupo encontrado por DBSCAN, compuesto en su mayor parte por partículas del halo, se considera mayormente como ruido. 93
- 6.12. Izq: Histograma de distribución de vecinos en función de la distancia para el experimento con solo velocidades. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN. Der: Amplificación del histograma. 94
- 6.13. Superior: Grupos encontrados por DBSCAN búsqueda en el cono, visto en un corte del espacio de velocidades v_x vs v_z . Inf: Grupos individuales encontrados por DBSCAN. 95
- 6.14. Izq: Histograma de distribución de vecinos en función de la distancia para el experimento en una zona coniforme utilizando las variables $v_x v_y v_z$ +densidad. La línea roja indica el bin de distancia donde se separan las partículas del satélite y del halo y que define el valor ϵ (eps) que se utilizará como hiperparámetro del algoritmo DBSCAN. Der: Amplificación del histograma. . . 96
- 6.15. Izq: Histograma de distribución de vecinos para el experimento con $v_x+v_y+v_z$ +densidad. Der: Histograma de distribución de vecinos para el experimento con v_r+v_{tan} +densidad. Observar que la señal tiene una forma distinta para cada caso. 97
- 6.16. Inspección de los tres casos de combinación de variables de velocidad para el experimento con velocidades+densidad. 97

- 6.17. Izq: Histograma de distribución de vecinos para solo el satélite del experimento con $v_{\text{tan}} + \text{densidad}$. Der: Histograma de distribución de vecinos para el experimento con $v_r + v_{\text{tan}} + \text{densidad}$ con todas las partículas (Halo+satélite). 98
- 6.18. Esta figura representa lo que DBSCAN clasifica como grupos al pasarle solo las partículas del satélite, sin las del halo, en el espacio de velocidad radial y velocidad tangencial. Los paneles superiores representan los grupos encontrados en el espacio de $v_{\text{rad}} + v_{\text{tan}}$ lado izquierdo y en el espacio de configuración x vs z lado derecho. Los paneles inferiores representan como se ve cada grupo en el espacio de configuración. 98
- 6.19. Izq: Histograma de distribución de vecinos en función de la distancia para el experimento solo utilizando las variables $v_{\text{rad}} + \text{densidad}$. Der: Histograma de distribución de vecinos en función de la distancia para el experimento solo utilizando las variables $v_{\text{tan}} + \text{densidad}$ 99
- 7.1. Modelos de evolución galáctica. Izq. Colapso Monolítico: El gas se enfría primero y colapsa en el halo oscuro. El gas se acomoda en el disco. Una galaxia espiral se forma con estrellas, gas y polvo en el disco. Si el gas primordial se convierte en estrellas prontamente, se forma una galaxia elíptica en lugar de una galaxia espiral. Der: Modelo jerárquico: El gas enfriándose y colapsando es seguido por la formación de los primeros discos. Varias de estas protogalaxias surgen a través de fusiones de dos galaxias espirales. El gas residual puede acomodarse en un disco delgado. Halo oscuro (Azul), gas (Amarillo), estrellas (rojo). Crédito de imagen y texto: Bertola (2002) . . . 102
- 7.2. Crédito de la imagen: (Vogelsberger et al., 2020). Representaciones visuales de simulaciones recientes de formación de estructuras y galaxias seleccionadas. Algunas son de gran volumen, lo que proporciona una amplia visión estadística de las galaxias, mientras que otras son de zoom-in y resuelven detalles más pequeños. También hay dos tipos de simulaciones: las de materia oscura (N-cuerpos) y las hidrodinámicas, que incluyen la materia oscura y los bariones. Las simulaciones de materia oscura han convergido en predicciones sobre la distribución a gran escala y la estructura de los halos de materia oscura. Las simulaciones hidrodinámicas coinciden sorprendentemente bien con observaciones reales en cuanto a la formación de galaxias, pero algunos detalles aún dependen de cómo se implementa la física de los bariones en el modelo. 104

7.3.	Densidad superficial de materia oscura (izquierda) y estelar de la simulación GARROTXA G.321, at $z = 0$. (derecha). Los círculos indican la posición y radio del virial de los halos que contienen estrellas, es decir, de la galaxia central y de sus galaxias satélites, obtenidos usando el buscador de halos Rockstar”(Behroozi et al., 2013).	108
7.4.	Proyecciones de la densidad superficial de gas en la galaxia principal del modelo GARROTXA. La resolución espacial de ~ 100 pc en el disco galáctico permite capturar la formación de estructuras espirales y las corrientes frías de gas que alimentan la formación estelar en la galaxia.	109
8.1.	Diagrama de flujo que explica de manera resumida el esquema para crear observaciones sintéticas a partir de las simulaciones	118
8.2.	El error relativo en la paralaje para las estrellas de la simulación remuestreada (izquierda) para estrellas más cercanas al plano galáctico, $ z < 300$ pc y más allá de $ z > 300$ pc (derecha). El sol se encuentra en la posición central debido a que durante el remuestreo la simulación esta centrada en la posición solar a diferencia de los mapas mostrados en los paneles superiores de la figura 8.3 Romero-Gómez et al. (2015).	120
8.3.	Imagen tomada directamente de Romero-Gómez et al. (2015) para efectos de comparación del modelo de errores de Gaia. El error relativo promedio en paralaje para las muestras RC-G20 (izquierda) y RC-RVS (derecha) para estrellas más cercanas al plano galáctico, $ z < 300pc$ (arriba) y más allá de $ z > 300pc$ (abajo). La estrella amarilla en todos los paneles muestra la posición del Sol, mientras que la línea sólida negra marca la posición y longitud de la barra galáctica.	121
8.4.	Izq: Diagrama color magnitud pesado por error en velocidad radial de acuerdo al modelo de errores de DESI. Der: Mismo diagrama construido directamente de las estrellas remuestreadas de la simulación. Las partes más tenues son consistentes con la densidad de estrellas en cada región del mapa que siguen la función de selección, como se muestra en la figura 8.5	122
8.5.	Mapa de densidad numérica de estrellas en el diagrama color magnitud construido a partir de las estrellas remuestreadas de la simulación.	123
8.6.	Comparación de los perfiles de densidad (en escala logarítmica) y también de la dispersión de velocidad total (radial+tangencial) tanto del sistema original como del sistema generado a través del remuestreo. Izquierda: Perfil de densidad. Derecha: Dispersion de velocidad total (radial+tangencial)	124
8.7.	Razones de ejes: semi-eje mayor c/a y semi-eje menor b/a entre el sistema original y el nuevo sistema generado, que muestran la triaxialidad del sistema. 125	

10.1. Ejemplo de como se ve la subestructura en el espacio fase a través de las simulaciones.	134
10.2. Para una de las simulaciones aplicamos el esquema “Cloud-in-cell” (Birdsall and Fuss, 1997) sobre una de las simulaciones con un perturbador o satélite vivo para detectar las diferentes partes de la respuesta que ocasiona. Izquierda: En la imagen se aprecian la respuesta tanto local y global. Derecha: Aplicación del esquema “Wavelets” implementado por P. Ramos para el mismo experimento.	135
D.1. Resumen visual del algoritmo de asignación partícula-halo. Crédito: (Behroozi, Wechsler, and Wu, 2013)	149

Índice de cuadros

4.1. Los modelos del halo y los satélites utilizados, incluyendo el caso de satélite rígido, satélite vivo compacto (cúmulo globular), el satélite de extensión intermedia (galaxia tipo enana), el satélite vivo extendido (galaxia temprana). De arriba a abajo: la masa total (M_{200}), el número de partículas (N_p), el radio de escala (r_s) del perfil de densidad tipo NFW, el radio a media masa ($r_{1/2}, M_{1/2}$), el radio de truncamiento (R_{tr}) y la longitud de suavizamiento (ϵ). Para el satélite rígido $M_{200}/M_{\odot} = 1 \times 10^{10}$, $\epsilon = 0.6$ kpc.	44
4.2. Condiciones iniciales para cada satélite en un experimento con 4 satélites cayendo hacia un halo masivo, estos 4 satélites coexisten en el mismo plano pero inicialmente se encuentran a diferentes distancias.	48
4.3. Condiciones iniciales para cada satélite en un experimento con 4 satélites cayendo hacia un halo masivo, solo que estos 4 satélites tienen sus vectores velocidad orientados en diferentes direcciones o planos.	53
4.4. Condiciones iniciales para cada satélite en un experimento con 4 satélites cayendo hacia un halo masivo, solo que estos 4 satélites tienen sus vectores velocidad orientados en diferentes direcciones y planos e inicialmente se encuentran a diferentes distancias.	55
6.1. Número de partículas del halo y del satélite presentes en el conjunto de datos final en el plano que contiene a las partículas del satélite.	80
6.2. Número de partículas del halo y del satélite presentes en el conjunto de datos final en el cono que contiene a las partículas del satélite.	84
6.3. Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera para el experimento $v_x+v_y+v_z+\rho$	89
6.4. Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera para el experimento con velocidad+densidad.	91
6.5. Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera para el experimento con $v_x+v_y+v_z$	93

6.6. Cantidad de partículas del halo y del satélite en diferentes grupos, separadas por puntos core y frontera.	95
7.1. Parametros de las simulaciones GARROTXA G.321, a $z=0$. Para más detalles ver Tabla 1 de (Roca-Fàbrega et al., 2016)	109
7.2. Resolución mínima dentro del conjunto de simulaciones Latte del proyecto FIRE2 (Wetzel et al., 2023), de las simulaciones dentro del proyecto Auriga (Grand et al., 2017) y del proyecto Garrotxa-HART y GARROTXA-RAMSES (Roca-Fàbrega et al., 2016)	110

... Hay un lugar allí que está alejado
de Belcebú por un inmenso hueco;
no se ve, pero se oye un riachuelo
que fluye por la brecha de una roca
que él mismo ha erosionado con su curso
y en círculos desciende suavemente.

Mi guía y yo por escondida senda
fuimos para volver al mundo claro;
y sin pensar siquiera en descansar,
subimos, él primero y yo después,
y por un agujero vi las cosas
bellas del cielo. Y por allí salimos
a contemplar de nuevo las estrellas.

—*Infierno*, Canto XXXIV,

Divina Comedia—

