



UNIVERSIDAD NACIONAL DE LA PLATA
Facultad de Ciencias Astronómicas y Geofísicas

**CREACIÓN DE CATÁLOGOS DE GALAXIAS
SIMULADAS PARA GRANDES RELEVAMIENTOS
EXTRAGALÁCTICOS**

Lic. Cristian Antonio Vega Martínez

Directora: Dra. Sofía A. Cora
Codirector: Dr. Nelson D. Padilla

Tesis presentada para optar por el grado de
DOCTOR EN ASTRONOMÍA

La Plata, marzo 2017

Revisión

Esta Tesis ha sido evaluada por un Jurado examinador compuesto por:

Dra. Claudia Scóccola (FCAG – UNLP),

Dr. Daniel D. Carpintero (FCAG – UNLP), y

Dr. Mario Abadi (OAC – Córdoba).

El Jurado ha revisado en detalle el contenido incluido, ha corregido su presentación y aprueba la publicación de esta Tesis.

Resumen

Los grandes relevamientos extragalácticos han sido una pieza fundamental en nuestra comprensión de la cosmología y la formación de galaxias en las últimas décadas. Actualmente, con el advenimiento de proyectos revolucionarios del conocimiento como *Euclid* y el *Gran Telescopio para Rastreos Sinópticos* (LSST), las simulaciones cosmológicas ocupan un rol fundamental tanto en el diseño de estos relevamientos como en la interpretación de los datos que serán recolectados. Estas simulaciones son requeridas para establecer las predicciones teóricas, para generar catálogos que permitan cuantificar efectos observacionales sistemáticos y de selección, para realizar análisis de datos y para optimizar las estrategias de observación. Entre los tipos de simulaciones utilizadas para comprender la evolución de las galaxias, los modelos semianalíticos de formación y evolución de galaxias adquieren una importancia estratégica dada su ventaja en tiempo de cálculo y predictibilidad teórica.

En esta Tesis nos enfocamos en el código semianalítico SAG, el cual hemos modificado y adaptado para poder permitir generar catálogos de galaxias de amplia envergadura, así como también hemos estudiado y mejorado aspectos particulares del modelo físico que nos permiten obtener poblaciones de galaxias con propiedades más consistentes con las observadas.

Específicamente, hemos introducido cambios técnicos globales en el código que incluyen: una adaptación a nuevos tipos de simulaciones con formato estándar, estrategias de optimización, rendimiento y paralelización para analizar grandes entradas de datos en supercomputadoras. Detallamos comparaciones entre las diferentes versiones del modelo, así como análisis del algoritmo del mismo y estadísticas de rendimiento.

También nos hemos enfocado en algunos tratamientos evolutivos aplicados a las galaxias satélites. Analizamos el método de integración de órbitas de las galaxias satélites huérfanas mediante el ajuste de la tasa de fusiones de la población de galaxias, implementamos correcciones para considerar la periodicidad de las simulaciones en esta integración, y desarrollamos un ajuste empírico para modelar la pérdida de material de las galaxias satélites debido a la presión de barrido ejercida por el medio.

Finalmente, utilizamos todos los cambios y mejoras anteriores para crear un catálogo de galaxias de gran envergadura aplicando el código SAG sobre la simulación pública MDPL2 del grupo *MultiDark*, en contexto de un proyecto de colaboración. Este catálogo será publicado mediante una base de datos abierta en línea. Detallamos todo el análisis necesario realizado para construir este catálogo, así como los resultados intermedios y las estadísticas generales de cálculo. Incluimos también análisis de los modelos utilizados en cada etapa de la construcción del catálogo final. El catálogo resultante consiste en una población de galaxias simuladas cuyas propiedades cumplen una serie de restricciones fundamentales basadas en observaciones de relevamientos extragalácticos y muestra predicciones consistentes. Esta muestra se compone de aproximadamente 194 millones de galaxias a $z = 0$, y considera un total de 125 instantáneas en su historia evolutiva.

Dedicada a María José

*If you look up at the Milky Way through the eyes of Carl Sagan,
you get a feeling in your chest of something greater than yourself,
And it is. But it's not supernatural.*

Richard Dawkins

Agradecimientos

Esta Tesis, fruto del trabajo de varios años llenos de aprendizaje, estudio y crecimiento personal, no hubiera podido concretarse sin el apoyo de varias personas que me acompañaron en este camino cumpliendo un rol fundamental en mi avance tanto académico como personal. A todos ellos les dejo un pequeño reconocimiento en esta sección.

En primer lugar agradezco a mi directora, Sofía Cora, responsable de mi mayor crecimiento académico y guía profesional durante estos años. Gracias a su paciencia y dedicación pude encausar y sacar provecho de mis mayores inquietudes y mejores habilidades personales, permitiéndome lograr gratificantes contribuciones tanto a nuestro grupo de trabajo como al área de investigación. También agradezco a Nelson Padilla, responsable de sembrar en mí el interés por la astrofísica extragaláctica desde el enfoque numérico e iniciarme en el área. Le agradezco además todo el empuje y apoyo brindado, particularmente durante los años anteriores de iniciar mi doctorado. A su vez, quiero agradecer al jurado examinador de esta Tesis: Claudia Scóccola, Mario Abadi y Daniel Carpintero por sus valiosos comentarios respecto al trabajo realizado y su presentación.

En segundo lugar quiero agradecer a aquellos seres queridos que acompañaron mi crecimiento personal. A los cuaches, quienes me enseñaron que vivir en base a la fraternidad y la búsqueda de la verdad es el camino a la felicidad. Agradezco a aquellos que me impulsaron a tomar este camino académico y me siguieron acompañando hasta hoy: Carola Carrasco, Paloma Gallardo, Luis Matsuda, Manuel Lagos y muchos más que llevo siempre en mis pensamientos. Muy especialmente agradezco a Cinthya Vargas por todo ese cariño incondicional y confianza que nos hizo crecer tanto durante estos años.

También quiero agradecer a aquellas hermosas personas que conocí en Argentina y con quienes compartimos inolvidables momentos: Florencia Teppa, Juan Pablo Caso, Olga Bracco, entre otros. Esos jueves que compartimos en torno a la música y el compañerismo me hicieron muy feliz, espero que nuestro vínculo dure para siempre. Entre ellos agradezco muy especialmente a Ignacio Gargiulo, no solamente por ser un excelente colega sino por ser un gran amigo que me brindó confianza y contención en todo momento. Además, agradezco a Johanna Torres y Ulises Gonzáles, con quienes compartí maravillosos momentos llenos de fraternidad, comprensión y cariño.

Finalmente agradezco a mi familia por su apoyo constante; a mis padres y mi hermano Gabriel. En especial, con todo mi corazón agradezco a mi pareja, María José Guzmán, a quien le dedico esta Tesis. Valoro de sobremanera el esfuerzo que significó aventurarse conmigo en otro país; las distancias que tuvimos que sortear; los infinitos viajes que tuvo que hacer a su lugar de trabajo para que pudiésemos vivir juntos; el apoyo y preocupación constante en mis peores momentos. Pero por sobre todo, valoro con lo más profundo de mi ser su amor puro e incondicional, absolutamente digno de reciprocidad plena y desinteresada. Por todo aquello e incontables detalles más, mi más profunda gratitud.

Índice general

Resumen	v
Agradecimientos	xi
Índice general	xiii
1. Introducción	1
1.1. Teoría de formación de galaxias	2
1.1.1. Simulaciones de N -cuerpos/hidrodinámicas	3
1.1.2. Modelos semianalíticos	4
1.1.3. Modelos de ocupación de halos	5
1.2. Formación y estructura de halos de materia oscura	6
1.2.1. Perturbaciones de densidad	6
1.2.2. Estructura interna de los halos	11
1.3. Halos en simulaciones de N -cuerpos	14
1.4. Relevamientos extragalácticos	15
2. Modelo de formación y evolución de galaxias SAG	19
2.1. Descripción del modelo SAG	20
2.1.1. Modelo básico	23
2.1.2. Modelo extendido	28
2.1.3. Cálculo de magnitudes	33
2.1.4. Ajuste de parámetros libres	34
2.1.5. Licencia	34
2.2. Análisis técnico del código	35
2.2.1. Formato de entrada	38
2.2.2. Formato de salida	39
2.3. Cambios introducidos en el código	40
2.3.1. Administración de versiones	40
2.3.2. Rendimiento	41
2.3.3. Paralelización multihilos con OPENMP	42
2.3.4. Adaptación a SUSSING	44
2.3.5. Nuevos formatos de salida	49
2.4. El proyecto de comparación nIFTy	50
2.5. Calibración de simulaciones SUSSING	55
2.5.1. Selección del subvolumen para la calibración	55
2.5.2. El paquete SUSSINGSCRIPTS	56

3. Modelo de galaxias satélites	59
3.1. Modelo de órbitas de galaxias satélites huérfanas	61
3.1.1. Adaptación a SUSSING	63
3.1.2. Ajuste de la tasa de fusiones de galaxias	64
3.1.3. Corrección por periodicidad	69
3.2. Presión de barrido sobre el gas	70
3.2.1. Descripción del ajuste original	71
3.2.2. Nuevo perfil de presión de barrido	74
4. Creación del catálogo MultiDark-SAG	79
4.1. Las simulaciones MultiDark	80
4.2. Recursos computacionales	82
4.3. Preprocesado de la simulación MDPL2	83
4.3.1. Órbitas de galaxias satélites: MDPL2	85
4.4. Calibración del modelo	88
4.5. Simulación con modelo simplificado	91
4.6. Segundo catálogo: uso de SEDs	96
4.7. Generación del catálogo público	103
4.7.1. Árboles de fusiones de las galaxias	108
4.7.2. Inyección en la base de datos	110
4.8. Lectura de salidas del código SAG	110
5. Conclusiones	115
Bibliografía	119
A. Propiedades de la salida reducida	129

Capítulo 1

Introducción

El modelo estándar de la cosmología describe la evolución del Universo a gran escala mediante el escenario Λ CDM con inflación, en el cual el Universo está gobernado por energía oscura representada por una constante cosmológica Λ y materia oscura fría (CDM, por sus siglas en inglés). Resultados recientes como los obtenidos por la misión *Planck* (Planck Collaboration: Ade et ál., 2016) estiman un valor para el parámetro de densidad asociado a la constante cosmológica de $\Omega_\Lambda = 0.692 \pm 0.012$ con una constante de Hubble $H_0 = 67.81 \pm 0.92 \text{ km s}^{-1} \text{ Mpc}^{-1}$, mientras que la densidad de materia correspondería a $\Omega_m = 0.308 \pm 0.012$, de la cual sólo alrededor del 3% estaría asociado a la materia bariónica tradicional ($\Omega_b h^2 = 0.02226 \pm 0.00023$, con $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$). A pesar de que observaciones recientes realizadas con el telescopio espacial *Hubble* para estimar H_0 mediante la medición de Cefeidas en el Universo cercano han entregado un valor alrededor de 3σ más alto que el detectado por el satélite *Planck*, obteniendo $H_0 = 73.24 \pm 1.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Riess et ál., 2016), el modelo cosmológico estándar es nuestra mejor y más precisa descripción del Universo a gran escala.

En este escenario actual, las grandes estructuras que componen el Universo siguen una estrategia de formación por agregación jerárquica, en donde las estructuras más pequeñas se fusionan a tiempos tempranos para formar las más grandes. Las estructuras se crean a partir de inestabilidades gravitacionales provocadas por el crecimiento lineal de las perturbaciones iniciales del campo de densidad de materia que llegan al valor crítico, revirtiendo el comportamiento expansivo del Universo y terminando por colapsar localmente, formando halos de materia que pueden constituir sistemas virializados. Estos halos continúan creciendo con el tiempo ya sea por acreción de masa del entorno o por la fusión con otros halos ya formados. Una amplia mayoría de los halos que se fusionan con otros preservan su estructura ligada creando subestructuras dentro de los halos que los contienen, los llamados subhalos o halos satélites.

Desde el punto de vista gravitacional, la materia oscura tiene un rol dominante en la definición de los pozos de potencial que propician la formación de galaxias. Específicamente, los halos de materia oscura son los lugares en donde la materia bariónica inicialmente se aglomerará para luego enfriarse radiativamente y condensarse dando lugar a la formación de discos gaseosos y la consecuente formación estelar. A su vez, la evolución dinámica de los halos de materia oscura tiene un rol dominante en las interacciones de galaxias, y consecuentemente, en las propiedades que tendrán las mismas a lo largo del tiempo (una buena revisión del área en términos globales se puede encontrar en los textos Peacock, 1999; Mo et ál., 2010).

La mayor parte de la información extragaláctica que podemos medir proviene efecti-

vamente de la componente bariónica de las galaxias y sus entornos, por lo que la relación entre estas y los halos de materia oscura demanda una buena comprensión que nos permita construir modelos cosmológicos suficientemente precisos para hacer predicciones e inferencias en base a nuestras observaciones. No obstante, la evolución de las galaxias a escalas temporales cosmológicas sigue un comportamiento bastante alejado del régimen lineal, y está particularmente dominada por procesos astrofísicos violentos tales como la realimentación energética y química, y los efectos de ambiente. En este punto es crucial la consideración de las simulaciones como herramienta teórica básica en la comprensión y caracterización de las interacciones bariónicas conductoras de la evolución de las galaxias (una buena revisión del tema se puede encontrar en [Benson, 2010](#)).

El objetivo de esta Tesis es hacer una revisión de los procesos físicos que conducen la formación y evolución de galaxias y sus predicciones en diferentes escalas espaciales. El método consiste en la creación de poblaciones de galaxias simuladas que describan conjuntos de observaciones obtenidas por grandes relevamientos extragalácticos, utilizando para ello un modelo semianalítico de formación y evolución de galaxias. En una primera instancia, el trabajo contempla el análisis y modificación del código del modelo semianalítico para que permitiese concretar el objetivo planteado. En una segunda instancia, esta Tesis se centra en el desarrollo de catálogos de galaxias con este nuevo modelo y en análisis específicos sobre procesos particulares de la evolución de las galaxias. Los catálogos generados constituyen un producto derivado de esta Tesis, se ofrecerán de forma pública mediante una base de datos en línea, y permitirán generar predicciones más avanzadas para futuros experimentos. Esta Tesis está organizada de la siguiente forma: en lo que resta de este Capítulo introductorio se detalla el escenario teórico en el cual está embebido el trabajo; en el Capítulo 2 se describe el modelo semianalítico utilizado junto a los mayores cambios realizados al modelo durante esta Tesis y sus primeros logros; en el Capítulo 3 se describen tratamientos referentes a las galaxias satélites que fueron estudiados y modificados durante el trabajo; y en el Capítulo 4 se detallan el proyecto de colaboración con el grupo MultiDark consistente en la creación de catálogos masivos de poblaciones de galaxias usando simulaciones de gran envergadura, y los avances que fueron parte de este trabajo de Tesis, donde también se describen los primeros catálogos públicos de galaxias generados incluyendo todas las modificaciones introducidas.

1.1. Teoría de formación de galaxias

El proceso de formación de galaxias involucra física en el régimen no lineal y un amplio conjunto de mecanismos de interacción físicos que abarcan una gran variedad de escalas tanto espaciales como energéticas. Debido a esto, es imposible solucionar de forma totalmente detallada estos procesos utilizando técnicas analíticas. Los modelos teóricos puramente analíticos sólo permiten realizar estudios restringidos de evolución aplicados a casos particulares (como los propuestos por [McKee y Ostriker, 1977](#); [White y Rees, 1978](#); [White y Frenk, 1991](#); [Efstathiou, 2000](#)).

Dos estrategias que han sido desarrolladas en las últimas décadas son las que generalmente se utilizan para atacar este problema. La primera corresponde al uso de simulaciones numéricas de N -cuerpos que incluyen tratamientos hidrodinámicos. Estas intentan solucionar numéricamente de forma directa el conjunto completo de ecuaciones no lineales que describen los procesos físicos inherentes a la formación de galaxias, a una escala lo suficientemente pequeña que permita resolver las galaxias como estructuras complejas.

La segunda estrategia utilizada corresponde al desarrollo de modelos semianalíticos, los cuales se basan en la construcción de un conjunto coherente de aproximaciones analíticas que trata de describir la misma física, pero a una escala espacial mayor que la resuelta por las simulaciones hidrodinámicas. Cada una de estas técnicas tiene ventajas y limitaciones y, dado que se enfocan en escalas diferentes, sus resultados se han alimentado mutuamente con los años para conseguir una mejor comprensión de los procesos astrofísicos estudiados.

Por otro lado, un tercer método que no permite entender los procesos físicos involucrados sino más bien reproducir propiedades de galaxias en una escala mayor es la estrategia empírica de ocupación de halos. Esta utiliza métodos estadísticos para asignar propiedades siguiendo distribuciones fenomenológicas.

A continuación describiremos en más detalle cada uno de estos métodos con el fin de dar un contexto más acabado del trabajo de Tesis.

1.1.1. Simulaciones de N -cuerpos/hidrodinámicas

El método computacional más preciso para resolver la formación de galaxias es a través de las simulaciones, donde las ecuaciones fundamentales de gravitación, hidrodinámica e incluso transferencia y enfriamiento radiativo son resueltas para un gran número de elementos puntuales o partículas. Estas partículas son unidades que discretizan el espacio y pueden ser organizadas en forma de grilla o seguir las trayectorias de flujo de fluidos.

La materia oscura, al ser entendida como una componente del Universo no colisional a grandes escalas, es relativamente simple de modelar mediante simulaciones dado que responde únicamente a la fuerza gravitacional. Para la evolución de las distribuciones de masa, velocidades y campos gravitacionales en el contexto de formación de estructuras y evolución de galaxias, la dinámica de Newton (no relativista) es suficiente si se considera la expansión del Universo de forma independiente. La evolución se resuelve a partir de condiciones iniciales donde la grilla de partículas de materia oscura sigue una distribución de densidad consistente con el espectro de potencias detectado en el fondo cósmico de radiación de microondas (CMB, por sus siglas en inglés).

Las simulaciones de N -cuerpos de mayor resolución, usualmente disponibles para uso general, han contribuido enormemente a nuestra comprensión de la formación de estructuras (por ejemplo [Springel et ál., 2005](#); [Boylan-Kolchin et ál., 2009](#); [Klypin et ál., 2016](#)); siendo la simulación Millennium-XXL ([Angulo et ál., 2012](#)) la de mayor número de partículas realizada a la fecha. Esta simulación contiene más de 303 mil millones de partículas (6720^3) de materia oscura con masa de $6.17 \times 10^9 h^{-1} M_{\odot}$ distribuidas en un volumen comóvil de $3 h^{-1} \text{Gpc}$ de lado.

En general, las simulaciones de materia oscura han sido exitosas en determinar la formación de halos y la estructura a gran escala del Universo, definiendo la existencia de la llamada “red cósmica”. Como resultado, las propiedades de correlación espacial y en velocidades de la materia oscura y los halos de materia oscura, junto a los perfiles de densidad, momento angular y estructura interna de los halos, han sido estudiados con una alta precisión (una buena compilación de resultados recientes en este tema se puede encontrar en el libro de [Mo et ál., 2010](#)).

Sin embargo, la materia oscura por sí sola es insuficiente para estudiar la formación de galaxias, de modo que el contenido bariónico debe ser incluido. Esto dificulta bastante el problema debido a que esto requiere que por lo menos sean calculadas las fuerzas de presión y la energía interna del fluido bariónico incluido. Se suelen utilizar varios

métodos para resolver esta interacción. Los métodos lagrangianos basados en partículas, en especial el método hidrodinámico de partículas suavizadas (SPH por sus siglas en inglés *Smoothed Particle Hydrodynamics*), han sido muy exitosos para estudiar componentes bariónicas que se comporten como gases ideales a gran escala (Springel, 2005). Asimismo, en los últimos años han sido desarrollados métodos hidrodinámicos eulerianos como la técnica de refinado de malla adaptativa (AMR, por sus siglas en inglés *Adaptive Mesh Refinement*), los cuales también han sido ampliamente adoptados (Springel, 2010). Sobre estos tipos de modelos, la inclusión de enfriamiento radiativo se puede realizar de forma simple mediante la tabulación de las tasas a las cuales el gas pierde energía dada su temperatura y densidad, excluyendo interacciones entre esta energía y el gas.

La inclusión de procesos físicos más allá de este nivel se vuelve desafiante. Muchos códigos actualmente son capaces de modelar la formación estelar y procesos de realimentación energética y química, incluso el crecimiento de los agujeros negros supermasivos de los centros de las galaxias. Ejemplos recientes son la simulación Illustris (Sijacki et ál., 2015) que utiliza AMR, o el conjunto de simulaciones EAGLE (Schaye et ál., 2015) que utilizan SPH. Debido a que el conjunto de todos estos procesos físicos a resolver es muy amplio, estas metodologías suelen ser extremadamente demandantes de poder de cálculo computacional, de modo que el tamaño espacial alcanzado por este tipo de simulaciones no suele superar las centenas de Mpc de lado para poder ser completadas en tiempos razonables. Por otro lado, es importante destacar que para simulaciones que pueden resolver el detalle interno de las galaxias, los procesos físicos reales que intervienen en la evolución ocurren realmente a escalas espaciales muy inferiores al límite de resolución utilizado en las simulaciones. Por ello, el tratamiento de los procesos astrofísicos es frecuentemente realizado en un nivel de “subgrilla”, en donde esencialmente se utilizan técnicas semianalíticas por debajo de cierto límite de resolución.

1.1.2. Modelos semianalíticos

La técnica del modelado “semianalítico” trata los procesos físicos asociados a la formación de galaxias usando técnicas analíticas aproximadas. Tal como en las simulaciones hidrodinámicas, el grado de aproximación varía considerablemente con la complejidad de la física incluida, lo cual conlleva incertezas que suelen ser controladas mediante los parámetros libres que son incluidos en las relaciones que modelan la evolución de galaxias.

Este tipo de modelos opera a una escala distinta de las simulaciones, dado que suelen tratar a las galaxias como fuentes puntuales cuyas propiedades son caracterizadas de forma general. No obstante, están enfocados principalmente en resolver la evolución del contenido bariónico de las galaxias excluyendo la caracterización y evolución de los halos de materia oscura que las contienen. Por este motivo la información relacionada a los halos de materia oscura que evolucionan en un contexto cosmológico, que puede ser construida de forma analítica o extraída de simulaciones, suele ser suministrada a los modelos semianalíticos que la usan como condición inicial para poder resolver la evolución de las galaxias acorde a ello.

La principal ventaja de los modelos semianalíticos es que tienen una demanda computacional mucho más baja que las simulaciones de N -cuerpos con hidrodinámica. Esto facilita la construcción de muestras de galaxias mucho mayores que las posibles utilizando puramente técnicas de N -cuerpos, así como también una exploración más rápida del espacio de parámetros libres. Por otro lado, es una técnica ideal para poner a prueba hipótesis sobre los procesos físicos que conducen la evolución de propiedades particula-

res durante la formación de galaxias, posicionándose como excelentes laboratorios para probar teorías astrofísicas a nivel extragaláctico. No obstante, estas ventajas son a costa de utilizar un mayor nivel de aproximación (una buena revisión de esta técnica se puede encontrar en [Baugh, 2006](#)). Una gran variedad de modelos, de forma independiente, han permitido llegar a un acuerdo sobre el conjunto de procesos físicos principales involucrados en moldear propiedades estadísticas de la población de galaxias observadas en grandes relevamientos extragalácticos. Incluso, recientemente se han realizado esfuerzos para congregarse a la gran mayoría de los grupos desarrolladores de modelos y comparar las diferencias entre los modelos semianalíticos actuales ([Knebe et ál., 2015](#)).

Esta Tesis se centra sobre el modelo semianalítico de formación de galaxias SAG, el cual en este trabajo ha sido modificado y aplicado sobre halos extraídos de simulaciones. Mayores detalles sobre el funcionamiento de este tipo de modelos, en especial sobre el SAG, serán descritos en extenso en el Capítulo 2.

1.1.3. Modelos de ocupación de halos

Si bien las dos técnicas descritas anteriores corresponden a modelos numéricos que buscan explicar la formación de las galaxias desde un punto de vista teórico, dada la complejidad de estos procesos físicos a veces es necesario tomar un enfoque más empírico del problema. Esto puede ser ventajoso tanto para relacionar observaciones con la estructura física del Universo que suponemos correcta, como para realizar predicciones basadas en extrapolaciones de relaciones conocidas.

Estos modelos de distribuciones de ocupación de halos (HOD por sus siglas en inglés *Halo occupation distributions*), originalmente propuestos por [Neyman y Scott \(1952\)](#), en su forma más simple se basan en especificar la probabilidad de encontrar N galaxias de un tipo particular en un halo de materia oscura con masa M , $P(N|M)$. Dada esta distribución de probabilidad, junto a las distribuciones de masas de los halos de materia oscura, distribuciones espaciales y adicionando suposiciones sobre las localizaciones de las galaxias satélites en halos, es posible construir una distribución de galaxias *ad hoc*. Este enfoque puede ser extendido para incorporar más propiedades a través de prescripciones simples, tales como la formación estelar con respecto a la masa de los halos o el corrimiento al rojo (*redshift*), pudiendo generar muestras de galaxias que ajustan una gran variedad de observables.

Otro enfoque más monotónico pero siguiendo la misma metodología anterior es la asignación por abundancia de subhalos (SHAM, por sus siglas en inglés *SubHalo Abundance Matching*), propuesto originalmente por [Vale y Ostriker \(2004\)](#). En este método, los (sub)halos de materia oscura se pueblan suponiendo una biyección entre las propiedades de las galaxias (como la masa estelar o la luminosidad) con propiedades de los halos (tales como la velocidad circular máxima). Al igual que el método anterior, permite obtener muestras de galaxias que ajustan un conjunto de propiedades, pero no entrega información alguna sobre los procesos físicos que conducen la evolución de las galaxias.

Al igual que un modelo semianalítico, las técnicas de ocupación de halos requieren *a priori* la información de una población de halos de materia oscura con consistencia cosmológica. Para ello, se puede utilizar como base la información de halos de materia oscura construida de forma analítica o extraída a partir de simulaciones de N -cuerpos.

En definitiva, todos los métodos utilizados ya sea para comprender los procesos físicos preponderantes en la evolución de galaxias (simulaciones hidrodinámicas y modelos semianalíticos) como aquellos usados para obtener distribuciones estadísticas sin conexión

física (utilizando HOD o SHAM), requieren la información de halos de materia oscura para poder poblarlos con galaxias. Estos halos, principalmente resultantes de simulaciones cosmológicas de N -cuerpos de materia oscura, son caracterizados por un conjunto de propiedades que deben ser consideradas por los modelos mencionados.

Dada la importancia de los halos de materia oscura en la formación y evolución de galaxias, en la siguiente sección describiremos de modo general las principales características de la formación así como las propiedades de estas estructuras de acuerdo al escenario cosmológico estándar.

1.2. Formación y estructura de halos de materia oscura

Los halos de materia oscura son estructuras de sobredensidad que pueblan el Universo en diferentes épocas, formados como consecuencia directa de las perturbaciones inherentes a la distribución del contenido de materia del Universo y que están presentes desde la época inflacionaria. La existencia de estas estructuras, así como su formación y evolución, emergen como una predicción directa al solucionar numéricamente la evolución de la distribución de materia del Universo con el modelo cosmológico estándar. Estos halos, al estar constituidos por una componente de materia que es dominante respecto a la materia bariónica, definen los pozos de potencial que permiten la acumulación de gas y su posterior enfriamiento que da lugar a la formación de galaxias.

Para una mejor comprensión de la evolución de estas estructuras, durante esta sección introduciremos brevemente la teoría de formación y caracterización de las perturbaciones que propician la formación de los halos, transcendental para la construcción de simulaciones cosmológicas de N -cuerpos. Luego de ello, caracterizaremos las principales propiedades de los halos que son medidas como resultado directo de estas simulaciones y son utilizadas por los modelos de formación de galaxias. En lo que resta de esta sección se incluyen resúmenes de los tópicos más relevantes para esta Tesis basados en los desarrollos de los libros *Cosmological Physics* de Peacock (1999, capítulo 16) y *Galaxy Formation and Evolution* de Mo et ál. (2010, capítulo 7), que contienen revisiones acabadas de la amplia literatura disponible del tema.

1.2.1. Perturbaciones de densidad

Como se describe en Peacock (1999), la formación de grandes estructuras en el Universo tiene su origen en las perturbaciones iniciales del campo de densidad de la materia generadas por fluctuaciones cuánticas aleatorias en el campo asociado al inflatón, y que luego se expandieron a escalas cosmológicas debido al proceso inflacionario. La materia oscura, al no mostrar interacción electromagnética con otras partículas ni consigo misma, carece de presión que inhiba su propio colapso gravitacional debido a la presencia de las perturbaciones, y a su vez permite el crecimiento de las mismas en el tiempo de forma casi lineal.

Para caracterizar la distribución de materia en un volumen cosmológico arbitrario, es conveniente introducir las coordenadas comóviles \mathbf{x} , definidas por

$$\mathbf{x}_f = a(t)\mathbf{x}, \quad (1.1)$$

donde $a(t)$ es el factor de escala que parametriza la expansión del Universo según el modelo cosmológico, y \mathbf{x}_f representa las coordenadas físicas. A su vez, el factor de escala se

expresa en términos del corrimiento al rojo según $a = 1/(1+z)$; de modo que $a = 1$ en el Universo local ($z = 0$). Luego, el campo aleatorio de perturbaciones puede ser caracterizado mediante la diferencia de densidad de materia respecto a la densidad promedio $\langle \rho \rangle$ en un instante determinado:

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle \rho \rangle}{\langle \rho \rangle}. \quad (1.2)$$

El conjunto de perturbaciones $\delta(\mathbf{x})$ debe cumplir los principios cosmológicos de isotropía y homogeneidad, de modo que el promedio (denotado por $\langle \dots \rangle$) se realiza sobre un conjunto de realizaciones del campo de perturbaciones aleatorias, o de forma equivalente sobre un volumen lo suficientemente grande. Al ser realizaciones aleatorias, todos sus momentos estadísticos deben ser invariantes ante rotaciones y traslaciones espaciales, de modo que estos constituyen un conjunto de variables útiles para caracterizar el espectro de perturbaciones que formarán los halos de materia oscura.

Por construcción, el primer momento estadístico es $\langle \delta(\mathbf{x}) \rangle = 0$, y la varianza del campo de perturbaciones queda definida por

$$\sigma^2 = \langle \delta^2(\mathbf{x}) \rangle. \quad (1.3)$$

Asimismo, se define el momento $\xi(\mathbf{r})$, conocido como función de autocorrelación o función de correlación de dos puntos del campo de densidad, de la forma

$$\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle, \quad (1.4)$$

valor que depende solamente de la distancia r entre los puntos respectivos, y además $\xi(0) = \sigma^2$.

Por otro lado, el comportamiento estadístico del conjunto de perturbaciones aleatorias presentes en un campo de densidad puede ser caracterizado por la descomposición de los contrastes de densidad $\delta(\mathbf{x})$ definidos por la Ecuación (1.2) en sus modos de onda. Si suponemos que el campo está contenido en un volumen cúbico de largo L con condiciones de borde periódicas, la expansión de Fourier resulta

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp(-i\mathbf{k} \cdot \mathbf{x}), \quad (1.5)$$

con $\delta_{\mathbf{k}}$ los coeficientes de Fourier y \mathbf{k} representando a los vectores de onda respectivos. Como estos últimos están contenidos en un cubo con volumen fijo L^3 , solo pueden tomar los valores $k_{x,y,z} = 2\pi n_{x,y,z}/L$, con $n_{x,y,z} = 1, 2, 3, \dots$. Expandiendo esta definición para un L arbitrariamente grande, la sumatoria de los contrastes de densidad en sus modos se puede aproximar a su forma integral, obteniendo así la relación entre estos δ y su transformada de Fourier $\delta_{\mathbf{k}}$:

$$\begin{aligned} \delta(\mathbf{x}) &= \left(\frac{L}{2\pi}\right)^3 \int \delta_{\mathbf{k}}(\mathbf{k}) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3\mathbf{k}, \\ \delta_{\mathbf{k}}(\mathbf{k}) &= \left(\frac{1}{L}\right)^3 \int \delta(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) d^3\mathbf{x}. \end{aligned} \quad (1.6)$$

De este modo, podemos definir el segundo momento estadístico de los contrastes de densidad, pero en el espacio de Fourier

$$P(k) \equiv \langle |\delta_{\mathbf{k}}|^2 \rangle, \quad (1.7)$$

el cual es conocido como el espectro de potencias de las fluctuaciones del campo de densidad y está relacionado con la función de correlación al ser su transformada de Fourier. Esto se puede verificar simplemente expresando la función de autocorrelación, definida por la Ecuación (1.4), en términos de la expansión de Fourier de los contrastes de densidad. Como los $\delta_{\mathbf{k}}$ son reales, podemos reemplazar uno de ellos por su complejo conjugado $\delta_{\mathbf{k}}^*$, obteniendo

$$\xi = \left\langle \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \delta_{\mathbf{k}}(\mathbf{k}) \delta_{\mathbf{k}'}^*(\mathbf{k}') e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{x}} e^{-i\mathbf{k} \cdot \mathbf{r}} \right\rangle. \quad (1.8)$$

Por otro lado, en cualquier campo real los modos con vectores de onda opuestos se relacionan entre sí según: $\delta_{\mathbf{k}}(-\mathbf{k}) = \delta_{\mathbf{k}}^*(\mathbf{k})$. Ahora, debido a que tenemos condiciones de borde periódicas, todos los términos cruzados en la suma (donde $\mathbf{k}' \neq \mathbf{k}$) promedian cero. Los términos restantes, al ser expresados de forma integral, quedan

$$\xi(\mathbf{r}) = \left(\frac{L}{2\pi} \right)^3 \int |\delta_{\mathbf{k}}|^2 e^{-i\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k}, \quad (1.9)$$

la cual corresponde precisamente a la transformada de Fourier del espectro de potencias definido anteriormente.

En un universo isotrópico, el espectro de perturbaciones de densidad no puede tener una dirección privilegiada, por lo que también debemos tener un espectro de potencias isotrópico, en el cual

$$\langle |\delta_{\mathbf{k}}|^2(\mathbf{k}) \rangle = |\delta_k|^2(k). \quad (1.10)$$

En este caso, la parte angular de la integral en el espacio de los modos \mathbf{k} puede resolverse directamente, quedando

$$\xi(r) = \left(\frac{L}{2\pi} \right)^3 \int P(k) \frac{\text{sen}(kr)}{kr} 4\pi k^2 dk. \quad (1.11)$$

Por conveniencia, se puede eliminar la dependencia con el volumen de esta relación expresando el espectro de potencias en una forma adimensional definida por

$$\Delta_{\mathbf{k}}^2(k) \equiv \frac{L^3}{2\pi^2} k^3 P(k) = \frac{2}{\pi} k^3 \int_0^\infty \xi(r) \frac{\text{sen}(kr)}{kr} r^2 dr, \quad (1.12)$$

la cual también se puede entender como el cambio de la varianza por intervalo logarítmico de k (es decir, $\Delta_{\mathbf{k}}^2 = d\langle \delta^2 \rangle / d \ln k$). Esta definición, al no tener unidades de distancia, otorga una forma más simple de visualizar el espectro de potencias, dado que $\Delta_{\mathbf{k}}^2(k) = 1$ implica la existencia de fluctuaciones de densidad del orden de la unidad para modos en el intervalo logarítmico alrededor del número de onda k . Por lo tanto, $\Delta_{\mathbf{k}}^2(k)$ corresponde a la elección más conveniente como contraparte en el espacio de Fourier de la cantidad adimensional $\xi(r)$.

En definitiva, el espectro de potencias es una cantidad medular al caracterizar las fluctuaciones de densidad en la distribución de materia del Universo, más allá de las realizaciones individuales que puedan generarse para un volumen cosmológico particular. Asimismo, la función de correlación puede ser medida para diferentes épocas del Universo. Por ejemplo, mediante mediciones del fondo de radiación de microondas (como han realizado las misiones *COBE*, *WMAP* o *Planck*) se puede recuperar la forma del espectro de potencias mediante la medición de la función de correlación proyectada de las diferencias de temperatura de esta radiación, y con ello obtener el espectro de potencias inicial que

caracteriza las fluctuaciones de densidad antes de que evolucionen debido al colapso de los halos y el desacoplamiento de los diferentes tipos de materia (una revisión temprana de esta técnica se puede encontrar en el libro de Partridge, 1995). Esta función también puede ser medida para diferentes épocas mediante estadísticas de galaxias obtenidas de grandes relevamientos, especialmente para bajo corrimiento al rojo (ver, por ejemplo Zehavi et ál., 2005); así como también a corrimientos al rojo altos mediante mediciones de los bosques Lyman- α (*Lyman- α forests*) presentes en los espectros de cuásares (ver, por ejemplo, Rollinde et ál., 2003).

En principio, no hay razones para pensar que el espectro inicial deba tener preferencias en escalas particulares para la formación de estructuras, por lo tanto puede ser descrito como una ley de potencias simple

$$\langle |\delta_k|^2 \rangle \propto k^{n_s}, \quad (1.13)$$

coincidente con lo que se espera de los modelos inflacionarios (ver, por ejemplo, Capítulo 4.5 de Mo et ál., 2010), y en donde n_s es comúnmente llamado índice espectral. En este caso, aplicando directamente la definición de la Ecuación (1.12), obtenemos que el espectro de potencias es

$$\Delta_k^2(k) \propto k^{3+n_s}. \quad (1.14)$$

Un caso de especial interés sucede cuando $n_s = 1$, el llamado espectro de Harrison-Zel'dovich, en donde nos queda $\Delta_k^2 \propto k^4$. Utilizando este valor al perturbar la ecuación de Poisson que define el potencial gravitacional ($\nabla^2\phi = 4\pi G\rho$, donde ϕ representa el potencial), se obtiene que el espectro de potencias asociado a este potencial es una constante que no depende de k . De este modo, esta elección de n_s hace que el espectro de perturbaciones sea invariante a las escalas espaciales, constituyendo así una estructura tipo fractal.

Para caracterizar cabalmente el espectro, además de su forma es necesario conocer también la amplitud que lo define en las diferentes épocas. Una forma de encontrar esto es mediante el cálculo de la varianza del campo de perturbaciones de densidad de alguna distribución particular de observables de interés (por ejemplo, utilizando galaxias). Teóricamente, es más simple tratar estas estadísticas mediante la aplicación de una estrategia que permita simplificar los cálculos evitando múltiples integraciones en todo el campo de interés. Para ello, por conveniencia se seleccionan funciones ventana de modo que el cálculo de las integrales quede acotado. Por ejemplo, en el caso más simple podemos utilizar una función tipo galera (*top-hat*) definida como

$$W_R(r) = \begin{cases} 3/(4\pi R^3) & r \leq R, \\ 0 & r > R, \end{cases} \quad (1.15)$$

en donde R es el tamaño de la esfera elegida para suavizar el cálculo. No obstante, la elección *ad hoc* más utilizada en la literatura es seleccionar funciones con forma gaussiana dado que introducen menos errores en las estimaciones a costa de obtener expresiones más complejas de solucionar. De este modo, los contrastes de densidad son convolucionados con una función ventana W_R , quedando

$$\delta(\mathbf{x}; R) = \int W_R(|\mathbf{x} - \mathbf{y}|)\delta(\mathbf{y})d^3\mathbf{y} = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \widetilde{W}_R(k) \exp(-i\mathbf{k} \cdot \mathbf{x}), \quad (1.16)$$

en donde $\widetilde{W}_R(k)$ es la transformada de Fourier de la función ventana, la cual en el caso galera vale

$$\widetilde{W}_R(k) = \frac{3(\text{sen}(kR) - kR \cos(kR))}{(kR)^3}. \quad (1.17)$$

Del mismo modo, la varianza convolucionada con una función ventana queda definida por

$$\sigma^2(R) = \frac{1}{2\pi^2} \int P(k) \widetilde{W}_R^2(k) k^2 dk. \quad (1.18)$$

El valor de $\sigma(R)$ obtenido con la distribución de galaxias ronda la unidad para $R = 8 h^{-1}\text{Mpc}$ (Davis y Peebles, 1983), de modo que el espectro de potencias teórico puede ser normalizado forzando que $\sigma(R = 8 h^{-1}\text{Mpc}) = 1$. No obstante, las galaxias no son trazadores precisos de las fluctuaciones en la distribución de masa debido a que presentan preferencias de formación según la densidad del medio. Esta diferencia puede ser considerada introduciendo un parámetro de sesgo (*bias*) como constante de proporcionalidad entre las fluctuaciones en la distribución de masa y las fluctuaciones en la distribución de galaxias

$$\delta_{\text{gal}} = b \delta_{\text{m}}. \quad (1.19)$$

El parámetro de sesgo b es una constante que depende de cómo las galaxias se han formado en el campo de densidad y no se conoce con precisión. De todos modos, en este caso tenemos

$$\sigma_{\text{m}}(8 h^{-1}\text{Mpc}) = \frac{\sigma_{\text{gal}}(8 h^{-1}\text{Mpc})}{b} \approx \frac{1}{b}. \quad (1.20)$$

Esto nos permite definir el parámetro $\sigma_8 \equiv \sigma_{\text{m}}(8 h^{-1}\text{Mpc})$, utilizado ampliamente para caracterizar la amplitud del espectro de potencias (White et ál., 1993). Este parámetro se estima evolucionando el espectro de potencias inicial hasta $z = 0$ siguiendo una teoría lineal, de modo que su valor difiere de la varianza real de la distribución de masa dado que las perturbaciones a estas escalas han evolucionado de forma altamente no lineal producto de la agregación jerárquica (ver, por ejemplo, Evrard, 1989; Frenk et ál., 1990). En este sentido, el uso de este parámetro se mantiene más por razones históricas y se incluye entre los parámetros libres del modelo cosmológico.

La evolución del espectro de potencias en el tiempo se puede estimar analíticamente mediante una aproximación lineal para todas aquellas perturbaciones que aún no colapsan formando halos. Para esto, debe introducirse un modelo para las perturbaciones (por ejemplo, mediante un campo de perturbaciones gaussianas). No obstante, cuando estas perturbaciones se equiparan a la densidad crítica $\rho_c(t)$ en un instante determinado, el crecimiento se revierte propiciando el colapso gravitacional de aquellas estructuras (Gunn y Gott, 1972), llevando el crecimiento de las perturbaciones con esa densidad a un régimen no lineal donde el aglomeramiento de los halos se incrementa debido a este colapso, siempre siguiendo una estrategia de formación jerárquica que se obtiene directamente del estudio de la evolución de las perturbaciones en un universo en expansión.

En definitiva, una vez definido el espectro de potencias inicial del campo de densidad a través del índice espectral n_s y su amplitud σ_8 , es posible crear condiciones iniciales tanto para realizar simulaciones de N -cuerpos que tracen el comportamiento evolutivo de la materia oscura formando los halos, como para desarrollar modelos evolutivos de los mismos usando técnicas de Montecarlo. Ambas técnicas permiten estudiar el comportamiento de las perturbaciones en el régimen no lineal de la evolución de los pozos de potencial que conducirán a la formación de las galaxias.

1.2.2. Estructura interna de los halos

A continuación detallaremos un conjunto de características físicas usadas para describir los halos de materia oscura. En la literatura, la caracterización de los halos se ha hecho mayoritariamente mediante la realización y análisis de simulaciones cosmológicas. Estas simulaciones siguen la evolución de la materia oscura desde el Universo temprano utilizando N -cuerpos, y consideran condiciones iniciales de materia perturbada con un espectro de sobredensidades aleatorio que tenga las mismas propiedades estadísticas que el observado en el fondo de radiación de microondas. En esta sección se describirán las propiedades más representativas de los halos de materia oscura que son relevantes para los modelos de formación y evolución de galaxias: definición de masa, perfil de densidad y momento angular. Para ello, el resto de la sección se basará en las secciones correspondientes del libro de Mo et ál. (2010).

Masa y perfil de densidad

Como primera aproximación, un halo de materia oscura se puede describir como un objeto esférico, cuya distribución interna de masa queda totalmente descrita por su perfil de densidad radial $\rho(r)$.

El modelo más simple que encaja con el desarrollo de colapso esférico (Gunn y Gott, 1972) considera que los halos se aproximan a esferas isothermas truncadas con $\rho(r) \propto r^{-2}$ para $r \leq R_h$, donde R_h define el radio límite de un halo de materia oscura. Este borde de los halos se puede definir según la distancia radial para la cual la densidad media del halo $\bar{\rho}_h$ alcance un factor Δ de la densidad media del entorno $\bar{\rho}$, es decir

$$\bar{\rho}_h = \Delta \bar{\rho} = \Delta \rho_c \Omega_m, \quad (1.21)$$

en donde ρ_c corresponde a la densidad crítica del Universo en ese instante. La densidad crítica varía según se expande el Universo, y está definida por

$$\rho_c = \frac{3H^2(t)}{8\pi G}, \quad (1.22)$$

en donde G corresponde a la constante de gravitación universal y $H(t)$ es el parámetro de Hubble para un instante t . El parámetro de Hubble se define en términos del factor de escala del Universo y su derivada temporal, según $H(t) = \dot{a}/a$, de modo que en el Universo local: $H(a = 1) = H_0$

Considerando esto, el perfil de densidad para un halo esférico isoterma queda definido por

$$\rho(r) = \frac{V_h^2}{2\pi G r^2}, \quad (1.23)$$

en donde $V_h = \sqrt{GM_h/R_h}$ es la velocidad circular del halo en R_h , comúnmente llamados velocidad virial y radio virial, respectivamente. Asimismo, M_h corresponde a la masa del halo encerrada dentro de R_h , el cual puede ser reescrito en términos de V_h utilizando las ecuaciones (1.21) y (1.22), quedando

$$R_h = \sqrt{\frac{200}{\Delta \Omega_m}} \frac{V_h}{10H(t)}. \quad (1.24)$$

Una elección *ad hoc* del factor de densidad para la determinación de $\bar{\rho}_h$ sería $\Delta = \Delta_{\text{vir}}$, correspondiente al factor necesario para que la sobredensidad definida por él pueda

colapsar y virializarse. Sin embargo, este valor sólo puede ser aproximado usando ciertas simplificaciones en modelos particulares de colapso esférico (tal como el obtenido en [Bryan y Norman, 1998](#)). Por otro lado, dado que el criterio de virialización no es estricto, otras elecciones para esta constante son ampliamente utilizadas en la literatura. Por ejemplo, muchos trabajos adoptan $\bar{\rho}_h = 200\bar{\rho}$, de modo que $\Delta = 200$; mientras que otros utilizan $\bar{\rho}_h = 200\rho_c$, de modo que $\Delta = 200/\Omega_m$. Cada una de estas definiciones implica distintas relaciones entre M_h , R_h y V_h , pero la conversión entre ellas es directa si se fija el perfil de densidad. Estas diferentes posibles elecciones para el factor Δ también originan a su vez distintas definiciones tanto de las masas M_h como de los radios R_h viriales con los cuales se caracterizan los halos de materia oscura. Según lo anterior, la relación más general entre estas cantidades se puede expresar de la forma

$$M_{\text{ref}}(< R_{\text{ref}}) = \Delta_{\text{ref}} \rho_{\text{ref}} \frac{4\pi}{3} R_{\text{ref}}^3, \quad (1.25)$$

donde el subíndice ref alude al criterio de densidad elegido para determinar el tamaño de halo (ver ecuación (1) de [Knebe et ál., 2015](#)). Por ejemplo, para el caso más común donde los halos son definidos cuando su densidad media alcanza un factor 200 de la densidad crítica, se tiene $\Delta_{\text{ref}} = 200$ y $\rho_{\text{ref}} = \rho_c$. Según el criterio elegido, los subíndices suelen reemplazarse por algún otro que pueda reflejar aquella elección. Sin embargo, no existe un acuerdo en la literatura que unifique la nomenclatura de estas definiciones. Por simplicidad, en esta Tesis se seguirá la nomenclatura que emplea M_{vir} , R_{vir} y V_{vir} para referirse a las masas, radios y velocidades viriales de los halos de forma genérica, y se especificará el criterio adoptado para definirlos cuando sea necesario.

Por otra parte, una descripción mejor del perfil de densidad fue propuesta por [Navarro et ál. \(1996, 1997\)](#), quienes en su primer trabajo mostraron que el perfil medido en halos obtenidos usando simulaciones cosmológicas de N -cuerpos de materia oscura fría presentaban pendientes más planas que r^{-2} hacia radios pequeños, y más pronunciadas a radios mayores. En el trabajo posterior mostraron la universalidad del perfil para halos de distinta masa en todos los escenarios donde la materia oscura fría es dominante en la formación. El perfil radial propuesto fue

$$\rho(r) = \rho_c \frac{\delta_c}{(r/r_s)(1 + r/r_s)^2}, \quad (1.26)$$

el cual ahora es ampliamente conocido como perfil de Navarro, Frenk y White (NFW). Esta expresión tiene dos parámetros libres que varían para cada halo: r_s conocido como radio característico y δ_c una sobredensidad característica. La pendiente logarítmica de este perfil $\gamma = d \ln \rho / d \ln r$ cambia gradualmente desde -1 cerca del centro hasta -3 a radios mayores, de modo que se asemeja al perfil de una esfera isoterma en $r \sim r_s$.

Una magnitud que permite caracterizar los halos adoptando el perfil NFW es el parámetro de concentración, el cual se define como

$$c \equiv \frac{R_h}{r_s}, \quad (1.27)$$

y permite medir el grado de concentración de masa en cada perfil. Por otro lado, la masa total encerrada dentro de un radio r se puede obtener integrando el perfil de densidad adoptando una definición para R_h

$$M(r) = 4\pi \bar{\rho} \delta_c r_s^3 \left[\ln(1 + cx) - \frac{cx}{1 + cx} \right], \quad (1.28)$$

donde $x \equiv r/R_h$. De este modo, el perfil de densidad de un halo particular queda totalmente definido por su masa M_h y su parámetro de concentración c cuando aplicamos una definición para el borde del halo. De este modo, se puede obtener también la relación entre la sobredensidad característica y el parámetro de concentración

$$\delta_c = \frac{\Delta}{3} \frac{c^3}{\ln(1+c) - c/(1+c)}. \quad (1.29)$$

Según lo obtenido por Navarro et ál. (1997), la sobredensidad característica está relacionada con el tiempo de formación de los halos, de modo que los que se forman antes son más concentrados. Estudios posteriores (como por ejemplo Zhao et ál., 2009) han encontrado que halos que han tenido fusiones mayores en su historia reciente típicamente tienen concentraciones más bajas que aquellos que han tenido épocas más largas de crecimiento suave. De este modo, el proceso de relajación violenta cumple un rol importante en la distribución de masa interna de los halos, y esta dependencia con el número de fusiones es sensible a la historia de formación, la cual es sensible a su vez a la masa característica de cada uno.

Aunque el perfil de NFW es ampliamente usado en la literatura, trabajos con simulaciones de alta resolución (como por ejemplo las simulaciones Aquarius presentadas en Springel et ál., 2008) han mostrado que los perfiles de densidad de los halos de materia oscura presentan variaciones pequeñas pero sistemáticas en relación al perfil de NFW, y son descritos de forma más precisa por el perfil propuesto por Einasto (1965)

$$\rho(r) = \rho_{-2} \exp \left[\frac{-2}{\alpha} \left\{ \left(\frac{r}{r_{-2}} \right)^\alpha - 1 \right\} \right], \quad (1.30)$$

en donde r_{-2} corresponde a la distancia donde la pendiente logarítmica γ de la distribución de densidad alcanza un valor de -2 , y $\rho_{-2} = \rho(r_{-2})$. Estas cantidades junto a α constituyen los tres parámetros libres que caracterizan este perfil, a diferencia de NFW que tiene solo dos. El tercer parámetro permite fijar el comportamiento de la derivada logarítmica, la cual se describe mediante una ley de potencias, es decir

$$\frac{d \ln \rho}{d \ln r} = -2 \left(\frac{r}{r_{-2}} \right)^\alpha. \quad (1.31)$$

Los valores encontrados del parámetro α típicamente se ubican en el rango $0.12 \lesssim \alpha \lesssim 0.25$ y parece incrementarse conforme aumenta la masa (ver, por ejemplo, Gao et ál., 2008). No obstante, si se fija este parámetro en un valor estándar (por ejemplo $\alpha = 0.17$), la función restante con sólo dos parámetros libres aún ajusta los perfiles de densidad de halos de simulaciones de una forma significativamente mejor que al usar NFW. Por otro lado, a pesar de que este perfil tiene una pendiente mucho más aplanada hacia el centro que el de NFW, aún sigue siendo muy brusco o puntiagudo en esta región. Si bien el valor obtenido del perfil en las zonas centrales de los halos tiende a ser mayor que lo esperado, hay claros indicios de una universalidad en la forma de los perfiles de densidad de los halos de materia oscura.

Momento angular

Otra propiedad importante de los halos de materia oscura es su momento angular, el cual es significativo en objetos que colapsan gravitacionalmente en este contexto cosmológico. Se suele utilizar un parámetro adimensional para caracterizarlo, el parámetro de

espín propuesto originalmente por Peebles (1969)

$$\lambda \equiv \frac{\sqrt{E}|\mathbf{J}|}{GM^{5/2}}, \quad (1.32)$$

donde J es el momento angular total, E la energía y M la masa del halo. El parámetro de espín así definido es aproximadamente la raíz cuadrada de la proporción entre la energía rotacional y la total del sistema, de modo que caracteriza la importancia general del momento angular en comparación con el movimiento aleatorio.

Suponiendo un halo con simetría esférica que rota, la energía total se puede estimar utilizando el teorema del virial $E = -K$, de modo que

$$E = -4\pi \int_0^{r_h} \frac{\rho(r)V_c^2(r)}{2} r^2 dr \equiv -\frac{M_h V_h^2}{2} F_E, \quad (1.33)$$

donde $V_c(r)$ es la velocidad circular del halo en r . F_E queda definido según la función de la distribución de densidad del halo utilizada. Por ejemplo, para el caso de una esfera isoterma: $F_E = 1$, mientras que para un perfil NFW resulta

$$F_E = \frac{c}{2} \frac{[1 - 1/(1+c)^2 - 2 \ln(1+c)/(1+c)]}{[c/(1+c) - \ln(1+c)]^2}. \quad (1.34)$$

No obstante, medir la energía total de un halo es una tarea compleja inclusive en simulaciones de N -cuerpos. Es por ello que en la literatura se suele utilizar una definición alternativa para el parámetro de espín propuesta por Bullock et ál. (2001b) la cual no depende explícitamente del valor de la energía

$$\lambda_B \equiv \frac{|\mathbf{J}|}{\sqrt{2} M V_h r_h}, \quad (1.35)$$

estando este parámetro de espín relacionado con el de Peebles mediante $\lambda_B = \lambda/\sqrt{F_E}$.

Además, estudios estadísticos sobre halos de simulaciones han mostrado que el parámetro de espín sigue una distribución de probabilidad de tipo log-normal (Barnes y Efstathiou, 1987), definida por

$$P(\lambda) = \frac{1}{\lambda \sigma_\lambda \sqrt{2\pi}} \exp\left(-\frac{\ln^2(\lambda/\lambda_0)}{2\sigma_\lambda^2}\right), \quad (1.36)$$

donde λ_0 y σ_λ definen la media y la desviación estándar del logaritmo de la variable. El mismo tipo de distribución se ha encontrado al considerar el parámetro de espín λ_B (Bullock et ál., 2001a). En este último caso, valores normalmente encontrados que caracterizan la distribución son $\lambda_0 \approx 0.035$ y $\sigma_\lambda \approx 0.5$, con una dependencia casi nula con la masa, el corrimiento al rojo y la cosmología.

1.3. Halos en simulaciones de N -cuerpos

Hemos visto que sólo es necesario fijar el espectro de potencias de las perturbaciones para crear condiciones iniciales de simulaciones de N -cuerpos que permitan resolver la evolución del contenido de materia oscura del Universo. Estas simulaciones se calculan en un contexto cosmológico, es decir, sus escalas espaciales se ven afectadas por la expansión

del Universo que se ve parametrizada en función del parámetro de expansión a , el cual evoluciona según las constantes que definen la cosmología utilizada. Siguiendo además los principios de isotropía y homogeneidad, estas simulaciones deben realizarse con condiciones de borde periódicas si es que se pretende representar una realización particular del Universo a gran escala sin agregar inconsistencias debido a la presencia de bordes que rompen la homogeneidad.

Por otro lado, la aplicación de modelos de formación y evolución de galaxias requiere de la información de los halos de materia oscura obtenidos preferentemente de una simulación. No obstante, el resultado directo de estas simulaciones consiste en distribuciones de partículas, caracterizadas por su masa, posición y velocidad, para un conjunto de instantáneas de la evolución en tiempo cósmico. El procedimiento para detectar y medir los halos de materia oscura en las colecciones de partículas es otro problema para el cual se han propuesto diversas estrategias.

La estrategia más utilizada para encontrar los halos de materia oscura en simulaciones de este tipo consiste en analizar la distribución de las partículas con códigos dedicados a este propósito (ver, por ejemplo, [Springel et ál., 2001](#); [Knollmann y Knebe, 2009](#); [Behroozi et ál., 2013b](#), entre los más utilizados actualmente). Esto se realiza mediante la búsqueda de sobredensidades de partículas en cada una de las instantáneas almacenadas, que además cumplan el criterio de ligadura energética necesario para constituir un sistema. Es común que muchos buscadores de halos utilicen como base estrategias que evalúan la cercanía de cada partícula con sus vecinas, tales como el algoritmo *amigos de los amigos* (*friends-of-friends*, FOF, [Huchra y Geller, 1982](#); [Davis et ál., 1985](#)), el cual suele ser refinado con distintas estrategias para definir un halo de forma consistente. Luego estas detecciones son caracterizadas mediante el perfil de densidad de las partículas que fueron asociadas a cada una, de modo que es necesario establecer un criterio para el borde del halo según la Ecuación (1.21) para poder determinar todas sus propiedades (una buena comparación entre muchos de los buscadores de halos disponibles se puede encontrar en [Knebe et ál. 2011](#) y en [Onions et ál. 2012](#)). Como resultado de esto se obtiene una lista con todos los halos detectados, junto a sus respectivas propiedades, para cada una de las instantáneas almacenadas de la simulación.

Posteriormente, la evolución de los halos debe ser establecida realizando la vinculación de cada uno de ellos con sus correspondientes progenitores, es decir, conectando halos de diferentes instantáneas para establecer la correspondiente evolución temporal. Esto se decide conforme al número de partículas compartidas entre pares de halos pertenecientes a instantáneas consecutivas. Estos vínculos históricos deben ser depurados de modo que sean consistentes en el tiempo, sin vacíos históricos, y cada halo debe tener asignado un único descendiente para que las historias tengan una evolución coherente (una buena comparación de los códigos utilizados para este propósito se puede encontrar en [Srisawat et ál., 2013](#)). De este modo, este proceso termina por generar los árboles de fusiones de halos, requeridos por los modelos semi-analíticos para establecer de forma consistente la evolución de las galaxias.

1.4. Relevamientos extragalácticos

Una vez presentado el marco teórico y las estrategias utilizadas para crear muestras teóricas de poblaciones de galaxias, terminaremos este capítulo introductorio mencionando la conexión que tienen estos resultados teóricos con los obtenidos de los grandes

relevamientos extragalácticos, así como con su utilización y diseño.

El rol que cumplen los grandes relevamientos extragalácticos en la determinación de la teoría de formación de galaxias es fundamental. Los relevamientos con telescopios dedicados como el *Sloan Digital Sky Survey* (SDSS, ver [SDSS Collaboration, 2016](#), para la última publicación de datos de este proyecto) han permitido medir un sinnúmero de propiedades de la población de galaxias de forma estadística. Por ejemplo, en el Universo local el comportamiento de propiedades de galaxias (colores, densidades de masa y concentraciones) han sido caracterizados por el SDSS como “bimodales” en torno a galaxias con una masa estelar total de $3 \times 10^{10} M_{\odot}$, mostrando una división clara entre las galaxias azules activas y las rojas brillantes pasivas ([Kauffmann et ál., 2003](#)). Todas estas medidas estadísticas de la población de galaxias global imponen restricciones importantes a los modelos teóricos de formación (una buena revisión de las propiedades observadas por relevamientos se puede encontrar en [Madau y Dickinson, 2014](#)).

Esta relación de importancia es en ambas direcciones. Tanto el entendimiento de las observaciones realizadas por relevamientos extragalácticos como el diseño y planificación de los mismos requieren colecciones de datos de galaxias simuladas que tengan consistencia teórica y permitan generar predictibilidad. Este último punto es crucial, pues implica que relevamientos más vastos en términos de tamaño angular del cielo observado o de la resolución abarcada requieren muestras de galaxias más extensas ya sea en volumen, resolución de masa o número de propiedades consistentes obtenidas. Si bien esta demanda se puede satisfacer construyendo catálogos de forma simple y rápida con los modelos de ocupación de halos, estos no son capaces de brindar información física relevante de los mecanismos involucrados, ni tampoco son confiables para realizar predicciones en los rangos de propiedades que no han sido totalmente observados aún. Esto cobra vital importancia en la actualidad, dado los ambiciosos relevamientos extragalácticos que se están diseñando para entrar en operación a principios de la próxima década.

Por un lado tenemos la misión espacial *Euclid*, en desarrollo por la Agencia Espacial Europea, cuya fecha de lanzamiento se estima para el año 2020, con un tiempo de vida estimado de 7 años ([Laureijs et ál., 2011](#)). El objetivo de este proyecto es adquirir una mejor comprensión de la materia oscura y la energía oscura mediante mediciones precisas del crecimiento de las estructuras que componen el Universo. La recolección de datos se realizará en dos tipos de relevamientos: uno ancho que abarcará entre 15 000 y 20 000 deg^2 del cielo con una sensibilidad de hasta 24.5 magnitudes en el visual, y otro profundo que abarcará alrededor de 40 deg^2 de cielo siendo hasta dos magnitudes más profundo que el anterior. Entre sus principales metas científicas está el medir la tasa de expansión con una precisión lo suficientemente alta que permita distinguir entre la Relatividad General y las teorías de Gravedad Modificada, además de probar el paradigma de la materia oscura fría para el crecimiento de estructuras, así como restringir el valor del índice espectral n_s del espectro de potencias.

Para alcanzar estos objetivos, el relevamiento se enfoca en realizar mediciones precisas del corrimiento al rojo de las galaxias abarcadas, lentes gravitacionales débiles y concentración de galaxias. Esto fuerza a que se tenga un control muy preciso de las fuentes de errores sistemáticos en cada una de estas variables, para lo cual es necesario contar con catálogos de galaxias lo suficientemente grandes para que permitan reproducir el patrón de obtención de datos del telescopio. Por ejemplo, los conos de luz construidos en base a catálogos de galaxias creados con modelos semianalíticos sobre la simulación Millennium ([Bower et ál., 2006](#)) fueron utilizados para medir la dependencia de la selección por densidad en muestras de 1 deg^2 . No obstante, el tamaño y la cantidad disponible

de simulaciones de galaxias siguen siendo una limitante en esta materia, por lo que son necesarios en la actualidad un mayor número de modelos aplicados a nuevas simulaciones de N -cuerpos.

Por otro lado, también tenemos el *Gran Telescopio de Sondeos Sinópticos* (LSST, por sus siglas en inglés, [Ivezic et ál., 2008](#)) un nuevo tipo de telescopio dedicado exclusivamente a este relevamiento que entrará en funcionamiento en el 2022 y tendrá un tiempo de vida de 10 años. El telescopio, de 8.4 metros, tiene un campo de visión de 9.6 deg^2 (alrededor de 40 veces el tamaño de la luna) e incorpora una cámara de 3200 Megapíxeles. El relevamiento abarcará un área de alrededor de $20\,000 \text{ deg}^2$ utilizando un sistema propio de seis bandas anchas *ugrizy*, y sus exposiciones serán consecutivas de 15 segundos cada una. Esto permitirá que el área de sondeo se pueda cubrir por completo en un par de noches, y que cada campo sea capturado alrededor de 2000 veces durante el tiempo de vida del telescopio. Seguirá una estrategia de análisis de datos en tiempo real, los cuales se harán públicos desde el momento de captura. La gran cantidad de capturas de cada campo permite estudiar fenómenos transitorios tanto de la vecindad local como de la propia Vía Láctea, pero la profundidad y el gran campo del relevamiento también permitirá mejorar nuestra comprensión de las componentes de energía oscura y materia oscura del Universo, así como de la formación de galaxias y de la cosmología en general ([LSST Science Collaboration, 2009](#)).

En el diseño del relevamiento LSST las simulaciones han tomado un rol importante, las cuales en términos generales se utilizan de cuatro formas. La primera es para obtener un cálculo preciso del comportamiento esperado de la materia oscura para un modelo cosmológico particular, para lo cual las simulaciones de N -cuerpos son suficientes. El segundo uso es obtener una predicción detallada de la población de galaxias observada y otros observables en el relevamiento, incluyendo todos los efectos bariónicos. Tercero, los catálogos de galaxias simuladas proveen una herramienta esencial en la planificación de la estrategia de observación del relevamiento (por ejemplo, [Awan et ál., 2016](#)), la cual es susceptible de ser alterada conforme avancen las observaciones dependiendo de la necesidad, y además son utilizados para poner a prueba los aspectos técnicos cruciales en este relevamiento dado que la publicación de datos se hará en tiempo real. El último uso trascendental de simulaciones en el LSST será entregar un marco de referencia para interpretar los datos recolectados en el contexto de los modelos tanto cosmológicos como de la física bariónica. En todos los casos, no sólo se requieren simulaciones del tamaño adecuado, sino también una amplia variedad de ellas con el fin de reducir al mínimo los efectos estocásticos y particularidades de la física utilizada por cada modelo.

La cantidad de datos que generará este último relevamiento será abrumadora, tanto que la capacidad computacional actual no es capaz de generar simulaciones que satisfagan cabalmente los puntos mencionados. Por ello, han de realizarse esfuerzos múltiples para la construcción de catálogos de galaxias en base a las mayores simulaciones disponibles hoy en día.

Es en este contexto donde se centra este trabajo de Tesis. El poder posicionar nuestro modelo semianalítico SAG como una fuente vasta y competitiva en la formación de galaxias, así como en la generación de catálogos de galaxias públicos. Esto último tiene un impacto trascendental no sólo en nuestros propios análisis científicos, sino también como una contribución al desarrollo de conocimiento global del área de formación y evolución de galaxias pronta a sufrir una nueva revolución con los nuevos datos que estarán disponibles en los años venideros gracias a estos nuevos relevamientos.

Capítulo 2

Modelo de formación y evolución de galaxias SAG

Para el desarrollo de esta Tesis hemos utilizado en extenso el modelo semi-analítico de formación y evolución de galaxias SAG, cuyo nombre proviene del acrónimo *Semi Analytic Galaxies*. Los investigadores que mantienen el código SAG que implementa este modelo lo han facilitado para uso exclusivo en este trabajo.

Este modelo semianalítico SAG tiene más de una década de desarrollo por el grupo de investigación del que he formado parte durante el desarrollo de este trabajo de Tesis, y en base a él se han realizado más de una decena de trabajos de investigación en el área. Está basado originalmente en el modelo semianalítico de formación y evolución de galaxias presentado por [Springel et ál. \(2001\)](#), el cual es comúnmente conocido como modelo de *Munich*. El código con el cual se implementó el modelo de *Munich* fue refinado mediante la implementación de un tratamiento más detallado del enriquecimiento químico de las diversas componentes de las galaxias debido a los procesos de realimentación ([Cora, 2006](#)), la inclusión del crecimiento de los agujeros negros centrales supermasivos de las galaxias mediante inestabilidades de discos, acreción de gas y realimentación debido a los núcleos galácticos activos ([Lagos et ál., 2008](#)), la consideración de la evolución detallada de los tamaños de los discos ([Tecce et ál., 2010](#)) y de cómo estos son afectados por los cambios de momento angular debido a la acreción de material ([Padilla et ál., 2014](#)), y a la inclusión de un tratamiento de brotes estelares extendidos en el tiempo ([Gargiulo et ál., 2015](#)). El código además se ha utilizado en combinación con simulaciones hidrodinámicas para estimar la contribución a la metalicidad del medio intracúmulo ([Cora et ál., 2008](#)) y estimar el efecto de presión de barrido sobre galaxias satélites ([Tecce et ál., 2010, 2011](#)). Se ha utilizado también para entender el impacto de considerar una función inicial de masa integrada galáctica en la evolución química de las galaxias ([Gargiulo et ál., 2015](#)).

Los parámetros libres utilizados en todas las relaciones incluidas en el código pueden ser ajustados mediante el algoritmo de optimización por enjambre de partículas (PSO, por sus siglas en inglés: *Particle Swarm Optimization*), imponiendo, como restricciones, estadísticas globales de propiedades observadas de galaxias.

Entre las opciones del modelo disponibles, el código SAG puede adicionalmente calcular las distribuciones de energía espectral de las galaxias de forma detallada, así como también obtener un conjunto de líneas de emisión de las galaxias en combinación con el código MAPPINGS-III ([Orsi et ál., 2014](#)).

Este código no sólo se ha utilizado para estudiar a nivel teórico procesos físicos que predominan en la evolución de las galaxias, sino que también se ha utilizado para generar

catálogos realistas utilizables para otros fines. Por ejemplo, estos catálogos se han utilizado para el desarrollo del método de *Count Matching* (Muñoz Arancibia et ál., 2015), estrategia propuesta para modelar galaxias con emisión submilimétrica que utiliza conos de luz construidos con catálogos de galaxias simuladas. También se han utilizado como base para alimentar programas destinados a estudiar estrategias de muestreo de cielo durante la planificación de relevamientos como el LSST (Awan et ál., 2016). No obstante, la publicación de catálogos masivos abiertos al uso general es un área en la que el modelo semianalítico SAG está comenzando a indagar mediante esta Tesis.

Durante el desarrollo de este trabajo de Tesis, el código y sus componentes (bloques de cálculo que lo constituyen) han sido modificados para lograr los objetivos planteados. Todos estos cambios, sus justificaciones y los resultados correspondientes serán detallados en extenso durante este capítulo y los siguientes.

El presente capítulo se organiza de la siguiente forma: se iniciará con una descripción de la física que modela el código semianalítico, para luego centrarse en una descripción técnica del mismo enfocada ya sea en la descripción del funcionamiento global del código como de su organización interna. Posteriormente, se detallarán los cambios más importantes realizados como parte de esta Tesis que involucran tanto adaptaciones a nuevos tipos de simulaciones como mejoras particulares en rendimiento de cálculo realizado por el código. Finalmente, se terminará describiendo las contribuciones anexas que se han desarrollado para el uso cabal del modelo semianalítico en estos nuevos tipos de simulaciones considerados.

2.1. Descripción del modelo SAG

El código semi-analítico SAG implementa un modelo numérico que resuelve la formación y evolución de galaxias, permitiendo construir muestras de galaxias con consistencia histórica. Esto último implica que las galaxias obtenidas para una instantánea determinada constituyen ya sea los progenitores o descendientes de las galaxias obtenidas para las otras instantáneas. Cada una de las galaxias está constituida por un grupo discreto de *componentes* (por ejemplo: disco, bulbo, gas frío, halo de gas caliente, etc.) que son caracterizadas individualmente por propiedades físicas únicas (por ejemplo: masa total, posición, radio característico, masas de elementos químicos, luminosidad, etc.), constituyendo este conjunto de datos las *propiedades* de cada galaxia. El modelo vincula las propiedades de las galaxias que calcula estableciendo relaciones físicas entre ellas mediante la resolución de ecuaciones diferenciales simples que se van integrando durante el tiempo cósmico. En general, estas relaciones son extraídas de resultados ya sea de observaciones u otros modelos teóricos numéricos. Por este motivo, algunos autores prefieren utilizar el término semiempírico para referirse a este tipo de modelos, mas la estrategia utilizada es la misma.

El código semianalítico se encarga de modelar los procesos físicos bariónicos partícipes en la formación y evolución de galaxias. No obstante, su margen de resolución no incluye la formación de halos de materia oscura, su crecimiento ni su evolución en un contexto cosmológico. Debido a esto, el modelo necesita que le sea suministrada *a priori* toda la información correspondiente a las propiedades de un conjunto de halos de materia oscura que evolucionan en un contexto cosmológico. Estas propiedades se utilizan como condiciones iniciales sobre las cuales se integran las relaciones que modelan el comportamiento de la componente bariónica correspondiente, es decir, las galaxias que deben poblar estos

halos.

Si bien es posible crear con Montecarlo un conjunto de halos de materia oscura que se comporten de forma aproximada a las predicciones del modelo cosmológico estándar mediante el formalismo de Press-Schechter Extendido (Press y Schechter, 1974; Bond et ál., 1991), incluyendo además sus historias de fusión de forma consistente (Lacey y Cole, 1993), el modelo SAG está acondicionado para utilizar únicamente la información de halos de materia oscura extraídos de una simulación cosmológica de N -cuerpos. En particular, el código utiliza por defecto las listas de halos e historias de fusiones que son construidas con el código de búsqueda de subhalos SUBFIND desarrollado y presentado por Springel et ál. (2001).

La información requerida por el código, correspondiente a los halos de materia oscura, suele ser agrupada en un conjunto de catálogos asociados a diversas instantáneas de la simulación. La elección del espaciado temporal dado por el corrimiento al rojo entre cada instantánea es libre, pero debe ser medianamente regular y lo suficientemente pequeño para no agregar demasiada incerteza numérica durante el cálculo. Esto debido a que el código SAG usa un número fijo de pasos temporales N_{steps} entre cada par de instantáneas para realizar las integraciones y evaluaciones correspondientes. Regularmente, se suelen utilizar entre 90 y 130 instantáneas extraídas de simulaciones, espaciadas temporalmente de forma logarítmica en términos del factor de expansión, y distribuidas entre $z = 0$ y un corrimiento al rojo lo suficientemente alto que permita trazar la formación de los primeros halos resueltos por la resolución de la simulación utilizada. Lamentablemente, la elección del espaciado temporal y, por lo tanto, del número de instantáneas quedan regularmente supeditados a las capacidades computacionales y necesidades que motivaron inicialmente la realización de la simulación de N -cuerpos. Esta última es la que traza el comportamiento de la materia oscura, siendo este proceso al menos un orden de magnitud más costoso en recursos que el procesamiento posterior mediante códigos de búsqueda de halos o modelos semianalíticos.

Adicionalmente a las listas de halos por instantánea, el modelo requiere la información de las historias evolutivas de cada uno de ellos, popularmente conocidos como *árboles de fusiones de halos* (*merger trees*). Estos árboles de fusiones son construidos mediante la asignación de conexiones históricas entre halos de instantáneas consecutivas, de modo que los más antiguos sean los *progenitores* de los más recientes, los cuales son llamados *descendientes*. Si bien un halo puede poseer una cantidad arbitraria de progenitores (incluso hasta carecer de ellos, como en el caso de los halos recién formados), todos deben tener asociado un único descendiente para que la formación de las galaxias tenga coherencia. Para lograr esto, un halo descendiente se define como aquel que comparte más del 50% de las partículas con su progenitor. Estos dos ingredientes en conjunto, las listas de halos y los árboles de fusiones, constituyen la única información que requiere el modelo SAG para resolver el contenido de materia bariónica de los halos.

Si bien durante este trabajo utilizaremos el término “halo” de forma genérica, es importante notar que podemos clasificarlos según su localización o jerarquía espacial dependiendo de cómo son detectados por los buscadores de halos. Según esto, los halos pueden clasificarse como:

- *halos principales* aquellas sobredensidades que no se encuentran embebidas dentro de otros halos, y pueden o no contener subhalos dentro de ellas, y
- *subhalos* o *halos satélites* aquellas sobredensidades que fueron detectadas dentro de otros halos.

La asignación de categoría de cada sobredensidad depende mucho del algoritmo utilizado por el buscador de halos. En el caso de SUBFIND, todos los sistemas de halos tienen asignado como mínimo un halo principal y un subhalo asociado a este. Además, todos los otros subhalos que puedan residir dentro del halo principal constituyen el conjunto de satélites del sistema. Esta jerarquización se debe a que el algoritmo que realiza la identificación se ejecuta en dos etapas: una primera donde se localizan los halos no anidados mediante el método *amigos de los amigos* (FOF, por su nombre en inglés *friends-of-friends*), y una segunda donde se identifican las subestructuras (subhalos) que residen en ellos. De este modo, siempre queda al menos un subhalo asignado en cada sistema inclusive cuando estos están compuestos por una sola sobredensidad. En la terminología de SUBFIND, a los halos encontrados en la primera etapa se les suele llamar *halos FOF*, mientras que los de la segunda simplemente *subhalos*. El subhalo más masivo en cada sistema es marcado como el *subhalo principal*.

Siguiendo este criterio, tal como fue definido por Springel et ál. (2001), son sólo los subhalos detectados en la segunda etapa de identificación los que hacen de anfitriones de las galaxias que les son modeladas por los códigos semianalíticos. Como es complicado mantener dos definiciones de tipos de halos que, en definitiva, responden a métodos fenomenológicos distintos, trataremos a los halos FOF y sus respectivos subhalos principales como constituyentes de una estructura única correspondiente a un halo principal. Asimismo, todos los otros subhalos detectados dentro del halo FOF constituirán el sistema de satélites correspondiente. De este modo, serán estas estructuras (halos principales y satélites) las que albergarán las galaxias del modelo.

El modelo se encarga de asignar una galaxia central por cada halo nuevo que aparezca en los árboles de fusiones utilizados, las cuales evolucionarán dependiendo del comportamiento de las propiedades de sus halos respectivos y sus historias de fusiones. Según lo anterior, en un instante cualquiera las galaxias que constituyen la población generada por el modelo pueden ser de tres tipos: las tipo 0 correspondientes a galaxias centrales de halos principales (que son las centrales de cada sistema), las tipo 1 que son galaxias centrales pero de halos satélites, y las tipo 2 correspondientes a las galaxias cuyo halo de materia oscura ha desaparecido debido a que se fusionó por coalescencia con otro halo o se destruyó, pero la galaxia aún sobrevive. Este último tipo de galaxias es llamado galaxias *huérfanas*, y tanto los halos a los que estas pertenecen como la galaxia alrededor de la cual orbitan son identificados de forma consistente por el modelo. La justificación de la persistencia de las galaxias huérfanas y su tratamiento serán presentados en detalle en el Capítulo 3.

Esto nos lleva a que todas las galaxias del modelo tengan asignado un halo de pertenencia directo y un halo principal. Las tipo 0 y 1 estarán además asociadas a un halo de materia oscura del cual serán centrales, y las tipo 1 y 2 tendrán asociada una galaxia central alrededor de la cual orbitarán. Esto se puede apreciar más claramente en el esquema presentado en la Figura 2.1, donde se muestra un conjunto de galaxias, cada una con su tipo correspondiente, que habitan un sistema de tres halos compuestos por uno principal (la circunferencia mayor) y dos satélites. Las líneas continuas marcan la asociación de las galaxias centrales tipo 0 y 1 con sus halos respectivos, las líneas de puntos marcan la pertenencia de cada galaxia con el halo principal (el mismo para todas), y las líneas a trazos con flecha indican la relación entre galaxias satélites y centrales.

En términos generales, el código incluye dos tipos de tratamientos para la evolución de galaxias: un *modelo básico* que considera las relaciones físicas características utilizadas por la mayoría de los semianalíticos que se han convertido en un estándar del modelo

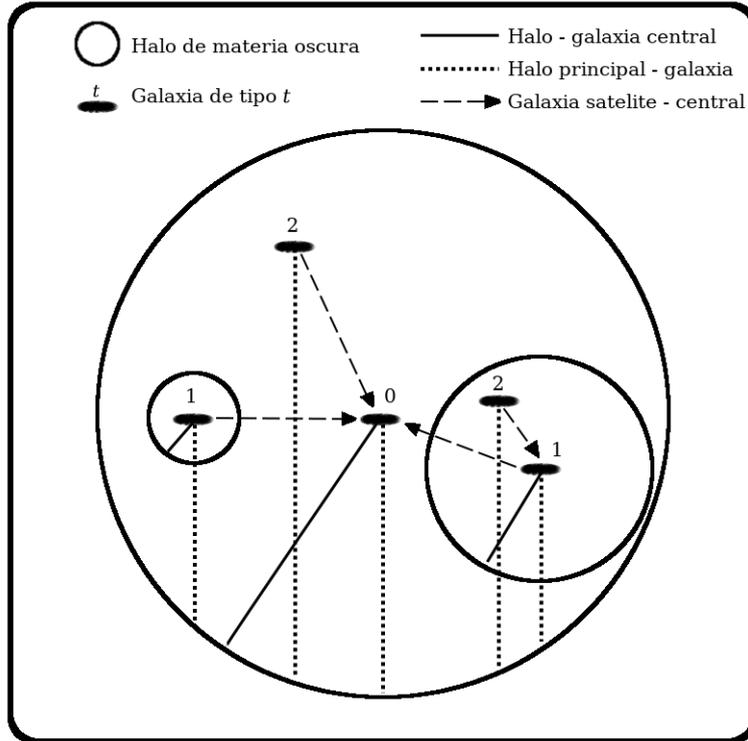


Figura 2.1: Esquema que representa los diferentes tipos de galaxias en un halo principal, con sus respectivas asociaciones: el halo principal de cada galaxia (línea de puntos), el halo directo para las tipo 0 y 1 (línea continua) y la galaxia central para las satélites 1 y 2 (líneas a trazos con flecha).

SAG, y un *modelo extendido* que considera un conjunto de tratamientos complementarios que pueden ser activados de forma opcional y pueden ya sea reemplazar relaciones del modelo básico, o agregar nuevas relaciones al paquete activado. En las siguientes secciones describiremos en detalle los procesos físicos considerados en cada uno de los casos.

2.1.1. Modelo básico

En el código semianalítico SAG las propiedades de las galaxias son determinadas según los procesos físicos incluidos en el modelado. La matriz básica de procesos que se consideran generalmente incluye: los efectos del enfriamiento del gas caliente debido a pérdidas radiativas, la formación estelar, la producción de metales, la realimentación energética y química por explosiones de supernovas, el crecimiento de agujeros negros centrales, la realimentación por núcleos galácticos activos, inestabilidades de disco y fusiones de galaxias.

Basado en las historias de fusión de los halos, la masa de gas caliente de un halo principal es calculada al comienzo de la evolución entre dos instantáneas consecutivas de la simulación. Se supone que el gas caliente siempre sigue una distribución semejante a la de la masa del halo de materia oscura en el que reside, cuya masa virial cambia de una instantánea a otra debido al crecimiento jerárquico de la estructura. Si consideramos que las galaxias que habitan este halo ya han formado estrellas, y tienen gas frío y agujeros

negros centrales, la masa total de gas caliente queda definida por

$$M_{\text{hot}} = f_b M_{\text{vir,tot}} - \sum_i \left[M_{\star}^{(i)} + M_{\text{cold}}^{(i)} + M_{\text{BH}}^{(i)} \right], \quad (2.1)$$

donde f_b es la fracción bariónica universal definida por la cosmología utilizada en la simulación, $M_{\text{vir,tot}}$ es la masa virial total del halo principal (considerando también sus satélites) obtenida del catálogo, y $M_{\star}^{(i)}$, $M_{\text{cold}}^{(i)}$ y $M_{\text{BH}}^{(i)}$ corresponden a las masas estelares, de gas frío y agujeros negros centrales, respectivamente, para cada galaxia i que reside en el halo de materia oscura. En este caso, se considera el modelo más simple donde las galaxias satélites no retienen su halo de gas caliente una vez que dejan de ser centrales aisladas y pasan a estar embebidas en otros sistemas, estrategia comúnmente conocida como “estrangulación instantánea”.

El gas caliente atrapado por los halos de materia oscura puede alcanzar temperaturas superiores a $T = 10^4$ K (White y Rees, 1978). Este gas sufre enfriamiento por pérdidas radiativas siguiendo una función de enfriamiento $\Lambda(T, Z)$ que depende fuertemente de su metalicidad Z y de la temperatura virial del halo $T = 35.9(V_{\text{vir}}/\text{km s}^{-1})^2$ K (White y Frenk, 1991). El modelo utiliza las tasas de enfriamiento estimadas por Foster et ál. (2012). Con ello, y adoptando una distribución isoterma para la distribución del gas caliente dentro del halo ($\rho_g(r) \propto r^{-2}$), se puede estimar el tiempo de enfriamiento local $t_{\text{cool}}(r)$ definido como la razón entre la energía térmica específica del gas y la tasa de enfriamiento por unidad de volumen como

$$t_{\text{cool}}(r) = \frac{3}{2} \frac{\bar{\mu} m_p k T}{\rho_g(r) \Lambda(T, Z)}, \quad (2.2)$$

donde $\bar{\mu} m_p$ es la masa media de las partículas y k la constante de Boltzmann. Con ello se puede definir un radio de enfriamiento r_{cool} como el radio para el cual t_{cool} es igual al tiempo en que el halo ha sido capaz de realizar enfriamiento radiativo de forma cuasi-estática. Este tiempo puede ser aproximado a la escala temporal dinámica del halo dada por $R_{\text{vir}}/V_{\text{vir}}$, donde R_{vir} y V_{vir} corresponden al radio y velocidad viriales, respectivamente. De este modo, la cantidad de gas enfriado en el tiempo queda definida como

$$\frac{dM_{\text{cool}}}{dt} = 4\pi \rho_g(r_{\text{cool}}) r_{\text{cool}}^2 \frac{dr_{\text{cool}}}{dt}. \quad (2.3)$$

El gas que se ha enfriado pasa a constituir los discos de gas de las galaxias, cuya formación y perdurabilidad depende del parámetro adimensional de espín de halos λ . Dado que sólo aquellas estructuras que son dinámicamente estables pueden sobrevivir en el tiempo, el radio característico (*scale radius*) $r_{\text{s,d}}$ que determina el tamaño de los discos puede ser calculado ya sea utilizando la estimación de Mo et ál. (1998) donde $r_{\text{s,d}} = \lambda R_{\text{vir}}/\sqrt{2}$, o estableciendo criterios de estabilidad más estrictos que también dependen de las masas de otras componentes del sistema (Tecce et ál., 2010).

La formación estelar suave (tipo *quiescent*) ocurre en los discos galácticos que han sido formados por este gas frío. En SAG se utiliza el modelo simple de Croton et ál. (2006) que establece una masa crítica mínima de gas frío necesaria para la formación estelar, definida como

$$M_{\text{cold,crit}} = 3.8 \times 10^9 \left(\frac{V_{\text{vir}}}{200 \text{ km s}^{-1}} \right) \left(\frac{3r_{\text{s,d}}}{10 \text{ kpc}} \right) M_{\odot}, \quad (2.4)$$

de modo que todo exceso de gas frío disponible puede formar estrellas según

$$\frac{dM_\star}{dt} = \alpha \frac{M_{\text{cold}} - M_{\text{cold,crit}}}{t_{\text{dyn}}}, \quad (2.5)$$

donde M_{cold} es el gas frío disponible en el disco, $t_{\text{dyn}} = V_{\text{vir}}/3r_{\text{s,d}}$ es el tiempo dinámico del disco de la galaxia y α es un parámetro libre que regula la eficiencia de este proceso.

La formación de esferoides o bulbos galácticos, por otro lado, se ve determinada por eventos que afectan la estabilidad de la galaxia a gran escala y generalmente ocasionan brotes de formación estelar intensos. Este mecanismo es gatillado por procesos tales como fusiones de galaxias, las cuales pueden clasificarse como *mayores* si la razón entre la masa de la galaxia satélite M_{sat} y la masa de la galaxia central M_{cen} cumple que $M_{\text{sat}}/M_{\text{cen}} > 0.3$, o *menores* en caso contrario. Si la fusión es mayor, ocasiona un brote estelar donde el gas de la galaxia remanente es consumido rápidamente, y el disco estelar se relaja y es transferido al bulbo. En cambio, las fusiones menores en el rango $0.05 < M_{\text{sat}}/M_{\text{cen}} \leq 0.3$ sólo desencadenan brotes estelares dependiendo del cociente entre el gas frío y la masa de disco (M_{disc}) de la galaxia central. Tal como se define en Lagos et ál. (2008), si $M_{\text{cold}}/M_{\text{disc}} > f_{\text{burst}}$ en la galaxia central, con $f_{\text{burst}} = 0.6$, la fusión menor también desencadena un brote de formación estelar. En el resto de las fusiones menores, sólo las estrellas de la galaxia satélite involucrada en el proceso son transferidas a la componente esferoidal de la galaxia central.

Otro mecanismo considerado en el modelo que puede producir brotes de formación estelar son las inestabilidades de disco (Lagos et ál., 2008). Durante el proceso evolutivo de las galaxias, se define si un disco pasa a ser inestable al igual que en Mo et ál. (1998), en base al criterio de Efstathiou-Lake-Negroponte (Efstathiou et ál., 1982). Este criterio permite cuantificar la estabilidad de la formación de barras en discos y es estimado en base a la comparación entre la autogravedad del mismo con la del halo que lo contiene. Según este criterio, definimos

$$\epsilon_d = \frac{V_{\text{máx}}}{(GM_{\text{disc}}/r_{\text{s,d}})^{1/2}}, \quad (2.6)$$

donde $V_{\text{máx}}$ es la velocidad circular máxima del disco, y M_{disc} la masa del disco considerando tanto el gas frío como las estrellas que lo componen. Dado que la curva de rotación se vuelve plana a grandes distancias, el valor de $V_{\text{máx}}$ se puede estimar en $r \sim 3r_{\text{s,d}}$ (Tecce et ál., 2010). Con ello, cuando $\epsilon_d \leq \epsilon_{\text{thresh}}$, con $\epsilon_{\text{thresh}} \simeq 1.0$ fijado como valor crítico, el disco se vuelve inestable y es susceptible a colapsar mediante un brote de formación estelar bajo la presencia de un perturbador externo tal como una galaxia satélite. Para mantener escepticismo sobre el nivel de acción del perturbador, el modelo incluye un parámetro libre adicional f_{pert} que permite cuantificar el impacto de un sistema de satélites que habitan un halo sobre un disco inestable, mediante el cálculo de la separación media entre las galaxias satélites \bar{d}_{sat} . De este modo, el colapso de los discos inestables sólo sucede si la separación media de las satélites cumple con la condición

$$\bar{d}_{\text{sat}} < f_{\text{pert}} r_{\text{s,d}}. \quad (2.7)$$

El modelo SAG no trata los brotes estelares como eventos instantáneos en el tiempo, sino que son tratados como eventos que tienen lugar en una escala temporal durante la cual el gas involucrado es llevado al centro de las galaxias y es consumido gradualmente en varios pasos con una tasa de formación estelar constante. Esto permite que el gas pueda

ser contaminado químicamente mientras es partícipe de este mecanismo (Gargiulo et ál., 2015). Este gas frío consumido en forma gradual constituye una nueva componente de la galaxia comúnmente asociada al gas frío del bulbo, pero que en este caso corresponde al gas en transición símil a una barra del disco. Cabe notar que, debido a la naturaleza y resolución de este tipo de modelos, no podemos definir con certeza la presencia de este tipo de estructuras como parte de las galaxias. Para efectos prácticos este gas sigue siendo considerado parte del disco aunque se lo asigne a una componente independiente.

El último ingrediente del modelo básico lo constituyen los mecanismos de realimentación, ya sea debido a las explosiones de supernovas (SNe) que se producirán según las poblaciones estelares formadas, o a los núcleos galácticos activos (AGN, por sus siglas en inglés *Active Galactic Nuclei*). Cada vez que se tiene formación estelar, esta sigue una función inicial de masa específica que permite estimar cuáles serán las contribuciones de los diversos rangos de masa estelares, y consecuentemente, la cantidad de SNe que estas provocarán ya sea de tipo Ia o de colapso de núcleo que agrupa los tipos Ib/c y II. Las SNe, además de reciclar el material que fue procesado en las estrellas, también inyectarán energía al medio provocando el recalentamiento del gas frío, reduciendo así el gas disponible para la formación estelar. La cantidad de gas recalentado ΔM_{reheat} depende de la cantidad n_{SN} de SNe generadas por una población estelar de masa ΔM_{\star} , según

$$\Delta M_{\text{reheat}} = \frac{4}{3} \epsilon \frac{n_{\text{SN}} E_{\text{SN}}}{V_{\text{vir}}^2} \Delta M_{\star}. \quad (2.8)$$

Aquí, $E_{\text{SN}} = 10^{51} \text{erg s}^{-1}$ es la energía liberada por una supernova, y ϵ un parámetro adimensional libre que regula la eficiencia de este mecanismo.

Dado que ambos tipos de SNe tienen tiempos de explosión y contribuciones químicas diferentes, en cada paso de integración se debe estimar la cantidad n_{SN} que producirá el recalentamiento de gas frío según la contribución instantánea de los dos tipos de SNe considerados en el modelo. En el caso particular de las SNe Ia, es necesario incluir otro parámetro que regule la cantidad de sistemas binarios que contribuyen con este tipo de supernovas según la cantidad de estrellas formadas por cada rango de masa en los diversos eventos de formación estelar. Este factor, usualmente llamado fracción de binarias f_{bin} , es considerado dentro de los parámetros libres del modelo. Por otro lado, el recalentamiento debido a las SNe asociadas a los eventos de formación estelar producidos en el gas frío del bulbo, no necesariamente tendrá la misma eficiencia que el del disco dado su estado de transición presumiblemente con distribución espacial inhomogénea. Por ello, para las SNe de esa componente se asigna un segundo parámetro libre ϵ_b en la Ecuación (2.8), para regular la eficiencia del proceso de forma independiente en esa componente.

Los AGNs también son un mecanismo de realimentación que afecta fuertemente el comportamiento de la formación estelar en el modelado de la evolución de galaxias, debido a que las eyecciones en forma de chorros que se producen pueden transportar grandes cantidades de energía hacia el gas caliente que rodea la galaxia y afectar el enfriamiento de ese gas, en especial en galaxias masivas. El principal conductor de este mecanismo es la acreción de masa sobre el agujero negro supermasivo (SMBH, por sus siglas en inglés *Super Massive Black Hole*) que se localizaría en el centro de las galaxias. Por consiguiente, en el modelo están consideradas estrategias de crecimiento de esta componente de las galaxias (Lagos et ál., 2008). Dado que hay dos mecanismos de creación de estrellas, mediante la formación suave y mediante brotes, el crecimiento de los SMBH también ocurre de forma distinta en los dos escenarios. Ambos mecanismos son incluidos en el modelo siguiendo los tratamientos descritos por Lagos et ál. (2008), basados en los propuestos por Croton

et ál. (2006), y son detallados a continuación.

En los procesos de formación estelar violentos como los desencadenados por fusiones de galaxias, suponemos que los SMBH de ambas se combinan de forma instantánea sin pérdida de energía, de modo que la masa del SMBH de la galaxia satélite es transferida al de la central. A su vez, una fracción del gas frío involucrado en el brote de formación estelar respectivo es depositada en el nuevo SMBH dependiendo de la razón entre las masas de las galaxias involucradas en la fusión, según

$$\Delta M_{\text{BH,mer}} = f_{\text{BH}} \left(\frac{M_{\text{sat}}}{M_{\text{cen}}} \right) \frac{M_{\text{cold,cen}} + M_{\text{cold,sat}}}{1 + (280 \text{ km s}^{-1}/V_{\text{vir}})^2}, \quad (2.9)$$

donde $\Delta M_{\text{BH,mer}}$ es la masa que contribuye al crecimiento del SMBH mediante este mecanismo, $M_{\text{cold,cen}}$ y $M_{\text{cold,sat}}$ corresponden a la masa de gas frío de las galaxias central y satélite respectivamente, y f_{BH} es un parámetro libre. De este modo, la masa que se transfiere al agujero negro central es proporcional a la cantidad total de gas frío partícipe en el proceso, pero su eficiencia disminuye en sistemas de menor masa (con $V_{\text{vir}} < 280 \text{ km s}^{-1}$), disminuye para fusiones desiguales y está globalmente regulada por f_{BH} . También debemos considerar que otro proceso capaz de desencadenar brotes estelares son las inestabilidades de disco, en cuyo caso el crecimiento del SMBH también es regulado por una relación análoga. Para ello, la masa que contribuye al SMBH por este mecanismo, $\Delta M_{\text{BH,DI}}$, la calculamos mediante la Ecuación (2.9) pero considerando $M_{\text{sat}}/M_{\text{cen}} = 1$. Este mecanismo de crecimiento del SMBH debido a brotes estelares es conocido como modo *cuásar*.

Por otro lado, durante los procesos de formación estelar suave, se supone que los SMBH pueden crecer mediante la acreción de gas frío durante la etapa de enfriamiento. Considerando que este proceso ocurre de forma continua y suave, la tasa de acreción de masa del SMBH se estima según

$$\frac{dM_{\text{BH}}}{dt} = \kappa \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \right) \left(\frac{M_{\text{hot}}}{10^{11} M_{\odot}} \right), \quad (2.10)$$

donde M_{BH} es la masa del SMBH de la galaxia y κ es un parámetro libre que regula la eficiencia de este mecanismo. Esta relación, propuesta por [Henriques et ál. \(2015\)](#), fue obtenida en base a un ajuste de la tasa de supresión de formación estelar esperada, siendo a su vez una revisión del modelo fenomenológico simple propuesto por [Croton et ál. \(2006\)](#) que tenía una eficiencia insuficiente para galaxias masivas en corrimientos al rojo bajos. El calentamiento mecánico del entorno generado por este proceso de acreción es conocido como la luminosidad del agujero negro L_{BH} , y puede ser expresada de la forma ([Soltan, 1982](#))

$$L_{\text{BH}} = \eta \left(\frac{dM_{\text{BH}}}{dt} \right) c^2, \quad (2.11)$$

donde c es la velocidad de la luz y η es la eficiencia estándar de producción de energía que ocurre en la vecindad del horizonte de eventos, usualmente fijada en $\eta = 0.1$. Este proceso puede inyectar energía suficiente al medio circundante llegando a reducir de forma considerable la eficiencia del enfriamiento del gas. Acorde con esto, la tasa de enfriamiento del gas se ve modificada por

$$\left(\frac{dM_{\text{cool}}}{dt} \right)' = \left(\frac{dM_{\text{cool}}}{dt} \right) - \frac{L_{\text{BH}}}{0.5V_{\text{vir}}^2}, \quad (2.12)$$

de modo que $(dM_{\text{cool}}/dt)'$ corresponde a la nueva tasa de enfriamiento, reducida debido a la energía inyectada al medio por el SMBH. Este proceso de realimentación por AGN es usualmente conocido como modo *radio*.

Todo este conjunto de mecanismos constituyen los ingredientes básicos del modelo semianalítico. El conjunto de opciones extra incluidas en el modelo permiten ya sea agregar procesos o funcionalidades por sobre los descritos en esta sección, o reemplazar mecanismos por unos más minuciosos. En la siguiente subsección detallaremos de forma general las opciones de modelado adicionales que se pueden activar en el código.

2.1.2. Modelo extendido

Durante los años de desarrollo del código SAG, varios colaboradores han contribuido en la inclusión de procesos, ya sea analíticos o fenomenológicos, con el fin de estudiar alguna propiedad particular de la población de galaxias resultante o para alcanzar mejores acuerdos con las observaciones de las predicciones generadas por el modelo. Durante esta sección detallaremos algunas de estas contribuciones, poniendo especial énfasis en aquellas que serán utilizadas para el desarrollo de este trabajo de Tesis.

Momento angular de discos

En la versión del código SAG presentado por Padilla et ál. (2014) se incluye un seguimiento simple de la evolución del momento angular de los discos de las galaxias, considerando cambios de los mismos producto de la acreción de material. Para ello, se asignan cambios de momento angular en cada paso de la evolución siguiendo distribuciones estadísticas medidas en la simulación Millennium II (Boylan-Kolchin et ál., 2009), suponiendo que los cambios de momento angular medidos en los subhalos de materia oscura también son sufridos por los bariones que residen en sus centros. En su prescripción original, cuando un disco sufre un cambio de momento angular entre dos instantes, medido en términos del ángulo de diferencia entre estos α_{flip} , el momento angular específico del disco se ve reducido según

$$\lambda_{\text{disc}} = \lambda_{\text{disc,old}} \cos(\alpha_{\text{flip}}), \quad (2.13)$$

donde $\lambda_{\text{disc,old}}$ y λ_{disc} corresponden al momento angular específico del disco antes y después del cambio de ángulo, respectivamente. Este ángulo α_{flip} acumularía los cambios que afectan a un disco individual de forma estadística, resultando en un decrecimiento continuo del valor de λ_{disc} a medida que la acreción de material ocurre, en comparación con el momento angular específico al momento de formación $\lambda_{\text{disc,ini}}$. De este modo, se define el factor $\eta_{\text{disc}} = \lambda_{\text{disc}}/\lambda_{\text{disc,ini}}$ que disminuye con cada evento de acreción. En caso de un brote de formación estelar, este parámetro es regresado al valor original $\eta_{\text{disc}} = 1$ debido a que la estructura completa del disco se ve alterada y el mecanismo se reinicia en su totalidad. Estos cambios son transferidos a las propiedades del disco mediante el cambio de tamaño del mismo producto de este mecanismo, según

$$r_{\text{s,d}} = \eta_{\text{disc}} r_{\text{s,d(orig)}} \quad (2.14)$$

medido en términos de radio característico del disco en el caso de no presentar estos cambios $r_{\text{s,d(orig)}}$. De este modo, este modelado permite impactar en la formación estelar y alterar la historia evolutiva de las galaxias.

Este tratamiento también requiere modificar el criterio para inducir brotes de formación estelar mediante fusiones, dado que establece un nuevo mecanismo para tratar este aspecto particular de la evolución de las galaxias. En este caso, los brotes de formación estelar que antes eran decididos según el criterio de masas entre las galaxias satélites y centrales, ahora son determinados por el criterio de inestabilidad de los discos. Esto se debe a que las fusiones ahora transfieren su impacto efectivo a los discos mediante el factor η_{disc} y los brotes resultantes por fusiones suceden ahora a través de las inestabilidades de discos. No obstante, esto se vuelve intrincado al momento de considerar los brotes extendidos en el tiempo en vez de instantáneos, dado que la sensibilidad temporal se vuelve un factor relevante al momento de asignar los cambios de momento angular. En este último caso, dos nuevos parámetros libres son introducidos con el fin de controlar esta nueva variable: una eficiencia f_η en el proceso de reducción del momento angular específico definido en la Ecuación (2.13), y una escala temporal característica η_τ en la que se aplicarán estos cambios según la distribución estadística utilizada para generarlos de forma aleatoria.

Efectos de ambiente en los satélites

Una mejora sustancial en el comportamiento de los colores de las galaxias, en particular de las fracciones de galaxias rojas y azules, se obtiene con la inclusión de los efectos de ambiente que se ejercen sobre las galaxias satélites tales como la presión de barrido y las fuerzas de mareas (Cora, Vega-Martínez et ál. en preparación). Para incluir estos efectos es necesario considerar dos cambios importantes al modelo básico: las galaxias deben ser capaces de retener su gas caliente una vez que se convierten en satélites de otro sistema, y las galaxias huérfanas deben tener posiciones bien modeladas dada la dependencia de estos efectos con la distancia radial al centro de los halos. Este último problema lo abordaremos en detalle en el Capítulo 3, por lo que en esta sección nos centraremos exclusivamente en las prescripciones asociadas a los efectos de ambiente.

En este nuevo escenario, la cantidad de gas caliente disponible para las galaxias centrales tipo 0 de los halos principales, que es calculado al principio de cada instantánea, ya no es obtenido mediante la Ecuación (2.1). En cambio, ahora se considera el gas caliente de las satélites, quedando

$$M_{\text{hot}} = f_b M_{\text{vir,tot}} - M_{\star,\text{cen}} - M_{\text{cold,cen}} - M_{\text{BH,cen}} - \sum_{i>2} (M_{\star}^{(i)} - M_{\text{cold}}^{(i)} - M_{\text{BH}}^{(i)} - M_{\text{hot}}^{(i)}), \quad (2.15)$$

donde $M_{\star,\text{cen}}$, $M_{\text{cold,cen}}$ y $M_{\text{BH,cen}}$ corresponden a la masa estelar, de gas frío y agujero negro central de la galaxia central del halo principal, y la suma se realiza sólo sobre las galaxias satélites del sistema incluyendo ahora la masa de gas caliente de cada una $M_{\text{hot}}^{(i)}$. Este reservorio de gas caliente de las galaxias satélites se irá modificando con el tiempo. Podrá reducirse paulatinamente tanto por el enfriamiento radiativo o por los efectos de ambiente que le harán perder masa. Asimismo, también podrá ser regenerado debido a los procesos de recalentamiento del gas frío en cada galaxia.

El modelo de pérdida de material por presión de barrido (RPS por sus siglas en inglés *Ram Pressure Stripping*) está basado en el utilizado por Font et ál. (2008), que usa el criterio de McCarthy et ál. (2008) según resultados de simulaciones hidrodinámicas. Esta presión de barrido afecta la componente gaseosa de las galaxias y es producida por el medio intragrupo o intracúmulo, el gas caliente y difuso contenido dentro del

profundo pozo de potencial de grupos y cúmulos. El halo de gas caliente de un satélite será arrancado si la presión de barrido (P_{ram}) excede la fuerza gravitacional restauradora por unidad de área, según

$$P_{\text{ram}} \equiv \rho_{\text{ICM}} v^2 > \alpha_{\text{RP}} \frac{GM_{\text{sat}}(r_{\text{sat}})\rho_{\text{hot}}(r_{\text{sat}})}{r_{\text{sat}}}, \quad (2.16)$$

donde ρ_{hot} es la densidad del gas del satélite, $M_{\text{sat}}(r_{\text{sat}})$ es la masa del satélite contenida en una esfera de radio r_{sat} , y α_{RP} es una constante geométrica usualmente elegida entre los valores 1 y 10 con el fin de ajustar los resultados de simulaciones hidrodinámicas. En el código SAG se adopta un valor intermedio para $\alpha_{\text{RP}} = 5$. Por otro lado, P_{ram} depende de la densidad del gas del medio intracúmulo ρ_{ICM} del halo anfitrión del satélite correspondiente, el cual se mueve con una velocidad relativa v a este medio. En el caso de simulaciones hidrodinámicas, estos valores pueden ser medidos directamente de las partículas de gas y la dinámica de las galaxias satélites, pero para simulaciones de materia oscura que carecen de esta información se puede utilizar un ajuste donde P_{ram} depende de la distancia del centro del halo, su masa y del corrimiento al rojo, como el propuesto por [Tecce et ál. \(2011\)](#).

Luego, la masa total de cada satélite queda definida por

$$M_{\text{sat}}(r_{\text{sat}}) = M_{\star} + M_{\text{cold}} + 4\pi \int_0^{r_{\text{sat}}} [\rho_{\text{hot}}(r) + \rho_{\text{DM}}(r)] r^2 dr \quad (2.17)$$

donde ρ_{DM} es el perfil de densidad de la materia oscura. Por simpleza, se pueden expresar ambos perfiles de densidad como esferas isotermas (donde $\rho(r) \propto r^{-2}$). Luego, combinando las Ecuaciones (2.16) y (2.17) se puede resolver numéricamente el radio del gas caliente debido al RPS, $r_{\text{hot}}^{\text{RPS}}$.

Por otro lado, para calcular la pérdida de material debido a fuerzas de marea (TS, por sus siglas en inglés *Tidal Stripping*), se supone que el gas se distribuye y comporta de manera análoga a la materia oscura. Considerando que el tamaño de los subhalos detectados y caracterizados ya incorpora este efecto naturalmente, podemos considerar el radio virial R_{vir} de los mismos como el radio de pérdida de gas $r_{\text{hot}}^{\text{TS}}$, de modo que todo el gas que yace fuera de este radio será arrancado. Si bien este valor está bien definido para las galaxias satélites tipo 1, las satélites huérfanas carecen de un halo directamente asociado que permita obtener este valor. Por ello, en este último caso se preserva el valor del radio virial que tenía el halo en el momento de convertirse en satélite.

De este modo, el actuar de los dos procesos en simultáneo, RPS y TS, permite definir el nuevo radio de gas caliente según

$$r_{\text{hot}}^{\text{new}} = \text{mín}(r_{\text{hot}}^{\text{RPS}}, r_{\text{hot}}^{\text{TS}}, r_{\text{hot}}^{\text{old}}) \quad (2.18)$$

donde $r_{\text{hot}}^{\text{old}}$ corresponde al valor del radio del halo de gas caliente previo a la acción de los procesos de ambiente, el cual inicialmente se fija en $r_{\text{hot}} = R_{\text{vir}}$.

Es importante destacar que, dado que ahora las galaxias satélites preservan su halo de gas caliente, el gas frío recalentado por procesos de realimentación en las mismas ahora puede ser retenido por sus propios halos de gas caliente en vez de terminar en el halo de la galaxia central del halo principal. Para regular esta circulación, se establece que las galaxias pueden alimentar su propio gas caliente sólo si este aún conserva una fracción $f_{\text{hot,sat}}$ de su masa bariónica, en caso contrario el material recalentado se expulsará a la fase caliente de su galaxia central. La reconstrucción del gas caliente de las galaxias

satélites se hace depositando la masa de gas recalentado, y su correspondiente contenido de metales, en forma proporcional a la fracción $f_{\text{hot,sat}}$; el resto del material se transfiere a la fase caliente de su galaxia central.

Una vez que el gas caliente de las galaxias satélites se reduce más allá que la fracción anterior, este sólo se ve afectado por el proceso de enfriamiento con lo cual disminuye hasta llegar a valores muy bajos. Cuando esto sucede, el disco de gas frío queda expuesto a la pérdida de masa por efectos de ambiente que son modelados de forma análoga al gas caliente. Para el caso de RPS, la fuerza gravitacional de restauración por unidad de área ahora se expresa en términos de las componentes del disco. Para esto seguimos el modelo presentado por [Tecce et ál. \(2010\)](#), cuyo criterio queda

$$P_{\text{ram}} \geq 2\pi G \Sigma_{\text{disc}}(r) \Sigma_{\text{cold}}(r) \quad (2.19)$$

donde Σ_{disc} y Σ_{cold} son la densidad superficial de masa del disco (incluyendo tanto las estrellas como el gas frío) y del gas frío, respectivamente. Ambas componentes pueden ser modeladas suponiendo un perfil de densidad superficial exponencial $\Sigma(r) = \Sigma_0 \exp(-r/r_{s,d})$, donde Σ_0 es la densidad superficial central de la componente particular. Utilizando esto junto al criterio de la Ecuación (2.19), el radio de gas frío debido a RPS, $r_{\text{cold}}^{\text{RPS}}$, queda definido por

$$r_{\text{cold}}^{\text{RPS}} = -0.5 r_{s,d} \ln \left(\frac{P_{\text{ram}}}{2\pi G \Sigma_{0,\text{disc}} \Sigma_{0,\text{cold}}} \right) \quad (2.20)$$

donde $\Sigma_{0,\text{disc}}$ y $\Sigma_{0,\text{cold}}$ son las densidades superficiales centrales del disco y sólo del gas frío, respectivamente.

Análogamente al caso del gas caliente, el efecto de las fuerzas de mareas es considerado mediante el radio del subhalo de materia oscura correspondiente, es decir, $r_{\text{cold}}^{\text{TS}} = R_{\text{vir}}$. De este modo, el gas frío perderá toda la masa cuya distancia radial cumpla que $r > \min(r_{\text{cold}}^{\text{RPS}}, r_{\text{cold}}^{\text{TS}})$. Asimismo, la escala característica del disco $r_{s,d}$ debe ser reducida apropiadamente.

Finalmente, las fuerzas de marea también pueden afectar la componente estelar de una galaxia, para lo cual hay que considerar el disco y el bulbo de forma separada. En el caso del disco, el criterio se reduce sólo a comparar el tamaño del disco estelar con R_{vir} por analogía a los casos anteriores, y se sigue el mismo procedimiento de reducción de tamaño que el aplicado al gas frío. No obstante, si hay bulbo presente es necesario considerar un perfil de densidad para el mismo con el fin de estimar la pérdida efectiva de masa del bulbo al momento de realizar la reducción. Para esto, en el modelo se considera que las estrellas del bulbo se distribuyen siguiendo el perfil propuesto por [Hernquist \(1990\)](#) y se replica el proceso aplicado a las otras componentes que se ven afectadas por estos mecanismos de pérdida de masa.

Nuevo modelo de realimentación

Con el fin de que el modelo pueda satisfacer restricciones a alto corrimiento al rojo, se han introducido modificaciones al modelo de realimentación ([Cora, Vega-Martínez et ál. en prep.](#)), tomando como referencia el trabajo de [Hirschmann et ál. \(2016\)](#) donde midieron el impacto de una serie de modelos de realimentación de SNe utilizando otro código semianalítico. Según ello, en el modelo SAG se incorporó una variante alternativa del tratamiento que muestra mejoras sustanciales en la evolución de la metalicidad de las galaxias. Esto gracias a la parametrización, dependiente del corrimiento al rojo, en el que

actúa el proceso de recalentamiento por las SNe, y a la modulación con la velocidad virial del halo en el que se encuentra la galaxia. Estas dependencias adicionales introducidas en el cálculo del material recalentado se desprenden de los ajustes presentados por [Muratov et ál. \(2015\)](#) en base al análisis de las simulaciones hidrodinámicas FIRE. Si bien tales ajustes dependen de la tasa de formación estelar en el corrimiento al rojo considerado, adaptamos los mismos a la estimación hecha en SAG que tiene en cuenta el número de SNe que explotan con cierto retardo de acuerdo al tiempo de vida del progenitor.

A partir de estas consideraciones, el nuevo modelo de realimentación utiliza una versión alternativa a la presentada en la Ecuación (2.8) para estimar la masa recalentada debido a la energía inyectada por las SNes, de la forma

$$\Delta M_{\text{reheat}} = \frac{4}{3} \epsilon \frac{\eta E_{\text{SN}}}{V_{\text{vir}}^2} (1+z)^{\beta_{\text{F}}} \left(\frac{V_{\text{vir}}}{60 \text{ km s}^{-1}} \right)^{\alpha_{\text{F}}(V_{\text{vir}})} \Delta M_{\star}, \quad (2.21)$$

donde el exponente $\alpha_{\text{F}}(V_{\text{vir}})$ toma los valores -3.2 y -1.0 para los valores de V_{vir} menores o mayores que 60 km s^{-1} , respectivamente. Por otro lado, si bien el ajuste obtenido por [Muratov et ál. \(2015\)](#) indica un valor para $\beta_{\text{F}} = 1.3$, en el modelo se deja este parámetro libre para permitir libertad en el comportamiento de esta relación con el tiempo.

El gas recalentado es transferido desde la componente de gas frío al halo de gas caliente, desde donde posteriormente retornará por el enfriamiento radiativo. No obstante, este tipo de modelado con circulación cerrada tiende a crear un exceso de masa estelar a alto corrimiento al rojo ([Guo et ál., 2011](#); [Henriques et ál., 2013](#); [Hirschmann et ál., 2016](#)), de modo que también se propone que parte del material que se ve afectado por este proceso debe ser eyectado de los halos. Según lo propuesto por [Guo et ál. \(2011\)](#), la cantidad de gas caliente que es eyectada ΔM_{ejec} se puede estimar mediante una comparación de las energías involucradas, según

$$\Delta M_{\text{ejec}} = \frac{\Delta E_{\text{FB}} - 0.5 \Delta M_{\text{reheat}} V_{\text{vir}}^2}{0.5 V_{\text{vir}}^2}, \quad (2.22)$$

donde ΔE_{FB} corresponde a la energía inyectada por las estrellas masivas, y se puede modelar de una forma similar a la masa recalentada:

$$\Delta E_{\text{FB}} = \frac{4}{3} \epsilon_{\text{ejec}} \frac{\eta E_{\text{SN}}}{V_{\text{vir}}^2} (1+z)^{\beta_{\text{F}}} \left(\frac{V_{\text{vir}}}{60 \text{ km s}^{-1}} \right)^{\alpha_{\text{F}}(V_{\text{vir}})} \Delta M_{\star} (0.5 V_{\text{SN}}^2). \quad (2.23)$$

Aquí, ϵ_{ejec} es la eficiencia específica de este proceso considerada como otro parámetro libre del modelo, y $0.5 V_{\text{SN}}^2$ es la energía cinética de la eyección por SNe por unidad de masa de estrellas formadas. Este valor se puede aproximar con el ajuste presentado por [Muratov et ál. \(2015\)](#), donde $V_{\text{SN}} = 1.9 V_{\text{vir}}^{1.1}$.

Finalmente, este nuevo reservorio de gas caliente eyectado de los halos puede aportar gas que retorna ya sea al halo de gas caliente de la galaxia que lo generó, si se considera el modelo que permite que las galaxias satélites retengan su gas caliente, o al halo de la galaxia central del halo principal, en caso contrario. Para esto, se considera que este material podrá ser reincorporado acorde a una escala temporal que depende inversamente con la masa M_{vir} del halo correspondiente, mediante

$$\Delta M_{\text{reinc}} = \gamma \Delta M_{\text{ejec}} \frac{10^{10} M_{\odot}}{M_{\text{vir}}}, \quad (2.24)$$

donde ΔM_{reinc} es la masa reincorporada y la constante γ regula la eficiencia del proceso que puede ser considerada como un parámetro libre del modelo o fijada en $\gamma = 0$ para desactivarlo. Esta relación, tal como es propuesta por [Henriques et ál. \(2013\)](#), permite reproducir la evolución de la función estelar de masa de las galaxias.

2.1.3. Cálculo de magnitudes

Además de las masas y metalicidades de cada componente de las galaxias obtenidas con el código SAG, el modelo puede calcular las luminosidades y respectivas magnitudes de cada componente mediante la estimación de la contribución energética de cada población estelar generada. Actualmente el código tiene implementados dos métodos distintos para realizar estos cálculos, ambos basados en modelos sintéticos de evolución espectral de poblaciones estelares ya sea de Bruzual y Charlot (2003), Maraston (2005) o Coelho et ál. (2007).

La primera estrategia incluida para realizar este cálculo es el modelo original tal como fue presentada en Lagos et ál. (2008), que consiste en estimar la contribución energética de cada población estelar formada para todas las instantáneas que se deseen calcular, según la edad correspondiente que las estrellas de la población tendrán al momento de la instantánea deseada. Para ello, se tabulan previamente las contribuciones energéticas según edad y metalicidad para un conjunto específico de once filtros fotométricos: los U , B , V , R , I , K del sistema Johnson y los filtros u , g , r , i , z del relevamiento SDSS, de modo que este método calcula este conjunto de magnitudes en su totalidad por defecto cuando es activado. Como la estimación de las contribuciones energéticas depende exclusivamente de las poblaciones estelares, se puede estimar independientemente la contribución de luminosidad de cada componente de la galaxia, en especial los discos y bulbos por separado. Cada una de estas contribuciones energéticas permite obtener la magnitud absoluta de cada componente de las galaxias para un observador en reposo, es decir, sin aplicar ninguna corrección debido al corrimiento al rojo de la instantánea respectiva.

El segundo método utiliza la misma estrategia que el anterior, pero en vez de leer los datos previamente tabulados para un conjunto particular de filtros, calcula la distribución de energía espectral de cada componente de las galaxias (SED, por sus siglas en inglés *Spectral Energy Distribution*), utilizando para esto un rango en longitud de onda y resolución a elección. Este método necesita remuestrear los modelos sintéticos en su totalidad sobre el rango energético deseado al inicio del cálculo. Por ejemplo, usualmente se suelen calcular espectros comprendidos entre los 90 Å y los 25 000 Å, utilizando 300 intervalos en longitud de onda. Si bien esta estrategia es más costosa en términos de cálculos, mediante la convolución de las SEDs con curvas de transmisión de filtros fotométricos se puede obtener la magnitud absoluta en cualquier filtro deseado, no sólo para un observador en reposo sino también para un observador en $z = 0$. Actualmente el código SAG incluye una extensa base de datos con más de 200 curvas de transmisión de filtros que se pueden utilizar para este propósito, siendo además un proceso trivial agregar nuevas. Por otro lado, el detalle de la distribución energética de las componentes de las galaxias permite a su vez obtener con precisión un conjunto de líneas de emisión de las galaxias con formación estelar mediante el método propuesto por Orsi et ál. (2014), que combina el código SAG con el código de fotoionización MAPPINGS-III (Dopita et ál., 2000; Levesque et ál., 2010). El conjunto de 13 líneas de emisión que se pueden estimar con este método incluye: H_α , H_β , $Ly-\alpha$, [N II] 205 μm , [N II] (654.8 y 658.4 nm), Ne III 387 nm, [O II] (372.7 y 372.9 nm), [O III] (495.9 y 500.7 nm), y [S II] (671.6 y 673.1 nm), cada una con su correspondiente cálculo de emisión en el espectro continuo.

Ambos tipos de modelos para calcular luminosidades, el clásico y el que utiliza las SEDs, pueden activarse tanto de forma individual como en conjunto. A su vez, a todas las luminosidades calculadas por el modelo se les puede incluir el efecto de la extinción debido al polvo siguiendo la implementación de De Lucia et ál. (2004).

2.1.4. Ajuste de parámetros libres

Los parámetros libres del modelo son ajustados mediante la comparación de las propiedades resultantes de la población de galaxias modelada contra un conjunto de observables aplicando una técnica estocástica conocida como optimización por enjambre de partículas (PSO, por sus siglas en inglés *Particle Swarm Optimization*). Este es un algoritmo de autoaprendizaje que utiliza un conjunto de agentes exploradores con conocimiento individual y colectivo para explorar y localizar regiones de probabilidad máxima en espacios multidimensionales de parámetros. La técnica resulta más eficiente en términos de costo computacional en comparación con otros métodos tales como los Montecarlo mediante cadenas de Markov (Ruiz et ál., 2015). Los datos utilizados para restringir los parámetros libres del modelo consideran relaciones fundamentales observadas entre las propiedades de las galaxias, tales como: la función de masa estelar en corrimientos al rojo $z = 0$ y $z = 2$, la relación entre la masa de los agujeros negros centrales y las masas de los bulbos de las galaxias, la función de la tasa de formación estelar versus la masa estelar, la fracción de gas frío, las funciones de luminosidad, entre otros. La restricción utilizada (o el conjunto de ellas) para calibrar depende del tipo de modelo que se quiera estudiar y la libertad que este tenga para reproducir las propiedades observadas.

Si bien el algoritmo que se utiliza es muy eficiente para encontrar el conjunto de parámetros que mejor reproduzca las restricciones impuestas, cada evaluación del modelo requerida en el proceso puede ser de un costo computacional variable dependiendo del tamaño de la simulación utilizada como entrada. Esto restringe la utilización de este método de ajuste sobre simulaciones completas a sólo cuando estas son muy pequeñas (a modo de referencia, una simulación pequeña en número de halos tendría del orden de hasta un par de decenas de miles de halos en la instantánea correspondiente a $z = 0$). Por ello, para poder encontrar el conjunto de parámetros óptimos en simulaciones mayores, se procede a encontrar una muestra volumétrica de la simulación que sea representativa de la misma con el fin de utilizarla para ejecutar el modelo en esta etapa y realizar el ajuste de parámetros sobre ella.

La selección de la muestra usualmente se hace buscando un subvolumen de la simulación cuya función de masa de halos tenga un comportamiento lo suficientemente similar al de la función de masa de la simulación completa, generalmente sobre la lista de halos detectados en la instantánea en $z = 0$. Tanto el tamaño de este subvolumen como el nivel de acuerdo de la función de masa del mismo quedan a criterio de elección libre dado que son decisiones que dependen fuertemente de la resolución de masa y espacial de la simulación original. Con ello, se procede a identificar todos los halos de las instantáneas anteriores que se localizan en el mismo volumen comóvil que los elegidos o que son parte de los árboles de fusiones de estos, con el fin de crear un nuevo conjunto de catálogos análogos a los originales pero sólo incluyendo estos halos.

Este nuevo conjunto de catálogos, usualmente llamado subvolumen o muestra de calibración, es utilizado únicamente para realizar el proceso de ajuste mediante el algoritmo PSO.

2.1.5. Licencia

El código semianalítico SAG es mantenido y administrado por la Dra. Sofía A. Cora, directora de esta Tesis doctoral. El código fuente es cedido libremente para el desarrollo científico sin restricciones de uso o modificación, únicamente bajo consentimiento de su desarrolladora.

Todas las modificaciones y usos del código SAG realizadas durante este trabajo se han efectuado bajo la autorización correspondiente.

2.2. Análisis técnico del código

Entre los productos derivados de este trabajo de Tesis está la generación de catálogos masivos de galaxias, con propósitos múltiples, mediante la aplicación del código SAG sobre simulaciones públicas que se destacan por ser de muy alta resolución en masa o que abarcan volúmenes excepcionalmente grandes. En términos prácticos, esto implica la necesidad de tener un código que pueda procesar datos de gran envergadura de una forma efectiva y eficiente, además de poder leer los tipos de catálogos que son suministrados por los creadores de estas simulaciones. Con este objetivo en mente, durante esta sección realizaremos un análisis técnico de funcionamiento del código SAG con el fin de detectar las limitaciones y modificaciones a introducir necesarias para concretar los objetivos planteados. Todas las mejoras y cambios realizados derivados de este análisis serán detalladas durante la sección siguiente.

El código semianalítico SAG sigue una estrategia de procesamiento de galaxias usualmente llamada “horizontal”, donde el cálculo se realiza mediante un bucle cuyos pasos procesan la evolución de la totalidad de la población de galaxias entre dos instantáneas consecutivas cada vez. En cada uno de los pasos, el análisis se realiza procesando cada galaxia de forma individual, manteniendo el criterio de que las centrales deben calcularse antes que sus respectivas satélites. Esto impone una restricción importante pues implica que todas las galaxias de una muestra deben ser analizadas de forma secuencial, lo cual hace que el tiempo de cálculo total crezca linealmente con el número de halos procesados.

La estrategia usada hasta ahora para poder realizar el cálculo sobre simulaciones de tamaño mediano, como las que han sido utilizadas en los últimos trabajos relacionados con el modelo SAG (ver, por ejemplo, [Ruiz et ál., 2015](#); [Gargiulo et ál., 2015](#)), es dividir el volumen de la simulación en una colección de n^3 subvolúmenes, de forma análoga a como se realiza la selección de la muestra para calibración. Este proceso consiste en tomar la instantánea a $z = 0$ y dividir sus lados en n partes, generando así n^3 subvolúmenes de aproximadamente la misma cantidad de halos cada uno. Luego, para cada uno de ellos se seleccionan todos los halos que están encerrados en el mismo volumen comóvil en cada instantánea, incluyendo los progenitores de estos que pertenezcan a otros subvolúmenes y excluyendo aquellos que están vinculados históricamente con halos de los subvolúmenes vecinos. Debido a la condición de borde periódica de las simulaciones utilizadas y la complejidad que pueden alcanzar las historias de fusiones de los halos, el proceso debe realizarse de forma cuidadosa y controlada con el fin de no cortar ningún árbol de fusiones, y de no repetir ni descartar halos. Cada subvolumen termina constituyendo un conjunto de catálogos separado, de modo que la simulación completa queda dividida en n^3 partes, las cuales pueden ser procesadas de forma independiente con copias del modelo que se ejecuten sobre ellas de forma simultánea.

Si bien este proceso permite efectivamente distribuir el cálculo y reducir el tiempo total de procesamiento en un factor n^3 , tiene la limitante de que no permite cambiar fácilmente el número de divisiones o procesos distribuidos, quedando esta cantidad fija una vez realizada la división. Esto a su vez impacta en el número de archivos de salida que generará el modelo, ya que al ser cada cálculo independiente con sus propios archivos de salida, la población completa terminará almacenándose en al menos n^3 archivos.

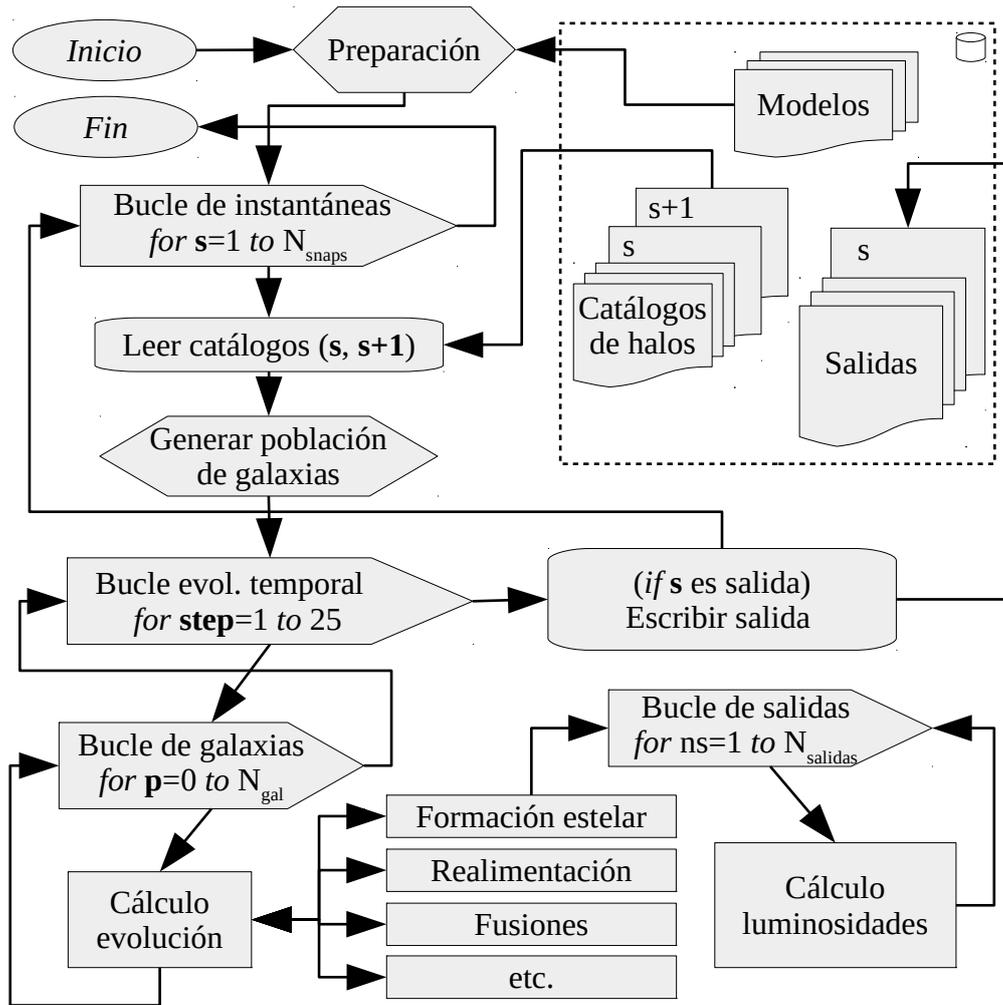


Figura 2.2: Diagrama de flujo de cálculo del código SAG en su versión 7.0, que engloba sólo las operaciones dominantes del procesamiento. En pentágonos se simbolizan los bucles en el cálculo, y en la línea punteada se encierran las operaciones de acceso a disco duro.

Para entender mejor la estrategia de cálculo del código, procedemos a estudiar el flujo global del cálculo dentro del mismo. Este análisis general es presentado en el diagrama de la Figura 2.2 en donde se indica mediante flechas el orden de las operaciones, las cuales se inician y terminan en la esquina superior izquierda del diagrama. Con rectángulos se simbolizan las operaciones de cálculo, con pentágonos los bucles y con hexágonos las operaciones de preparación. Con la línea punteada se encierran las operaciones de acceso (tanto lectura como escritura) a disco duro.

Como se puede apreciar en el diagrama, el flujo está dominado por tres grandes bucles que organizan la forma en que el cálculo es realizado. El primero es un bucle sobre todas las instantáneas de la simulación, partiendo desde la de corrimiento al rojo más alto hasta la correspondiente a $z = 0$. Al iniciar cada uno de estos pasos se procede a cargar los catálogos de halos de la simulación correspondiente a ambas instantáneas consecutivas involucradas, información que permite construir las historias de las galaxias y generar la población descendiente de las que han sido calculadas previamente. Luego de ello, se da lugar al segundo bucle correspondiente al avance temporal entre ambas instantáneas. Aquí, cada paso resuelve las relaciones conductoras de la evolución de galaxias avanzando

temporalmente una fracción $1/25$ veces el tiempo transcurrido entre las dos instantáneas correspondientes, dado que se suele usar un número fijo de pasos igual a 25. El tercer bucle anidado es el que recorre todas las galaxias a las que se les calcula su evolución. Este último nivel mantiene un orden jerárquico dada la prioridad que tienen las galaxias centrales por sobre sus respectivas satélites.

En el diagrama también se muestra que la evolución de las galaxias está dominada por varios procesos cuya preponderancia dependerá de las condiciones de cada una. No obstante, es importante notar que cada vez que se da lugar a la formación estelar (ya sea suave o como parte de un brote), se activa un nuevo bucle que estima la contribución energética de la población estelar formada para todas las instantáneas en el futuro para las cuales la información sobre luminosidades desea ser obtenida. Dado que esta estimación es costosa puesto que requiere interpolar datos de las tablas previamente cargadas, la preponderancia de este bucle respecto al tiempo de cálculo total es muy sensible a la cantidad de salidas que se quiere obtener.

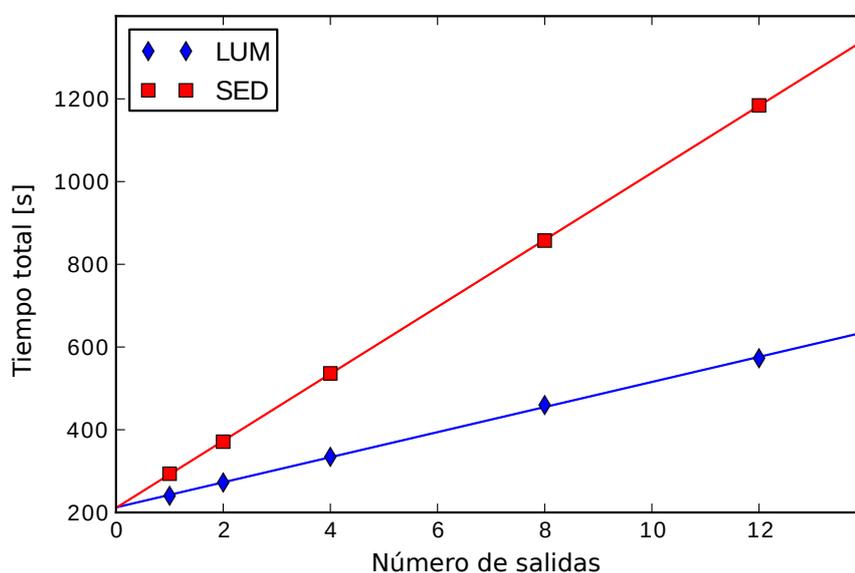


Figura 2.3: Comportamiento del tiempo de cálculo del código dependiendo del número de instantáneas que se consideren para calcular las magnitudes de las galaxias, para ambos métodos disponibles en el código.

Como evidencia del impacto del bucle encargado del cálculo de las luminosidades de las galaxias en el rendimiento del código, en la Figura 2.3 se muestra una medición del tiempo de cálculo total requerido para procesar una simulación de prueba pequeña, pero con diferentes números de salidas (corrimientos al rojo) generadas en cada caso. En la figura se presentan ambos métodos incluidos en el código para realizar este cálculo: el clásico representado por los diamantes azules (indicado como “LUM” en la leyenda), y el que utiliza las SEDs mediante cuadrados rojos. En la figura se puede apreciar claramente el crecimiento lineal del tiempo de cálculo según el número de instantáneas de salida calculadas, y que ambos métodos convergen al mismo valor cuando no se incluye ninguno cálculo de luminosidad.

Modelando este comportamiento con rectas con la misma ordenada en el origen, se puede ver fácilmente que se requieren 8 instantáneas con el método clásico para duplicar el

tiempo de cálculo original, en cambio sólo 2 con el método de los SEDs. Esto impone una limitación importante si se quisiesen obtener muchas salidas con luminosidad utilizando este último método para simulaciones de muchos halos como las que analizaremos luego en el desarrollo de esta Tesis.

Finalmente, durante todo el desarrollo de este proyecto fueron realizadas varias mediciones del perfil de tiempos de cálculo del código (proceso conocido como *profiling*). Estas mediciones, realizadas con el software GNU GPROF¹, consisten en cuantificar el tiempo que utiliza cada una de las rutinas definidas en el código con respecto al tiempo total, estimación útil para entender cuáles son los bloques de cálculo preferentes para optimizar o revisar, y cuál es el impacto de los cambios realizados con este propósito.

2.2.1. Formato de entrada

El código semianalítico SAG está originalmente adaptado para leer los catálogos de halos de simulaciones que fueron analizadas con el código de búsqueda de subhalos SUBFIND (Springel et ál., 2001). No obstante, este algoritmo recibió algunas modificaciones por parte del el grupo de investigación local con el tiempo, de modo que el formato de los catálogos generados y utilizados actualmente no son exactamente los creados por el código original.

El conjunto de archivos que componen estos catálogos se clasifica en dos grupos, uno que contiene la información relacionada con los halos principales (FOF, en este caso) y otro grupo que almacena de forma independiente la información referente a los subhalos detectados por SUBFIND. Cada tipo de catálogo se organiza teniendo un archivo por cada instantánea de la simulación, y todos vienen escritos en formato binario.

El grupo de catálogos referente a los halos principales incluye: una lista de los halos principales encontrados, una lista de las partículas de la simulación asociadas a cada uno de ellos, una lista de progenitores y una lista con las propiedades medidas de cada uno. Esta última es generada analizando todas las partículas asociadas a cada halo, las cuales se utilizan para medir: la masa M_{200} y radio R_{200} definidos según la Ecuación (1.25) utilizando la densidad crítica y $\Delta_{\text{ref}} = 200$, la masa M_{vir} y radio R_{vir} utilizando la densidad crítica y $\Delta_{\text{vir}}(t)$, el centro geométrico, la posición de la partícula más ligada, la posición y velocidad del centro de masas, las energías cinética y potencial, el parámetro de espín λ , el momento angular \vec{L} y la velocidad circular máxima $V_{\text{máx}}$ de cada halo.

El grupo de catálogos correspondiente a los subhalos detectados incluye: una lista de la cantidad de subhalos por cada halo FOF, una lista con el número de partículas asociado a cada uno (lo cual permite estimar su masa), otra lista independiente con las partículas asignadas, una lista de progenitores y una de propiedades de cada subhalo. A diferencia de la anterior, las propiedades estimadas para los subhalos encontrados con SUBFIND sólo incluyen la posición y velocidad de cada uno y el módulo de la velocidad peculiar con respecto al halo principal. Esto se debe a que el método utilizado por este buscador de subhalos utiliza un criterio de ligadura de las partículas que termina aislando las que pertenecen a cada uno, de modo que solo estas propiedades quedan correctamente definidas en este nivel.

Además de los catálogos anteriormente descritos, se agregan dos adicionales creados para la correcta utilización del SAG sobre estos: una lista con la estimación del parámetro de concentración de cada halo principal detectado, y una lista de consistencia resultante de la revisión de la estabilidad de las historias de fusión de cada halo que indica cuáles

¹<http://sourceware.org/binutils/docs/gprof/>

halos principales o subhalos no deben ser utilizados por presentar inconsistencias en sus historias evolutivas.

En definitiva, el conjunto de catálogos de halos que lee el código SAG para construir la población de galaxias respectiva se compone de 11 archivos distintos por cada una de las instantáneas de la simulación, todos escritos en formato binario.

2.2.2. Formato de salida

El formato de salida del código semianalítico SAG se compone de un conjunto de N archivos, todos en formato HDF5². El número de archivos está determinado por $N = n_{z,\text{out}} n^3$, con n^3 el número de subvolumenes en que fue dividida la simulación original y $n_{z,\text{out}}$ el número de instantáneas que se le solicitó escribir al código, valor que es elegido al momento de compilación y puede valer entre 1 y el número de instantáneas originales del catálogo de halos. De este modo, cada uno de estos archivos contiene una fracción $\sim 1/n^3$ de la simulación completa a un corrimiento al rojo determinado.

Los objetos almacenados en los archivos con formato HDF5 poseen nombres o etiquetas que permiten su acceso individual, sin requerir leer el resto del archivo. Estos objetos pueden ser de tres tipos distintos:

- Conjuntos de datos (*datasets*): corresponden a arreglos de variables de un tamaño fijo en donde todos sus elementos tienen el mismo tipo de dato.
- Grupos: ambientes internos del formato útiles para agrupar y jerarquizar la distribución de los elementos almacenados, análogo a los directorios de un sistema operativo. Todo archivo HDF5 posee al menos un grupo raíz “/”, dentro del cual se pueden crear subgrupos.
- Atributos: variable particular que almacena un único valor de cualquier tipo de dato. Pueden estar asociados a cualquier conjunto de datos o grupo del archivo.

Si bien el número final de galaxias calculadas por el modelo depende de la resolución en masa y volumen de la simulación, suele ser considerablemente mayor que el número de propiedades de las galaxias que se escriben. Por ello, las variables son almacenadas en un conjunto de arreglos donde cada uno se encarga de almacenar una propiedad en particular, y cuyo largo corresponde al número de galaxias obtenidas.

Siguiendo lo anterior, el código SAG almacena en forma de *conjuntos de datos* los arreglos correspondientes a cada propiedad de las galaxias. Genera *grupos* para ordenar conjuntos de propiedades que comparten alguna característica en particular (por ejemplo, los grupos `/Histories` o `/Magnitudes`); y utiliza los *atributos* para almacenar variables descriptivas o informativas ya sea de cada arreglo de propiedades o del código en general tales como los parámetros de entrada, opciones de compilación, unidades, corrimiento al rojo, etc., estando estos últimos asociados al grupo raíz del archivo. Los atributos son utilizados para caracterizar la procedencia y propiedades generales de un archivo en particular, entregando además la información necesaria para el correcto entendimiento de los datos almacenados. Es importante destacar que las unidades físicas de las propiedades almacenadas en los conjuntos de datos varían según la simulación procesada por el código.

²Para más información sobre este tipo de formato de almacenamiento de datos, visitar <https://portal.hdfgroup.org/display/HDF5/HDF5>

Por ello, la lectura de los atributos que definen estas unidades es importante para un correcto entendimiento de los datos almacenados.

Para permitir un mayor nivel de análisis de las galaxias creadas por el modelo, los archivos generados suelen contener una gran cantidad de propiedades útiles en un conjunto de más de 100 conjuntos de datos. Las dimensiones de cada uno de estos varía según la propiedad que tenga almacenada, de modo que puede valer entre: $N_{\text{gal}} \times 1$, con N_{gal} el número de galaxias del archivo, para aquellos que almacenan propiedades físicas unidimensionales como la masa de alguna componente particular; $N_{\text{gal}} \times 3$ para almacenar las propiedades tridimensionales como la posición o la velocidad; $N_{\text{gal}} \times 10$ para almacenar por separado masas de diferentes elementos químicos que el código calcula para las diferentes componentes de las galaxias; y $N_{\text{gal}} \times n_{\text{snap}}$, con n_{snap} el número de instantáneas originales de la simulación, para almacenar la evolución histórica de alguna propiedad particular de interés. Usualmente, este último tipo de arreglos suele ser incluido exclusivamente en las salidas correspondientes a la instantánea de $z = 0$.

Si bien este conjunto de archivos entrega una información muy detallada y aprovechable de la población de galaxias modelada, conlleva la limitación de requerir mucho espacio disponible en disco duro para ser almacenados en su totalidad. El formato actual es adecuado para simulaciones pequeñas o incluso de tamaño mediano, pero impone una limitante si se quisiera usar con simulaciones mayores en especial cuando se desean obtener las salidas de todas las instantáneas utilizadas como entrada.

A modo de referencia, consideramos que una simulación es mayor o de gran envergadura cuando permite producir del orden de las decenas de millones o hasta cientos de millones de galaxias en una instantánea particular. Esto se puede alcanzar con diversas combinaciones de resolución de masa y volúmenes, debido al comportamiento de ley de potencias de la función de masa de halos, creciente hacia masas menores (ver, por ejemplo, [Tinker et ál., 2008](#); [Klypin et ál., 2016](#)). De este modo, una simulación de materia oscura cuyo volumen tenga un par de centenas de Mpc/h de lado y con alta resolución en masa ($\text{mín}(M_{\text{halo}}) \sim 10^7 h^{-1} M_{\odot}$) puede alcanzar la misma cantidad de galaxias que una simulación con volumen de $1 \text{ Gpc}/h$ de lado con menor resolución en masa.

2.3. Cambios introducidos en el código

En la sección anterior se describió desde una perspectiva técnica el conjunto de limitaciones que el código semianalítico SAG presenta para ser utilizado en la generación de catálogos de gran envergadura. Algunas de ellas están relacionadas con rendimiento y demandan reformular parte del algoritmo del código, mientras que otras están relacionadas con la especificidad de las entradas (tipos de simulaciones que se pueden utilizar) y salidas (formato de escritura) de datos.

En esta sección describiremos los cambios más importantes realizados al código para atacar todas estas limitaciones, siempre apuntando a conseguir un modelo más versátil y adecuado para concretar los proyectos planteados en este trabajo de Tesis.

2.3.1. Administración de versiones

El código fuente del modelo SAG suele ser revisado y modificado por varios investigadores en simultáneo. Luego, en vista de que los cambios que se describirán en esta sección afectan de forma importante el flujo de cálculo original del modelo, es importante

mantener un registro ordenado y consecutivo de todas las alteraciones que se le realicen, sus justificaciones, fechas y autorías.

Para este propósito hemos introducido el uso de un administrador de versiones que realiza el seguimiento de todos los cambios acontecidos. Dada la naturaleza descentralizada que suelen tomar las modificaciones del código, se ha optado por MERCURIAL³ como administrador principal. Este software construye un repositorio de versiones que queda embebido con el código fuente circulante, sin requerir acceso a un repositorio centralizado.

MERCURIAL utiliza revisiones para almacenar todos los cambios que se desean registrar en el repositorio. Cada una de estas debe contener información básica del autor de los cambios, la fecha y una descripción breve de las modificaciones realizadas en relación a la revisión anterior. A cada revisión se le asigna un código identificador alfanumérico único, y un número natural correspondiente a la cantidad de revisiones almacenadas hasta el momento, partiendo desde cero.

La última versión interna del código, numerada según selección manual antes de la introducción del administrador de versiones, fue la número 7 (llamada SAG-7). Para mantener consistencia, todos los indicadores numéricos de las revisiones registradas con MERCURIAL se asignaron como sufijos en esta nomenclatura. De este modo, la versión SAG-7.0 fue registrada el 25 de abril de 2014, y la última concerniente a esta Tesis SAG-7.128 fue registrada el 28 de octubre de 2016; siendo cinco autores los que han registrado modificaciones en estas 129 versiones del código desde que se inició el registro con el administrador.

2.3.2. Rendimiento

En base a las mediciones del perfil del tiempos de cálculo realizadas al código fuente del SAG, se efectuaron varios cambios que impactaban en el rendimiento. Muchos de ellos fueron menores, tales como desanidamiento de algunos cálculos simples o reformulaciones. No obstante, cabe mencionar tres de estas modificaciones que efectivamente fueron importantes en el funcionamiento interno del código o que tienen consecuencias importantes en el modelado incluido.

- Un gran porcentaje del cálculo de la evolución de las galaxias ($\gtrsim 50\%$) se gastaba en interpolación de tablas, en particular en aquellas utilizadas en la asignación del radio característico de los discos $r_{s,d}$, según el modelo presentado por [Tecce et ál. \(2010\)](#). Esto sucedía porque el cálculo del nuevo tamaño de los discos debía realizarse cada vez que este sufre un cambio en su contenido de masa, efecto resultante de la mayoría de los procesos de evolución presentes en el modelo. Si bien este tipo de estimación permitía obtener tamaños de discos más realistas y adecuados al estado de las galaxias, el utilizar aproximaciones más simples que no involucren lectura e interpolación de datos tabulados (como la de [Mo et ál., 1998](#)) redujo considerablemente el tiempo de cálculo global del código. De este modo, para simulaciones grandes preferimos utilizar este último método por sobre el anterior.
- Tal como fue presentado en el esquema de la Figura 2.2, al inicio de cada paso del bucle sobre instantáneas de la simulación, los catálogos de ambas instantáneas involucradas eran leídos del disco duro. Esto producía que el conjunto de catálogos de halos en general pasaran por dos procesos de lectura, asignándose los valores correspondientes en arreglos distintos cada vez con el fin de tener la libertad de carecer de

³Más información sobre este administrador en su sitio web oficial <https://www.mercurial-scm.org/>

consecutividad si fuese necesario. No obstante, el procedimiento estándar de cálculo de poblaciones de galaxias requiere la utilización de todas las instantáneas disponibles de la simulación de forma consecutiva. Según esta característica, procedimos a reducir la lectura de catálogos a sólo una por cada paso de este bucle, de modo que, al final de cada uno, todos los arreglos que contenían la información de los halos de la instantánea $s + 1$ eran adaptados para ser ocupados en el paso siguiente, sin realizar modificación alguna a la localización de los datos en la memoria.

- El último cambio importante que se introdujo respecto al rendimiento fue la reducción de las variables leídas desde los catálogos de halos al mínimo necesario para calcular la evolución de las galaxias. Esto es debido a que originalmente los catálogos se cargaban en su totalidad en cada proceso de lectura. Para esto, se procedió a seguir todas las propiedades leídas de los halos dentro del modelo y ver cómo influían en la evolución con el fin de dirimir cuáles eran prescindibles y así evitar su lectura. En particular, todas las variables relacionadas con partículas fueron descartadas en este proceso.

Este último análisis, a su vez, permitió identificar todas las propiedades que necesita el modelo como entrada para construir y calcular la evolución de galaxias para un conjunto de halos de materia oscura, más allá del formato en que estas se suministren.

2.3.3. Paralelización multihilos con OPENMP

Es sabido que la estrategia más impactante para reducir el tiempo de cálculo de un código es organizar el algoritmo global de modo que el cálculo se pueda distribuir en diferentes procesos que puedan comunicarse, por ejemplo a través de MPI. Pero en el caso del SAG, la complejidad del algoritmo de evolución y la cantidad de variables compartidas utilizadas hace que una solución de este estilo requiera reescribir gran parte del flujo del código. A pesar de ello, se puede aplicar una solución parcial en secciones trascendentales del modelo.

Como se mostró en el análisis del código detallado anteriormente, el tiempo de cálculo total tiene una dependencia muy fuerte con el método que se ocupe para estimar las luminosidades de las galaxias, principalmente con el número de salidas que se consideren para este cálculo. Este incremento está dado por un bucle adicional desencadenado por los eventos de formación estelar, donde la contribución en luminosidad en los diferentes rangos energéticos de la población estelar generada es calculada para todas las salidas deseadas.

Este cálculo tiene la particularidad que cada uno de sus pasos es totalmente independiente del anterior, no alterando variables que son leídas por los otros, de modo que es ideal para ser paralelizado utilizando la estrategia de multihilos. Este tipo de paralelización es transitoria, y al ser aplicada a bucles distribuye cada uno de sus pasos en diferentes núcleos de la computadora teniendo un control estricto sobre cuáles de las variables utilizadas por el código son privadas en cada hilo de ejecución o se mantienen compartidas.

Para realizar esta paralelización hemos utilizado la librería OPENMP. Se incluyó la instrucción respectiva en cada uno de los bucles de cálculo de luminosidades existentes en el código para los dos métodos mencionados, siempre cuidando de definir las variables privadas y compartidas en cada uno de los casos. Con esta librería, el número de nú-

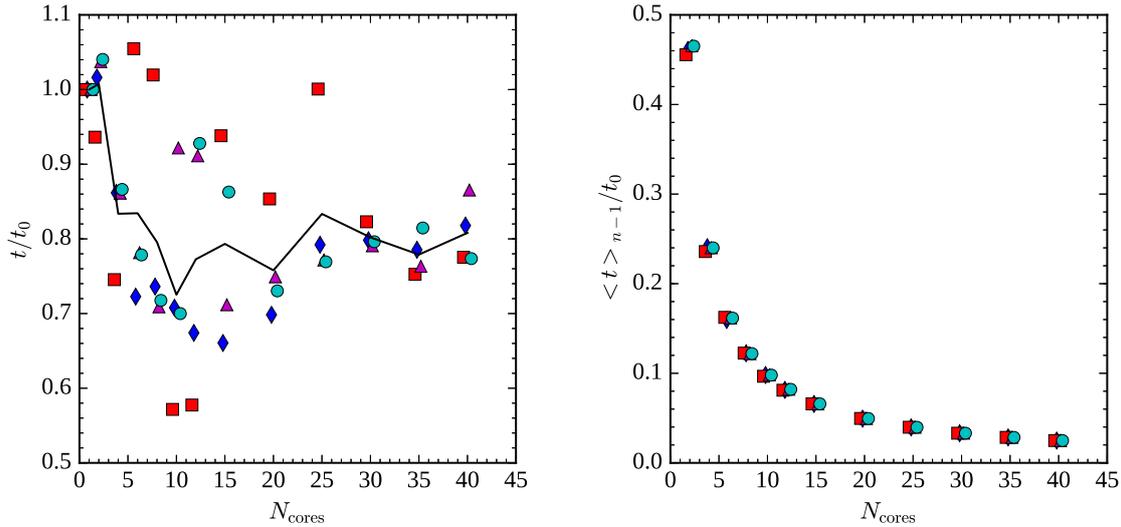


Figura 2.4: Ejemplo de rendimiento de paralelización con OPENMP con una simulación de prueba para diferentes valores del número de núcleos utilizados en la paralelización del cálculo de luminosidades. *Panel izquierdo*: tiempo total efectivo en comparación con el tiempo original con $N_{\text{cores}} = 1$. *Panel derecho*: rendimiento de los procesadores extra durante el cálculo, medido como la fracción del tiempo total en que estos estuvieron activos. En ambos casos, cada conjunto de símbolos agrupa una prueba diferente de rendimiento realizadas en días y horarios distintos.

cleos utilizado en paralelo por defecto usualmente toma el valor máximo definido por la computadora donde se esté ejecutando, pero también puede ser seleccionado por el usuario mediante la definición de la variable de ambiente `OMP_NUM_THREADS` igual al número de núcleos deseados. Este último comportamiento es ideal para trabajo en supercomputadoras que utilizan administradores de recursos para la ejecución de los trabajos.

Es importante destacar que la eficiencia de este tipo de paralelización es bastante susceptible tanto al número de núcleos que se utilicen como al hardware que ejecute el proceso. Por otro lado, el proceso de activar y desactivar temporalmente el ambiente de multihilos genera un gasto en tiempo de cálculo que puede llegar a superar la ganancia dada por la paralelización si el número de núcleos es muy alto. Esto se debe a que la administración de variables privadas y compartidas debe ser realizada en cada ocasión en que el ambiente multihilos es activado.

Un ejemplo de este comportamiento se puede ver en la Figura 2.4, en cuyo panel izquierdo se muestra el tiempo de ejecución efectivo con respecto al tiempo de cálculo original (utilizando un solo núcleo) de una simulación pequeña de prueba, para la cual se calcula un número elevado de salidas con luminosidades, versus diferentes elecciones de la cantidad de núcleos utilizados en la paralelización de este cálculo N_{cores} . Se utilizaron varias ejecuciones en cada caso. Utilizamos el método clásico de cálculo de luminosidades por ser el menos sensible a la ganancia de rendimiento por la paralelización debido a la menor cantidad de cálculos que se realiza en el mismo. Cada ejecución se muestra con un símbolo distinto, y mediante la línea negra el promedio de cada conjunto de ejecuciones para un mismo número de N_{cores} . Se puede apreciar que el tiempo de cálculo es bastante aleatorio debido a que esta medición fue realizada en una supercomputadora cuyo uso compartido con otros usuarios es bastante heterogéneo, pero las numerosas ejecuciones

realizadas permiten apreciar la tendencia global que presenta. Su comportamiento muestra una abrupta disminución en el tiempo de cálculo cuando se utilizan menos de ~ 5 núcleos, alcanzando un mínimo alrededor de los 10, y un aumento suave para valores mayores indicando la pérdida de rendimiento del proceso ya mencionada.

Una forma más efectiva de medir la pérdida de rendimiento, independiente de las condiciones de la supercomputadora, es estimar la fracción de tiempo total efectivo que estuvieron trabajando los procesadores adicionales que se utilizan únicamente durante el bloque en paralelo $N_{\text{cores}} - 1$, la cual se muestra en el panel derecho de la Figura 2.4. El comportamiento de este valor es totalmente suave, e indica con claridad que la tasa de pérdida de rendimiento por aumentar el número de procesadores es independiente de las condiciones externas que podrían alterar el tiempo de cálculo efectivo en una ejecución particular.

Independientemente de la pérdida de rendimiento por usar un número excesivo de núcleos, la utilización de un número bajo ya permite reducir el tiempo total hasta en un 40 %, ganancia muy importante para cálculos de gran envergadura.

2.3.4. Adaptación a SUSSING

Una de las mayores limitaciones que tenía el modelo en su versión SAG-7.0 era su poca versatilidad para leer catálogos de halos que no siguieran el formato específico para el cual fue originalmente diseñado. Esta no sólo constituía una simple restricción técnica, sino también condicionaba los procesos físicos modelados debido a que todo el tratamiento implementado estaba específicamente diseñado y adaptado para el tipo de propiedades que se generaban con SUBFIND, y que fueron detalladas en la Sección 2.2.1.

Entre los métodos de búsqueda de halos más utilizados además de SUBFIND, podemos mencionar los códigos libres *AMIGA halo finder*, AHF (Knollmann y Knebe, 2009), y ROCKSTAR (Behroozi et ál., 2013b), los cuales son muy atractivos para procesar simulaciones de gran tamaño debido a que están implementados para realizar los cálculos en paralelo.

A diferencia de SUBFIND que tiene una jerarquía definida de dos niveles en la identificación de estructuras, estos otros códigos se caracterizan por utilizar el método de refinamiento de malla adaptativo (AMR por sus siglas en inglés *Adaptive Mesh Refinement*), cuyo algoritmo recursivo para describir la distribución de las partículas permite encontrar sobredensidades localizadas dentro de otras, sin restricción en el número de anidamiento de las mismas. Esto último fuerza a que la definición de un halo deba ser generalizada, independientemente si este es un halo principal o un subhalo satélite, y lo mismo con las propiedades que los caracterizan. Un ejemplo importante de esto lo constituye la masa de estas estructuras, que en el caso de estos métodos corresponden a masas de halos que incluyen todas las sobredensidades que fuesen encontradas dentro de ellos, independientemente del nivel jerárquico que estas tengan. Si bien estos códigos sólo se centran en la búsqueda de halos en las distintas instantáneas de una simulación, también incluyen sus propias herramientas anexas para la construcción de los árboles de fusiones de halos y creación de catálogos en forma apropiada: MERGERTREE que viene incluido junto a AHF, y CONSISTENTTREES asociado al código ROCKSTAR.

Durante los últimos años, una serie de proyectos de comparación colaborativos se han dado lugar en diferentes reuniones de trabajo en las que se ha analizado la mayoría de los métodos o códigos disponibles para buscar halos en simulaciones cosmológicas (Knebe et ál., 2011), subhalos (Onions et ál., 2012) y constructores de árboles de fusiones (Sri-

sawat et ál., 2013). Estas reuniones han impulsado a que los diferentes autores adapten sus códigos para escribir formatos estandarizados y así poder concretar los proyectos de comparación. Incluso se ha llegado a proponer un estándar de almacenamiento y distribución de catálogos de halos y árboles de fusiones (Thomas et ál., 2015). No obstante, el formato más común para catálogos de halos utilizado hoy en día es el llamado SUSSING (Srisawat et ál., 2013), en referencia a la reunión donde se popularizó que se realizó en la Universidad de Sussex en el Reino Unido, y corresponde a una variante del formato estándar de AHF. Por este motivo, y apuntando a la participación del modelo SAG en estas colaboraciones de comparación, como parte de este trabajo de Tesis hemos modificado el código semianalítico para que pueda leer y procesar este tipo de formato. Esta adaptación considera las diferencias de las variables que son incluidas en estos archivos en comparación con las originalmente extraídas de SUBFIND, modificando las relaciones internas del modelo de evolución de galaxias en consecuencia.

El formato SUSSING original consiste en una colección de archivos en ASCII de sólo dos tipos: listas de halos por cada instantánea y una lista de progenitores. Las listas de halos se organizan almacenando las propiedades de un halo por cada línea del archivo, de modo que cada propiedad constituye una columna. Además, el archivo tiene una línea comentada de encabezado con el nombre de las columnas incluidas en él. Cada halo suele tener asignado un número identificador (ID) que se incluye entre las propiedades del mismo y es único en toda la simulación, de modo que no existen ID repetidos entre halos pertenecientes a instantáneas diferentes. La lista de progenitores, por el contrario, se organiza de una forma más heterogénea. Se inicia con tres líneas de encabezado y luego se compone por bloques de datos que almacenan la información correspondiente a un halo con sus respectivos progenitores mediante sus ID. Estos bloques tienen siempre al menos una línea con dos valores: el identificador del halo respectivo y el número de progenitores que este tiene en la instantánea anterior, los cuales están indicados en las siguientes líneas a través de sus respectivas ID, una por cada halo progenitor.

Las propiedades de halos que son incluidas en las listas de este formato son detalladas en la Tabla 2.1, la cual sigue el mismo orden con que aparecen en los archivos. En la tabla se incluye: el nombre del encabezado, una descripción breve, sus unidades y un indicador de si son utilizadas por el código SAG. Una de las principales diferencias de estas propiedades con las leídas de catálogos de SUBFIND es la estandarización de unidades. En este nuevo estándar, las unidades en las que se almacena cada propiedad están fijas por el formato a diferencia de los otros donde son dependientes de la simulación analizada. Por otro lado, la jerarquía múltiple hace que se incluyan tanto halos principales y satélites dentro de la misma lista, de modo que sus propiedades son homólogas. En particular, las masas de los halos son inclusivas, es decir, siempre incluyen las masas de los halos que son detectados dentro de ellos independientemente de la jerarquía a la que pertenezcan, a diferencia de SUBFIND donde sólo los halos principales tenían masas inclusivas. Por ello, la lectura de estas propiedades debe realizarse con extremo cuidado, así como las Ecuaciones (2.1) y (2.15) para estimar las masas de gas caliente en cada sistema deben ser implementadas de forma consistente.

Cabe destacar tres cambios importantes realizados en el modelo para trabajar con este tipo de catálogos:

- Por su definición inclusiva, se utilizaron las propiedades de los halos principales de forma análoga a las propiedades leídas de los halos tipo FOF anteriormente utilizados. No obstante, para mantener consistencia del modelado entre ambos tipos de catálogos, se restringió al máximo el uso de masas o radios de sistemas completos

Nro.	Nombre	Descripción	Unidad	SAG
1	ID	ID del halo	–	Sí
2	hostHalo	ID del halo anfitrión	–	Sí
3	numSubStruct	Número de subhalos	–	Sí
4	M200c	Masa del halo	M_{\odot}/h	Sí
5	npart	Número de partículas	–	No
6	Xc	Posición comóvil x	kpc/h	Sí
7	Yc	Posición comóvil y	kpc/h	Sí
8	Zc	Posición comóvil z	kpc/h	Sí
9	VXc	Velocidad peculiar x	km/s	Sí
10	VYc	Velocidad peculiar y	km/s	Sí
11	VZc	Velocidad peculiar z	km/s	Sí
12	R200c	Radio del halo	kpc/h	Sí
13	Rmax	Pos. curva de rotación máxima	kpc/h	No
14	r2	Pos. del máximo de ρr^2	kpc/h	No
15	mbp_offset	Dist. a la partícula más ligada	kpc/h	No
16	com_offset	Dist. al centro de masas	kpc/h	No
17	Vmax	Curva de rotación máxima	km/s	Sí
18	v_esc	Velocidad de escape en R_{vir}	km/s	No
19	sigV	Dispersión de velocidades 3D	km/s	No
20	lambda	Parámetro de espín λ_{B}	–	No
21	lambdaE	Parámetro de espín λ clásico	–	Sí
22	Lx	Momento angular x	$ \vec{L} = 1$	Sí
23	Ly	Momento angular y	$ \vec{L} = 1$	Sí
24	Lz	Momento angular z	$ \vec{L} = 1$	Sí
25	cNFW	Param. de concentración NFW	–	Sí

Tabla 2.1: Propiedades de los halos incluidas en cada lista del formato SUSSING.

(halos principales) en los tratamientos de evolución de las galaxias.

- Como el código SAG poblaba únicamente subhalos con galaxias, de los cuales había uno asociado a los halos principales (subhalo principal en SUBFIND), utilizamos las propiedades de los halos principales de los catálogos SUSSING de forma análoga a los subhalos principales de SUBFIND.
- Dado que el modelo fue construido originalmente pensando en sólo dos niveles de anidación de halos (principales y satélites), todos los niveles extra de anidamiento encontrados en los catálogos SUSSING se colapsan a sólo dos por defecto. De este modo, los halos satélites de satélites se consideran como satélites del principal, en la misma jerarquía que sus anfitriones.

A su vez, dado que hay propiedades que no se incluían en estos nuevos catálogos pero sí en los anteriores, tales como las estimaciones de energía cinética y potencial de los halos, optamos por excluir de las relaciones internas del código cualquier utilización de estos valores forzando así a que el comportamiento evolutivo de las galaxias no tuviese diferencias entre la utilización de distintos tipos de catálogos. Esta decisión se aplicó en especial a este tipo de variables que jugaban un papel menor en el modelado y no están siempre disponibles.

Por otro lado, la lectura y utilización de esta nueva lista de progenitores de halos no genera ninguna diferencia en el modelo. No obstante, fue necesario establecer una estrategia precisa de carga y búsqueda de halos debido a que la información de los árboles de fusiones ahora viene incluida en un único archivo y no para cada instantánea por separado como en el caso anterior. Esto provoca que la cantidad de datos almacenados en esta lista de progenitores sea considerablemente mayor que en el resto de los catálogos que contienen solamente listas de halos, lo que fuerza a utilizar técnicas de ordenamiento y búsqueda eficientes al momento de cargar los progenitores de cada uno.

Extensión para CONSISTENTTREES

Si bien el formato SUSSING es genérico y puede ser generado por variados códigos de búsqueda de halos y árboles de fusiones, durante esta adaptación dimos especial énfasis a aquellos catálogos de halos que son generados por CONSISTENTTREES (Behroozi et ál., 2013). La ventaja de estos catálogos radica en que tanto los halos como los árboles de fusiones son organizados de una forma muy conveniente y aprovechable por los códigos semianalíticos como el SAG.

En este contexto, introducimos el concepto de *bosque* (*forest*) de halos, como un conjunto de halos de una simulación que tienen conexiones ya sea históricas o espaciales, a diferencia de un árbol de fusiones que sólo considera las conexiones históricas. En este sentido, un bosque es más amplio que un árbol de fusiones particular dado que puede estar constituido por varios árboles de fusiones que hayan estado parcialmente vinculados en el pasado, dando lugar a más de un halo principal a $z = 0$. Cada una de estas estructuras constituye un conjunto de halos que puede ser analizado por el modelo semianalítico de forma totalmente independiente del resto de la simulación.

La mayoría de los bosques de una simulación suelen estar constituidos por la evolución histórica de un único halo principal debido a que este tipo de halos se vuelve cada vez más frecuente hacia masas menores, las que a su vez van ganando importancia numérica al disminuir la masa. No obstante, aquellos que albergan los halos más masivos suelen estar constituidos por numerosos halos y albergar más de un árbol de fusiones. Al igual

que los halos, a cada bosque se le asigna un identificador único en toda la simulación, generalmente coincidente con la ID de un halo perteneciente al mismo localizado en la última instantánea disponible.

Concerniente a los catálogos de halos, aquellos que son generados por CONSISTENT-TREES incluyen un archivo adicional que proporciona la información necesaria para caracterizar y localizar cada bosque. Este archivo es igualmente una lista en formato ASCII que se inicia con una línea comentada de encabezado y luego contiene la información respectiva a los bosques en cada línea siguiente, la que incluye: la ID de cada uno, el número de halos total que poseen y el número de halos que tienen en cada instantánea de los catálogos. El orden en que aparecen en esta lista es respetado tanto en las listas de halos como en la lista de progenitores, de modo que la lectura ordenada de estas cantidades permite extraer cualquier bosque por completo desde el conjunto de catálogos, permitiendo lecturas parciales autoconsistentes de la historia evolutiva de los halos.

Dada la ventaja que tiene esta estrategia de ordenamiento, la adaptación que realizamos en el código SAG incluyó por defecto este tipo de lectura, agregando además la opción de procesar de forma parcial cualquier conjunto de catálogos suministrados en este formato. Para esto, modificamos el diagrama de flujo del código mostrado en la Figura 2.2 agregando un bucle nuevo antes del ciclo de instantáneas, que recorre los bosques que se han de leer durante el proceso. Asimismo, tanto la lectura de catálogos como la escritura de salidas se repite en cada uno de sus pasos sin crear o leer archivos nuevos, sino actualizando los que ya están cargados desde el paso anterior.

Incluir esta forma de procesamiento por bosques, estrategia que podemos denominar como procesamiento “vertical”, resuelve el problema del número fijo de procesos que podían trabajar en simultáneo para procesar una simulación determinada con el fraccionamiento de catálogos. En efecto, ahora es posible ejecutar en simultáneo una cantidad arbitraria de procesos independientes del SAG que lean y calculen en paralelo diferentes secciones de la simulación. En vista de que este es el comportamiento deseado para procesar grandes simulaciones, el método de distribución de cálculo se automatizó e incluyó por defecto dentro del paquete de códigos del SAG.

ASCII como limitación

Si bien la capacidad de procesar una simulación en paralelo por una cantidad arbitraria de ejecuciones de cálculo distribuido propicia una ganancia importante en el tiempo requerido para completar el cálculo de una simulación, el formato ASCII de este tipo de catálogos genera una limitación adicional importante.

La forma de acceso y lectura usual para archivos ASCII es leer los caracteres almacenados uno a uno comenzando desde el inicio del archivo, forzando a realizar una interpretación de los mismos que permita obtener las variables numéricas almacenadas. Cuando esto es realizado en simultáneo por varios procesos, se genera una cantidad importante de llamados de acceso a los sistemas de disco duro, lo cual agrega una latencia bastante importante en la ejecución. En especial para aquellos procesos que deben saltarse una porción importante del archivo para alcanzar un bosque particular de la simulación. Como el rendimiento es un factor importante en este proyecto, durante su desarrollo implementamos dos soluciones a este problema.

La primera solución consistió en permitir que cada paso del bucle que recorre los bosques a procesar no se restringiera a sólo uno por vez, sino que se leyeran y procesaran grupos de estos en cada oportunidad. Si bien se requiere invertir algo de tiempo de cálculo

en rutinas que realizan ordenamientos y búsquedas en arreglos debido a la complejidad que adquiere la lectura de los progenitores, el tiempo ganado en lectura es más importante, por lo que es una inversión conveniente para el rendimiento general. Este comportamiento quedó implementado por defecto cuando el modelo es aplicado sobre catálogos SUSSING, en cuyo caso el número de bosques a ser leídos en simultáneo es elegido por el usuario como parámetro de entrada del código.

La segunda solución planteada consistió en modificar los catálogos ASCII directamente, reescribiéndolos en su totalidad en formato binario evitando así todo el proceso de interpretación de las variables numéricas almacenadas. Con esta forma de almacenamiento el proceso de lectura es mucho más rápido y permite además realizar saltos controlados en la lectura dentro de los archivos sin necesidad de interpretar las variables saltadas. La conversión en este caso se realiza manteniendo la estructura y el orden de los archivos originales, pero descartando un conjunto menor de propiedades que no están bien definidas por ROCKSTAR y definiendo de forma estricta los tipos de datos con que se almacena cada variable. Con ello, la conversión tanto de ASCII a binario como viceversa es un proceso absolutamente transparente y reversible. Al igual que la solución anterior, el código SAG por defecto tratará de leer la versión en binario de los archivos SUSSING, pero también incluimos de forma opcional la habilidad de leer los originales si se requiere.

Toda esta adaptación le ha dado acceso al programa para procesar una infinidad de nuevas simulaciones, ampliando enormemente el impacto del mismo en comparación con otros códigos semianalíticos utilizados por la comunidad astronómica tales como GALFORM (Gonzalez-Perez et ál., 2014) o SAGE (Croton et ál., 2016) que están a la fecha trabajando en el proceso de adaptación a los catálogos creados con CONSISTENTTREES. También ha abierto varios proyectos nuevos tanto de colaboración como los que surgen como ideas originales de nuestro grupo de investigación.

2.3.5. Nuevos formatos de salida

Con el objetivo de concretar la participación del modelo SAG en las reuniones y trabajos de comparación específicos para modelos semianalíticos, siguiendo los esfuerzos comparativos realizados anteriormente (como por ejemplo Knebe et ál., 2011; Onions et ál., 2012; Srisawat et ál., 2013), se implementó un nuevo formato de salida en el código.

El formato implementado corresponde al utilizado en la colaboración de comparación de modelos semianalíticos nIFTy (Knebe et ál., 2015) que será detallada en la sección siguiente. Si bien aún no constituye un estándar en la comunidad, fue originalmente propuesto como una colección de archivos ASCII, uno por cada instantánea de la simulación, en donde se almacenan las propiedades de las galaxias resultantes en cada línea. El conjunto de propiedades es fijo y, siguiendo la filosofía del formato SUSSING, las unidades de estas propiedades también son estándar y definidas por el formato en vez de depender de la simulación analizada. Por otro lado, el orden en que las galaxias se organizan en el archivo también debe seguir una estructura jerárquica de modo que las galaxias centrales aparezcan antes que sus respectivas satélites, y las correspondientes a un mismo halo deben quedar consecutivas. Cada archivo debe además incluir un encabezado no comentado que tenga los valores para metalicidad y luminosidad solares utilizados por los cálculos internos del código.

Entre las 55 propiedades que componen este formato se incluyen: ID de los halos de pertenencia de cada galaxia, posiciones y velocidades, masas de cada componente y sus

masas de metales en $h^{-1} M_{\odot}$, radios característicos tanto de discos como de bulbos en $h^{-1} \text{kpc}$ y un conjunto de magnitudes absolutas en reposo de variados sistemas con y sin incluir la extinción debido al polvo.

Si bien este tipo de salidas estándar con un número bajo de variables a escribir tiene gran utilidad en proyectos comparativos, el hecho de que sea en ASCII impone una limitación importante si quisiera ser utilizada para proyectos de gran envergadura. Por este motivo, decidimos implementar una versión alternativa a este tipo de salida pero almacenando los datos en archivos con formato HDF5 en vez de texto. Las mayores diferencias entre este nuevo formato de salida y el original del código SAG son el número reducido de variables incluidas, la estandarización de unidades y la restricción de que todos los arreglos correspondientes tengan dimensiones $N_{\text{gal}} \times 1$, con N_{gal} el número de galaxias incluido en el archivo correspondiente. Este nuevo formato, al cual nos referiremos en adelante como “salida reducida”, ha sido creado con el fin de convertirse en el estándar para compartir propiedades de galaxias en proyectos de gran envergadura o de colaboración en general en los que el modelo SAG participe.

2.4. El proyecto de comparación nIFTy

Gracias a las adaptaciones presentadas en la sección anterior, el código semianalítico SAG representado por sus desarrolladores, fue partícipe del proyecto de comparación de modelos de formación de galaxias nIFTy. Esta gran colaboración a la fecha ha elaborado una publicación de presentación del proyecto (Knebe et ál., 2015) y otra dedicada específicamente a estudiar la distribución espacial y funciones de correlación de las galaxias modeladas (Pujol et al., enviado). El proyecto, iniciado con una reunión de trabajo realizada en Madrid, España, en junio de 2014, reunió 14 modelos de formación de galaxias: 12 códigos semianalíticos y 2 modelos de ocupación de halos (HOD por sus siglas en inglés *Halo Occupation Distribution*). Estos últimos utilizan estrategias estadísticas para poblar los halos de la simulación con galaxias sin conexión histórica. El propósito de este proyecto fue congregarse la mayor cantidad posible de modeladores de formación de galaxias para analizar el comportamiento de estos en un ambiente común (una misma simulación y catálogos de halos) e investigar un enfoque estándar para calibrar los tratamientos de los procesos físicos implementados en cada uno.

Los catálogos utilizados para esta comparación fueron construidos a partir de 62 instantáneas de una simulación de materia oscura solamente, caracterizada por los parámetros cosmológicos WMAP7 (Komatsu et ál., 2011, $\Omega_{\text{m}} = 0.272$, $\Omega_{\Lambda} = 0.728$, $\Omega_{\text{b}} = 0.0455$, $\sigma_8 = 0.807$, $h = 0.7$, $n_s = 0.96$), constituida por 270^3 partículas de masa $9.31 \times 10^8 h^{-1} M_{\odot}$ cada una, en un volumen cúbico de $62.5 h^{-1} \text{Mpc}$ de lado. Los halos de la simulación fueron identificados con SUBFIND, pero los árboles de fusiones fueron generados con el paquete MERGERTREE, de modo que los catálogos correspondientes fueron suministrados en el formato SUSSING.

Debido a que el conjunto de modelos congregados provienen en su mayoría de grupos independientes, desarrollados en ambientes con estándares distintos, excepcionalmente se incluyeron en los catálogos un amplio conjunto de valores de masas y radios siguiendo distintas definiciones para los halos, de modo que cada modelo pudiese utilizar aquella que más se adecuaba a los tratamientos implementados de forma nativa. Según ello, se incluyeron cinco definiciones de masas en estos catálogos: la masa de los halos FOF, M_{FoF} , y de los subhalos, M_{bnd} , definidas nativamente por SUBFIND; las masas utilizando

Nombre	Masa	Calibración		Referencia
GALACTICUS	M_{BN98}	SMF	LF	Benson (2012)
GALICS-2.0	M_{bnd}	SMF	–	Cattaneo et al. (en prep.)
MORGANA	M_{FoF}	SMF	–	Monaco et ál. (2007)
SAG	M_{bnd}	–	LF	Gargiulo et ál. (2015)
SANTACRUZ	M_{BN98}	SMF	–	Somerville et ál. (2008)
YSAM	M_{200c}	SMF	–	Lee y Yi (2013)
GALFORM-GP14	M_{bnd}	–	LF	Gonzalez-Perez et ál. (2014)
GALFORM-KB06	M_{bnd}	–	LF	Bower et ál. (2006)
GALFORM-KF08	M_{bnd}	–	LF	Font et ál. (2008)
DLB07	M_{200c}	–	LF	De Lucia y Blaizot (2007)
LGALAXIES	M_{200c}	–	LF	Henriques et ál. (2013)
SAGE	M_{200c}	SMF	–	Croton et ál. (2006)
MICE	M_{FoF}	–	LF	Carretero et ál. (2014)
SKIBBA	M_{200c}	–	LF	Skibba y Sheth (2009)

Tabla 2.2: Modelos participantes del proyecto nIFTy. Los doce superiores corresponden a los semi-analíticos y los dos inferiores a los HOD que participaron en la comparación. La columna de calibración indica si el modelo ajusta sus parámetros usualmente mediante funciones de masa estelar (SMF) o funciones de luminosidad (LF).

un contraste de densidad constante $\Delta = 200$ con respecto a la densidad media, M_{200m} , y con respecto a la densidad crítica, M_{200c} ; y la masa M_{BN98} que utiliza un factor de contraste variable Δ_{vir} según la estimación de Bryan y Norman (1998). Además, no se le solicitó a los modeladores que recalibraran los parámetros libres de sus respectivos códigos. Por el contrario, se instó el uso de un conjunto de parámetros ajustados para las simulaciones con las que suelen trabajar e indicando las restricciones que suelen imponer para calibrar. La lista de modelos participantes se detalla en la Tabla 2.2, donde se muestra el nombre de los mismos, la masa que usaron, la restricción principal para la calibración y la referencia correspondiente.

En el caso del modelo SAG, la versión del código utilizada para el proyecto correspondió a SAG-7.42. Se utilizó el tratamiento básico para modelar la evolución de galaxias, con un paradigma de estrangulación instantánea (donde los satélites no retienen su gas caliente); y dado que la simulación es de pequeñas dimensiones se aplicó además el cálculo detallado del tamaño de los discos de galaxias (Tecce et ál., 2010). Para la comparación, utilizamos una serie de parámetros ajustados mediante la técnica PSO con la misma simulación y restricciones utilizadas en Gargiulo et ál. (2015) pero con el tratamiento más simple del modelo. En este caso particular, además, utilizamos las masas definidas por SUBFIND con el fin de preservar los tratamientos utilizados en las publicaciones anteriores.

Este primer proyecto de comparación en general mostró que las predicciones de todos modelos semianalíticos utilizados en una simulación para la cual no estaban calibrados eran estables, pero emergieron diferencias importantes. Un ejemplo de este comportamiento se puede apreciar en la Figura 2.5. En el panel superior, correspondiente a la Figura 4 de Knebe et ál. (2015), se muestra la función de masa estelar Φ a $z = 0$, correspondiente a la densidad numérica de galaxias por cada rango logarítmico de masa estelar, obtenida por todos los modelos participantes. Aunque existen un par de excepciones, los comportamientos generales de los modelos siguen tendencias análogas. En el rango de

bajas masas, antes del quiebre de la función, existe un acuerdo entre los modelos dentro de un factor ~ 3 . Sin embargo, las diferencias tienden a acrecentarse de forma importante a masas mayores.

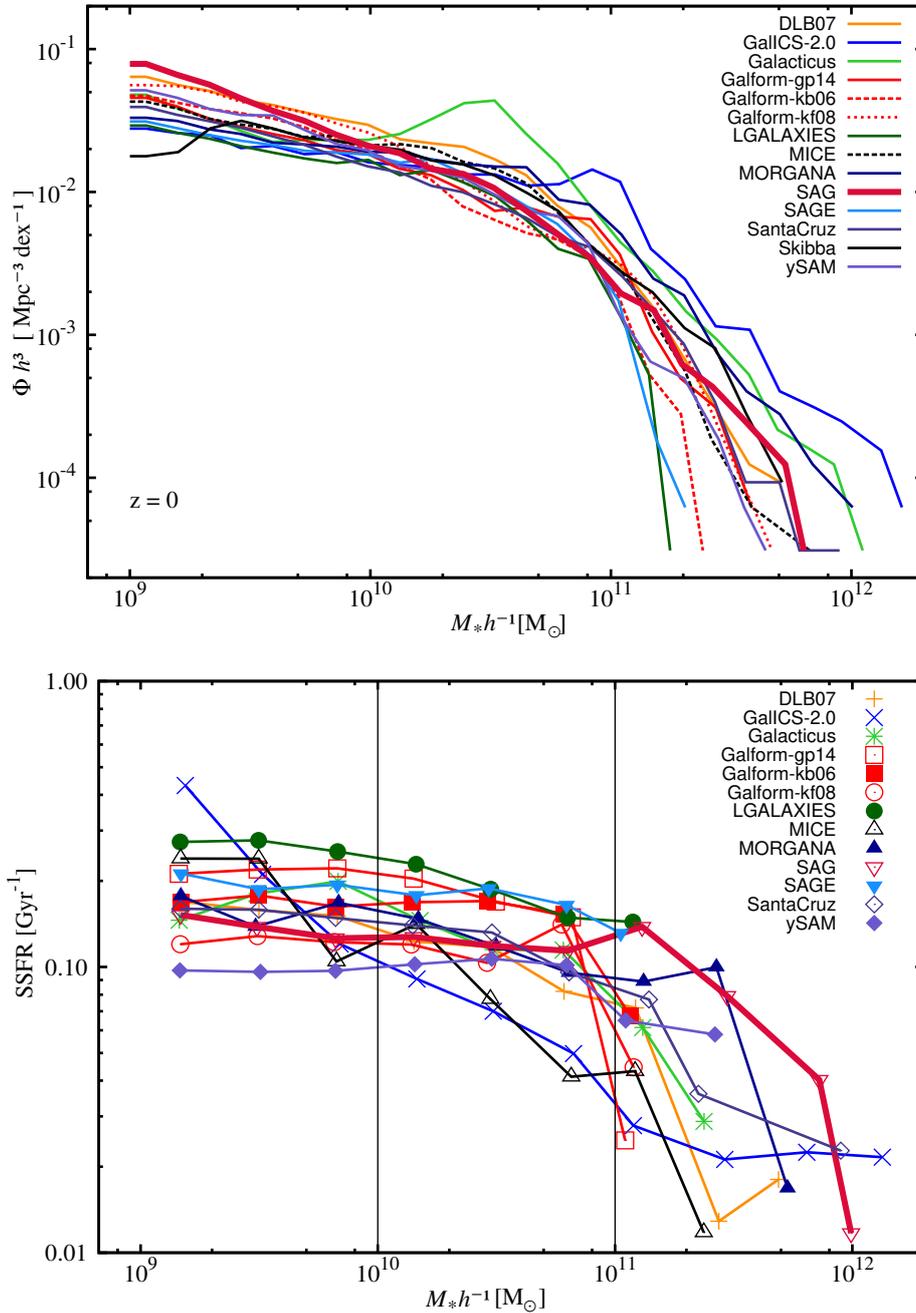


Figura 2.5: Predicciones a $z = 0$ de los modelos participantes en el proyecto de comparación nIFTy. *Panel superior*: función de masa estelar Φ . *Panel inferior*: tasa de formación estelar específica SSFR como función de la masa estelar.

Por otro lado, en el panel inferior de la Figura 2.5, correspondiente al panel izquierdo de la Figura 9 de Knebe et ál. (2015), se muestra la tasa de formación estelar específica (SSFR, por sus siglas en inglés), definida como la tasa de formación estelar por unidad de masa de las galaxias en cada rango de masa, que incluye sólo las galaxias activas (cuya tasa de formación estelar específica tenga un valor mayor que 0.01 Gyr^{-1}), igualmente para

$z = 0$. En la figura, los puntos representan la media de los intervalos considerados, y las líneas conectan los puntos obtenidos para cada modelo para mostrar las diferencias entre las distintas secuencias. Esta medición de la eficiencia de la creación de estrellas también muestra una dispersión importante entre los modelos que es transversal a todas las masas de las galaxias, lo cual también refleja una importante diferencia entre las edades (y los tiempos de formación de su contenido estelar) para las galaxias de un mismo rango de masas entre los diferentes modelos.

Las diferencias reportadas en las predicciones de las propiedades de las galaxias calculadas por los modelos pueden originarse por varias razones.

- Restricciones observacionales impuestas a cada modelo. Si bien se ha reportado que algunos códigos semianalíticos son capaces de satisfacer tanto restricciones de masa a través de las funciones de masa estelar, como de magnitudes a través de funciones de luminosidad o fracciones de galaxias rojas (Henriques et ál., 2013, 2015), no es un estándar la satisfacción de las distintas restricciones en simultáneo. Las restricciones basadas en mediciones de magnitudes agregan dos grados de libertad en el cálculo debido a su dependencia con los modelos sintéticos de evolución estelar y la extinción. No se puede descartar la contribución a la dispersión por preferir una restricción por sobre otra, dado que cada modelador suele adicionalmente incluir más restricciones que las mencionadas con el fin de explorar y restringir el espacio de sus parámetros libres.
- Una contribución adicional puede provenir de la diferencia entre las cosmologías utilizadas por cada modelo en su ambiente nativo. Si bien el proceso de recalibración puede llegar a absorber las diferencias originadas por este factor, en este caso estamos comparando sin efectuar este procedimiento.
- Diferentes ambientes nativos en cada modelo (árboles de fusiones). Se ha mostrado previamente en la literatura que el cambio del algoritmo de construcción de árboles de fusiones tiene un impacto no despreciable en las galaxias obtenidas por modelos semianalíticos cuando estos no son recalibrados (Lee et ál., 2014), de modo que utilizar un modelo en un ambiente distinto al que usualmente es aplicado tiene un impacto no despreciable en sus predicciones. Un ejemplo más detallado de este tipo de diferencias se puede encontrar en la Figura 6 de Knebe et ál. (2015), donde se evalúa cómo cambia la función de masa estelar de los modelos según se apliquen a su ambiente nativo versus el resultado del proyecto nIFTy.

En el trabajo también se exploró si las diferentes elecciones de definiciones de masa de halos o de funciones iniciales de masa para la formación podrían tener una contribución en esta dispersión, encontrando que homologando estas características no se presentan cambios en los comportamientos reportados, descartando de este modo la contribución dada por estas diferencias.

Si bien con los análisis realizados se puede esperar que una comparación de modelos calibrados considerando un conjunto de observables común sobre los mismos catálogos de halos y árboles de fusiones podría eliminar estas diferencias, el proceso requiere un trabajo adicional intenso de parte de los modeladores. No obstante, esta propuesta se ha iniciado como continuación del proyecto de comparación nIFTy en donde se utilizarán dos conjuntos de restricciones para la recalibración de todos los modelos participantes. Esto permitirá evaluar la factibilidad de predecir propiedades similares de la población

de galaxias para conjuntos estándar de observables a pesar de las distintas implementaciones de los procesos físicos introducidos en los códigos semianalíticos (Knebe et ál., en preparación).

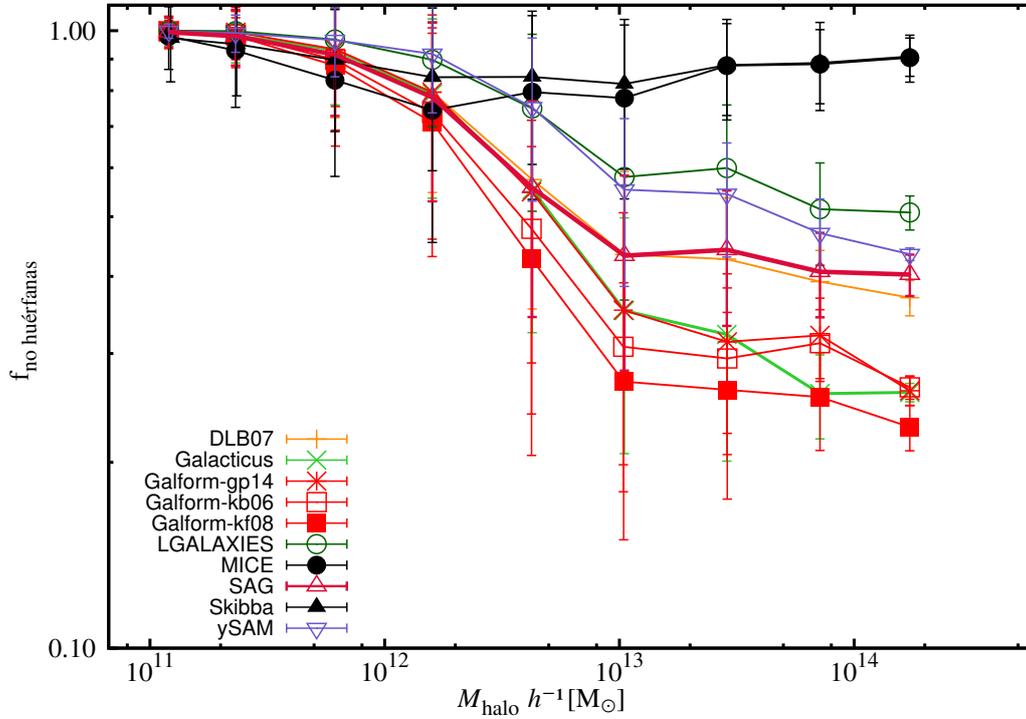


Figura 2.6: Predicciones a $z = 0$ de los modelos participantes en el proyecto de comparación nIFTy. Fracción numérica de galaxias no huérfanas $f_{\text{no huérfanas}}$ con masa estelar $M_{\star} > 10^9 h^{-1} M_{\odot}$ como función de la masa del halo principal

Otro resultado interesante del proyecto de comparación es la incidencia de las galaxias huérfanas en el modelado de la evolución de las galaxias. En la Figura 2.6, correspondiente al panel superior de la Figura 12 de Knebe et ál. (2015), se muestra la fracción del número de galaxias no huérfanas como función de la masa de los halos principales a $z = 0$ para todos los modelos que incluyen galaxias huérfanas en su tratamiento; esto es, excluyendo a los códigos: GALICS-2.0, MORGANA, SAGE y SANTACRUZ, que carecen de este tipo de galaxias en sus modelos. Se puede observar cierta bimodalidad en el comportamiento, donde los dos modelos HOD (SKIBBA y MICE) presentan las menores fracciones de galaxias huérfanas en los halos principales de mayores masas, mientras que en todos los modelos semianalíticos que utilizan galaxias huérfanas (como el SAG), estas forman la población dominante constituyendo entre el 40 y el 75 por ciento de todas las galaxias a este corrimiento al rojo. Inclusive, esta tendencia se mantiene para corrimientos al rojo más altos. Este comportamiento nos remarca la importancia que tiene la población de galaxias huérfanas en el modelado cuando se considera la evolución de la población de galaxias, aspirando a obtener propiedades con comportamientos consistentes para diferentes épocas. Esto justifica los esfuerzos para concretar un modelado más acabado y consistente de las galaxias satélites como el que trataremos en el Capítulo 3.

2.5. Calibración de simulaciones SUSSING

Como se ha mencionado en la sección anterior, la primera etapa del proyecto de comparación de modelos semianalíticos consistió en congregarse modeladores que ejecutaran sus códigos sobre un conjunto común de catálogos. Se mostró el impacto de no recalibrar los parámetros libres cuando se ejecutan sobre catálogos de halos y árboles de fusiones construidos con métodos distintos a los que los modelos utilizan de forma nativa. Esto justifica y motiva la segunda parte del proyecto de comparación, consistente en reiterar la ejecución de los modelos en catálogos comunes, pero requiriendo que los modelos, a su vez, estén calibrados específicamente con estos catálogos en base a un conjunto común de observaciones.

Para esta etapa, iniciada con una nueva reunión de trabajo realizada en los observatorios Carnegie en Pasadena, EE.UU., en julio de 2015, se suministró a los modeladores una nueva simulación de mayor volumen y número de partículas que la anterior. Esta simulación, también conocida como *Euclid*, fue construida utilizando los parámetros cosmológicos de Planck-2013 (Planck Collaboration: Ade et ál., 2014, $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_b = 0.048$, $\sigma_8 = 0.829$, $h = 0.677$, $n_s = 0.96$), utilizando 512^3 partículas de materia oscura con masa $1.24 \times 10^9 h^{-1} M_\odot$ cada una, en un volumen comóvil de $125 h^{-1}$ Mpc de lado. La evolución se almacenó en 108 instantáneas, sobre las que se construyeron los catálogos de halos utilizando ROCKSTAR, y sus correspondientes árboles de fusiones con CONSISTENTTREES.

El modelo SAG utiliza la técnica PSO para ajustar sus parámetros libres, en donde la exploración del espacio de parámetros se realiza utilizando una selección volumétrica de la simulación original. Como parte de la adaptación del código a este tipo de formatos, fue necesario desarrollar códigos específicos y reutilizables que permitieran la manipulación de forma genérica de conjuntos de catálogos almacenados en el formato SUSSING (en la versión obtenida a partir de CONSISTENTTREES). Este paquete de códigos es útil para poder concretar tanto el proceso de calibración del modelo como para realizar las fragmentaciones y conversiones de formato que ya hemos descrito en las secciones anteriores.

En esta sección describiremos el procedimiento utilizado para realizar la selección del subvolumen de calibración utilizando este tipo de catálogos, y luego terminaremos describiendo el paquete de códigos genérico que se ha desarrollado para este propósito, así como las vías de adquisición y el uso general.

2.5.1. Selección del subvolumen para la calibración

El proceso de selección del subvolumen utilizado para calibrar debe ser análogo al descrito en la Sección 2.1.4. No obstante, en este caso deben considerarse los bosques de halos, y el ordenamiento de los catálogos nuevos debe preservar la consistencia establecida por estas estructuras.

El proceso de selección se da inicio de forma análoga al original, es decir, tomando los halos correspondientes a la instantánea a $z = 0$ de la simulación y dividiéndolos según su ubicación espacial en n^3 subvolumenes. A cada uno de estos se le calcula la función de masa de halos, y esta se compara con la obtenida con el volumen completo con el fin de seleccionar aquel subvolumen que muestre un mejor acuerdo.

Luego de ello, para cada uno de los halos pertenecientes a este volumen seleccionado se debe identificar el bosque al cual pertenece, mediante la utilización de la lista de bosques creada por CONSISTENTTREES. De este modo, a partir de una lista de halos

se construye una lista de bosques que caracteriza el subvolumen elegido. Muchos de los halos originales podrían eventualmente pertenecer a un mismo bosque, con lo cual se deben descartar todos aquellos que estén repetidos en la lista. Por otro lado, una fracción de estos bosques seleccionados puede estar constituida además por halos que originalmente no estaban en el volumen seleccionado, los cuales deben ser agregados a la muestra para mantener la consistencia del catálogo. Dado que este proceso puede llegar a añadir muchos halos, agregamos un requerimiento adicional para la selección consistente en que la función de masa de la muestra no se vea alterada de forma significativa al incluir estos nuevos halos, es decir, exigiendo que el porcentaje de halos nuevos no supere aproximadamente el 1% del número de halos original. En caso contrario, el volumen es descartado y se procede a repetir el procedimiento con el siguiente cuya función de masa muestre un buen acuerdo.

Una vez obtenida la lista de bosques definitiva, se procede a reconstruir el conjunto de catálogos correspondiente a este subvolumen. Para ello se extrae cada bosque de las listas originales y se concatena a un nuevo conjunto de catálogos mediante la actualización de todos los archivos de forma simultánea, con el fin de preservar el orden estructurado de cada bloque.

Adicionalmente, en caso de que la selección para calibración tenga un número excesivo de halos (como puede suceder para simulaciones de gran tamaño, como sucedió con la simulación Euclid), el subvolumen elegido se puede particionar en un número arbitrario de fragmentos con el objetivo de realizar el procesamiento en simultáneo de ellos con el código SAG. Gracias a que en este tipo de catálogos tenemos disponible la lista de bosques que organiza los bloques de halos, utilizarla para realizar esta fragmentación es más simple que realizar particiones cúbicas del volumen, dado que estas últimas requieren un proceso adicional de búsqueda de halos y sus correspondientes bosques. Por lo tanto, implementamos además una estrategia de fragmentación de catálogos SUSSING que consiste en distribuir todos los bosques de una simulación en un conjunto de catálogos independientes de forma homogénea. Dada la autoconsistencia del formato elegido, esta fragmentación se puede aplicar tanto a simulaciones como a secciones de ellas tales como los subvolúmenes de calibración.

Finalmente, y dado que el rendimiento tiene una alta prioridad durante el proceso de calibración, todos los catálogos a utilizar para este proceso se almacenan preferentemente en formato SUSSING en binario, descrito en la Sección 2.3.4.

2.5.2. El paquete SUSSINGSCRIPTS

Como se ha descrito en diversas secciones de este capítulo, se implementó un conjunto de estrategias de lectura y manipulación de catálogos de halos y árboles de fusiones en formato SUSSING como parte del procesamiento del código semianalítico SAG. Si bien el fin de la creación de estos códigos ha sido principalmente resolver problemáticas relacionadas exclusivamente con el modelo, su utilización puede ser más generalizada que el uso particular que le hemos dado nosotros en este proyecto. Por este motivo, todos los códigos utilizados para la manipulación de estos catálogos los hemos agrupado en un único paquete llamado SUSSING_SCRIPTS, el cual se ha puesto a disposición pública en el repositorio oficial del autor⁴, mediante la licencia pública general GNU versión 3.

Debido a que el paquete se administra con MERCURIAL, en un sistema operativo tipo Unix puede ser descargado utilizando este software directamente desde la línea de

⁴https://bitbucket.org/cnvega/sussing_scripts/

comandos ejecutando:

```
1 hg clone https://bitbucket.org/cnvega/sussing_scripts
```

lo cual crea un directorio con el nombre del paquete y descarga el código fuente por completo. El paquete no tiene ninguna dependencia adicional más allá del compilador, por lo tanto se puede compilar directamente:

```
1 cd sussing_scripts
2 make
```

Si bien la mayoría de los ejecutables creados al momento de compilación se pueden utilizar directamente luego de la misma, algunos de ellos (detallados en la documentación adjunta en el paquete) requieren que el usuario edite sus códigos fuente para ser utilizados apropiadamente, en caso contrario estos no realizan cálculo alguno.

Entre los ejecutables creados más importantes del paquete podemos mencionar:

- `select_forests` que realiza la selección de los bosques a partir de una lista de halos;
- `extract_forests` que utilizando la lista creada por el programa anterior, extrae los bosques y crea un nuevo conjunto de catálogos con ellos;
- `extract_forests_block` que permite extraer un grupo consecutivo de bosques desde un conjunto de catálogos a partir de un bosque específico, y
- `split_forests` que permite tanto fragmentar una simulación en un número arbitrario de partes utilizando la lista de bosques, como hacer conversiones de formato entre ASCII y binario.

Además de ellos, se incluye un par códigos extra para realizar algunas pruebas de consistencia así como lecturas rápidas de archivos en binario.

Finalmente, en el paquete se incluye una pequeña documentación con instrucciones más detalladas de uso de cada uno de los códigos. Es importante notar que tanto esta ayuda como todo el código fuente del paquete estarán en constante actualización por el autor o los colaboradores voluntarios que quieran contribuir al desarrollo del mismo.

Debido a su licencia, toda versión derivada de este paquete debe respetar la política de código abierto, de modo que tal como este código fuente es público, garantizando su libertad de uso y de poder ser compartido, todas sus copias derivadas modificadas deben contar con las mismas libertades que el código original.

Capítulo 3

Modelo de galaxias satélites

Como fue detallado en el capítulo anterior, el código SAG es un código semianalítico de formación y evolución de galaxias que permite simular la evolución del contenido bariónico de un conjunto de halos de materia oscura mediante la estimación de las componentes de las galaxias que los habitan. Para esto utiliza catálogos de halos de materia oscura y sus correspondientes árboles de fusiones extraídos de una simulación de N -cuerpos cosmológica de materia oscura.

Para este propósito, hemos mencionado que el modelo no sólo utiliza los halos principales detectados en las simulaciones para asignarles sus respectivas galaxias centrales, sino que también utiliza los halos satélites como anfitriones de las galaxias satélites. Es decir, todo halo que se convierte en satélite de otro durante su evolución histórica preservará su galaxia central correspondiente mientras este se siga detectando.

Si se plantea el modelo más simple posible, donde el tiempo de vida de las galaxias se considera igual al de sus halos anfitriones, se obtendría un comportamiento donde el tiempo de supervivencia de las galaxias satélites en sistemas adquiere una dependencia importante con la resolución de masa de los halos detectados. Esta resolución, a su vez, depende tanto de la resolución de masa de la simulación utilizada, es decir, de la masa de las partículas de materia oscura, como de la cantidad mínima de partículas requerida por el buscador de halos para detectar una sobredensidad. Si se realiza el ejercicio de cambiar este último valor ligeramente en el proceso de detección de halos, implicaría que los tiempos de vida de las galaxias también varíen para una misma simulación. Este comportamiento es contrario a lo que se espera dado que la tasa de fusiones de galaxias y sus respectivos tiempos de supervivencia como satélites deberían ser independientes del límite de resolución de halos empleado.

Las variaciones en los tiempos de vida de galaxias satélites tienen un impacto importante en la evolución de la población completa resultante. Las galaxias satélites que viven más preservan por más tiempo el gas frío que se puede contaminar de forma progresiva por la realimentación química mientras acontece la formación estelar suave, también contribuyendo a la contaminación química del medio intragrupo/cúmulo en el cual se encuentran a través de la realimentación energética de las SNe. Estos mismos procesos tienen lugar en las correspondientes galaxias centrales, pero pueden verse alterados por la ocurrencia de fusiones más tempranas que aceleran la formación estelar mediante la aparición de brotes, adelantando la formación de los bulbos galácticos. Esto afecta no solo las relaciones morfológicas de las galaxias, sino también la evolución química, los colores y edades estelares tanto de las galaxias satélites como de sus respectivas galaxias centrales con las que se fusionan.

Por otro lado, las galaxias satélites atraviesan diferentes zonas de sus halos anfitriones mientras orbitan alrededor de sus respectivas galaxias centrales. Se ven expuestas a pérdidas de material debido a los efectos de ambiente tales como la presión de barrido (RPS) y las fuerzas de mareas (TS); este último proceso afecta igualmente a los halos de materia oscura satélites en los que residen las galaxias satélites. Ambos mecanismos de pérdida de material son incluidos entre los tratamientos extendidos del código SAG, detallados en la Sección 2.1.2.

Estos procesos dependen tanto de la densidad de material que se esté atravesando en la órbita, es decir, de la densidad de materia oscura y del gas que dependen de la distancia radial, como de las características del satélite respectivo (masa y radio). El impacto de estos efectos se verá pesado por la forma en que el tiempo de vida de las galaxias satélites es estimado, de modo que es importante establecer un buen criterio para definir este último.

En este sentido, la existencia de las galaxias huérfanas en el modelo semi-analítico constituye una buena solución para este problema, dado que este tipo de galaxias permiten compensar el tiempo de fusión faltante debido al límite de resolución que tiene cada simulación. Así, cada galaxia orbitará en torno a su galaxia central respectiva y vivirá durante una escala temporal que dependerá principalmente de sus condiciones iniciales. De este modo, si su halo anfitrión desaparece, la galaxia satélite pasa a ser huérfana y se continúa trazando su evolución hasta que se fusione o se destruya por los efectos de ambiente que provocan pérdida de masa.

No obstante lo anterior, una galaxia huérfana carece de la información tanto de la posición como de la masa del halo que la contenía dado que aquel ha dejado de ser detectable en los catálogos. Por lo tanto, para mantener la consistencia en el tratamiento evolutivo que reciben todas las galaxias satélites, sean huérfanas o no, en especial en lo que concierne a los efectos de ambiente, las órbitas de las galaxias huérfanas deben estimarse de forma analítica desde el momento en que sus halos anfitriones desaparecen hasta que se cumple el criterio de fusión o se destruya. Para esto se realiza una extrapolación analítica de la situación orbital y las propiedades de los halos de materia oscura que han dejado de ser detectados en la simulación a partir del instante de desaparición hasta completarse el ciclo evolutivo de las galaxias asociadas a ellos. Esta estimación debe considerar no sólo la cinemática orbital, sino también la dinámica del sistema incluyendo fricción dinámica y la correspondiente pérdida de masa por efectos de mareas. Esto es así debido a que el cálculo se realiza con el fin de reproducir el comportamiento de los halos de materia oscura satélites no detectados, los cuales deben evolucionar de forma análoga a los que sí lo son.

Todos estos cálculos orbitales y de pérdida de masa toman tiempo, y si son incluidos dentro del código SAG el tiempo de cálculo requerido para ejecutar el modelo se incrementa significativamente. Para evitar este problema, aprovechamos que la predicción orbital de estos halos que albergarán galaxias huérfanas depende sólo de los árboles de fusiones y no cambia para diferentes ejecuciones con el modelo semianalítico. Entonces, todo el proceso de identificación de subhalos de los catálogos, cálculo de la evolución orbital y determinación de la pérdida de masa por TS de aquellos halos que eventualmente dejarán de ser detectados, lo efectuamos en una etapa de procesado previa a la ejecución del código SAG. Esto se realiza con un código independiente llamado SUBORBITFINDER, el cual una vez completado almacena para cada instantánea la posición resultante de todos los halos (incluyendo los integrados pues dejan de ser detectados), y la masa estimada, indicando cuáles deberían fusionarse según el criterio de integración aplicado. Esta información es

leída por el código SAG en conjunto con los catálogos de halos para realizar el cálculo de la evolución de las galaxias.

En este capítulo nos centraremos en estudiar algunos tratamientos específicos aplicados a las galaxias satélites, los cuales se han analizado y reformulado para mejorar las predicciones analíticas implementadas. En la siguiente sección comenzamos describiendo el modelo analítico utilizado por el código para realizar la integración de las órbitas, para luego detallar los análisis que hemos llevado a cabo como parte de esta Tesis para ajustar ciertas libertades analíticas del modelo que afectan bastante la tasa de fusión y número de galaxias huérfanas predichas, así como mejoras técnicas introducidas en el mismo. Para finalizar el capítulo, en la Sección 3.2 analizaremos el impacto del ajuste de la dependencia de la presión de barrido con la masa del halo y el corrimiento al rojo que se aplica a todas las galaxias satélites modeladas por el código SAG, proponiendo una nueva funcional que estima de forma más consistente el impacto de la presión de barrido sobre el gas de las galaxias en base a resimulaciones de cúmulos de galaxias.

3.1. Modelo de órbitas de galaxias satélites huérfanas

El código SUBORBITFINDER encargado de modelar las posiciones de las galaxias satélites utiliza el mismo núcleo de rutinas que el código SAG, de modo que el flujo de cálculo funciona de forma análoga al presentado en la Sección 2.2. No obstante, se omite todo el cálculo de la evolución bariónica y se calcula en cambio la evolución dinámica de todos los halos que albergarán las galaxias satélites huérfanas de la población que posteriormente será simulada por el código SAG. De este modo, toda la lectura de catálogos y asignación de la población de galaxias debe ser recreada por este modelo de forma análoga a como lo hace el modelo semianalítico. Esto fuerza a que ambos códigos deban ser siempre consistentes entre sí para poder utilizar este modelado.

Mientras un halo satélite orbita dentro de su halo anfitrión pierde masa por TS, lo cual afecta su evolución orbital. Para las galaxias satélites tipo 1, donde aún se identifica el halo de materia oscura anfitrión, el efecto de la fricción dinámica y TS es estimado de forma consistente mediante el seguimiento de la evolución del halo detectado, de modo que sus propiedades medidas de posición, velocidad, masa y radio ya incorporan estos efectos, y el desprendimiento de material se puede estimar mediante el uso de estas variables. Entonces, para estas galaxias el código SUBORBITFINDER simplemente preserva las propiedades detectadas de sus halos anfitriones respectivos.

En el caso de las galaxias tipo 2, para las cuales el halo ha dejado de ser detectado, tomamos las últimas propiedades del halo de materia oscura detectado como condiciones iniciales de la integración de las órbitas de forma numérica. Esta integración, basada en los trabajos de Gan et ál. (2010) y Kimm et ál. (2011), se realiza mediante la utilización del método de salto de rana (por su nombre en inglés *leapfrog*) donde se actualizan progresivamente las posiciones y velocidades de los halos. Para esto, consideramos que el subhalo se puede aproximar como una fuente puntual orbitando en un pozo de potencial caracterizado por un perfil de densidad isoterma, medio que le imprime una fuerza de fricción dinámica. Según la aproximación de Chandrasekhar (1943), la desaceleración producto de la fricción dinámica para un satélite de masa M_{sat} , resulta

$$\frac{d\vec{v}}{dt} = -\frac{GM_{\text{sat}}}{r^2}u^{-2} \ln \Lambda_C \left[\text{erf}(u) - \frac{\sqrt{\pi}}{2}u \exp(-u^2) \right] \frac{\vec{v}}{v}, \quad (3.1)$$

donde r es la posición del satélite relativa a su galaxia central correspondiente, \vec{v} la

velocidad orbital, $u \equiv v/V_c$ con V_c la velocidad circular del halo que no depende de r , y $\ln \Lambda_C$ es el logaritmo de Coulomb que depende del parámetro de impacto de la interacción entre satélite y el entorno (Binney y Tremaine, 1987).

Aunque el valor del logaritmo de Coulomb en este contexto suele ser aproximado por

$$\ln \Lambda_C \simeq \ln(1 + M_{\text{cen},i}/M_{\text{sat},i}), \quad (3.2)$$

donde $M_{\text{cen},i}$ y $M_{\text{sat},i}$ son las masas de los halos central y satélite, respectivamente, definidos al momento de la acreción del satélite (como por ejemplo en Springel et ál., 2001), varios autores suelen utilizar otras definiciones que les permiten reproducir mediciones resultantes de simulaciones de N -cuerpos (ver por ejemplo Hashimoto et ál., 2003; Zentner y Bullock, 2003; Jiang et ál., 2008; Gan et ál., 2010; Petts et ál., 2015; Ogiya y Burkert, 2016), no habiendo un criterio único acordado en la literatura. En nuestro modelo nos inclinamos por usar un logaritmo de Coulomb variable como el definido por Hashimoto et ál. (2003), dado por

$$\ln \Lambda = \begin{cases} \ln(d_{\text{sat}}/bR_{\text{sat}}) & d_{\text{sat}} > bR_{\text{sat}}, \\ 0 & d_{\text{sat}} \leq bR_{\text{sat}}, \end{cases} \quad (3.3)$$

donde d_{sat} es la distancia del satélite al centro del halo anfitrión, R_{sat} es el radio virial del satélite y b es una constante que vale 1.4. Preferimos esta elección porque permite retardar el efecto de circularización y exceso de pérdida de masa de satélites que se reporta al comparar estos modelos con simulaciones de N -cuerpos (ver por ejemplo Fujii et ál., 2006). En nuestro modelo tratamos b como un parámetro libre debido a su impacto en esta materia (Vega-Martínez et ál., en preparación).

Durante la integración de la órbita también calculamos la masa que el halo satélite pierde debido a TS. El radio de desprendimiento de material estimado r_t corresponde a la distancia radial por sobre la cual la fuerza externa de mareas provocada por el anfitrión exceden la fuerza de ligadura del subhalo. Este valor puede ser aproximado por

$$r_t^3 = \frac{GM_{\text{sat}}(< r_t)}{\omega^2 + G[2M_{\text{cen}}(< r)/r^3 - 4\pi\rho(r)]}, \quad (3.4)$$

para un subhalo orbitando con velocidad angular ω a una distancia r del centro del halo anfitrión caracterizado por el perfil de densidad $\rho(r)$ que sigue la distribución de una esfera isoterma (Taylor y Babul, 2001; Zentner y Bullock, 2003). Esta estimación no considera las posibles interacciones del subhalo con otros halos satélites.

Por otro lado, también definimos el radio efectivo que contiene toda la masa del subhalo, r_{DM} , el cual inicialmente equivale al último radio del subhalo detectado R_{vir} . En cada paso temporal Δt de la integración se estima el valor actual de r_t , el cual se compara con el r_{DM} registrado en ese momento. Si $r_t < r_{\text{DM}}$, toda la masa contenida en esa zona es susceptible a ser perdida por el efecto de mareas. No obstante, como el proceso no es instantáneo, se remueve solamente una fracción $\Delta t/t_{\text{dyn,h}}$ de esta masa, donde $t_{\text{dyn,h}} = R_{\text{vir}}/V_{\text{vir}}$ es el tiempo dinámico del halo. Luego de la remoción, r_{DM} es actualizado suponiendo que el perfil de densidad de materia oscura no se ve alterado por este mecanismo.

Si bien es posible estimar con antelación el tiempo de vida de órbita de una galaxia hasta que esta se fusiona con su respectiva galaxia central (como por ejemplo los definidos por la Ecuación (3.1) o por las aproximaciones de Boylan-Kolchin et ál., 2008; Jiang et ál., 2008, que discrepan entre sí), dada la autoconsistencia del proceso de evolución orbital

calculado preferimos establecer criterios más adecuados a la dinámica para definir la finalización de la integración de la órbita. En el modelo se incluyen dos posibles criterios a evaluar: uno en base a la distancia mínima alcanzada con respecto al centro del halo, en donde se define una fusión cuando la órbita del satélite alcanza un factor 0.1 en términos del radio virial; y otro relativo a la pérdida de momento angular específico del satélite durante su evolución. En este último se define la fusión de las galaxias cuando sus halos integrados pierden más de un 99 % del momento angular específico inicial.

En las siguientes secciones presentaremos los cambios realizados a este modelo para cumplir los objetivos de la Tesis, y también el análisis realizado para definir un criterio consistente de término de integración de estas órbitas en base a los grados de libertad presentes en esta formulación analítica.

3.1.1. Adaptación a SUSSING

Debido a la similitud de estructura del código SUBORBITFINDER con el código SAG, muchos de los cambios presentados en la Sección 2.3 realizados en el modelo semianalítico debieron ser replicados en el integrador de órbitas. El más importante de ellos consistió en recrear la adaptación del código a la lectura y utilización de catálogos de halos y árboles de fusiones almacenados en el formato SUSSING.

La adaptación a este tipo de formato realizada en este código incluyó todas las características ya descritas en el capítulo correspondiente a este formato. El flujo de cálculo debió ser replanteado para permitir realizar un procesamiento *vertical* de los datos, en donde se calcula por árboles de fusiones completos, en vez del *horizontal* donde se calcula por instantáneas. Para esto se agregó un bucle extra que recorre todos los bosques leídos de los catálogos. Asimismo, se preparó el código para que en cada uno de los pasos del bucle de bosques fuese posible procesar en simultáneo más de una de estas estructuras, de modo que fue necesario incluir todas las estrategias de lectura y reordenamiento de datos que se implementaron en el código SAG, lo cual tiene un impacto importante en la reducción del número de accesos a disco duro que esto produce y la latencia de tiempo de cálculo asociada.

Si bien en el código SAG el orden final en que se almacenan las galaxias no necesita ser estrictamente fijo, la sincronía que debe existir entre este modelo y el integrador de órbitas exige que se preste crucial atención en este aspecto, en especial considerando que ambos se habilitaron para realizar su procesamiento respectivo mediante grupos de bosques de tamaño arbitrario. Por ello, fue necesario definir un criterio de ordenamiento estándar que se aplicara a ambos por igual, de modo que las lecturas de las posiciones de las galaxias fueran consistentes siempre y además fuese posible preservar la opción introducida en el modelo semianalítico que permite realizar cálculos parciales de una simulación. La estrategia de ordenamiento de galaxias que implementamos para este propósito se guía según los siguientes criterios:

- *i)* las galaxias centrales serán siempre seguidas por sus satélites priorizando las huérfanas, es decir, cada galaxia tipo 1 precederá a sus satélites tipo 2, y las tipo 0 serán seguidas por sus propias tipo 2 y luego por sus satélites tipo 1;
- *ii)* para galaxias centrales tipo 0 con varias satélites tipo 1, el orden entre estas últimas se establece según las masas de sus respectivos halos, de mayor a menor;
- *iii)* las galaxias pertenecientes a un mismo bosque se agruparán de forma consecutiva formando grupos y seguirán el mismo orden que los bosques originalmente

leídos de la lista correspondiente; y

- *iv)* dentro de cada grupo anterior, en caso de que exista más de un halo principal, estos se ordenarán según el identificador de sus halos correspondientes, de menor a mayor.

Este criterio queda establecido en la instancia de lectura de halos, por lo que no sólo se preserva en la escritura de las salidas sino también durante el cálculo realizado tanto por el código semianalítico como por el integrador de órbitas.

Todas estas modificaciones se realizaron con compatibilidad tanto para los catálogos SUSSING originales en ASCII, como para la versión en binario que hemos desarrollado, sin existir diferencia entre los resultados obtenidos en cada caso.

3.1.2. Ajuste de la tasa de fusiones de galaxias

En el modelo analítico de integración detallado en la introducción de la Sección 3.1 persisten dos grados libertad importantes que pueden alterar bastante el resultado del proceso de integración. Por una parte, la definición utilizada para el logaritmo de Coulomb que modula la intensidad de la desaceleración debido a la fricción dinámica presentada en la Ecuación (3.1), y altera de forma importante la rapidez con que las órbitas pierden energía y se circularizan. Órbitas con fricción dinámica sobreestimada tienden a reducir sus radios apocéntricos más rápidamente, de modo que pasan mayor tiempo con distancias radiales menores, afectando el impacto de los efectos de ambiente que dependen de esta distancia. Por otro lado, el criterio de fusiones de galaxias para dar término a la integración determina de forma directa el tiempo orbital de las galaxias satélites, con lo cual cada criterio utilizado debe ser justificado apropiadamente para evitar agregar resultados con tendencias sistemáticas en el modelado.

Una forma de determinar si el integrador tiene un comportamiento consistente con el resultante de simulaciones, es seleccionar un conjunto de halos satélites de una simulación particular que tengan una historia lo suficientemente extensa para luego intentar replicar su evolución mediante el modelo analítico. No obstante, tal como sucede con la estimación del tiempo de fusión de galaxias propuesta por [Boylan-Kolchin et ál. \(2008\)](#), la especificidad de la muestra podría generar un resultado que no es aplicable a poblaciones completas de halos en un contexto cosmológico ([McCavana et ál., 2012](#)). Por ello, este ejercicio se puede realizar como una prueba de funcionamiento pero carece de utilidad predictiva.

Otra forma propuesta para ajustar el modelo es mediante el estudio de la tasa de fusión de las galaxias satélites y compararla con datos obtenidos de observaciones de muestras grandes de galaxias (por ejemplo [Bell et ál., 2006](#); [Wen et ál., 2009](#); [De Propriis et ál., 2010](#); [Jiang et ál., 2012](#)). No obstante, dado que este tipo de estudios requiere establecer un criterio de proyección espacial para asignar una interacción entre galaxias comparable a las observaciones y poder determinar las fusiones, así como también requiere especificar algún tipo de fusión de galaxias, se obtiene como resultado una amplia dispersión entre las tasas de fusión (una buena discusión al respecto se puede encontrar en [Hopkins et ál., 2010a,b](#)). Una correcta comparación con las tasas de fusiones reportadas requeriría aplicar todo el tratamiento bariónico de evolución de galaxias de la población obtenida, el cual ya incluye de forma intrínseca más grados de libertad mediante los parámetros libres utilizados en el modelado. Los bariones, en efecto, tienen un efecto importante en las estimaciones de tiempo de fusión de galaxias satélites ([Zhang et ál., 2016](#)). Sólo utilizando un modelo

calibrado sobre una simulación de gran tamaño que permita estimar pares proyectados de galaxias, nos permitiría obtener tasas de fusión de galaxias que puedan ser comparadas con estos resultados (por ejemplo, la comparación realizada por [Patton y Atfield, 2008](#)).

En este sentido, nuestro modelo de integración propuesto no debe depender de la evolución bariónica de las galaxias. Es decir, tanto galaxias que se terminarán fusionando de forma efectiva con sus respectivas centrales, como aquellas que se destruyen antes de llegar a fusionarse producto de la pérdida de masa debida a los efectos de ambiente que actúan durante su evolución, deben ser tratadas por igual por esta integración. Se debe tener en cuenta que en este paso no se sabe cuál será el destino final de cada galaxia y, por consiguiente, su tiempo orbital real. Si bien debemos estimar el tiempo orbital de las galaxias que habitan estos halos, el destino final de cada una (fusión o destrucción) debe ser fijado por el modelo SAG y no por el integrador que está enfocado en la evolución de los halos de materia oscura. Por lo anteriormente planteado, hemos optado por construir una estimación teórica de la tasa de fusión de galaxias para todo corrimiento al rojo que nos permita ajustar nuestro modelo de evolución orbital. Para esto, hemos utilizado la simulación cosmológica de materia oscura *Euclid* mencionada en la Sección 2.5, consistente en un volumen comóvil de $125 h^{-1}$ Mpc de lado con resolución de masa $1.24 \times 10^9 h^{-1} M_{\odot}$.

Para este propósito, utilizamos una versión modificada del código de integración que decide el tiempo orbital de las galaxias centrales de los halos satélites mediante una estimación analítica al momento de acreción de los mismos. En base al estudio comparativo presentado por [McCavana et ál. \(2012\)](#), nos inclinamos por el tiempo de fricción dinámica definido por [Jiang et ál. \(2008\)](#) en vez de la estimación de Chandrasekhar ([Binney y Tremaine, 1987](#)) utilizada en el modelo básico de SAG hasta el momento, debido al mejor acuerdo alcanzado por el primero en un contexto cosmológico. Este tiempo estima la escala temporal de las galaxias en órbita hasta el momento de su fusión, y se define según

$$t_{\text{DF}} = \frac{0.94\eta_c^{0.60} + 0.60}{2C} \frac{M_{\text{cen}}}{M_{\text{sat}}} \frac{1}{\ln \Lambda} \frac{R_{\text{vir}}}{V_c}, \quad (3.5)$$

donde C es una constante que vale aproximadamente 0.43, el logaritmo de Coulomb $\ln \Lambda$ es definido según la Ecuación (3.2) y η_c es el parámetro de circularidad de la órbita del satélite. Este último es definido como la razón entre el momento angular específico de la órbita j y el momento angular de una órbita circular caracterizada por tener la misma energía $j_c(E)$, es decir

$$\eta_c \equiv j/j_c(E), \quad (3.6)$$

y puede ser expresado en términos de la excentricidad e según $\eta_c = (1 - e^2)^{1/2}$.

En el trabajo de [Jiang et ál. \(2008\)](#), donde se analiza la evolución de satélites en una simulación cosmológica de N -cuerpos con hidrodinámica, el tiempo de fricción dinámica equivale al tiempo que transcurre para una galaxia desde que se convierte en satélite hasta que se fusiona con su correspondiente galaxia central en torno a la cual orbita. Por lo tanto, en nuestro modelo lo calculamos para cada halo de materia oscura apenas se convierte en satélite, para luego asignar la fusión de su galaxia asociada con la central respectiva una vez transcurrido este tiempo. En caso de que el halo deje de ser detectado en los catálogos antes de haber transcurrido este tiempo, su galaxia pasa a ser huérfana por el tiempo restante. Por el contrario, en caso que el halo satélite siga siendo detectado habiendo transcurrido esta estimación, la fusión se asigna sólo cuando dicho halo deja de ser detectado de modo que su galaxia correspondiente no pasa por la etapa de huérfana.

Este tiempo es recalculado debidamente para toda galaxia que cambie su galaxia central respectiva para preservar la consistencia en todos los casos posibles. Además, como nuestro resultado es netamente estadístico, utilizamos un valor medio para la circularidad $\eta_c = 0.5$ (siguiendo los resultados de Tormen, 1997), así evitamos calcular numéricamente este valor en cada caso.

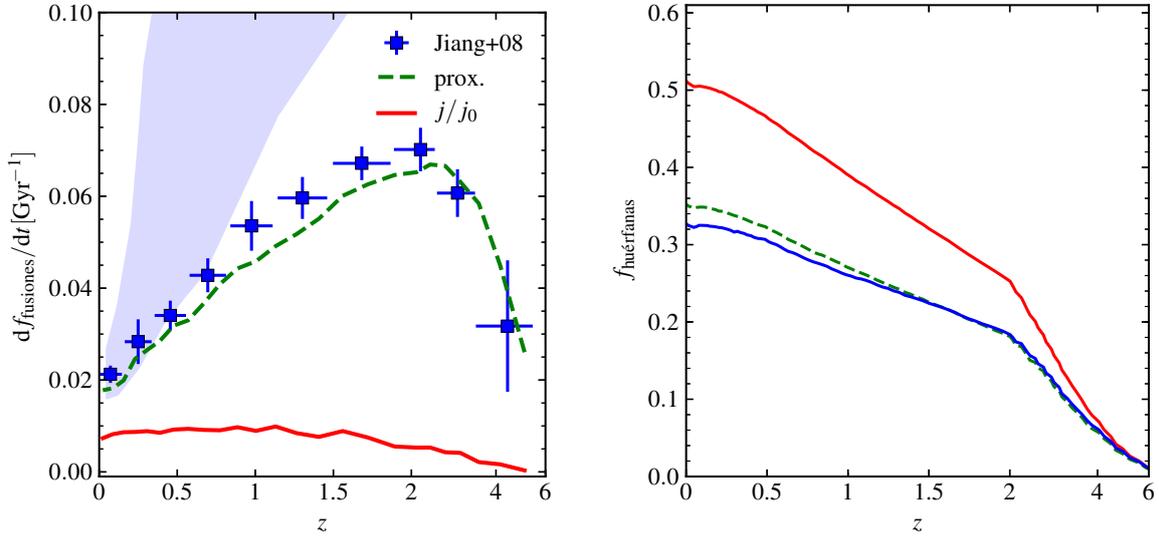


Figura 3.1: *Panel izquierdo*: evolución de tasas de fusión de galaxias. Se representa con cuadrados azules la tasa teórica resultante utilizando el tiempo de fricción dinámica definido por Jiang et ál. (2008). Con las líneas se muestran las tasas resultantes de utilizar el integrador considerando dos criterios para definir las fusiones de galaxias: por pérdida de momento angular específico (línea continua roja) y por proximidad (línea a trazos verde). La zona coloreada muestra el rango permitido por las observaciones recopiladas por Hopkins et ál. (2010a). *Panel derecho*: evolución de la fracción de galaxias satélites huérfanas resultante con cada uno de los modelos anteriores (se usa la misma consigna en cuanto al color de las líneas).

La tasa de fusiones resultante df_{fusiones}/dt la estimamos como el número de fusiones por unidad de tiempo, respecto al total de galaxias que componen la muestra en ese instante. Este resultado es mostrado en el panel izquierdo de la Figura 3.1, donde los cuadrados azules representan la tasa de fusiones obtenida utilizando el tiempo de fricción dinámica de Jiang et ál. (2008). El cálculo preciso de la tasa de fusiones debe realizarse en cada instantánea para mantener consistencia con el número total de galaxias en cada momento. Debido a que tenemos más de cien instantáneas en la simulación, cada punto representa la tasa de fusión media de un conjunto de 10 instantáneas cada uno. Las barras verticales indican la desviación de un sigma de este cálculo, y las horizontales el rango de corrimiento al rojo en cada caso. La zona coloreada muestra el rango aproximado permitido por las observaciones de fusiones mayores compilado por Hopkins et ál. (2010a), que aunque no es directamente comparable con lo que calculamos en este ejercicio, nos sirve como buena referencia del orden de magnitud que se debe alcanzar en este cálculo. Aun así, se puede apreciar que la predicción realizada de la tasa de fusión de galaxias para esta simulación es consistente con el rango observacional en corrimientos al rojo bajos, pero es progresivamente subestimada hacia corrimientos al rojo mayores. Esta diferencia

se podría acrecentar una vez aplicado el tratamiento de evolución de galaxias con el código SAG, dado que muchas galaxias podrían destruirse antes de llegar a fusionarse de forma efectiva. Es importante destacar que nuestro cálculo también permite establecer una predicción para $z > 2.5$, rango que está fuera de los datos de [Hopkins et ál. \(2010a\)](#). Estas diferencias pueden ser atribuidas a distintas razones, que pueden ir desde la baja resolución de masa de la simulación utilizada hasta la elección del tiempo de fricción dinámica elegido. No obstante, si se compara con la dispersión conocida que muestran otros modelos semianalíticos de formación de galaxias (ver por ejemplo la Figura 2 de [Hopkins et ál., 2010b](#)), el buen acuerdo alcanzado en el universo local es suficiente para ajustar los parámetros libres de nuestro modelo de órbitas.

En base a este cálculo, procedimos a evaluar el impacto de dos criterios de fusión implementados en el modelo. El primero considera una fusión cuando el momento angular específico j de la órbita del satélite se reduce un 99 % en comparación con el valor que tenía en el instante de acreción, j_0 , es decir $j/j_0 = 0.01$. El segundo criterio define una fusión cuando la distancia pericéntrica de la órbita alcanza una distancia menor al 10 % del radio virial del halo anfitrión en base a las estadísticas presentadas por [Shen et ál. \(2003\)](#). Para realizar esta evaluación ejecutamos el integrador utilizando la definición del logaritmo de Coulomb dada por la Ecuación (3.3) con el valor de la constante $b = 1.0$ a modo de referencia, y aplicando los dos criterios distintos para determinar las fusiones. Ambos resultados son incluidos en la Figura 3.1 mediante la línea continua roja y la línea a trazos verde respectivamente. Si bien la forma de estas curvas así como el valor en los máximos puede variar ligeramente con otras elecciones tanto de la constante b que interviene en la definición adoptada del logaritmo de Coulomb según la Ecuación (3.3), como del porcentaje de pérdida de momento angular específico que define la fusión, el comportamiento general de la tasa de fusión de galaxias versus el corrimiento al rojo es característico de cada uno de los criterios, incluso ante cambios en la definición del logaritmo de Coulomb. En particular, utilizar el momento angular como criterio de fusión provoca una persistente subestimación de la tasa de fusiones a bajo z , y no permite reproducir el máximo marcado que es predicho alrededor de $z \sim 2.5$. En el panel derecho de la figura se muestra la fracción de galaxias huérfanas resultantes en ambos modelos, incluyendo también la predicción en base al tiempo de fricción dinámica de [Jiang et ál. \(2008\)](#) (línea llena azul). Aquí se puede notar de forma mucho más clara el impacto del criterio utilizado para las fusiones, en donde el basado en el momento angular específico genera una población de galaxias compuesta de hasta 50 % de huérfanas a $z = 0$, mientras que los otros no superan el 35 %.

Según esta comparación, el criterio de fusión basado en la pérdida de momento angular específico, principal mecanismo que se venía utilizando hasta el momento, es ineficiente para determinar de forma consistente el tiempo orbital de las galaxias satélites, por lo cual en adelante preferiremos el basado en la proximidad de la órbita respecto al centro. No obstante, la estimación del 10 % de los radios viriales extraída de la estadística realizada por [Shen et ál. \(2003\)](#) sólo es válida para el universo local, de modo que esta cota no es confiable para corrimientos al rojo altos. Por ello, nuestro criterio debería ser refinado mediante el estudio de la evolución de los radios de las galaxias en comparación con los radios de los halos que las contienen, utilizando para ello modelos de formación de galaxias en simulaciones de muy alta resolución en masa, trabajo que queda planteado a futuro.

Logaritmo de Coulomb

Una vez establecido el criterio para terminar la integración, es menester determinar si nuestra elección del logaritmo de Coulomb es idónea, así como también ajustar las posibles constantes libres que estas definiciones incluyan. Para ello procedimos a ejecutar el modelo completo de integración de órbitas pero utilizando las dos definiciones presentadas en este capítulo para el logaritmo de Coulomb según las Ecuaciones (3.2) y (3.3), siempre utilizando el criterio de fusión definido en la sección anterior.

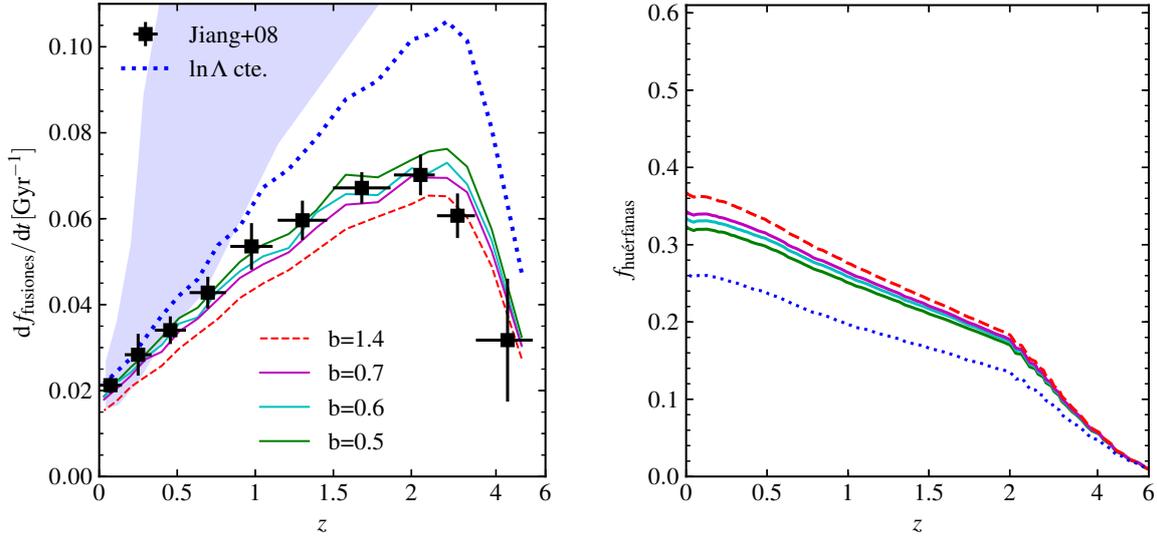


Figura 3.2: Comparación de resultados obtenidos a partir de diferentes elecciones para $\ln \Lambda$ en el integrador en base a un mismo criterio de fusión. *Panel izquierdo*: evolución de la tasa de fusiones comparada con la obtenida utilizando el tiempo de fricción dinámica dado por Jiang et ál. (2008) (representada por cuadrados negros). La línea punteada azul representa los resultados obtenidos de aplicar la definición dada por la Ecuación (3.2), mientras que las líneas llenas verde, celeste y magenta, y la de trazos rojos corresponden a distintos valores de la constante b utilizada en la definición del logaritmo de Coulomb dada por la Ecuación (3.3). *Panel derecho*: fracción de galaxias huérfanas obtenidas en cada caso presentado.

Esta comparación se presenta en la Figura 3.2, en cuyo panel izquierdo se muestra la evolución de la tasa de fusiones de galaxias obtenida, comparada con la resultante utilizando el tiempo de fricción dinámica de Jiang et ál. (2008), y en el panel derecho la proporción de galaxias huérfanas resultantes de cada modelo. En esta comparación hemos elegido cinco estimaciones para el logaritmo de Coulomb: una siguiendo la definición de la Ecuación (3.2) que se calcula al momento de acreción de cada halo y se mantiene constante durante la evolución (línea punteada azul), y otras cuatro estimaciones utilizando el formalismo variable dado por la Ecuación (3.3) pero con diversas elecciones para la constante b , la original de Hashimoto et ál. (2003) donde vale 1.4 (línea roja a trazos) y también con $b = 0.7$, 0.6, y 0.5 (líneas continuas magenta, celeste y verde, respectivamente). Si bien la obtenida con el logaritmo de Coulomb constante se acerca más al rango observacional, a su vez se aleja excesivamente de la predicha por nuestra estimación realizada en base al tiempo de fricción dinámica de Jiang et ál. (2008) para esta simulación. Por otro lado, la formulación original del logaritmo de Coulomb variable con $b = 1.4$

subestima la tasa de fusión global a todo z . Como la elección de la constante b en varios trabajos siempre ha sido efectuada mediante inspección visual, incluimos más pruebas con valores menores con el fin de encontrar un valor que mejor ajuste la predicción de las tasas de fusión. De las mostradas en la figura, nos inclinamos por la que utiliza $b = 0.6$ debido a su mejor comportamiento en relación con las otras, en especial alrededor del máximo y para altos corrimientos al rojo. Por otro lado, entre las posibles elecciones para b , la fracción de galaxias huérfanas obtenida no varía de forma significativa.

Dado el impacto que la elección de un logaritmo de Coulomb variable tiene en la estimación de circularización y en evitar un desprendimiento apresurado de material de los halos integrados, optamos por utilizar en nuestro integrador esta definición en vez de la definición constante. No obstante, esta comparación debe ser recreada en caso de querer estudiar una simulación que tenga otra cosmología, tipos de catálogos o una resolución en masa muy distinta a la de la simulación *Euclid* que hemos utilizado en este análisis.

3.1.3. Corrección por periodicidad

Otra modificación importante que le realizamos a este modelo de integración fue agregarle soporte para poder calcular las órbitas de las galaxias satélites huérfanas en simulaciones con volúmenes periódicos, característica que hasta entonces aún no había sido implementada en el código.

En este tipo de simulaciones con condiciones de borde periódicas, es común encontrar halos de materia oscura que se localizan en las cercanías de los bordes de las simulaciones, estando sus partículas compartidas entre un extremo y otro de una coordenada espacial determinada, o las tres inclusive (para aquellos que se encuentran en una esquina de la simulación). Esto también desencadena que un halo pueda contener satélites cuyos centros se localicen en los extremos opuestos de una coordenada particular.

Debido a que el integrador resuelve la evolución orbital en términos de coordenadas relativas para cada halo principal, este procedimiento deja de ser consistente al momento de realizar la estimación numérica para este tipo de halos satélites que quedan en extremos opuestos de la simulación en comparación con el centro de sus respectivos halos principales. El cálculo genera que las velocidades relativas queden mal expresadas, aplicándole impulsos ficticios enormes a las galaxias integradas, provocando que se alejen distancias importantes de la simulación. Al no tener una corrección por periodicidad consistente, aquellas galaxias cuyas posiciones son integradas de forma incorrecta pueden terminar por fuera de los bordes originales de la simulación en cualquier instante de la integración. Un ejemplo de este tipo de comportamiento se puede apreciar en el panel izquierdo de la Figura 3.3, donde se muestran las coordenadas espaciales x versus y solamente de las galaxias huérfanas que se localizan por fuera de los bordes de la simulación (marcado con las líneas de color rojo), para una integración de órbitas aplicada a la simulación *Euclid*.

El proceso implementado para corregir este comportamiento sin perder consistencia de la integración procede de la siguiente manera:

1. En cada paso del bucle de instantáneas, se verifican los satélites de todos los halos principales de la instantánea que da inicio a la integración. Si la distancia de una coordenada espacial entre el centro de un satélite y el de su halo principal es mayor que la mitad del lado del volumen de la simulación $0.5L_{\text{sim}}$, se corrigen todos sus satélites (sumando o restando L_{sim} a la coordenada correspondiente) de modo que queden en el mismo borde de la simulación que el centro del halo principal.

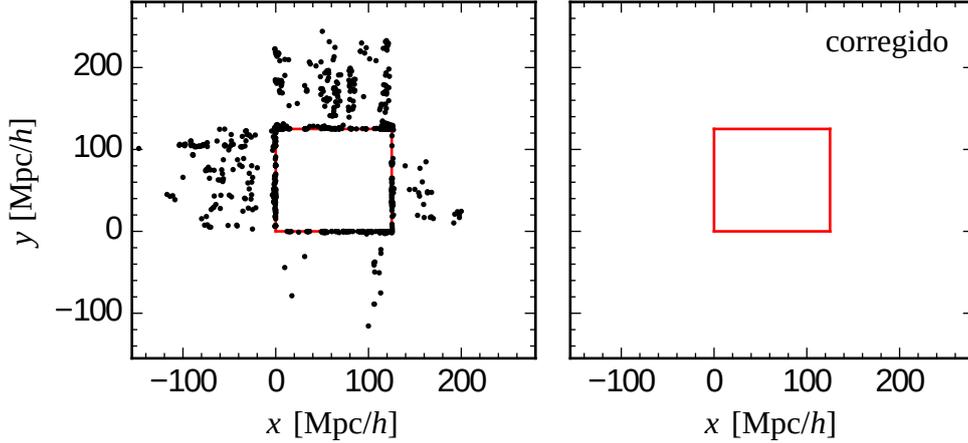


Figura 3.3: Ejemplo de la corrección por periodicidad en las galaxias huérfanas que luego de la integración terminan fuera de la simulación, cuyo borde es marcado con líneas de color rojo. *Panel izquierdo*: integración orbital sin la corrección. *Panel derecho*: integración orbital con la corrección aplicada.

2. Luego se verifican las posiciones de todos los halos progenitores con respecto a las de sus respectivos descendientes (pertenecientes a la instantánea final de la integración en este paso). Si la distancia entre ellos es mayor que $0.5L_{\text{sim}}$, la misma corrección anterior es aplicada a los descendientes. Esto garantiza que no hayan saltos de coordenadas durante todo el proceso de integración.
3. Luego de integrar, se corrigen todos los halos que quedaron fuera de los límites de la simulación en alguna coordenada particular, sumándoles o restándoles L_{sim} cuando corresponda.

En la panel derecho de la Figura 3.3 se muestra el resultado del proceso de integración con el mismo criterio de selección que en el otro panel, pero este modelo tiene activado el proceso de corrección por periodicidad. Como se puede apreciar, ante una misma simulación este algoritmo limpia todas las galaxias que inicialmente eran tratadas de forma incorrecta por el integrador de órbitas, agregando la consistencia necesaria para poder procesar una simulación completa con este integrador de galaxias huérfanas.

3.2. Presión de barrido sobre el gas

Entre los efectos de ambiente que afectan la evolución de las galaxias satélites tenemos la presión de barrido ejercida por el medio intragrupo/cúmulo sobre la componente gaseosa de las galaxias. Como se describió en la Sección 2.1.2, la presión de barrido (P_{ram}) provoca la pérdida de masa cuando supera a la fuerza gravitacional por unidad de área de la componente respectiva (halo de gas caliente o gas frío del disco).

El valor de P_{ram} está determinado tanto por la posición de la galaxia satélite en el halo de materia oscura anfitrión como de la órbita que tiene la misma, dado que se define según

$$P_{\text{ram}} \equiv \rho_{\text{ICM}} v^2, \quad (3.7)$$

donde ρ_{ICM} es la densidad del medio intracúmulo en el halo anfitrión y v es la velocidad relativa entre el satélite y el medio, ambas dependientes de la distancia al centro del halo. Debido a esta dependencia con la distancia radial, es importante que la evolución orbital de las galaxias huérfanas esté bien trazada con el fin de tener un tratamiento consistente para este efecto de ambiente en todas las galaxias satélites de los sistemas.

Para estimar el valor de esta presión tenemos dos métodos posibles. El primero, propuesto por [Tecce et ál. \(2010\)](#), es válido cuando se utilizan simulaciones cosmológicas de N -cuerpos de materia oscura que también tracen la evolución del gas mediante una componente hidrodinámica. En ellas, y conociendo con antelación las posiciones y velocidades orbitales de las galaxias satélites, la presión efectiva que estas sentirían se puede calcular directamente a partir de la distribución de partículas de gas de la simulación. Tal como la integración de órbitas, esto se puede hacer previamente a calcular la evolución de galaxias con el código SAG, de modo que estas estimaciones se lean desde el mismo para obtener con precisión el impacto de este efecto. No obstante, esto no es aplicable a simulaciones de materia oscura solamente.

El otro método, presentado por [Tecce et ál. \(2011\)](#), consiste en utilizar un ajuste fenomenológico de la distribución radial de presión de barrido que es sentida por los satélites. Esto fue definido en base a un conjunto de resimulaciones de N -cuerpos hidrodinámicas de cúmulos de galaxias, y la presión de barrido resultante depende de la distancia al centro de cada halo anfitrión, de la masa del mismo y del corrimiento al rojo.

En esta sección analizaremos el ajuste para el perfil radial de la presión de barrido propuesto por [Tecce et ál. \(2011\)](#) identificando sus falencias, y terminaremos proponiendo un nuevo perfil que permite reflejar de mejor forma el comportamiento de la presión de barrido para diferentes masas de halos anfitriones.

3.2.1. Descripción del ajuste original

El ajuste original para el perfil radial de la presión de barrido ([Tecce et ál., 2011](#)) se realizó combinando el código SAG con el conjunto de resimulaciones hidrodinámicas/ N -cuerpos de cúmulos de galaxias presentada por [Dolag et ál. \(2005, 2009\)](#). Estas consisten en resimulaciones de regiones centradas en cinco cúmulos con masa $\sim 10^{14} h^{-1} M_{\odot}$ (nombradas originalmente como g1542, g3344, g6212, g676 y g914) y tres cúmulos con masa $\sim 10^{15} h^{-1} M_{\odot}$ (g51, g1 y g8). Para todas ellas se utilizó una cosmología caracterizada por $\Omega_{\text{m}} = 0.3$, $\Omega_{\Lambda} = 0.7$, $\Omega_{\text{b}} = 0.039$, $h = 0.7$ y $\sigma_8 = 0.9$. La masa de las partículas de materia oscura utilizada para todas es $1.13 \times 10^9 h^{-1} M_{\odot}$, y las partículas de gas $1.69 \times 10^8 h^{-1} M_{\odot}$ (más detalles se pueden encontrar en [Dolag et ál., 2009](#)). Toda la identificación de halos así como la obtención de propiedades y árboles de fusiones se realizó con SUBFIND.

Tomando las posiciones de todas galaxias satélites calculadas por el código SAG que habitan los halos de las resimulaciones, se procedió a medir la presión de barrido efectiva que siente cada una de ellas en base a las propiedades de las partículas de gas contenidas en las resimulaciones mediante el método propuesto por [Tecce et ál. \(2010\)](#). Esto genera una muestra de valores de P_{ram} para diferentes tipos de halos a distintos corrimientos al rojo, aunque para el ajuste sólo se utilizaron las instantáneas comprendidas en el rango de $0 \leq z \leq 3$. Para las galaxias satélites tipo 1, las posiciones fueron establecidas por la ubicación de sus halos. No obstante, como el integrador de órbitas aún no había sido desarrollado al momento del ajuste, para las satélites huérfanas (satélites tipo 2), se utilizó simplemente la posición de la partícula más ligada del halo al que estas galaxias pertenecían la última vez que el mismo fue detectado. En [Tecce et ál. \(2010\)](#) se mostró que la presión de barrido

tendía a incrementarse con el corrimiento al rojo y la masa del halo anfitrión. Por ello, para este ajuste se separó la muestra de datos en 10 intervalos logarítmicos en masa, con ancho $\Delta \log(M_{200} [h^{-1} M_{\odot}]) = 0.5$ en el rango $12.5 \leq \log(M_{200} [h^{-1} M_{\odot}]) < 15.35$, de modo que cada intervalo en masa utilizado está superpuesto con sus adyacentes. En cada intervalo de masa se construyó un perfil de presión de barrido combinando todas las mediciones versus la distancia radial al centro del halo anfitrión en términos del valor de R_{200} de cada uno.

En [Tecce et ál. \(2011\)](#) se exploró un conjunto de formas funcionales comúnmente utilizados en la literatura para modelar los perfiles radiales de la presión de barrido obtenidos en estas resimulaciones. De todos los modelos propuestos para ajustar estos perfiles, el que arrojó mejores resultados es el beta completo, definido por

$$P_{\text{ram}}(M, z) = P_0(M, z) \left[1 + \left(\frac{r}{r_s(M, z)} \right)^2 \right]^{-\frac{3}{2}\beta(M, z)}, \quad (3.8)$$

donde el valor central P_0 , el radio característico r_s y el exponente β son parámetros libres que dependen tanto de la masa del halo anfitrión como del corrimiento al rojo. Utilizando este perfil, procedieron a ajustar estos tres parámetros mediante minimizaciones de χ^2 en cada uno de los intervalos en masa creados, y en cada instantánea del rango de corrimiento al rojo seleccionado. De este modo, obtuvieron un conjunto de valores para los tres parámetros libres del perfil para cada muestra en masa y en corrimiento en rojo.

Para obtener la dependencia de los parámetros obtenidos con la masa y z , los autores propusieron que todos evolucionan de forma lineal en función del parámetro de expansión del universo a_{exp} , según

$$\begin{aligned} \log \left(\frac{P_0}{10^{-12} h^2 \text{ dyn cm}^{-2}} \right) &= A_P + B_P(a_{\text{exp}} - 0.25), \\ \frac{r_s}{R_{200}} &= A_r + B_r(a_{\text{exp}} - 0.25), \\ \beta &= A_\beta + B_\beta(a_{\text{exp}} - 0.25), \end{aligned} \quad (3.9)$$

donde los coeficientes A_i y B_i , con $i = P, r, \beta$, dependen todos, en principio, de la masa virial del halo anfitrión. Los nuevos parámetros que definen estas relaciones fueron encontrados igualmente mediante minimizaciones de χ^2 , utilizando para ellos los valores encontrados en el ajuste anterior. Finalmente, la dependencia con la masa que tiene cada uno de estos seis parámetros fue determinada con un último ajuste. En él se utilizaron relaciones lineales de la forma

$$a + b(\log M_{200} - 12), \quad (3.10)$$

con a y b dos parámetros libres, duplicando con ello el número de parámetros a encontrar. El resultado numérico de estos doce coeficientes nuevos es mostrado en detalle en la Ecuación 6 de [Tecce et ál. \(2011\)](#). Para todos los valores encontrados, el error asociado a cada uno de ellos es bastante variable, alcanzando como máximo un 25% en un par de casos. Como resultado también de este ajuste, se determinó que tanto A_β como B_β no tienen dependencia con la masa del halo anfitrión, de modo que β sólo es función del corrimiento al rojo. De este modo, todos los perfiles de presión de barrido, dependientes de la distancia radial al centro del halo, quedan totalmente descritos por 10 constantes numéricas que definen la sensibilidad tanto temporal como con respecto a la masa del halo anfitrión para cada galaxia satélite a la que se le calcule esta cantidad.

Limitaciones del ajuste

Si bien cada una de las partes del ajuste presentado anteriormente fue realizada de forma correcta, el conjunto completo parámetros encontrados se caracteriza por predecir perfiles de la presión de barrido con un comportamiento mucho más suave que el originalmente medido utilizando las partículas de gas. Esto lo hemos determinado mediante la comparación entre la predicción de la presión de barrido obtenida con el ajuste de [Tecce et ál. \(2011\)](#) y la determinada por la distribución de las partículas en las mismas resimulaciones utilizadas para realizar el ajuste. Esta discrepancia se debe a que en cada ajuste realizado, la tendencia de los parámetros involucrados en el proceso se ve considerablemente suavizada y homogeneizada, al punto de perder la dependencia original que la presión de barrido tiene con la masa de los halos anfitriones.

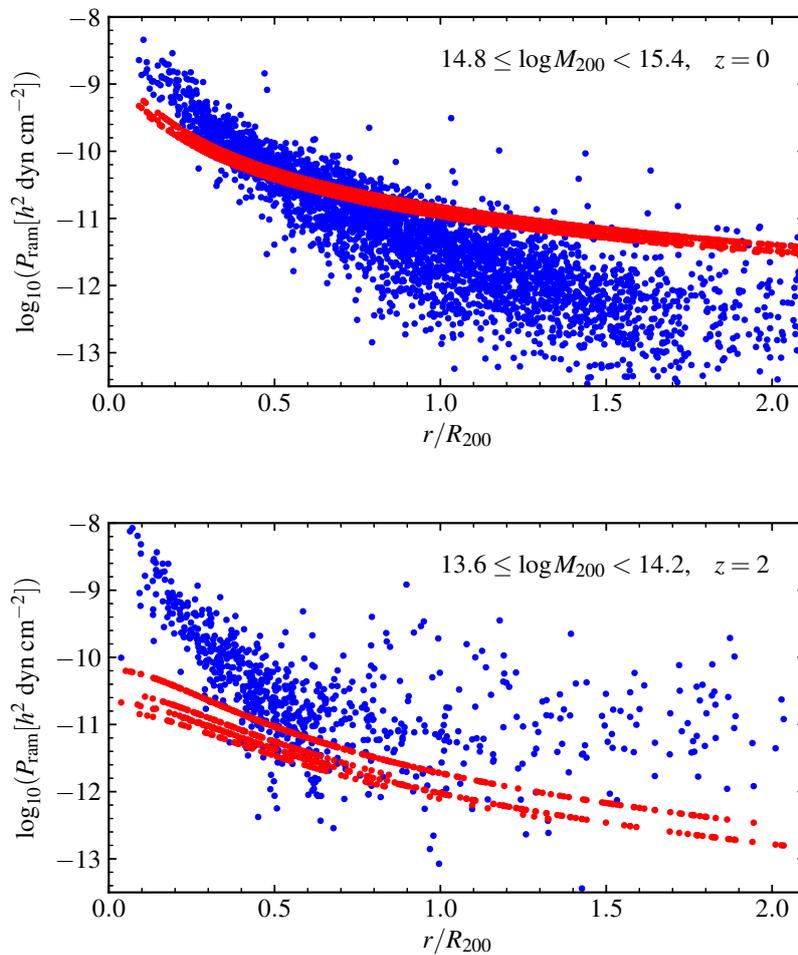


Figura 3.4: Comparación de la presión de barrido medida en una resimulación (en símbolos azules) con el ajuste presentado por [Tecce et ál. \(2011\)](#) para los mismos satélites (en puntos rojos). *Panel superior*: selección de los halos principales más masivos de la instantánea a $z = 0$ con $14.8 < \log(M_{200}[h^{-1} M_{\odot}]) < 15.4$. *Panel inferior*: selección de los halos principales más masivos de la instantánea a $z = 2$ con $13.6 < \log(M_{200}[h^{-1} M_{\odot}]) < 14.2$.

Un ejemplo de esto se puede apreciar en la Figura 3.4, en la que se muestran dos comparaciones entre los valores de P_{ram} versus la distancia radial del satélite r relativa al radio virial de los halos anfitriones R_{200} . Los valores de la presión de barrido medidos

en las partículas de gas de las resimulaciones de cúmulos de galaxias son representados con puntos azules, mientras que los obtenidos mediante el ajuste para las mismas ubicaciones son representados con puntos rojos. En la figura se muestran dos instantáneas: a $z = 0$ en el panel superior y a $z = 2$ en el panel inferior. Las mediciones corresponden únicamente a los halos satélites (galaxias tipo 1) de todos los halos principales no contaminados con partículas de baja resolución que se pudieron extraer del conjunto completo de resimulaciones. En ambos casos, se seleccionó el rango más masivo de halos principales que pudiera reflejar el comportamiento con pendiente más abrupta hacia el centro de los halos que caracteriza a la presión de barrido. Como se puede apreciar en ambos casos, la presión ejercida a distancias cercanas al centro es subestimada por el ajuste por hasta tres órdenes de magnitud, tendencia que es persistente y creciente a corrimientos al rojo mayores. De ello podemos concluir que si bien el perfil propuesto por [Tecce et ál. \(2011\)](#) según la Ecuación (3.8) es capaz que reproducir el comportamiento radial de la presión de barrido para un corrimiento al rojo y rango de masa de halo principal particular, la generalización de este modelo mediante las Ecuaciones (3.9) y (3.10) no produce resultados consistentes. En general, en el rango de corrimientos al rojo considerado, el comportamiento del modelo ajustado por [Tecce et ál. \(2011\)](#) tiende a predecir valores para la presión de barrido similares a los de la población menos masiva de halos, que resulta ser más numerosa y con mayor dispersión. No obstante, esta población carece de la tendencia creciente hacia distancias radiales menores mostrada en la Figura 3.4, la cual se acentúa en los halos anfitriones más masivos. A su vez, el ajuste tiende a subestimar de forma general los valores de la presión de barrido para corrimientos al rojo altos.

Esta inconsistencia resultante en la predicción se explica estudiando el efecto que tienen los ajustes realizados sobre el parámetro P_0 . Este factor del perfil se encarga de controlar la normalización general del ajuste, y en consecuencia, modula también el máximo que alcanza P_{ram} hacia las zonas más interiores de los halos, teniendo una fuerte dependencia con la masa. Debido a que fue ajustado primeramente con el factor de expansión del universo utilizando rectas, la tendencia con la masa quedó naturalmente suavizada. Este efecto incluso se puede apreciar en el panel superior de la Figura 4 de [Tecce et ál. \(2011\)](#), donde se muestra un ejemplo del ajuste temporal de este parámetro para un rango de masa específico. Si bien la recta superpuesta en aquella figura sigue la tendencia general, el valor de P_0 original llega a ser hasta un orden de magnitud más alto que el obtenido por el ajuste lineal para el universo local. Por otro lado, la tendencia creciente a subestimar los valores predichos para corrimientos al rojo altos puede asociarse al rango limitado de instantáneas utilizado durante el ajuste.

La propensión a la subestimación hacia los centros de los halos implica que, para diferentes valores de masa de halos principales, la presión de barrido utilizada posteriormente en la evolución de las galaxias satélites tenga un comportamiento tendencioso a valores similares entre sí, aminorando considerablemente el impacto de la distancia radial de este efecto de ambiente.

3.2.2. Nuevo perfil de presión de barrido

Como primer intento de corregir las diferencias presentadas en la sección anterior intentamos rehacer el ajuste en tres instancias, tal como fue propuesto originalmente, pero teniendo en mente las limitaciones mencionadas. Sin embargo, esto no fue suficiente para alcanzar resultados concluyentes. La cantidad de parámetros libres y el número de ajustes realizados hicieron que las degeneraciones existentes entre los parámetros generen

dispersiones de los resultados lo suficientemente altas como para imposibilitar ajustes concluyentes y predictivos. Esto sucede incluso si se invierten las dependencias de los parámetros, es decir, ajustando primero P_0 , r_s y β con respecto a la masa de los halos anfitriones y luego estableciendo su evolución temporal.

Para resolver este problema fue necesario tomar un enfoque distinto al del ajuste originalmente propuesto. Las limitaciones encontradas fuerzan a que se requiera un método para estimar la P_{ram} que utilice la menor cantidad de parámetros libres posible, y que reduzca al mínimo el número de ajustes realizados. Para ello, nos centramos en buscar una forma funcional que, mediante la variación de sólo uno de sus parámetros, permitiera reproducir tanto el comportamiento de los halos más masivos con tendencia creciente hacia el centro, como el de los menos masivos donde la presión de barrido tiende a ser más homogénea. De todas las opciones probadas, la que arrojó mejores resultados correspondió a una ley de potencias simple definida como

$$P_{\text{ram}} = P_0(z) \left[\frac{1}{\xi(z)} \left(\frac{r}{R_{200}} \right) \right]^{-\frac{3}{2}\alpha(M,z)}, \quad \text{donde} \quad (3.11)$$

$$\alpha(M, z) = \alpha_M(z) \log \left(\frac{M_{200}}{M_\odot} \right) + \alpha_N, \quad \text{y} \quad (3.12)$$

$$\alpha_N = -5.5. \quad (3.13)$$

Al igual que el modelo de la Ecuación (3.8), este perfil tiene en principio tres parámetros libres: $P_0(z)$ en unidades de $10^{-12} h^2 \text{ dyn cm}^{-2}$, una constante adimensional $\xi(z)$ y el nuevo exponente $\alpha(M, z)$. En reemplazo de r_s incluimos la constante adimensional ξ para definir la distancia radial, en términos del radio virial, a la cual el perfil tiende a aplanarse para un corrimiento al rojo particular. Por otro lado, en esta definición hemos encapsulado la dependencia con la masa de los halos anfitriones exclusivamente en el exponente $\alpha(M, z)$, el cual se expresa mediante un único parámetro libre dependiente del tiempo, $\alpha_M(z)$ y es linealmente dependiente con el logaritmo de la masa. Esta definición nos entrega un perfil de presión de barrido que, utilizando sólo tres parámetros libres, permite recuperar el comportamiento de todas las mediciones realizadas para diferentes masas de halos en una misma instantánea, como veremos enseguida.

En un principio, el parámetro α_N fue considerado como un cuarto parámetro libre, pero al ajustar detectamos una degeneración importante entre este y los demás. Esta degeneración provoca que existan múltiples combinaciones entre los valores de los parámetros que permiten reproducir el comportamiento de las mediciones, lo cual nos impide establecer luego una correcta dependencia temporal de los mismos en base a un segundo ajuste. Debido a esto, decidimos fijar α_N en un valor característico encontrado, con el fin de establecer mínimos absolutos de los demás parámetros en el proceso de minimización.

Con esta forma funcional propuesta procedimos a realizar un ajuste de los tres parámetros para cada una de las instantáneas mediante minimizaciones de χ^2 , considerando en simultáneo todas las mediciones realizadas en cada una de ellas. Estos ajustes abarcaron todo el rango de masa de halos disponible, y se consideraron sólo las posiciones de los halos satélites (galaxias tipo 1) para definir cuáles mediciones de presión de barrido de las resimulaciones utilizar. Al contrario de [Tecce et ál. \(2011\)](#), abarcamos todo el rango de corrimiento al rojo disponible en las resimulaciones para expandir la validez de esta estimación.

La baja degeneración entre estos tres parámetros se puede apreciar en la Figura 3.5, donde se muestra con puntos azules el resultado obtenido del proceso de ajuste de ellos en

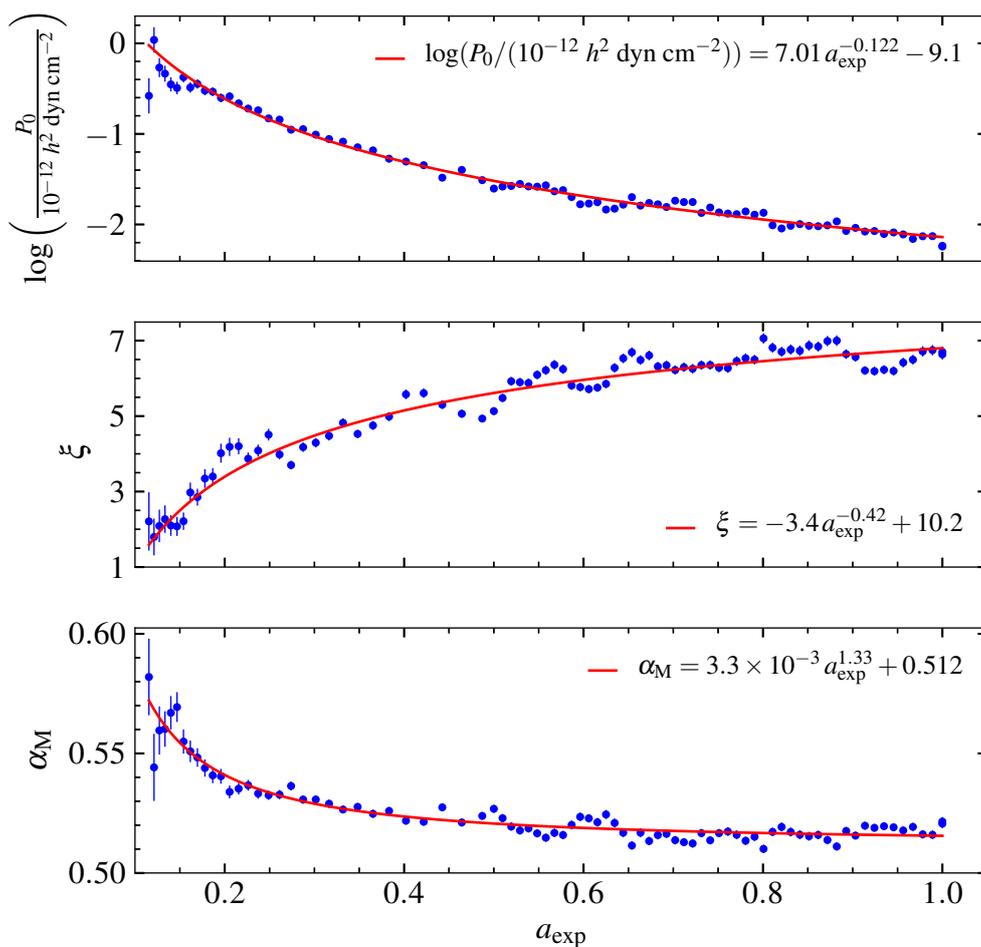


Figura 3.5: Ajuste de la evolución temporal de los parámetros P_0 (*panel superior*), ξ (*panel central*) y α_M (*panel inferior*) presentes en la nueva aproximación propuesta para la P_{ram} . Mediante puntos azules se muestra el valor obtenido para estos parámetros en cada instantánea analizada, y mediante la línea roja el ajuste realizado a cada uno.

Param.	λ_1	$\delta\lambda_1$	λ_2	$\delta\lambda_2$	λ_3	$\delta\lambda_3$
$\log(P_0)$	7.01	3.0	-0.122	0.047	-9.1	3.02
r_s	-3.4	1.4	-0.42	0.12	10.2	1.4
α_M	3.3×10^{-3}	1.1×10^{-3}	1.33	0.18	0.512	1.6×10^{-3}

Tabla 3.1: Valores de los coeficientes λ que describen la evolución temporal de los parámetros P_0 , r_s y α_M que intervienen en la aproximación para P_{ram} . También se incluyen los errores extraídos de la matrices de covarianza.

cada una de las instantáneas de la simulación. Cada conjunto de valores es mostrado en el corrimiento al rojo correspondiente a cada instantánea, pero en términos del factor de expansión del universo a_{exp} . El error reportado en cada punto fue extraído directamente de los elementos de la diagonal de las matrices de covarianzas obtenidas. Para todos los casos, se encontró una evolución muy suave respecto al parámetro de expansión del universo, justamente el tipo de comportamiento que permite recuperar la predictibilidad de la aproximación que buscamos para P_{ram} .

Finalmente, ajustamos estas evoluciones temporales con leyes de potencia simples con respecto al parámetro de expansión, y aplicando un desfase en el eje de las ordenadas, de la forma

$$\mathcal{P} = \lambda_1 a_{\text{exp}}^{\lambda_2} + \lambda_3,$$

siendo \mathcal{P} cualquiera de los parámetros P_0 , ξ , α_M . El resultado de estos tres ajustes es incluido también en la Figura 3.5 mediante las líneas rojas, cuyas tendencias siguen de forma bastante satisfactoria la evolución de los parámetros en el tiempo. El valor final obtenido para cada uno de los coeficientes λ es detallado en la Tabla 3.1, junto a los respectivos errores asociados a cada uno.

En base al orden de magnitud de los errores y del resto de los coeficientes obtenidos en las matrices de covarianza que no se incluyen en la tabla, pudimos determinar que suele existir cierta degeneración entre los coeficientes λ_1 y λ_3 para los tres parámetros ajustados. No obstante, dado que estos valores los usamos directamente sin volverlos a ajustar en otra relación, las degeneraciones en este nivel no alteran los resultados de P_{ram} . A modo de ejemplo del buen comportamiento obtenido con este nuevo ajuste, mostramos en la Figura 3.6 la comparación entre las estimaciones de la presión de barrido extraídas de las resimulaciones y las obtenidas con este nuevo ajuste, para diferentes rangos masa de halos anfitriones en la instantánea a $z = 0$. Con los puntos grises se muestran todas las mediciones obtenidas de la instantánea, mientras que los azules marcan el conjunto de ellas correspondiente al rango de masa indicado en cada panel. La predicción obtenida por el ajuste propuesto para el conjunto de posiciones asociadas a las mediciones de cada rango de masa son mostradas mediante los puntos rojos en cada caso. Es importante notar que la dispersión de los puntos rojos exhibida en cada rango de masa se produce debido a que la muestra está compuesta por diferentes halos principales en cada caso, cuyas masas viriales alteran levemente el exponente α del modelo definido por la Ecuación 3.11, cambiando así la pendiente de los perfiles de la presión de barrido.

El comportamiento del ajuste sigue exactamente la tendencia que fue planificada, en donde una única forma funcional es capaz de entregar una buena aproximación de la presión de barrido para una instantánea en particular considerando todo el rango de masas de halos anfitriones. A partir de la versión del modelo semianalítico SAG-7.96, esta implementación quedó habilitada por defecto para la estimación del impacto de la

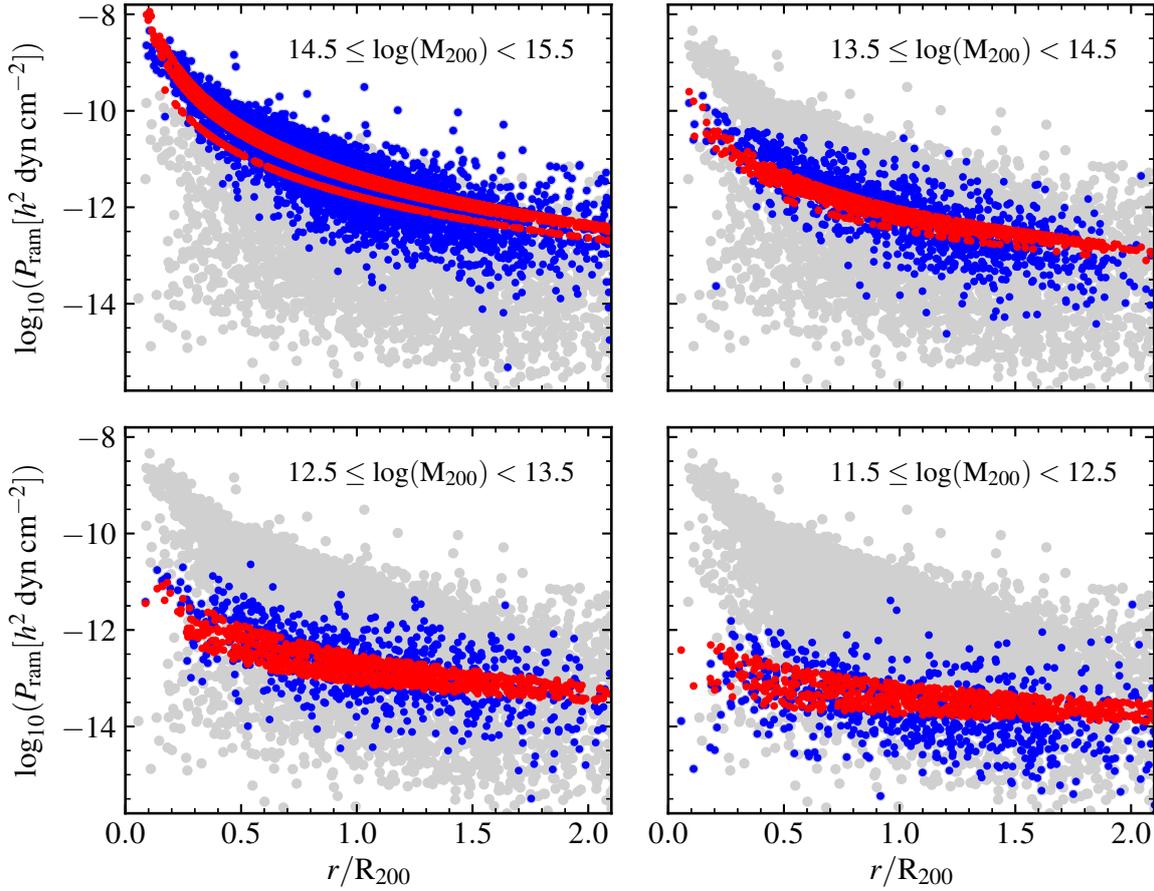


Figura 3.6: Comparación de las mediciones de la presión de barrido obtenidas a $z = 0$ con el nuevo ajuste implementado para diferentes rangos de masa. En puntos grises se muestran todas las mediciones de la instantánea, en azules aquellos que pertenecen al rango de masa específico de cada gráfico, y en rojo la estimación obtenida con el ajuste para ese conjunto de posiciones.

presión de barrido cada vez que el modelo utilice los tratamientos extendidos en el cálculo de la evolución de galaxias en base a catálogos de halos y árboles de fusiones extraídos de simulaciones de N -cuerpos únicamente.

Capítulo 4

Creación del catálogo MultiDark-SAG

Durante las últimas décadas, los relevamientos astronómicos han adquirido una vasta cantidad de datos mediante observaciones fotométricas y espectroscópicas de galaxias, lo que ha permitido incrementar nuestro entendimiento de los procesos de formación y evolución de galaxias. La tasa de adquisición de datos durante los años ha ido en aumento, y se estima que crecerá exponencialmente en los años venideros debido a los proyectos que se mantienen operativos hoy en día, así como aquellos que están siendo planeados tales como el LSST, tal como fue descrito en la Sección 1.4.

Para escalas espaciales grandes, el modelo estándar de la cosmología basado en el paradigma Λ CDM nos brinda un buen entendimiento de las observaciones. En este modelo, las estructuras son gobernadas por la materia oscura y se forman a través de la agregación jerárquica durante la evolución cósmica. No obstante, a escalas espaciales menores las interacciones bariónicas de los procesos astrofísicos dominan la evolución de galaxias de una forma más compleja que la evolución de la materia oscura. En este sentido, las simulaciones numéricas cumplen un rol fundamental en el modelado de las interacciones que gobiernan la evolución de las galaxias como se vio en detalle en la Sección 1.1. Este enfoque numérico para construir teorías debe ser capaz de producir datos o catálogos que se asemejen a las observaciones y que sean lo suficientemente grandes como para ser comparados con la cantidad de información que es recolectada por los relevamientos o será recolectada en el futuro. De este modo es posible poner a prueba nuestro entendimiento de la cosmología y de la evolución de galaxias, así como también poder probar la precisión de los relevamientos. Asimismo, se debe tener la capacidad de realizar predicciones teóricas para las grandes colecciones de datos futuras.

Es en este contexto que presentamos la creación del primer catálogo de galaxias simuladas para propósitos múltiples (de acceso público) realizado con nuestro código semianalítico SAG, utilizando como base las simulaciones de materia oscura del proyecto español MultiDark¹ disponibles de forma pública. Este trabajo está embebido en un proyecto de colaboración directa y de largo plazo entre el equipo local de investigadores a cargo del modelo semianalítico, y el grupo español que genera las simulaciones respectivas ofreciéndolas de forma abierta a través de una base de datos de acceso libre mediante los protocolos estándar de Observatorios Virtuales (VO, por sus siglas en inglés)².

En el desarrollo de este capítulo se presentará el procedimiento de creación de este catálogo, las muestras obtenidas y las estadísticas generales relacionadas a las restricciones físicas y predicciones de cada una de las poblaciones de galaxias simuladas.

¹<http://projects.ift.uam-csic.es/multidark/>

²Más detalles sobre los protocolos VO se pueden encontrar en la página oficial www.ivoa.net.

4.1. Las simulaciones MultiDark

Con objetivo de poder estudiar la distribución de materia del Universo a grandes escalas, las simulaciones de materia oscura solamente requieren la aplicación de modelos que permitan trazar la conexión entre las propiedades observadas y la distribución de materia oscura. Por ello, un producto importante de la colaboración consiste en poblar las simulaciones Multidark con galaxias mediante la utilización de diversos modelos semianalíticos, entre los que se encuentra el código SAG. Como ya se ha detallado en los capítulos anteriores, este tipo de códigos se aplica sobre catálogos de halos de materia oscura a diferentes corrimientos al rojo junto a sus correspondientes historias de fusión, información que es extraída directamente de los resultados de la simulación de materia oscura respectiva. En este caso particular, los catálogos de halos a utilizar provienen de las simulaciones del proyecto Multidark que fueron presentadas por [Klypin et ál. \(2016\)](#). En este trabajo se presentó un conjunto de simulaciones cosmológicas de materia oscura solamente, la mayoría con 3840^3 partículas y cuyos tamaños van desde los $250 h^{-1}\text{Mpc}$ de lado hasta los $2500 h^{-1}\text{Mpc}$, considerando resoluciones de masa desde $\sim 10^8$ hasta $\sim 10^{10} h^{-1}M_{\odot}$. De estas simulaciones, incluso aquellas con la resolución más baja, correspondientes a las simulaciones con $2500 h^{-1}\text{Mpc}$ de lado, permiten resolver un gran número de halos que potencialmente son capaces de hospedar galaxias similares a la Vía Láctea con $\gtrsim 250$ partículas. Sólo dos simulaciones de este conjunto fueron elegidas para iniciar nuestra colaboración: una de resolución de masa intermedia llamada MDPL2 (por *MultiDark PPlanck*, versión 2), y la de mayor resolución de masa llamada SMDPL (por *Small MultiDark PPlanck*). No obstante, la primera no corresponde exactamente a una simulación presentada por [Klypin et ál. \(2016\)](#), sino que es una nueva versión análoga a la MDPL original descrita en aquel trabajo, pero con una condición inicial diferente, preservando todas sus otras propiedades. En adelante, nos referiremos a estas dos simulaciones como “simulaciones MultiDark”.

Estas dos simulaciones fueron integradas utilizando el código GADGET-2 ([Springel, 2005](#)) siguiendo un tratamiento cosmológico de N -cuerpos comenzando a alto corrimiento al rojo ($z = 120$). Ambas simulaciones son caracterizadas por una cosmología definida por los parámetros obtenidos por el satélite *Planck* ([Planck Collaboration: Ade et ál., 2016](#)), entre los cuales la constante de Hubble vale $H_0 = 100 h \text{ km s}^{-1}\text{Mpc}^{-1}$ con $h = 0.6777$, y las densidades de energía oscura, materia y bariones valen $\Omega_{\Lambda} = 0.692885$, $\Omega_{\text{m}} = 0.307115$ y $\Omega_{\text{b}} = 0.048206$, respectivamente. Además, el espectro de potencias utilizado para generar las condiciones iniciales está descrito por un valor de $\sigma_8 = 0.829$ y con un índice espectral $n_s = 0.96$. Otros detalles específicos se listan en la Tabla 4.1. Asimismo, se puede encontrar información adicional sobre el proyecto Multidark en la página web oficial³, en particular sobre estas simulaciones, las presentadas por [Klypin et ál. \(2016\)](#) y otras más disponibles de forma pública.

Una vez completada una simulación cosmológica de materia oscura, el producto final obtenido consiste en una colección de instantáneas de la distribución de partículas y sus correspondientes velocidades a diferentes épocas. No obstante, el modelado de la evolución de galaxias requiere específicamente la información de los halos de materia oscura que se constituyen en cada una de estas, así como sus correspondientes árboles de fusiones.

En este caso particular, las simulaciones MultiDark estaban procesadas con el código de identificación de halos ROCKSTAR ([Behroozi et ál., 2013b](#)), y los árboles de fusiones fueron construidos con CONSISTENT TREES ([Behroozi et ál., 2013](#)). Los catálogos resul-

³<https://www.cosmosim.org>

Nombre	N_{part}	L_{box}	m_{p}	ϵ	N_{snap}	$z_{\text{máx}}$
MDPL2	3840^3	$1.0 h^{-1}\text{Gpc}$	$1.51 \times 10^9 h^{-1}M_{\odot}$	$5.0 h^{-1}\text{kpc}$	126	14.95
SMDPL	3840^3	$0.4 h^{-1}\text{Gpc}$	$9.63 \times 10^7 h^{-1}M_{\odot}$	$1.5 h^{-1}\text{kpc}$	115	18.56

Tabla 4.1: Propiedades principales de las simulaciones de materia oscura MultiDark: el número de partículas de materia oscura utilizadas N_{part} , el tamaño de lado del volumen comóvil L_{box} , la masa de cada partícula m_{p} , la resolución de fuerzas expresada en términos de la distancia de amortiguamiento gravitacional ϵ , el número de instantáneas almacenadas de la simulación N_{snap} , y el corrimiento al rojo de la instantánea más antigua disponible $z_{\text{máx}}$.

tantes de este conjunto cumplen los estándares mínimos requeridos para ser utilizados de forma consistente por los modelos semianalíticos. De este modo, ambos catálogos de halos y de árboles de fusiones constituyen la información proporcionada por el grupo MultiDark para este proyecto.

Todo el conjunto de catálogos fue provisto en el formato SUSSING, detallado en la Sección 2.3.4, considerando que las masas de los halos determinadas por ROCKSTAR en este formato se definen en función de un contraste de densidad constante $\Delta_{200} = 200$ relativo a la densidad crítica del universo ρ_c . Según lo anterior, utilizando la Ecuación (1.25), la definición de masa para estos halos es

$$M_{200} = \Delta_{200}\rho_c \frac{4\pi}{3} R_{200}^3, \quad (4.1)$$

donde R_{200} corresponde al radio desde el centro del halo detectado hasta donde se alcanza una densidad media $\bar{\rho}_{\text{h}} = \Delta_{200}\rho_c$. Además, debemos recordar que estas masas tienen una definición inclusiva, es decir, las masas de todos los halos incluyen a su vez la contribución de todas las masas de los subhalos detectados dentro de ellos, sin hacer distinción alguna entre halos y subhalos para definir las propiedades de cada estructura encontrada.

En ambas simulaciones, se utilizó como límite un mínimo de 20 partículas de materia oscura para la detección de los halos, lo cual define un límite de confiabilidad y completitud en masa en cada simulación que depende de la masa de las partículas de materia oscura utilizadas. A pesar de la gran diferencia en tamaño y resolución de masa, las densidades numéricas de halos detectados por intervalo de masa de las simulaciones MDPL2 y SMDPL que son parte de este proyecto, convergen de forma suave en un amplio rango de masas. Específicamente, coinciden en el rango de masas de halos $11 \lesssim \log_{10}(M/M_{\odot} h^{-1}) \lesssim 15$, lo cual se puede apreciar en la Figura 4.1 en donde se muestran de forma superpuesta las funciones de masas de halos para ambas simulaciones a $z = 0$. En la figura, la línea continua de color naranja corresponde a la función de masa de la simulación MDPL2, en cambio la línea a trazos de color púrpura corresponde a la función de masa de la simulación SMDPL que tiene mejor resolución de masa que la anterior. Las líneas verticales punteadas marcan la masa mínima de detección para cada caso, correspondiente al criterio de un halo compuesto por 20 partículas de materia oscura.

Es importante destacar que por debajo de la cota de resolución, en ambas simulaciones se reporta una población persistente de halos con masas de hasta un orden de magnitud menor al límite de detección establecido, pero estos tienen una densidad numérica que decrece y se aleja rápidamente de la tendencia esperada. Estos halos son incluidos en los

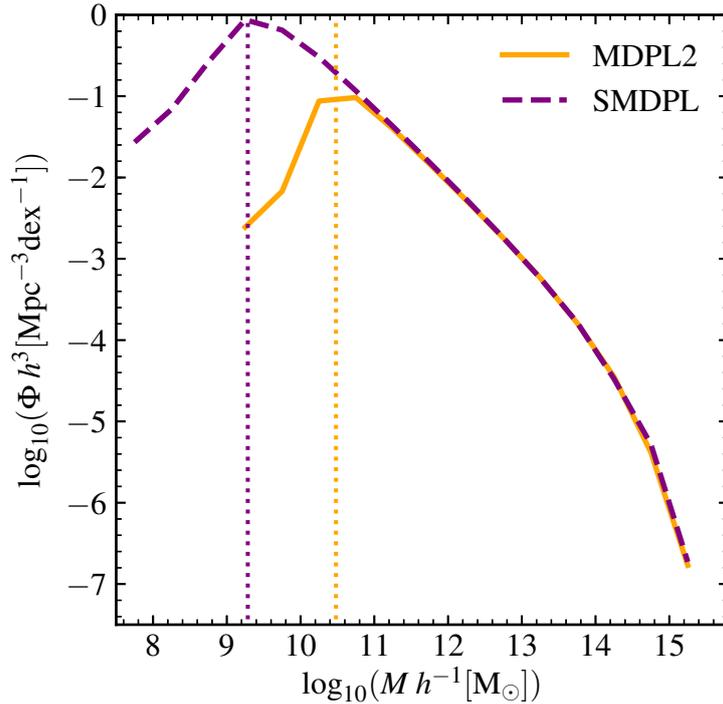


Figura 4.1: Funciones de masa de halos Φ para las simulaciones MDPL2 (línea naranja) y SMDPL (línea púrpura a trazos) a $z = 0$. Las líneas verticales marcan el límite de completitud dado por el criterio de masa mínima de 20 partículas de materia oscura por halo.

procesos de depuración de los árboles de fusiones realizados por CONSISTENTTREES posteriores a la detección. Esto se realiza para que las historias de los halos sean consistentes en el tiempo y carezcan de vacíos históricos, algo común en los halos muy cercanos al límite de resolución. Estos halos con masa menor al límite de detección de cada simulación deben ser tratados con sumo cuidado, pues debido a su consistencia histórica han de ser considerados como reales. No obstante, al modelar luego la población de galaxias que los habiten (en especial si son centrales), estas no constituirían en ningún caso una muestra con completitud volumétrica en esos rangos de masa.

A pesar de que en el proyecto de colaboración se propuso la utilización de ambas simulaciones, MDPL2 y SMDPL, los resultados incluidos en esta Tesis sólo comprenden la primera etapa de este proyecto que sólo contempla el análisis de la primera simulación mencionada. La SMDPL se programó para una segunda etapa dentro de los próximos años, una vez que los primeros catálogos con la MDPL2 estén disponibles y en uso. De este modo, en lo que resta de este capítulo nos centraremos en detallar todo el proceso de análisis, procesado y cálculo realizado sobre la simulación MDPL2, finalizando con la descripción de los catálogos creados y su inyección en la base de datos pública.

4.2. Recursos computacionales

Tanto almacenar como procesar estas simulaciones es una tarea que demanda recursos computacionales grandes. Los datos usados como entrada y las salidas resultantes

abarcan del orden de Terabytes de información. El cálculo ha de realizarse usando múltiples procesos ($\gtrsim 100$) para que sea completado en tiempos razonables, y usualmente el almacenamiento temporal de variables para cada galaxia durante el proceso de cálculo requiere una gran cantidad de memoria disponible, valor que puede ir desde las decenas de Gigabytes hasta los Terabytes, dependiendo de la cantidad de propiedades consideradas y el tipo de modelo utilizado.

Debido a lo anterior, para la concreción del proyecto comprendido en este capítulo hemos utilizado los recursos computacionales disponibles en el Centro de Astro-Ingeniería de la Pontificia Universidad Católica de Chile (AIUC), lugar con una vasta experiencia en las áreas de Astroinformática, específicamente en la computación de alto rendimiento. Las dos supercomputadoras disponibles, dedicadas principalmente a cálculos astronómicos, se detallan a continuación.

- GERYON. Consistente en un arreglo de 64 nodos con un total de 128 procesadores de tipo Intel Xeon Quad-Core (512 núcleos) de la familia E5405 de 2.00 GHz, alcanzando 1.0 Terabyte de memoria RAM en total. Los nodos usan sistemas Linux con arquitectura de 64-bit y se comunican entre sí mediante una red tipo Ethernet con velocidad de hasta 10 Gbit/s.
- GERYON2. Consistente en un arreglo de 15 nodos con un total de 584 procesadores de tipo Intel Xeon de la familia E7, modelos 4820 o 4850 dependiendo del nodo, alcanzando 5.47 Terabytes de memoria RAM en total. Al igual que el anterior, los nodos usan sistemas Linux con arquitectura de 64-bit, pero están conectados mediante una red Infiniband con una tasa de transferencia de datos de hasta 50 Gbit/s. Esta supercomputadora cuenta además con un arreglo de discos de estado sólido SSD de alta velocidad de 58 Terabytes, los cuales están destinados específicamente a minimizar las latencias producidas por los cálculos que tienen una alta demanda de lectura y/o escritura de datos.

Adicionalmente, ambos recursos comparten una capacidad de almacenamiento en discos duros de hasta 250 Terabytes, ideal para la amplia envergadura de los datos manipulados en este proyecto.

Las dos supercomputadoras descritas fueron utilizadas para el almacenamiento, análisis y procesamiento de las dos simulaciones MultiDark referidas. Los detalles de utilización de cada uno de estos recursos serán especificados en cada sección respectiva.

4.3. Preprocesado de la simulación MDPL2

Los catálogos de halos y árboles de fusiones de la simulación MDPL2 comprenden una colección de 128 archivos de texto ASCII en formato SUSSING, de los cuales 126 son catálogos de halos detectados en cada una de las instantáneas de la simulación, un archivo contiene la información de los árboles de fusiones, es decir, la lista de progenitores para cada halo, y el último archivo contiene la lista de bosques respectiva. El conjunto en su totalidad demanda una gran capacidad de almacenamiento de disco duro, dado que en total suman alrededor de 2.6 Terabytes de datos a utilizarse como entrada en el cálculo.

Las listas de halos de la colección suelen identificarse con números enteros consecutivos. En este caso, la más antigua en tiempo cósmico, marcada como la instantánea número 000, está asociada al corrimiento al rojo $z = 14.949$. Por su parte la más reciente, marcada como la instantánea número 125, está asociada al universo local con $z = 0.0$.

La lista de bosques determina los conjuntos de halos que están conectados ya sea espacial o históricamente, de modo que cada una de estas listas de árboles de fusiones constituye un grupo autocontenido de la simulación. En total, la simulación MDPL2 está compuesta por aproximadamente 103 millones de bosques, los cuales permiten que el cálculo de la población de galaxias se pueda distribuir en distintos hilos de cálculo independientes entre sí. No obstante, dado que cada uno de estos grupos de halos debe ser procesado de forma secuencial, el número de halos que contiene cada bosque fija la carga de trabajo que alcanzará cada uno de estos hilos de cálculo, y el que contenga más halos fijará a su vez la carga de cálculo máxima.

La distribución del trabajo en los diversos hilos de cálculo depende de la cantidad de bosques que se componen, a su vez, de diferentes números de halos. Esto último se puede estudiar mejor calculando \bar{n}_{hpi} , el número medio de halos por instantánea que tiene cada bosque, y obteniendo la distribución numérica de estos para diferentes rangos de \bar{n}_{hpi} . Esta estimación se muestra en la Figura 4.2, en donde se puede apreciar que un porcentaje importante de los bosques de la simulación se caracterizan por tener un valor relativamente bajo de halos por instantánea, con $\bar{n}_{\text{hpi}} \lesssim 10$. Incluso algunos de estos están compuestos por un único halo que no está presente en todas las instantáneas de la simulación ($\bar{n}_{\text{hpi}} < 0$). El número de bosques decae de forma constante en escala logarítmica hacia mayores densidades numéricas de halos, lo cual es consistente con el comportamiento observado de la función de masa de halos. Esto se debe a que asociamos los halos principales de mayor masa con sistemas que tienen mayor número de satélites e interactúan más con sus vecinos, de modo que definen sistemas de muchos halos ligados que deben ser agrupados de forma conjunta en bosques densos. No obstante, a diferencia de la función de masa que tiene un límite máximo relativamente poco sensible a la resolución de la simulación, la distribución de los bosques según el número de halos por instantánea depende fuertemente de la masa de la partícula de materia oscura utilizada. Esta dependencia se debe a que la cantidad de satélites poco masivos crece hacia masas de halos principales mayores, los que suelen estar asociados a sistemas más numerosos. En el caso de la MDPL2, su baja resolución de masa produce una cantidad razonable de bosques con $\log_{10}(\bar{n}_{\text{hpi}}) \gtrsim 3.5$, con $\max(\bar{n}_{\text{hpi}}) \simeq 6 \times 10^3$. Estos valores nos otorgan cierta libertad para separar y reagrupar bloques de la simulación sin preocuparnos demasiado en generar un cuello de botella en el cálculo según los recursos computacionales disponibles.

Si bien hemos implementado en el código semianalítico SAG una estrategia para poder procesar diferentes grupos de bosques con un conjunto de hilos independientes de cálculo, el tamaño de estos catálogos particulares hace que la demanda de acceso a disco sea demasiado alta y genere una reducción considerable en el rendimiento cuando todo el cálculo se realiza de forma simultánea. Debido a esto, para procesar la simulación completa de forma satisfactoria, optamos por reescribir los catálogos de la simulación utilizando los códigos de manipulación de formato SUSSING que hemos desarrollado para este propósito⁴ y que fueron detallados en la Sección 2.5.2. Con ellos, aplicamos dos cambios que ayudan a aligerar el acceso a disco duro.

- Cambiamos el formato de almacenamiento de ASCII original a nuestra versión personalizada de SUSSING en binario (Sección 2.3.4). Esto reduce el espacio de disco utilizado en más de un 45 % quedando en 1.3 Terabytes de datos, y también elimina la sobrelectura que realizan los procesos para avanzar a puntos particulares

⁴Este conjunto de códigos se pueden descargar libremente de https://bitbucket.org/cnvega/sussing_scripts

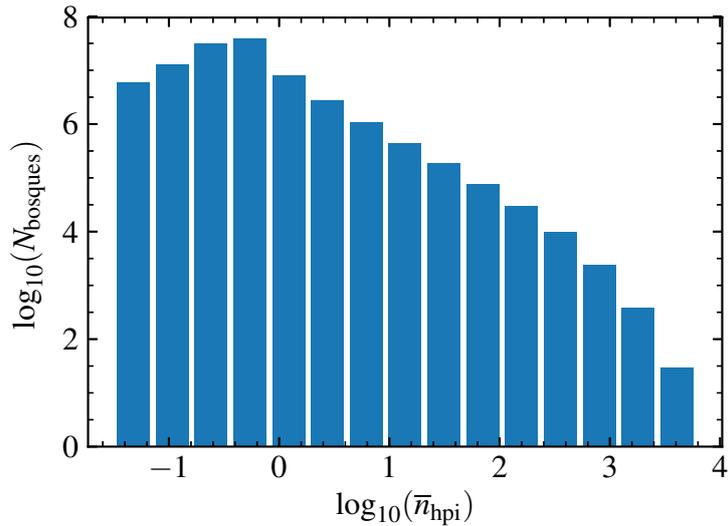


Figura 4.2: Histograma de la cantidad de bosques N_{bosques} de la simulación MDPL2 según el número medio de halos por instantánea que tiene cada uno.

de los archivos.

- Fragmentamos los catálogos originales en 128 partes distintas utilizando para esto una repartición homogénea en el número de bosques contenidos en cada una, sin tomar en consideración su distribución espacial. De este modo, cada una de estos subvolumenes puede ser analizado en sistemas computacionales independientes, agregando versatilidad en el transporte de datos o en la copia remota.

A pesar de las ventajas en rendimiento que la división en 128 subvolumenes nos brinda, este procedimiento acarrea la limitación de acrecentar la cantidad de archivos tratados, ya sea de entrada como de salida, en dos órdenes de magnitud. En función de ello, para la generación de catálogos de galaxias usando esta simulación, en especial para leer y analizar los resultados obtenidos, se deben considerar estrategias que permitan tratar archivos múltiples de forma eficiente. Trataremos este tema en detalle al final de este capítulo, una vez descritos los catálogos de galaxias generados.

4.3.1. Órbitas de galaxias satélites: MDPL2

Como fue descrito en detalle en el Capítulo 3, durante el proceso de cálculo de la evolución de las galaxias que habitan los halos de la simulación, emerge una población no despreciable de galaxias huérfanas. Estas galaxias corresponden a todas aquellas que alguna vez fueron galaxias centrales de halos de materia oscura bien definidos, pero que luego desaparecieron al fusionarse con otros o simplemente dejaron de ser detectables por el buscador debido a que su masa disminuyó hasta llegar al límite de detectabilidad (por ejemplo, en el caso de satélites que han sufrido pérdida de masa por efectos de ambiente). En ambos casos, debido a que el tiempo de fricción dinámica o fusión de las galaxias es usualmente mayor que el de los halos que las contienen, una vez que el halo desaparece la evolución de la galaxia se sigue calculando hasta que esta se fusione (o se destruya) de forma efectiva. A estas galaxias que han perdido una asociación directa con un halo de

materia oscura se les llama huérfanas, y continúan existiendo como satélites de galaxias que están asociadas a halos definidos hasta que desaparecen.

La determinación de la posición de estas galaxias, así como su tiempo de supervivencia en órbita, es un problema netamente dinámico y lo resolvemos con el código dedicado SUBORBITFINDER. Este se ejecuta como un preprocesado al modelo de evolución de galaxias, y realiza sólo la integración analítica de las galaxias huérfanas que emergen de los árboles de fusiones respectivos.

Para ejecutar este cálculo utilizamos la versión de la simulación ya dividida y en formato binario, de modo que los archivos de entrada sean los mismos que se usan en el código semianalítico. Se aprovechó la división realizada para ejecutar 128 procesos independientes en paralelo. A pesar de que el cálculo realizado por cada uno de estos procesos es menor en comparación con todas los cálculos que se realizan en el SAG, medir su rendimiento a través de la contabilización del tiempo de ejecución en el sistema es un buen punto de referencia para todo el subsecuente procesamiento.

Para este primer cálculo utilizamos la supercomputadora GERYON2, detallada en la Sección 4.2. Si bien el recurso computacional tiene la capacidad de ejecutar los 128 procesos en forma simultánea, y tiene discos sólidos que no generan latencia en las operaciones de escritura y lectura, estos se comunican con el sistema mediante una red que restringe el número de transferencias en simultáneo, sin bloqueos, solamente a 32. Esto es suficiente para procesos usuales que requieren mucho cálculo para múltiples usuarios inclusive, pero en nuestro caso particular estamos lidiando con procesamiento de datos caracterizados por una muy alta demanda de lectura y escritura de datos. El límite impuesto por la red genera una restricción importante en la cantidad de procesos que se pueden ejecutar en simultáneo antes que la eficiencia de procesamiento, medida como la razón entre el tiempo de procesador utilizado y el tiempo efectivo total, comience a disminuir de forma drástica y a su vez comience a entorpecer el correcto uso de la supercomputadora para los demás usuarios.

Considerando los puntos anteriores, procedimos a ejecutar el código SUBORBITFINDER para realizar la integración analítica de las órbitas de las galaxias satélites huérfanas en cada uno de los subvolumenes respectivos, considerando siempre mantener activos en forma simultánea entre 20 y 32 procesos, dependiendo de la actividad global de la supercomputadora en cada momento de uso. Con ese criterio, el cálculo total requirió aproximadamente 4272.5 horas de procesador para completarse, consumadas en aproximadamente 4436.2 horas de reloj efectivas. Esto implica que cada proceso requirió en promedio 33.4 horas de uso de procesador, con una eficiencia promedio del 96%, valor bastante aceptable y en concordancia con lo esperado según nuestras estimaciones previas de latencia por acceso a disco.

Un ejemplo particular del resultado efectivo de este procesado se puede apreciar en la Figura 4.3, donde se muestra la distribución espacial proyectada de galaxias satélites en sus respectivos halos de materia oscura para un sistema a $z = 0$. Su halo principal tiene una masa de $M = 2.69 \times 10^{13} h^{-1} M_{\odot}$, y se caracteriza por estar muy al borde de la simulación con una coordenada $x = 3.9 h^{-1} \text{ kpc}$, de modo que debido a las condiciones de borde periódicas, su estructura es compartida por ambos extremos de la coordenada respectiva. En la figura, las circunferencias negras marcan los centros de los halos de materia oscura los cuales también son asignados como posiciones de sus respectivas galaxias centrales. Cada una de estas circunferencias está rodeada por un círculo lleno de color claro cuyo radio corresponde a $r = R_{200}$ de cada halo. El círculo de mayor radio (de color amarillo claro) es el correspondiente al halo principal, y los menores (color verde claro)

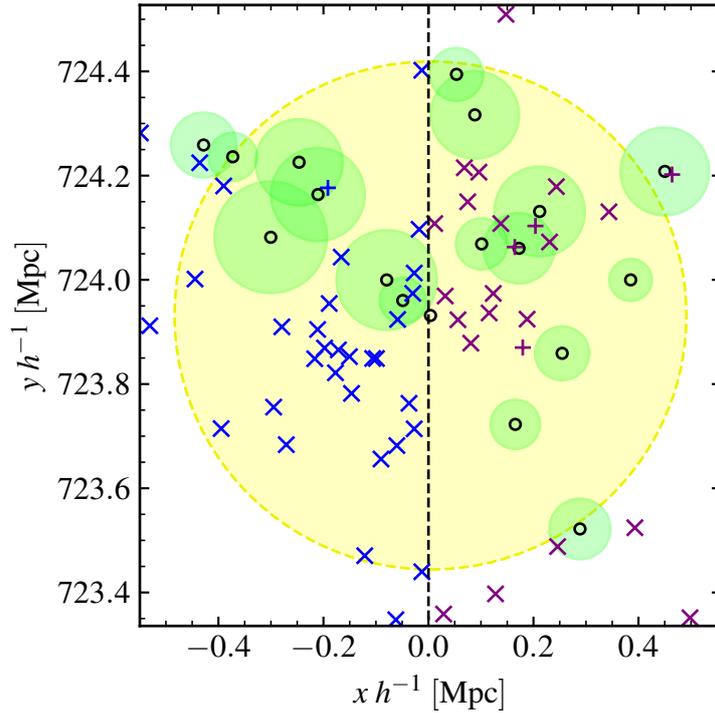


Figura 4.3: Ejemplo de halo de materia oscura poblado con galaxias a $z = 0$, correspondiente a un halo principal de $M = 2.69 \times 10^{13} h^{-1} M_{\odot}$ ubicado en el borde de la simulación para la coordenada $x \simeq 0$. Más detalles de la simbología se encuentran en el texto.

son los correspondientes a los halos satélites a este. Con cruces se muestra el resultado al término de la integración de las órbitas para galaxias satélites huérfanas: con \times se marcan las posiciones resultantes de las galaxias satélites de la galaxia central del halo principal, y con $+$ se marcan las posiciones de las galaxias satélites de las galaxias centrales de los halos satélites (es decir, las satélites de satélites). En la simulación, todas las coordenadas espaciales son positivas, por lo que la línea vertical a trazos marca el borde efectivo de la caja. Por ello, todas las galaxias pertenecientes a este halo principal, cuyas posiciones se encontraban en el otro extremo de la simulación en la coordenada x , es decir, con $x > 500 h^{-1} \text{Mpc}$, fueron corregidas mediante la substracción del largo de la caja simulada para incluirlas en esta figura. De este modo, las cruces de color azul corresponden a aquellas posiciones que fueron corregidas y las de color púrpura identifican las que no fueron modificadas. Mostrado de esta forma se puede apreciar claramente que la corrección por periodicidad introducida en el modelo de órbitas (Sección 3.1.3) es sumamente efectiva, no produce galaxias que se salgan de las coordenadas de la caja original y preserva de forma satisfactoria la estructura del sistema de satélites integrado.

Es importante mencionar que en la figura se pueden apreciar halos satélites que están hacia los bordes o incluso fuera del radio R_{200} del halo principal al cual pertenecen. Este efecto es común, y es debido a que el algoritmo que encuentra los halos e identifica las conexiones espaciales entre los mismos es distinto al que asigna las masas y radios respectivos. El primero se realiza en una etapa inicial en la cual se definen las jerarquías de pertenencia espaciales, mientras que el segundo utiliza esta información de forma posterior y ajusta perfiles de densidad sobre la distribución de partículas utilizando el

criterio de densidad de la Ecuación (4.1), el cual es una elección particular de entre las diversas definiciones de masa existentes para los halos de materia oscura. Lo mismo sucede para las galaxias satélites huérfanas cuya posición es estimada analíticamente. El hecho de que tengamos galaxias a distancias mayores a R_{200} del centro de sus halos correspondientes no implica que estas pierdan la relación de pertenencia o ligadura con las respectivas galaxias centrales. El borde efectivo de los halos es un problema que no está bien definido de forma única en la literatura, habiendo autores que incluso proponen que el límite real de los halos depende más bien de la tasa de acreción de masa, y que puede extenderse hasta incluso $\approx 2R_{200}$, considerando que los efectos de ambiente pueden actuar a distancias mucho mayores que las que resultan de las definiciones usualmente utilizadas para fijar radios viriales (por ejemplo, [More et ál., 2015](#)).

Es importante destacar que todo el proceso de cálculo de órbitas de galaxias huérfanas fue realizado dos veces en esta simulación. Esto fue debido a que en el primer procesamiento aún no se realizaban los análisis descritos en la Sección 3.1, de modo que el código para realizar este cálculo sólo incluía la adaptación a los catálogos SUSSING. Aunque no existieron diferencias significativas entre ambas ejecuciones, las estadísticas presentadas en esta sección corresponden específicamente a la última de ellas.

Con este cálculo completado, sólo resta encontrar un conjunto de parámetros del modelo que permita satisfacer las restricciones de propiedades de galaxias observadas, proceso que explicaremos en detalle en la siguiente sección.

4.4. Calibración del modelo

Como describimos en detalle en la Sección 2.1, el modelo semianalítico utiliza parámetros libres que regulan características de los procesos modelados. Entre otros, estos parámetros pueden controlar eficiencias, índices en leyes de potencias, tasas de procesos particulares o umbrales de corte en la definición de algunos de estos procesos. Cuando se imponen restricciones de propiedades de galaxias observadas que deben ser satisfechas por las modeladas, cada uno de estos parámetros adquiere valores diferentes para cada simulación con cosmología y resolución específicas. No obstante, cada uno suele preservar valores dentro de un rango numérico determinado. Encontrar el conjunto de parámetros que satisfaga el criterio impuesto en la mayoría de los casos es un procedimiento bastante complejo, que además exige como mínimo un muestreo general del comportamiento del modelo en el espacio de los parámetros que se desea ajustar.

A partir de simulaciones que tienen un número de halos mayor a 50 000 (en la instantánea a $z = 0$), ya no es conveniente utilizar la simulación en su totalidad para realizar el análisis del comportamiento del modelo en cada uno de los puntos explorados del espacio de parámetros debido al gran tiempo de cálculo que esto conlleva. Para estas situaciones es conveniente seleccionar un subconjunto de la simulación que sirva como muestra representativa de la misma. Para ello podemos calcular estadísticas globales independientes y proporcionales al volumen, tal como la función de masa de halos, es decir, la densidad numérica de halos para diferentes rangos de masa. Se procede entonces a encontrar un subvolumen de la simulación cuya función de masa de halos sea lo más similar posible a la obtenida con la simulación completa, garantizando de esta forma la representatividad.

Para el caso de la MDPL2, hemos procedido dividiendo cada lado de la simulación en 9 partes, generando así 9^3 subvolúmenes de $111.1 h^{-1}$ Mpc de lado. A pesar de que cada una de estas muestras constituye alrededor de sólo el $\sim 0.137\%$ de la simulación

completa, las mismas son lo suficientemente grandes como para ser comparables a una simulación independiente de baja resolución en masa. Por lo tanto, una vez generada la función de masa de halos para cada una de ellas, la comparamos con la función de masa obtenida con la simulación completa, y simplemente elegimos mediante inspección visual aquella que mostraba mayor similitud.

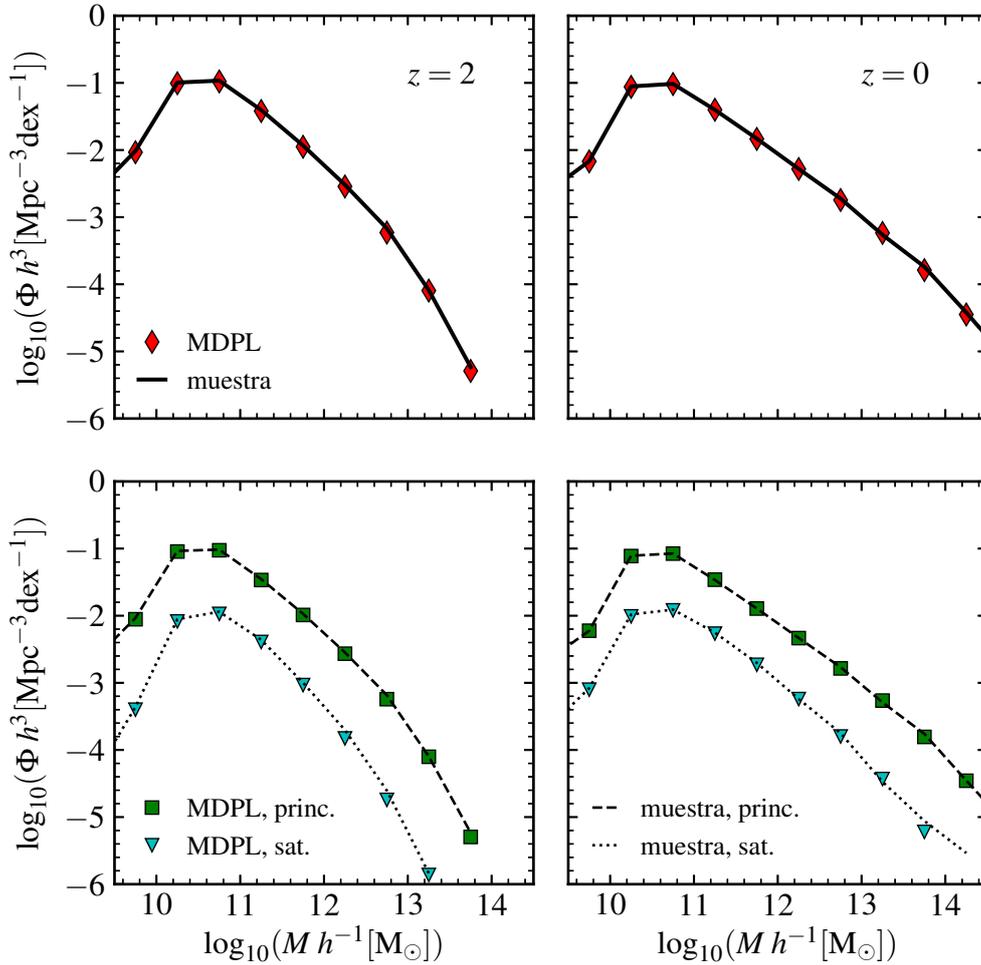


Figura 4.4: Funciones de masa Φ de halos del subvolumen de calibración elegido (líneas) comparadas con las obtenidas con la simulación completa (símbolos). Cada columna corresponde a un corrimiento al rojo distinto: $z = 2$ a la izquierda y $z = 0$ a la derecha. Los paneles superiores consideran la población completa de halos, mientras que los inferiores muestran por separado halos principales con subhalos satélites.

Las funciones de masa de la muestra elegida, en adelante “subvolumen de calibración”, son mostradas en la Figura 4.4. Cada uno de los paneles de la figura compara, para distintos casos, funciones de masa de halos de la simulación completa, representada mediante símbolos, y del subvolumen de calibración, identificada mediante líneas. En los paneles superiores se muestran las funciones de masa considerando todos los halos, y en los paneles inferiores discriminando los halos principales de los subhalos satélites. A su vez, los paneles de la izquierda corresponden a una instantánea de la simulación a alto corrimiento al rojo ($z = 2$), mientras que los paneles de la derecha comparan funciones de masa de halos detectados en la instantánea correspondiente al universo local ($z = 0$). En

todos los casos, el acuerdo es suficientemente bueno para todo rango de masa como para permitir la utilización de esta selección como muestra representativa de la simulación completa y, por consiguiente, para realizar la búsqueda de parámetros del modelo.

A pesar de que el subvolumen de calibración comprende un porcentaje muy pequeño de la simulación, esta selección es igualmente una colección bastante numerosa de halos debido a la gran dimensión de la MDPL2. En particular, la muestra elegida está constituida por 140 804 bosques, los cuales contienen 176 177 halos a $z = 0$. Este número es alrededor de ~ 3.5 veces más grande que la cota inicial propuesta para simulaciones que requerían este tratamiento de muestreo, imposibilitando entonces utilizar el subvolumen completo para ajustar los parámetros del modelo mediante corridas secuenciales del mismo.

Para evitar tener que buscar una selección más pequeña a riesgo de perder convergencia en las funciones de masas de halos, optamos por fragmentar el subvolumen de calibración en ocho partes utilizando la misma estrategia aplicada a la simulación completa. Se procedió a fragmentar repartiendo equitativamente los bosques contenidos en estos catálogos sin tomar en consideración la distribución espacial de los mismos ni el número de halos que tiene cada uno, y además convirtiendo el formato a SUSSING en binario. De este modo, todas las partes se pueden procesar en simultáneo reduciendo el tiempo requerido para conseguir cada uno de los puntos en la exploración del espacio de parámetros del modelo. Es importante destacar que este tipo de división sólo permite versatilidad en la estrategia de procesamiento de las simulaciones, pero no sirve para obtener resultados realizando análisis individuales de cada una de sus partes dado que estas constituyen muestras aleatorias y parciales de los bosques que inicialmente constituyen el conjunto de catálogos. De forma análoga al procesado que se le realizó a la simulación MDPL2 completa, tanto la extracción del subvolumen de calibración desde los catálogos originales como el fraccionado posterior se realizó mediante la utilización de los códigos de manipulación de catálogos SUSSING desarrollados.

Como el proceso de calibración debe realizarse con un modelado análogo al que se utilizará en el cálculo de la simulación MDPL2 completa, es necesario aplicarle a esta muestra todo el mismo preprocesado descrito en la sección anterior. Por ello, procedimos asimismo a calcular la integración de las órbitas de las galaxias satélites huérfanas esperadas en el modelo en esta muestra mediante la ejecución del código SUBORBITFINDER en cada una de sus partes. A diferencia de la simulación completa, este cálculo no es tan demandante y puede realizarse en cualquier computadora personal con un par de horas de cálculo dedicado, por lo que no realizamos estadística alguna de rendimiento en esta etapa.

Finalmente, utilizando el conjunto seleccionado de catálogos de halos separados y sus correspondientes órbitas de galaxias satélites integradas es posible ajustar los parámetros del código semianalítico SAG por cada tipo de modelo deseado. Para ello utilizamos el método PSO descrito en detalle en la Sección 2.1.4, para lo cual usamos la supercomputadora GERYON debido al menor requerimiento de recursos de este proceso en comparación con el cálculo de la simulación completa, y la mayor disponibilidad de núcleos a utilizar.

Con toda esta preparación realizamos en total tres calibraciones de modelos distintos durante el transcurso de este proyecto con sus correspondientes generaciones de catálogos incluyendo diferentes propiedades:

- Una prueba inicial utilizando la versión más simplificada del modelo sin efectos ambientales ni integración de órbitas de galaxias satélites huérfanas, y usando las recetas clásicas de realimentación.

- Un segundo procesamiento utilizando todo el cálculo de órbitas, efectos de ambiente, tratamiento de evolución de discos por momento angular y estimación de magnitudes utilizando los SEDs. No obstante, acá utilizamos una versión del modelo previa a los cambios descritos en el Capítulo 3 y antes del desarrollo del nuevo modelo de realimentación de SNe presentado en la Sección 2.1.2, lo que junto a incompatibilidades entre los tratamientos físicos utilizados produjo resultados insatisfactorios.
- Un procesamiento final con el modelo que incluye casi todos los cambios descritos en los capítulos anteriores, que nos permitió reproducir un buen conjunto de restricciones observacionales, y cuyos resultados se harán públicos durante la primera mitad de 2017.

Los procedimientos y resultados de cada uno de estos cálculos, incluyendo los correspondientes a los procesos de calibración, al modelo utilizado en cada caso y a los parámetros ajustados, son detallados en las siguientes tres secciones de este capítulo.

4.5. Simulación con modelo simplificado

A modo de prueba de procedimiento y rendimiento, un primer cálculo que involucrase tanto realizar una calibración como generar el catálogo se realizó luego que el subvolumen de calibración fue seleccionado y preparado. Para este cálculo se utilizó el código SAG en su versión más simple disponible.

Los procesos físicos incluidos en este cálculo corresponden a los descritos en la Sección 2.1.1. En términos generales, se consideró: el tratamiento estándar de realimentación y enriquecimiento químico por supernovas, realimentación debido a núcleos galácticos activos, inestabilidades de disco y el tratamiento de brotes extendidos de formación estelar presentado en [Gargiulo et ál. \(2015\)](#). El modelo además sigue una estrategia de estrangulación instantánea de gas caliente de las galaxias satélites, de modo que estas son despojadas de la totalidad de su gas caliente una vez que se convierten en satélites de un sistema. También se sigue una estrategia de retención, en donde el gas frío recalentado producto de la realimentación de las supernovas va directo a la fase caliente de las galaxias centrales de los halos principales de cada sistema, sin considerar que parte de este gas pueda ser expulsado de los halos para ser reincorporado en el futuro.

Por otro lado, en esta versión tampoco se utilizan las órbitas analíticas de las galaxias satélites huérfanas, de modo que las posiciones de estas galaxias son asignadas de forma arbitraria. Para ello, se emula una órbita ficticia que sigue un movimiento espiral hacia el centro del halo partiendo del radio virial R_{200} respecto del centro, y cuya duración antes de la fusión es estimada analíticamente por el tiempo de fricción dinámica obtenido de la aproximación de Chandrasekhar ([Binney y Tremaine, 1987](#)). Esta simplificación también impide que se estimen adecuadamente efectos de ambiente tales como arrancamiento de material por presión de barrido o fuerzas de mareas, de modo que estos efectos tampoco son incluidos en el modelado.

Es una consecuencia conocida en la literatura que al utilizar estrategias de estrangulación instantánea, que representa un modelado extremo de los efectos de ambiente, y de retención, dan lugar a un exceso de galaxias rojas a bajas masas estelares con respecto a las observaciones (ver, por ejemplo [Font et ál., 2008](#)). No obstante, el modelo es igualmente capaz de cumplir restricciones básicas para la calibración tales como la función de masa estelar de galaxias a $z = 0$. Por lo tanto, la aplicación de esta versión del código

Parám.	Valor	Descripción
α	0.0785	Eficiencia en la formación estelar, Ec. (2.5).
ϵ	0.3117	Eficiencia de la realimentación de SNe, Ec. (2.8).
ϵ_b	0.8577	Eficiencia de la realimentación de SNe en bulbo, Ec. (2.8).
κ	1.57×10^{-4}	Eficiencia de la realimentación por AGN, Ec. (2.10).
f_{pert}	79.2	Distancia relativa para perturbación de discos, Ec. (2.7).
f_{bin}	0.0294	Fracción de binarias, Sec. 2.1.1, p. 26.
f_{BH}	0.0547	Eficiencia de acreción de masa en BHs en brotes, Ec. (2.9).

Tabla 4.2: Parámetros ajustados de la versión básica del modelo SAG-7.58 sobre la simulación MDPL2.

SAG nos fue de utilidad para confirmar que el procedimiento utilizado pudiese entregar resultados satisfactorios, así como para generar las primeras mediciones de rendimiento del código.

La versión del modelo utilizada es la asociada a la revisión SAG-7.58 del código fuente (registrada con fecha 18 de mayo de 2015) según el repositorio interno del grupo. Para calibrar la misma, se dejaron libres siete parámetros para ajustar y se impusieron dos restricciones observacionales a las propiedades resultantes: la función de masa estelar a $z = 0$ correspondiente a los datos compilados por [Henriques et ál. \(2013\)](#), y la relación entre las masas de los bulbos de las galaxias y sus agujeros negros supermasivos centrales según los datos presentados por [Kormendy y Ho \(2013\)](#) y [McConnell y Ma \(2013\)](#). Ambos conjuntos de observaciones constituyen relaciones fundamentales que permiten caracterizar las propiedades de las galaxias. La función de masa estelar mide la cantidad de galaxias que se observan en un volumen cosmológico determinado por cada intervalo de masa. Permite estimar la eficiencia de formación de estrellas en halos de materia oscura con diferentes masas al establecer una relación unívoca entre estos últimos y las galaxias. Por otro lado, el logaritmo de la masa de los agujeros negros que se detectan en los centros de las galaxias muestra una estrecha relación con el logaritmo de la masa de los bulbos de las galaxias en las que estos habitan. Esta relación permite conectar los procesos de formación de bulbos con el crecimiento de los agujeros negros centrales, los cuales son los principales reguladores del mecanismo de realimentación por AGN.

En el proceso de ajuste de parámetros libres del modelo se empleó la técnica PSO utilizando 50 agentes exploradores del espacio de parámetros en simultáneo. Se ejecutó en la supercomputadora GERYON en donde una buena convergencia de la minimización de los residuos de la comparación entre las propiedades de la población de galaxias obtenida y las observadas se logra en un rango temporal entre 2 y 6 días, dependiendo de la carga de trabajo global del recurso computacional. Los parámetros ajustados en esta calibración son mostrados y descritos en la Tabla 4.2, junto a los valores numéricos resultantes del proceso.

Una vez obtenidos los parámetros procedimos a procesar la simulación completa con el código SAG. Por ser la primera prueba de rendimiento, optamos por escribir solamente una instantánea de salida, la correspondiente al universo local ($z = 0$), para la cual también incluimos el cálculo del conjunto de magnitudes absolutas definidas por el modelo estándar de cálculo de luminosidades (que incluye los filtros U, B, V, R, I, K del sistema Johnson y los filtros u, g, r, i, z del SDSS). La cantidad de propiedades de las galaxias escritas en los archivos de salida así como los nombres con que estas fueron almacenadas, siguieron el

formato estándar de archivos del semianalítico SAG, en donde cada archivo con formato HDF5 está compuesto por 125 conjuntos de datos con dimensiones $(N_{\text{gal}}, n_{\text{dim}})$, donde N_{gal} corresponde al número de galaxias del archivo y n_{dim} al número de dimensiones del arreglo, el cual puede valer 1, 3, $n_{\text{elem}} = 10$ o $n_{\text{inst}} = 126$ dependiendo del tipo de propiedad almacenada.

Para este procesamiento utilizamos un núcleo por cada uno de los 128 subvolumenes de la simulación y, a diferencia del proceso de calibración, el cálculo completo fue ejecutado en la supercomputadora GERYON2. Cada uno de estos procesos requirió alrededor de 16.5 horas de procesador para terminar, completando un total de 2200 horas de procesador para realizar los cálculos sobre la simulación completa. No obstante, fue precisamente durante esta ejecución cuando advertimos la limitación de comunicación de red existente en el uso en simultáneo de los discos. Muchos de estos procesos tuvieron una importante reducción en su rendimiento llegando incluso al 50% de eficiencia en algunos casos, dependiendo de cuántos se estuviesen ejecutando en simultáneo en cada instante. La elección del número de procesos ejecutados en simultáneo en este primer ejercicio fue medianamente arbitraria pero no excesiva en cantidad dado que *a priori* preveíamos que iba a constituir una limitación en el cálculo debido a la inmensa cantidad de datos trabajados. En efecto, cada archivo HDF5 resultante ocupa alrededor de 14 Gigabytes de disco, de modo que la colección completa, compuesta de 128 archivos, requiere alrededor de 1.7 Terabytes para ser almacenada. Considerando que esto corresponde sólo a una de las 126 instantáneas que necesitamos generar para el catálogo definitivo, la creación y utilización del formato alternativo (reducido) de archivos de salida del código semianalítico descrito en la Sección 2.3.5 es crucial para completar este proyecto.

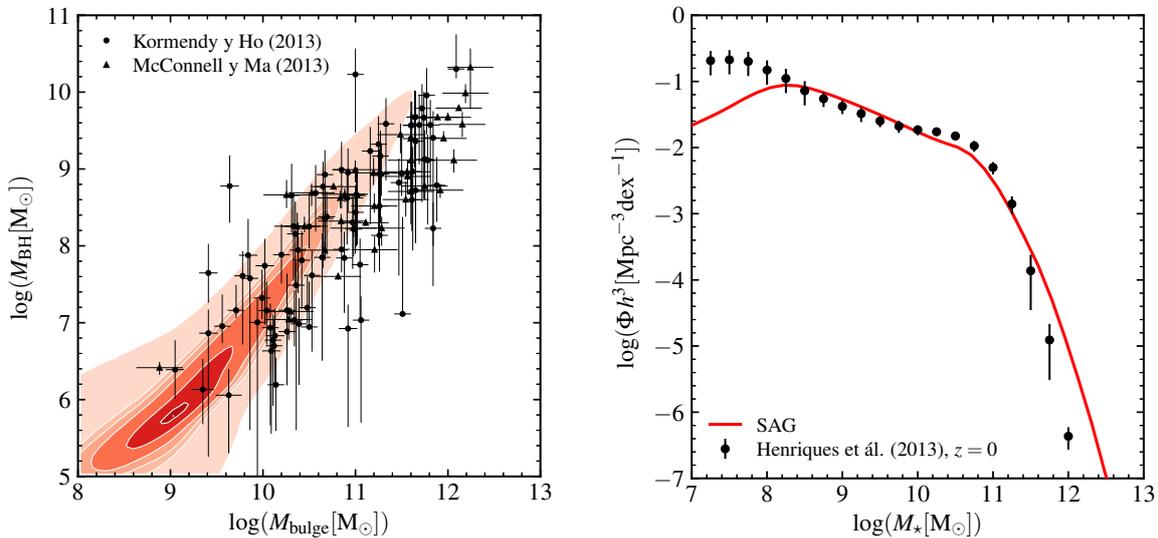


Figura 4.5: Resultados de las restricciones observacionales aplicadas para calibrar la versión básica del modelo usando el código SAG-7.58. *Panel izquierdo*: Relación entre las masas de agujeros negros centrales M_{BH} versus masas de bulbos M_{bulge} de las galaxias a $z = 0$, donde los contornos corresponden a los niveles: 0.01, 0.19, 0.26, 0.38, 0.68, 0.95 y 0.997 con respecto al máximo. *Panel derecho*: función de masa estelar, también a $z = 0$.

La población de galaxias resultante de este modelo quedó constituida por alrededor de 178 millones de galaxias, de las cuales el $\sim 28.76\%$ corresponden a galaxias satélites

huérfanas y el resto a galaxias centrales de halos ya sea principales o satélites. Las propiedades de la población de galaxias correspondientes a las restricciones utilizadas en el proceso de calibración son mostradas en la Figura 4.5. En el panel izquierdo se muestra la relación de la masa de los agujeros negros supermasivos de las galaxias M_{BH} versus la masa de sus bulbos respectivos M_{bulge} . El mapa de contornos representa la densidad del número de galaxias obtenidas con el modelo en el plano del gráfico. A partir del más externo, los contornos corresponden a los niveles: 0.01, 0.19, 0.26, 0.38, 0.68, 0.95 y 0.997 con respecto a la máxima densidad numérica de galaxias obtenida. Se puede apreciar que la tendencia se comporta de forma consistente a lo esperado, es decir, en buen acuerdo con los datos utilizados en la calibración (McConnell y Ma, 2013; Kormendy y Ho, 2013), representados con distintos símbolos de color negro. En el panel derecho se presenta la función de masa estelar obtenida con el modelo Φ , la cual es mostrada con línea continua, y es comparada con los datos observacionales (Henriques et ál., 2013), identificados mediante puntos. En este último gráfico podemos ver que, a partir de $\log_{10}(M_{\star}[M_{\odot}]) \sim 8.5$, la función de masa estelar adquiere un comportamiento consistente con lo esperado, valor que está acorde además al límite de completitud de masa de la simulación. Por otro lado, el extremo masivo muestra una sutil pero persistente sobreestimación del número de galaxias de mayor masa, las cuales constituyen un porcentaje muy pequeño del total de galaxias. Este exceso es un reflejo de las limitaciones de la física modelada para reproducir de forma precisa lo observado a grandes escalas espaciales. Este comportamiento es persistente, pero con menor amplitud, cuando se trabaja con simulaciones de tamaños menores ($\lesssim 250 h^{-1}$ Mpc), como venía sucediendo con el modelo semianalítico SAG en los trabajos previos. Las simulaciones de menor tamaño espacial estadísticamente presentan un menor número de halos principales de alta masa, de modo que las galaxias modeladas a partir de ellos no son suficientes para poblar aquella zona de la función de masa de estelar y generar un comportamiento suave de la función en el extremo respectivo. Sin embargo, la simulación MDPL2 no presenta este problema dado su gran tamaño espacial de $1 h^{-1}$ Gpc de lado, de modo que las diferencias entre la función de masa estelar obtenida y la observada en el límite de altas masas emergen de forma más notoria al utilizar este tipo de simulaciones de mayor resolución. Es importante destacar que, si bien para la calibración hemos utilizado un conjunto de datos particular para describir la función de masa estelar observada, en la literatura existen trabajos que discrepan con estos valores mostrando un extremo masivo de la función estelar más alto (ver, por ejemplo Bernardi et ál., 2017). Debido a que la estimación de las masas de las galaxias observadas tiene una alta sensibilidad a los modelos de síntesis de poblaciones estelares utilizados en los análisis, este factor en combinación con la baja estadística de estas galaxias producen un mayor desacuerdo en el comportamiento observado del extremo masivo de la función de masa estelar.

Más allá de estas diferencias, la consistencia del ajuste es lo suficientemente buena como para permitirnos validar la metodología utilizada para conseguir el mejor conjunto de parámetros en el modelado. En particular, esta elección básica de procesos incluidos en el modelo permiten, en efecto, alcanzar un grado de convergencia bastante razonable de las restricciones impuestas. No obstante, entre las predicciones obtenidas con el modelo hay algunas que muestran discrepancias importantes; justamente ciertas propiedades fundamentales que son requeridas para generar el catálogo definitivo que formará parte de la colaboración, y que en este caso no coinciden con lo observado debido a falencias de la física incluida en el modelado. A modo ilustrativo, en el panel izquierdo de la Figura 4.6 se muestra el diagrama color $u - r$ versus magnitud absoluta M_r resultante, el cual nos

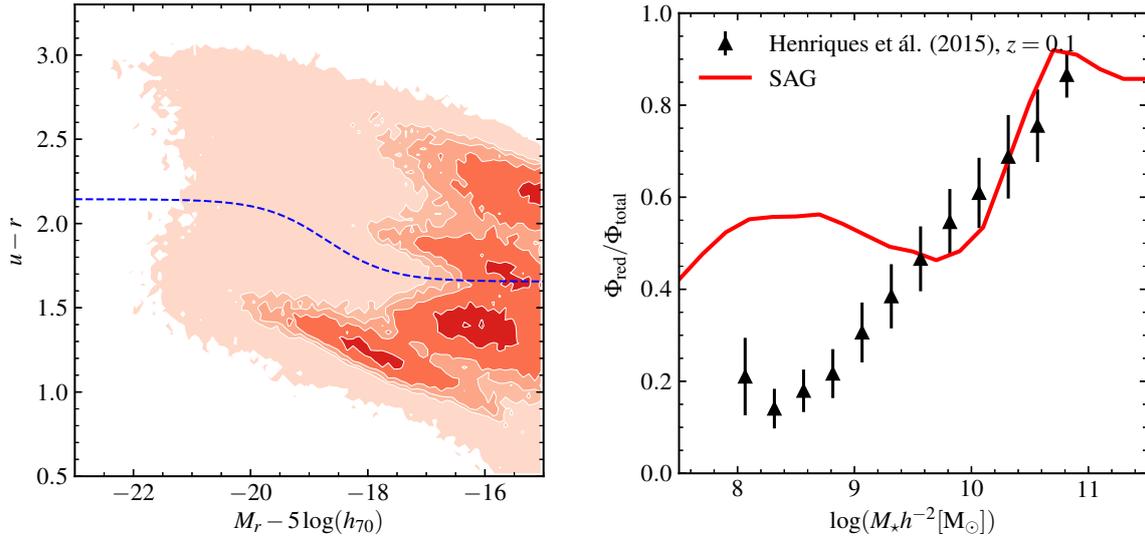


Figura 4.6: Predicciones obtenidas con el código SAG-7.58 utilizando el conjunto de tratamientos básicos. *Panel izquierdo*: diagrama color $u - r$ versus magnitud M_r de las galaxias obtenidas, representadas con el mapa de contornos. *Panel derecho*: Fracción de galaxias rojas obtenida, mostrada con la línea continua roja, comparada con los datos de Henriques et ál. (2015) con triángulos.

permite ver la distribución de colores de las galaxias para diferentes magnitudes. Para esta figura consideramos sólo la muestra obtenida con el subvolumen para calibración⁵. Con fines comparativos con otros trabajos de la literatura, en el eje de las abscisas la magnitud M_r incluye la dependencia explícita con la cosmología en términos del factor adimensional h_{70} , el cual está definido por $h_{70} = H_0/(70 \text{ km s}^{-1} \text{ Mpc}^{-1})$. El mapa de contornos muestra la densidad numérica de galaxias obtenida con el modelo, utilizando para ello los mismos niveles relativos al máximo que la Figura 4.5: 0.01, 0.19, 0.26, 0.38, 0.68, 0.95 y 0.997. La distribución de galaxias esperada debería tener un comportamiento bimodal, con una clara secuencia roja creciente hacia las galaxias más brillantes acompañada con una nube más difusa de galaxias azules (ver, por ejemplo Baldry et ál., 2004). No obstante, en nuestro gráfico podemos apreciar que la distribución de las galaxias obtenida no sigue un comportamiento bimodal definido como se espera, en especial para las galaxias rojas que no presentan una secuencia definida y parecen mezclarse de forma homogénea con la nube azul. A su vez, la zona de galaxias menos brillantes parece estar constituida por varias poblaciones intermedias, con exceso de galaxias rojas. A pesar de lo anterior, podemos aplicar igualmente un criterio de separación de poblaciones tal como la parametrización encontrada por Baldry et ál. (2004) a partir de datos del SDSS. En ella, el color de separación entre las poblaciones C_{ur} depende de M_r y está definido por

$$C_{ur}(M_r) = p_1 - 0.244 \tanh \left[\frac{M_r + p_2}{1.09} \right], \quad (4.2)$$

⁵Utilizamos sólo el subvolumen de calibración para producir estos gráficos debido a que las magnitudes de las galaxias de la muestra completa original tenían un error sistemático que causaba inconsistencias en su cálculo, y lo detectamos luego de completar el proceso. No procedimos a recalcular las magnitudes debido a que en este catálogo de prueba nos enfocamos principalmente en el comportamiento de las masas de las componentes de las galaxias.

en donde p_1 y p_2 son parámetros libres. Hemos corregido levemente estos parámetros respecto a los valores encontrados por [Baldry et ál. \(2004\)](#), ($p_1 = 2.06$ y $p_2 = 20.07$), con el fin de ajustar la curva de separación a la distribución encontrada a partir de los resultados del modelo utilizando la inspección visual como criterio. En el diagrama color versus magnitud de la población obtenido (mapa de contornos) se muestra la separación que elegimos para nuestra muestra mediante la línea a trazos, la cual quedó definida con $p_1 = 1.9$ y $p_2 = 18.7$. Utilizando esta separación se puede estimar la fracción de galaxias rojas definida como la densidad numérica de galaxias rojas Φ_{red} dividido por la densidad numérica total Φ_{total} en cada rango de masa. Este resultado se muestra en el panel derecho de la Figura 4.6 mediante la línea continua, y se compara con los datos recopilados por [Henriques et ál. \(2015\)](#) mostrados en triángulos. A pesar de que el corrimiento al rojo de los datos utilizados para comparar ($z = 0.1$) es diferente al de la muestra que hemos generado, la tendencia de la fracción de galaxias rojas es semejante durante distintas épocas, de modo que sirven como una buena referencia para este análisis. Como se puede apreciar, la muestra que hemos generado no sigue la tendencia esperada sino que muestra un marcado exceso de galaxias rojas a menores masas, comportamiento que ha sido reportado en trabajos previos de otros modelos semianalíticos que carecen de modelado acabado de los efectos de ambiente o no incluyen eyección de material de las galaxias (ver Figura 5 de [Henriques et ál., 2015](#), para una buena exposición del tema a diferentes corrimientos al rojo).

Acorde a esta prueba, es necesario calcular una nueva población de galaxias que incluya el tratamiento de las galaxias huérfanas y los efectos de ambiente de modo que las propiedades de la población de galaxias satélites obtenidas sean más consistentes con las observaciones.

4.6. Segundo catálogo: uso de SEDs

Luego de realizada la primera prueba que nos ayudó a entender los requerimientos mínimos necesarios para completar un procesado de la simulación completa con el modelo semianalítico SAG, procedimos a generar un segundo catálogo aprovechando las mejoras referentes a rendimiento, presentadas en el Capítulo 2, pero antes de que se realizaran los cambios detallados en el Capítulo 3. El modelo utilizado para este catálogo correspondió a la revisión SAG-7.86 del código fuente, registrada con fecha 10 de octubre de 2015 según el repositorio.

El modelado en esta ocasión incluyó como base los tratamientos estándar anteriores: realimentación y enriquecimiento químico por supernovas y núcleos galácticos activos, inestabilidades de disco y brotes extendidos de formación estelar. Pero adicionalmente también se permitió que las galaxias satélites preservaran su propio halo de gas caliente que se ve reducido paulatinamente por el enfriamiento radiativo, y que sufre pérdidas de material por efectos de ambiente tales como presión de barrido y fuerzas de mareas. Como ya fue mencionado, estos dos últimos procesos dependen de la posición y velocidad de las galaxias satélites, por lo que se utilizó el integrador de órbitas para galaxias huérfanas presentado en el Capítulo 3. No obstante, de los cambios presentados en ese capítulo, el modelo sólo incluyó la adaptación a los catálogos SUSSING pero no los resultados de los análisis de la tasa de fusión de galaxias debido a que ellos se realizaron con posterioridad al cálculo de este catálogo. Por lo tanto, esta versión del integrador consideró el criterio de fusión establecido únicamente por la pérdida de momento angular específico de la órbita y

un logaritmo de Coulomb variable con constante $b = 0.5$. De este modo, la tasa de fusión de galaxias en este modelo quedó subestimada y con una evolución histórica inadecuada. Cabe notar también una falencia inherente a la presión de barrido ejercida por el medio intracúmulo sobre las galaxias satélites que estuvo presente en esta versión de SAG. La dependencia de esta presión con la masa del halo en el que se encuentran las galaxias y con el corrimiento al rojo fue modelada mediante el ajuste de [Tecce et ál. \(2011\)](#) y no con el propuesto en la Sección 3.2 que fue desarrollado con posterioridad a la generación de este catálogo. El ajuste utilizado en este caso arroja un comportamiento homogéneo para todo rango de masa, sin la dependencia radial mostrada por simulaciones hidrodinámicas.

Un aspecto distintivo de la versión de SAG utilizada para generar este catálogo es la inclusión del modelado de la disminución del tamaño de los discos de las galaxias debido a los cambios del momento angular por la acreción de material, detallado en la Sección 2.1.2. También considera una condición distinta para determinar las fusiones de galaxias donde los brotes de formación estelar son gatillados predominantemente por las inestabilidades de disco ([Padilla et ál., 2014](#)). De este modo, se incluyó toda la física disponible en el código a la fecha de ejecución de este catálogo.

Con esta configuración, se procedió a calibrar el modelo utilizando la misma estrategia que en la primera prueba, es decir, mediante la técnica PSO y utilizando 50 agentes exploradores del espacio de parámetros en la supercomputadora GERYON. Asimismo, como restricciones a las propiedades de la población de galaxias resultantes se impusieron nuevamente la reproducción de la función de masa estelar a $z = 0$ recopilada por [Henriques et ál. \(2013\)](#) y la relación de masa de los agujeros negros supermasivos centrales con la masa de los bulbos de las galaxias según los datos presentados por [Kormendy y Ho \(2013\)](#) y [McConnell y Ma \(2013\)](#). Sin embargo, para esta ocasión se dejaron libres un total de nueve parámetros en el ajuste, seis que corresponden a los mismos utilizados en la calibración anterior relacionados con los procesos básicos de realimentación en el modelado, y tres que controlan características de los nuevos tratamientos activos. Dos de estos últimos corresponden a los parámetros libres introducidos en el modelo de evolución de discos según los cambios de momento angular ([Padilla et ál., 2014](#)), más uno nuevo, f_{rec} , que permite controlar la fracción de masa reciclada por las estrellas debido a la realimentación por supernovas que se queda en el gas frío con respecto a la que regresa a la fase caliente. Este grado de libertad opcional es habilitado en el código junto con el modelo extendido (Sección 2.1.2). Por otro lado, el parámetro que controla la fracción de binarias se dejó fijo en $f_{\text{bin}} = 0.04$, siguiendo resultados de pruebas realizadas por el grupo. Con esto se consiguió evitar exceder la cantidad de parámetros libres en el proceso de minimización, el cual con más de nueve parámetros ya se vuelve inestable. La descripción de cada uno de los parámetros considerados así como los valores encontrados durante este proceso se muestran en la Tabla 4.3.

Para la generación de este catálogo decidimos utilizar la estrategia de obtención de magnitudes en base al cálculo de los SEDs de las galaxias, para así obtener más filtros que los incluidos con el método estándar y calcular además el conjunto de líneas de emisión de cada galaxia según el modelo de [Orsi et ál. \(2014\)](#). Cada uno de los espectros se trazó usando 300 intervalos entre los 91 Å y los 25 000 Å. El conjunto de magnitudes consideradas estuvo compuesto por 19 bandas: los u, g, r, i, z del SDSS, el conjunto J, H, K de infrarojo cercano, los ultravioleta NUV y FUV del telescopio GALEX, el conjunto u, g, r, i, z, y de la cámara DECam (relevamiento DES, [The Dark Energy Survey Collaboration, 2005](#)) y los J, H, K_s del telescopio VISTA (relevamiento VVV, [Minniti et ál., 2010](#)). Por otro lado, el conjunto de líneas de emisión estuvo compuesto por

Parám.	Valor	Descripción
α	0.0017	Eficiencia en la formación estelar, Ec. (2.5).
ϵ	0.4225	Eficiencia de la realimentación de SNe, Ec. (2.8).
ϵ_b	0.0161	Eficiencia de la realimentación de SNe en bulbo, Ec. (2.8).
κ	4.85×10^{-5}	Eficiencia de la realimentación por AGN, Ec. (2.10).
f_{pert}	83.906	Distancia relativa para perturbación de discos, Ec. (2.7).
f_{BH}	0.01	Eficiencia de acreción de masa en BHs en brotes, Ec. (2.9).
f_{rec}	0.1049	Fración de material reciclado que queda en el gas frío, p. 97.
f_{η}	1.9334	Eficiencia de transferencia de momento angular en discos debido a acreción, Sec. 2.1.2, p. 28.
η_{τ}	3.5192	Escala temporal de evolución de los discos por acreción, en unidades de 100 Myr, Sec. 2.1.2, p. 28.

Tabla 4.3: Parámetros ajustados en la calibración del código SAG-7.86 con todos los tratamientos disponibles en esa versión, para la simulación MDPL2.

las 13 líneas incluidas por defecto: H_{α} , H_{β} , Ly- α , [N II] 205 μm , [N II] (654.8 y 658.4 nm), Ne III 387 nm, [O II] (372.7 y 372.9 nm), [O III] (495.9 y 500.7 nm), [S II] (671.6 y 673.1 nm). Todas estas cantidades son absolutas, es decir, se expresan para un observador en reposo sin aplicar ninguna corrección por corrimiento al rojo. Además, cada una incluyó versiones con y sin la correspondiente corrección por polvo incluida en el modelo.

En vista de que el proyecto de colaboración considera generar salidas para cada una de las 126 instantáneas de la simulación, esta vez procedimos a utilizar el formato reducido de salidas del código SAG basándonos en los resultados del primer catálogo generado. No obstante, al momento de comenzar la ejecución del modelo se nos presentó una limitación importante en este contexto: la cantidad de memoria requerida para el procesamiento de los espectros de las galaxias crece linealmente según el número de salidas que se desean. Esto se debe a que es necesario almacenar de forma temporal un espectro completo por cada instantánea requerida durante el cálculo, para cada una de las galaxias. Considerando el número de instantáneas requeridas, el valor de memoria alcanza niveles excesivos. Específicamente, la memoria mínima requerida por cada proceso era de alrededor de 6.5 GB de datos, valor que se incrementaba en 0.28 GB por cada instantánea que se incluyera en la salida. De este modo, para ejecutar un proceso configurado para guardar las 126 salidas, implicaba un uso de alrededor de 40 GB de memoria disponible mientras era ejecutado, demanda que excedía los recursos computacionales disponibles para completar este proyecto.

Debido a esta limitación, y recordando el comportamiento del tiempo de cálculo según el número de salidas para este tipo de modelado presentado en la Figura 2.3, optamos en este caso por seleccionar sólo un conjunto reducido de instantáneas para este cálculo. Este conjunto quedó constituido por 18 corrimientos al rojo de salida, los cuales están distribuidos de manera aproximadamente homogénea entre la instantánea 075 correspondiente a $z = 2$, hasta la 125 correspondiente al universo local ($z = 0$). De este modo, si obteníamos resultados satisfactorios con esta primera muestra, procederíamos a completar posteriormente el conjunto de instantáneas restantes con nuevas ejecuciones completas.

Con esta configuración procedimos entonces a ejecutar el modelo en los 128 fragmentos de la simulación en bloques simultáneos de entre 10 y 30 procesos, dependiendo de la demanda de la supercomputadora en cada instante. Para cada uno de estos procesos se

utilizaron cuatro hilos de cálculo (núcleos) mediante la paralelización en base a OPENMP desarrollada para el cálculo de espectros y magnitudes. Para completarse, cada fragmento de la simulación requirió en promedio 173.2 horas de reloj durante las cuales se utilizaron efectivamente un promedio de 483.8 horas de CPU repartidas en los cuatro núcleos que se distribuyen el cálculo de cada proceso, lo cual implica una eficiencia aproximada del 70 % por cada núcleo utilizado. Globalmente, para el procesamiento de la simulación completa generando sólo 18 salidas, se requirieron aproximadamente 60 000 horas de CPU, distribuidas en 512 núcleos, los cuales habrían necesitado alrededor de 1.5 TB de memoria disponible si se hubieran ejecutado todos en simultáneo.

La demanda computacional descrita anteriormente impuso una restricción importante, debido a que si proyectamos el requerimiento necesario para la generación de las 126 salidas necesarias para la publicación del catálogo en base a estos valores, obtenemos que necesitaríamos alrededor de 420 000 horas de CPU repartidas en 512 núcleos, lo cual equivale a más de un mes de uso continuo de gran parte de los recursos de la supercomputadora, o alrededor de cuatro meses si ejecutamos los procesos en grupos de 32, sin margen de error para contingencias.

Las galaxias resultantes quedaron almacenadas en un conjunto de 2268 archivos con formato HDF5, habiendo 128 por cada una de las 18 instantáneas calculadas. Cada archivo contiene 113 conjuntos de datos con iguales dimensiones ($N_{\text{gal}}, 1$), y el conjunto completo de ellos requiere alrededor de 1.5 Terabytes para almacenarse. Si bien esto implica una reducción de aproximadamente un factor 20 en el tamaño de disco utilizado, este factor podría ser mayor si no se considera la gran cantidad de magnitudes y líneas que incluimos en este cálculo particular, las cuales constituyen un 66 % del contenido de datos de cada archivo. No obstante, esta cantidad de información es abordable con un número reducido de salidas como lo hemos hecho para esta muestra.

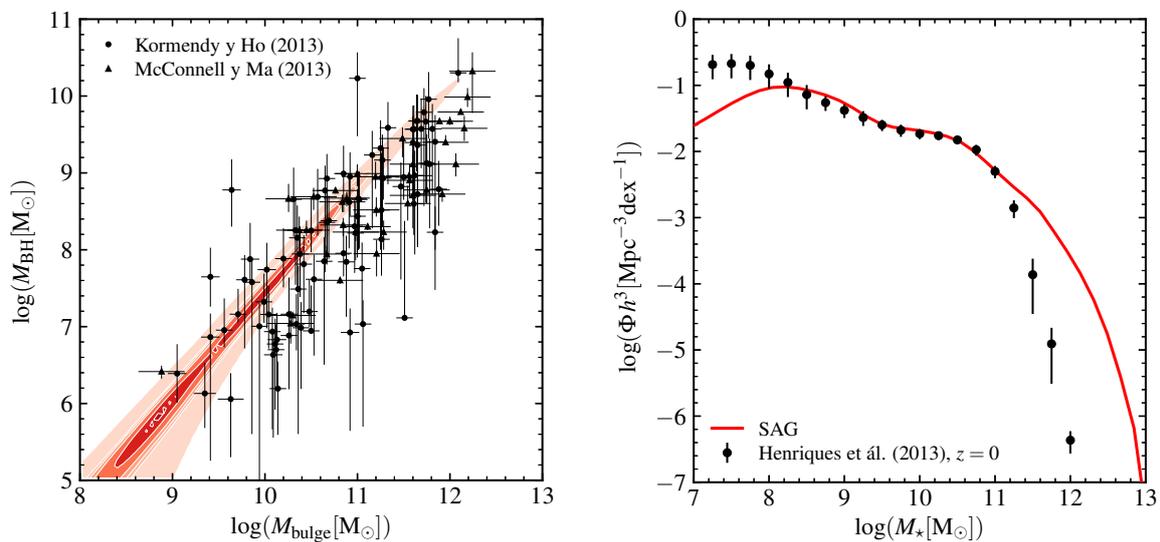


Figura 4.7: Resultados de las restricciones en la calibración aplicadas al código SAG-7.86, con toda la física disponible activada en la versión, ambas a $z = 0$. *Panel izquierdo*: Relación entre las masas de agujeros negros centrales M_{BH} versus masas de bulbos M_{bulge} resultante, utilizando los mismos contornos que en la Figura 4.5. *Panel derecho*: función de masa estelar.

La población de galaxias obtenida en esta ocasión es ligeramente más numerosa que la

resultante del modelo anterior, alcanzando esta vez alrededor de 196 millones de galaxias, de las cuales aproximadamente el 35.1% corresponde a galaxias satélites huérfanas y el resto a galaxias centrales ya sea de halos principales o de subhalos. El exceso de galaxias en comparación con el modelo anterior está compuesto exclusivamente de galaxias satélites huérfanas, de modo que el tiempo de supervivencia en órbita resultó ser mayor que lo que se puede estimar analíticamente con el tiempo de fricción dinámica según la aproximación de Chandrasekhar (Binney y Tremaine, 1987) utilizada en el modelo simplificado. Esta subestimación concuerda con el comportamiento analizado en la Sección 3.1.2 debido al uso incorrecto del criterio de fusión, el cual en este caso fue establecido según la pérdida de momento angular específico.

Contrario a la planificación original, luego de completada esta muestra no procedimos a continuar con el cálculo de las instantáneas restantes debido a dos motivos importantes: la cantidad de recursos computacionales necesarios para completarla excedía los disponibles en aquel momento, y las propiedades de las galaxias que componían la población generada no cumplieron cabalmente las restricciones impuestas, siendo los resultados insatisfactorios para la publicación de este catálogo. El comportamiento de ambas restricciones utilizadas en la calibración se muestra en la Figura 4.7, ambas correspondientes a $z = 0$. En el panel izquierdo de la figura se muestra la relación de la masa de los agujeros negros supermasivos centrales de las galaxias versus la masa de los bulbos de las mismas al igual que en el modelo anterior. Se indica mediante puntos los datos observacionales presentados por Kormendy y Ho (2013) y McConnell y Ma (2013), mientras que el mapa de contornos representa la densidad de galaxias obtenidas, utilizando los mismos niveles que en la Figura 4.5. El acuerdo de esta relación es igual de consistente que en el modelo básico presentado en la sección anterior; inclusive en esta ocasión se obtiene un comportamiento con una menor dispersión en la relación entre estas masas. Por otro lado, en el panel derecho de la Figura 4.7 se muestra la función de masa estelar comparada con los datos recopilados por Henriques et ál. (2013) y utilizados en el proceso de calibración de los parámetros libres del modelo. Esta relación tiene un comportamiento bien particular; el extremo de masas bajas e intermedias se ajusta mucho mejor a los datos mencionados, reproduciendo de forma excepcional el comportamiento de cambio de concavidad de la curva alrededor de $\log_{10}(M_{\star}[M_{\odot}]) \sim 9.5$. No obstante, el modelo produce una población excesiva en el extremo masivo generando galaxias con masas de hasta un orden de magnitud más alto que lo observado en el universo local. Este es un resultado directo del proceso de minimización de parámetros, de modo que el método PSO utilizado para calibrar es incapaz de encontrar un conjunto de ellos que permita mejorar este comportamiento a altas masas manteniendo el buen acuerdo en el resto de las propiedades consideradas en el proceso de ajuste.

El comportamiento observado de la función de masa estelar alrededor de $\log_{10}(M_{\star}[M_{\odot}]) \sim 9.5$ se debe a la utilización del modelo detallado de evolución del tamaño de los discos respecto a los cambios del momento angular provocados por la acreción de material. Este modelo establece que las inestabilidades de discos son el mecanismo principal en la aparición de brotes de formación estelar durante las fusiones de galaxias, y una de las consecuencias de este modelado es la aceleración de la formación estelar especialmente en las galaxias de baja masa a todo corrimiento al rojo. Esto en general permite que las galaxias alcancen masas estelares altas más rápidamente que en los modelos que no consideran este tipo de tratamiento (Padilla et ál., 2014). La utilización de este tratamiento en el modelado permite la aparición del pequeño cambio de concavidad observado alrededor de $\log_{10}(M_{\star}[M_{\odot}]) \sim 9.5$, a diferencia de la pendiente plana característica resultante con

el modelo básico. Por otro lado, la inclusión de los efectos de ambiente sobre las galaxias satélites tales como la remoción de material por presión de barrido o fuerzas de mareas tienden a su vez a acelerar la reducción del gas caliente de las galaxias satélites, de modo que al producirse el enfriamiento radiativo del gas hay una menor cantidad de gas frío disponible para formar estrellas en los discos. La combinación de estos dos procesos físicos produce una extinción excesiva de los discos de las galaxias favoreciendo la formación de bulbos, y provocando que las poblaciones de galaxias resultantes tiendan a estar dominadas mayoritariamente por galaxias enrojecidas y con morfologías elípticas. Como este efecto es persistente, durante la calibración de parámetros del modelo se tiende a preferir un conjunto de ellos que ajusten los datos para masas intermedias y bajas a costa de perder el ajuste en las masas mayores, dado que la física incluida no permite una convergencia correcta que reproduzca la función de masas estelar en ese rango de masa.

Cabe destacar que este comportamiento resultante es a su vez intensificado por la baja tasa de fusiones que presenta el modelo debido a la sobre-estimación de los tiempos de fusión de las galaxias satélites descrito anteriormente. Bajo este tratamiento, galaxias que deberían haberse fusionado en épocas tempranas continúan orbitando sus respectivas galaxias centrales, ganando más tiempo para destruir sus discos de gas frío mediante inestabilidades y enrojeciéndose.

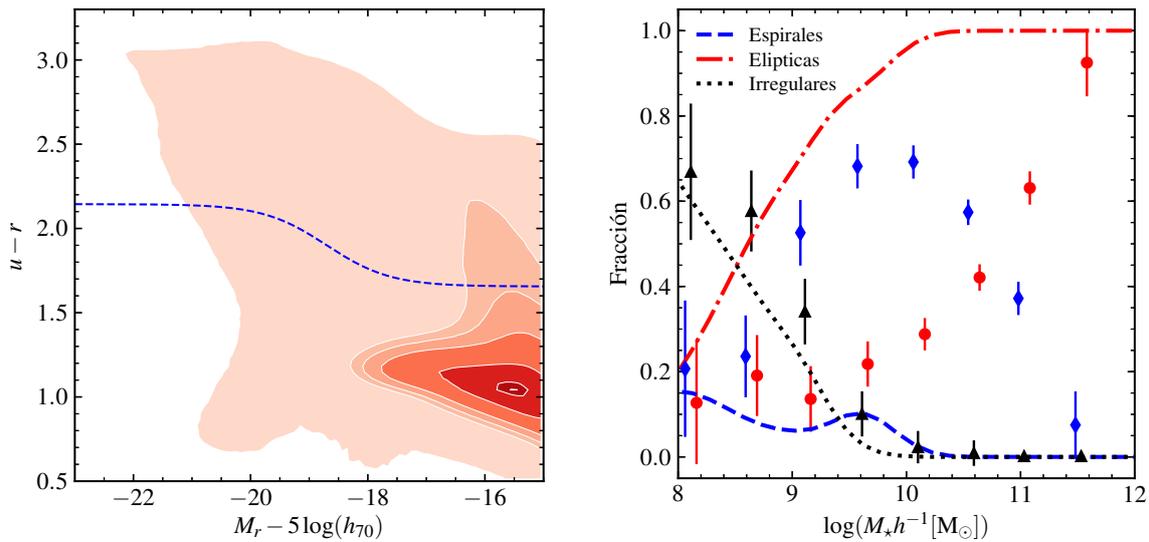


Figura 4.8: Predicciones a $z = 0$ del código SAG-7.86, activando toda la física disponible en la versión. *Panel izquierdo*: diagrama color $u - r$ versus magnitud M_r de las galaxias obtenidas, representadas en contornos. *Panel derecho*: Fracción de galaxias de distintas morfologías de la población de galaxias obtenida (en líneas) comparada con los resultados de Conselice (2006). En rojo se muestra la fracción de galaxias elípticas, en azul la de espirales y en negro la de irregulares.

El comportamiento descrito anteriormente se puede visualizar más explícitamente observando la Figura 4.8. En su panel izquierdo se muestra el diagrama color $u - r$ versus magnitud absoluta M_r de la muestra resultante completa. También se incluye como referencia la separación entre las poblaciones azul y roja definida con la Ecuación (4.2), con los mismos valores de los parámetros p_1 y p_2 definidos para el modelo anterior con fines comparativos. Cabe destacar que este modelo, a diferencia del anterior, permite

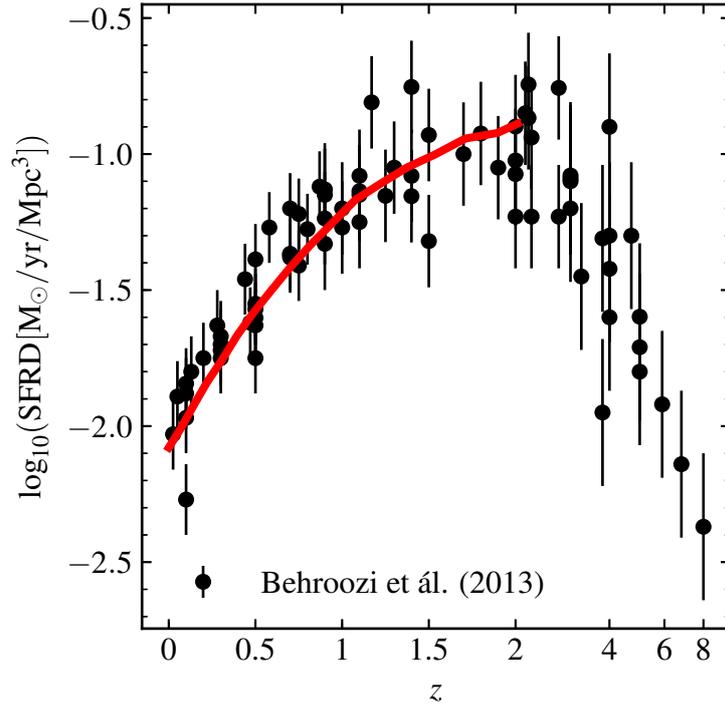


Figura 4.9: Evolución de la tasa de formación estelar cósmica del código SAG-7.86, con toda la física disponible activada en esa versión, la cual es mostrada mediante la línea continua, comparada con las observaciones presentadas por Behroozi et ál. (2013a) mostradas con círculos negros.

recuperar una muy débil secuencia roja la cual no es apreciable en la figura debido a los niveles de contorno utilizados. Las galaxias de esta secuencia muestran un color al menos 0.5 magnitudes más enrojecido que las reportadas, en especial en la zona de altas masas donde debería converger a valores cercanos a 2.5 (Baldry et ál., 2004). No obstante, esta secuencia roja no muestra una división tan definida con respecto a la nube azul, la cual en este caso alberga la mayoría de las galaxias en el rango de bajas masas donde está en mayor acuerdo con las observaciones. En el panel derecho de la figura además se muestra el comportamiento de la fracción de galaxias con distinta morfología versus la masa estelar de las galaxias obtenidas en diferentes líneas, en comparación con las observaciones presentadas por Conelice (2006) en símbolos. Para esto, utilizamos el mismo método de Bertone et ál. (2007) realizando la división entre los tres tipos morfológicos presentados mediante el cálculo del parámetro b_t . Este parámetro es definido como la fracción de masa estelar del bulbo de la galaxia respecto a la masa estelar total. Para realizar una clasificación general, elegimos un valor intermedio entre los utilizados por Bertone et ál. (2007) y Lagos et ál. (2008), de modo que las galaxias con $b_t > 0.85$ son clasificadas como galaxias elípticas, en cambio las con $0 < b_t \leq 0.85$ como galaxias espirales. Además, aquellas sin bulbo ($b_t = 0$) son clasificadas como galaxias irregulares. Como se puede apreciar en la figura, la población de galaxias tipo disco se ve muy reducida en comparación con lo esperado en el rango de masas donde debería dominar, estando la población de galaxias de este modelo principalmente dominada por galaxias de tipo temprano.

No obstante lo anterior, es importante destacar que entre las propiedades de las gala-

xias que componen esta población generada, la evolución de la tasa de formación estelar cósmica es bastante consistente con las observaciones, justamente en el rango que hemos considerado para producir las salidas de las instantáneas del catálogo, es decir desde $z = 2$ hasta $z = 0$. La tasa de formación estelar cósmica, también conocida como la densidad de formación estelar (SFRD, por sus siglas en inglés *Star Formation Rate Density*), es medida como la cantidad de estrellas formadas por unidad de tiempo y por unidad de volumen comóvil. La evolución de esta cantidad es mostrada en la Figura 4.9, donde se compara con los datos compilados por Behroozi et ál. (2013a). Es interesante el hecho de que tanto la pendiente como la normalización de este valor son consistentes con los datos, hecho que no es muy frecuente tanto en los modelos semianalíticos como en las simulaciones hidrodinámicas. Los modelos usualmente tienden a predecir pendientes más planas o subestimar el valor a $z = 0$ como se puede apreciar en la Figura 6 del trabajo de Guo et ál. (2016). En las Figuras 7 y D1 de Knebe et ál. (2015), como resultado del proyecto de comparación nIFTy, se aprecia con claridad la diversidad de tendencias que pueden resultar en este tipo de modelos semianalíticos.

Si bien no incluimos los resultados de instantáneas de más alto corrimiento al rojo al momento de ejecutar el modelo, recreando la ejecución en el subvolumen de calibración hemos visto que la tasa de formación estelar se comporta como una envolvente superior de los datos para $z > 2$. Sin alejarse de los datos observacionales, posee un máximo de formación estelar marcado hacia z cercano a 3, antes del esperado por las observaciones, comportamiento que podemos vincular con la sobregeneración de galaxias masivas que observamos a $z = 0$. Sin embargo, el buen acuerdo general de la tasa de formación estelar cósmica a su vez nos indica que este exceso de masa está asociado principalmente a las componentes esferoidales de las galaxias, preservando así cierta consistencia de la evolución de la cantidad de gas frío disponible a nivel global en el modelo, la cual permite la formación estelar en los discos de las galaxias. Gracias a esto, las propiedades de las galaxias asociadas al gas frío y la formación estelar instantánea pueden ser igualmente utilizadas para explorar relaciones útiles en la construcción y análisis de relevamientos extragalácticos. A modo de ejemplo, utilizaremos esta muestra para explorar la relación entre la tasa de formación estelar instantánea y la luminosidad asociada a la línea de emisión de [O II], y estudiar cómo esta relación evoluciona con el corrimiento al rojo (Favole et ál., en preparación).

Si bien el catálogo descrito en este capítulo no estará disponible de forma pública en la base de datos del proyecto original, podrá ser solicitado libremente de forma personalizada para ser utilizado a discreción, siempre considerando las limitaciones mencionadas durante esta sección.

4.7. Generación del catálogo público

Gracias a toda la experiencia adquirida con los modelos anteriores y las limitaciones descubiertas, se realizó un tercer cálculo sobre la simulación MDPL2 completa con el fin de convertirse en el modelo que genere el catálogo definitivo a ser inyectado en la base de datos del proyecto de colaboración.

El modelo que se utilizó en este caso correspondió a la versión SAG-7.128 del código fuente, registrada con fecha 28 de octubre de 2016 según el repositorio. Entre los tratamientos utilizados se mantuvieron las prescripciones del modelo básico relativas a la evolución química, realimentación de núcleos galácticos activos, inestabilidades de disco y

brotos extendidos. Además, se incluyeron varios tratamientos del modelo extendido descritos en las Sección 2.1.2. Entre estos, todo el modelado de los efectos de ambientes para galaxias satélites: desprendimiento de gas tanto de la fase caliente como la fría debido a la presión de barrido, y desprendimiento de material por fuerzas de mareas tanto sobre el gas como las estrellas. Para el ajuste de la fuerza de barrido sobre el gas en función de la masa del halo y del corrimiento al rojo se utilizó el descrito en la Sección 3.2. Asimismo, se incluyó la corrección por periodicidad de las posiciones y la integración de las órbitas de galaxias huérfanas con el criterio de fusión y ajuste que permite recuperar una tasa de fusiones más consistente. Para esto último, considerando la similitud tanto de la cosmología como de la resolución de masa entre las simulaciones *Euclid* y la MDPL2, se utilizó la misma definición del $\ln \Lambda$ y de la constante b encontrados en la Sección 3.1.2.

Para favorecer un mejor comportamiento de los colores de las galaxias satélites a costa de sacrificar el buen acuerdo de la función de masa estelar alrededor de $\log(M_{\star}[M_{\odot}]) \sim 9.5$, no utilizamos el tratamiento de evolución del momento angular de los discos de las galaxias según la acreción de masa, así como también simplificamos la estimación de tamaños de los mismos mediante el ajuste de Mo et ál. (1998) en vez del tratamiento más realista presentado por Tecce et ál. (2010), con el objetivo de ganar rendimiento. Finalmente, el nuevo modelo de realimentación por supernovas fue incluido en reemplazo del tratamiento original, considerando además eyección y reincorporación de material con el fin de poder conseguir un mejor acuerdo de las funciones de masa a corrimientos al rojo más altos.

Procedimos a calibrar este modelo de forma análoga a los dos casos anteriores, es decir, mediante la técnica PSO usando 50 agentes exploradores del espacio de parámetros en la supercomputadora GERYON. No obstante, como este procedimiento lo realizamos posteriormente a la segunda reunión del proyecto nIFTy en la que se definieron las calibraciones a aplicar en los modelos participantes, optamos por seguir el estándar establecido en ella que considera más restricciones al momento de ajustar. Este conjunto de restricciones incluyen cinco propiedades de la población de galaxias modelada: la función de masa estelar a $z = 0$ según los datos de Baldry et ál. (2008), Li y White (2009) y Baldry et ál. (2012); la función de masa estelar a $z = 2$ según los datos de Domínguez Sánchez et ál. (2011), Muzzin et ál. (2013), Ilbert et ál. (2013) y Tomczak et ál. (2014); la relación entre las masas de los agujeros negros supermasivos y las masas de los bulbos de las galaxias según Kormendy y Ho (2013) y McConnell y Ma (2013) a $z = 0$; la fracción entre el contenido de gas frío de las galaxias versus sus masas estelares según Boselli et ál. (2014), también a $z = 0$; y la función de la tasa de formación estelar según Gruppioni et ál. (2015) a $z = 0.15$. La tasa de formación estelar (SFR, por sus siglas en inglés *Star Formation Rate*) se define como la masa estelar formada por unidad de tiempo en cada galaxia de forma instantánea, de modo que la última restricción mencionada corresponde a la distribución del número de galaxias que presentan distintos valores de SFR.

Para el proceso de calibración se dejó un total de nueve parámetros libres, cuatro de los cuales regulan el tratamiento nuevo de realimentación, mientras que el resto de los parámetros regulan los tratamientos estándar del modelado. Cabe mencionar que, al igual que en el caso anterior, el parámetro que controla la fracción de binarias quedó fijo en $f_{\text{bin}} = 0.05$ para no aumentar en demasía el número de parámetros a ajustar dado que para más allá de nueve parámetros libres la convergencia se vuelve muy inestable. Los valores resultantes de la calibración para cada parámetro, así como su descripción breve, son mostrados en la Tabla 4.4.

Luego de realizada la calibración, procedimos a ejecutar los 128 fragmentos de la

Parám.	Valor	Descripción
α	0.0402	Eficiencia en la formación estelar, Ec. (2.5).
κ	3.02×10^{-5}	Eficiencia de la realimentación por AGN, Ec. (2.10).
f_{pert}	14.557	Distancia relativa para perturbación de discos, Ec. (2.7).
f_{BH}	0.0605	Eficiencia de acreción de masa en BHs en brotes, Ec. (2.9).
f_{rec}	0.2774	Fracción de material reciclado que queda en el gas frío, p. 97.
ϵ	0.3299	Eficiencia de la realimentación de SNe, Ec. (2.21).
ϵ_{ejec}	0.02238	Eficiencia de eyección debido a SNe, Ec. (2.23).
β_{F}	1.9973	Exponente de la dependencia temporal en la realimentación por SNe, Ecs. (2.21) y (2.23).
γ	0.0555	Eficiencia de la reincorporación de masa, Ec. (2.24).

Tabla 4.4: Parámetros ajustados del código SAG-7.128 completo, sobre la simulación MDPL2.

simulación en bloques de aproximadamente 25 procesos en simultáneo. A su vez, cada proceso utilizó cuatro núcleos en la paralelización con OPENMP encargada del cálculo de las luminosidades. Para este propósito, preferimos utilizar el método clásico que produce un conjunto fijo de filtros en vez de la estimación a través de las SEDs, debido al excesivo costo computacional que este último demanda según fue determinado durante la realización del catálogo descrito en la sección anterior. Con ello, optamos por incluir las 126 instantáneas de la simulación en las salidas, para las cuales se guardaron tanto las propiedades de las galaxias como las magnitudes obtenidas con el método utilizado. Asimismo, se utilizó el formato de salida reducida en HDF5.

Para completar la ejecución, cada proceso requirió aproximadamente 125.07 horas de reloj de las cuales se utilizaron en promedio 345.05 horas de CPU efectivas repartidas en los cuatro núcleos utilizados, lo cual equivale a una eficiencia del 68.9%. En términos globales, se requirieron 44 167.5 horas de CPU para completar el cálculo de todas las instantáneas de la simulación utilizando para ello 512 núcleos. Como cada proceso reservaba alrededor de 20 Gigabytes de memoria durante el cálculo, la ejecución de todos en simultáneo hubiera requerido alrededor de 1.3 Terabytes de memoria disponible en la supercomputadora. Estas estadísticas son equivalentes e incluso mejores que las obtenidas con el modelo anterior, a pesar de estar escribiendo 7 veces más salidas. Esto no sólo se puede atribuir al cambio del método de cálculo de luminosidades sino también a las diversas mejoras de rendimiento que se fueron introduciendo continuamente en el código. No obstante, la eficiencia alcanzada es muy similar al caso anterior, lo que nos muestra que la paralelización tiene una eficiencia análoga en ambos métodos donde fue implementada.

La población de galaxias resultante quedó almacenada en un conjunto de 16 128 archivos (128 fragmentos por 126 instantáneas) con formato HDF5 reducido. Cada uno de ellos contiene 64 conjuntos de datos con dimensiones $(N_{\text{gal}}, 1)$ y requieren en total 4.0 Terabytes de disco duro para ser almacenados. La lista de propiedades que se almacenaron en estos archivos está detallada en la Tabla A.1 incluida en el Apéndice A, en donde se detalla el nombre de cada conjunto de datos y sus respectivas descripciones.

La población de galaxias obtenida a $z = 0$ está constituida por aproximadamente 194 millones de las galaxias, de las cuales el 34% corresponde a galaxias huérfanas. Estos valores son ligeramente menores que los obtenidos en el modelo anterior que subestimaba

la tasa de fusiones, pero no constituye una diferencia tan importante. A su vez, tampoco se alcanzan valores tan reducidos como los obtenidos en el primer catálogo calculado. Sin embargo, estas estadísticas no son un método efectivo para medir las variaciones en la tasa de fusiones de diferentes modelos como se analizó en la Sección 3.1, de modo que sólo nos sirven con fines comparativos generales.

El comportamiento de las propiedades de la población de galaxias obtenida resultó ser bastante consistente con las restricciones impuestas en el proceso de calibración, las cuales se muestran la Figura 4.10. El nuevo modelo de realimentación permite efectivamente reproducir la evolución de la función de masa estelar de las galaxias mediante el ajuste de esta en dos corrimientos al rojo distintos, como se aprecia en los paneles superiores de la figura. Esto implica que la física considerada permite encontrar un conjunto de parámetros del modelo que ajusta no solo la función de masa estelar local ($z = 0$, panel superior izquierdo), sino que además ajusta la función de masa estelar a alto corrimiento al rojo ($z = 2$, panel superior derecho). En el panel central izquierdo se muestra la fracción de masa de gas frío con respecto a la masa estelar de las galaxias a $z = 0$ mediante el mapa de contornos, junto a líneas que indican la media de la muestra (línea gruesa) y una desviación estándar (líneas delgadas). Esto es comparado con los datos de [Boselli et ál. \(2014\)](#) mediante símbolos, observándose que la media de las galaxias obtenidas con el modelo se ajusta correctamente en todo el rango definido por ellos. En el panel central derecho de la figura se muestra la relación entre la masa de los agujeros negros centrales y las masas de los bulbos de las galaxias mediante el mapa de contornos. Ambos paneles centrales consideran las mismas curvas de nivel que la Figura 4.5: 0.01, 0.19, 0.26, 0.38, 0.68, 0.95 y 0.997 con respecto al máximo. En el panel inferior se muestra la función de la tasa de formación estelar a $z = 0.15$ obtenida por el modelo, la cual es representada con la línea continua. Se compara con los datos de [Grupponi et ál. \(2015\)](#) en símbolos, observándose un buen ajuste entre las galaxias modeladas y las observaciones.

Como se puede apreciar en la figura, en el modelo aún persiste una sobreestimación en el extremo masivo de la función de masa estelar obtenida a $z = 0$. No obstante, esta no es tan excesiva como la obtenida en el modelo presentado en la Sección 4.6, sino que está concentrada hacia masas más altas donde el número de galaxias es sustancialmente menor y mantiene el comportamiento de decaimiento que se espera para este observable. Este exceso ha sido persistente en todos los tratamientos utilizados, por lo tanto emerge como una limitación de los procesos físicos considerados por el modelo. El modelo puede ser ajustado para reproducir una gran cantidad de observables pero a costa de sobreestimar la masa estelar de las galaxias que habitan los halos más masivos, los cuales sólo son analizables con simulaciones de gran tamaño como la MDPL2. Dejaremos como trabajo a futuro estudiar en detalle la naturaleza física de este exceso, considerando además las incertezas en las observaciones descritas por [Bernardi et ál. \(2017\)](#) que incrementan considerablemente el número de galaxias observadas en el rango masivo en comparación con los datos compilados por [Henriques et ál. \(2013\)](#) utilizados en nuestro ajuste.

Más allá de las restricciones impuestas para calibrar el modelo, la población obtenida también se caracterizó por tener un amplio conjunto de buenas predicciones en las propiedades de las galaxias. Por ejemplo, en la Figura 4.11 se muestra la relación entre la masa estelar y la masa de los halos de materia oscura de las galaxias mediante el mapa de contornos y los símbolos, los cuales se comparan con el ajuste obtenido por [Moster et ál. \(2010\)](#) utilizando modelos de ocupación de halos (línea continua). Los niveles utilizados en el mapa de contornos corresponden a los mismos de la Figura 4.5, mientras que los símbolos representan la media en intervalos de masa de halo, con líneas horizontales

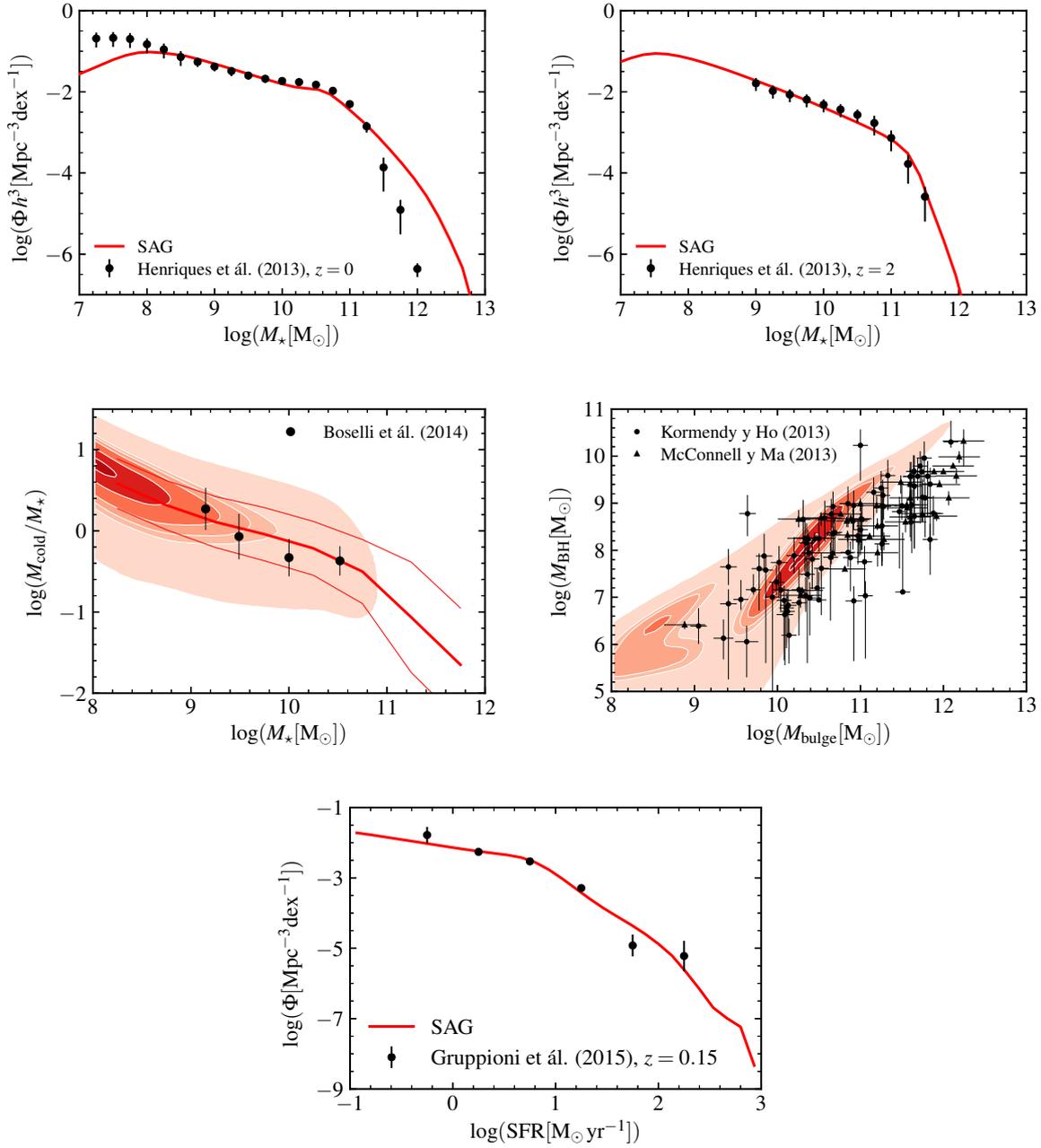


Figura 4.10: Resultados de las restricciones aplicadas al código SAG-7.128, que incluye el nuevo tratamiento de realimentación y efectos de ambiente. *Paneles superiores*: función de masa estelar a $z = 0$ (izquierda) y a $z = 2$ (derecha). *Paneles centrales*: fracción de masa de gas frío M_{cold} con respecto a la masa estelar M_* (izquierda) y relación de masa de agujeros negros centrales M_{BH} y masa de bulbos M_{bulge} (derecha) a $z = 0$. *Panel inferior*: función de la tasa de formación estelar a $z = 0.15$.

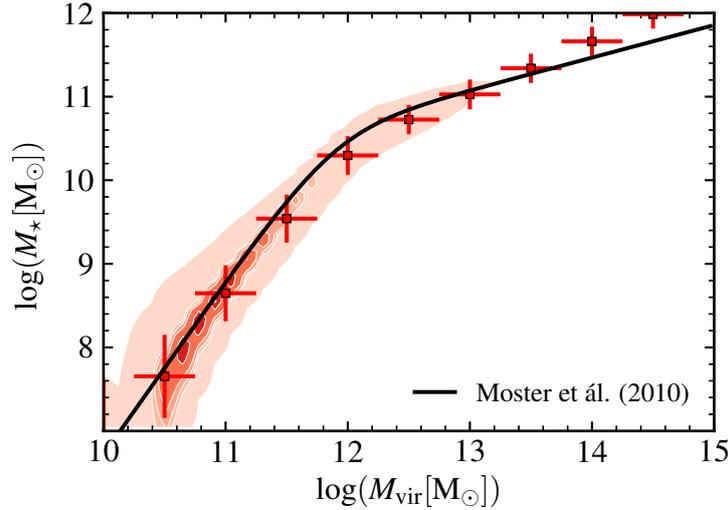


Figura 4.11: Relación entre la masa estelar y la masa del halo de cada galaxia central, obtenida con el código SAG-7.128 aplicado sobre la MDPL2. El mapa de contornos y los símbolos muestran la densidad de galaxias obtenidas, mientras que la línea continua el ajuste obtenido por [Moster et ál. \(2010\)](#).

indicando el ancho de cada rango utilizado, y líneas verticales indicando una desviación estándar en torno a la media. Para este cálculo sólo se consideran las galaxias que son centrales de halos principales, es decir, las tipo 0. Es importante destacar la correcta tendencia que muestra la población de galaxias que hemos obtenido, indicando que el poblamiento de los halos ocurre como es esperado, exceptuando el exceso de masa estelar ya mencionado para los halos más masivos. Análisis más detallados de las tendencias generales que muestra este catálogo serán presentados con prontitud (Cora, Vega-Martínez et ál. en preparación).

Gracias a que todas las restricciones impuestas son satisfechas en un grado de acuerdo bastante razonable, y las predicciones básicas del modelo tienen los comportamientos deseados, la población de galaxias obtenida con este modelo particular es apta para constituir un catálogo que sea inyectado en la base de datos pública de las simulaciones MultiDark.

4.7.1. Árboles de fusiones de las galaxias

Una característica importante que deben tener los catálogos de galaxias que se hacen públicos en una base de datos es que tengan la información necesaria para realizar solicitudes parciales de los mismos. En particular, el poder extraer conjuntos de galaxias que tienen conexión histórica desde diferentes instantáneas. Por ello, al generar este catálogo hemos incluido la información relevante para poder identificar los progenitores de todas las galaxias, en vista de que en esta ocasión hemos almacenado las propiedades obtenidas para todas las instantáneas originales de la simulación.

Para este propósito, la primera variable que preservamos de la lectura de los catálogos e incluimos en las salidas del código SAG es el identificador del bosque al cual pertenece cada galaxia. De esta forma, en caso de que se desee la información histórica de una galaxia en particular, solicitando a la base de datos todas las galaxias que pertenezcan

al mismo bosque que ella se pueden obtener todas las galaxias pertenecientes a los halos que han tenido un vínculo histórico o espacial con la misma, así como sus progenitores y descendientes para todo corrimiento al rojo. Si bien esto es suficiente para descargar listas de progenitores para galaxias que habitan bosques poco poblados donde domina una única línea evolutiva, no es suficiente para aislar historias de galaxias que habitan en bosques que tienen más de una línea histórica, tales como los que albergan los halos más masivos y tienen más de un halo principal a $z = 0$.

La segunda característica que incluimos para solucionar este problema particular es definir un nuevo identificador exclusivo de las galaxias, que sea único en toda la simulación pero estático en el tiempo. De este modo, este identificador resulta ser de naturaleza distinta a los identificadores utilizados para los halos, los cuales son asignados al momento de la localización en cada instantánea antes de realizar la conexión histórica entre los mismos. Las galaxias, por el contrario, son estructuras que naturalmente persisten en el tiempo y tienen una historia consistente. Por ello, nuestro identificador es estático para las galaxias, es decir, es asignado a cada galaxia nueva que es modelada y es preservado sin modificaciones por la misma hasta que desaparece o termina la evolución. De este modo, diferentes instancias de una misma galaxia registradas en distintas instantáneas tendrán el mismo identificador.

Esta variable, que hemos llamado *identificador estático de galaxias* S_{ID} , corresponde a un entero de 64 bits y lo definimos como

$$S_{ID} = s_{\text{form}} \times 10^{14} + p_{\text{box}} \times 10^{10} + i_{\text{gal}}, \quad (4.3)$$

donde s_{form} corresponde al número de la instantánea cuando nació la galaxia, p_{box} es el número del fragmento de la simulación que el proceso está analizando con el fin de no repetir identificadores entre diferentes fragmentos, e i_{gal} es el número de galaxias formadas en el proceso al momento de su nacimiento. Estas dos últimas variables garantizan la unicidad de este identificador en toda la simulación, y la primera se agrega simplemente para poder estimar la cantidad de instantáneas en que ha existido cada galaxia. Realizando solicitudes a la base de datos mediante este identificador es posible extraer la evolución histórica de una galaxia de forma individual.

Finalmente, para poder rastrear las galaxias satélites que se fusionaron con alguna central cuya historia se esté extrayendo, se incluye en las salidas la información referente tanto a los progenitores como a los descendientes de las galaxias, de modo que sea posible hacer búsquedas en ambas direcciones temporales. Para esto se utilizaron listas de identificadores estáticos.

Cada galaxia se caracteriza por tener un único descendiente en la instantánea siguiente, por lo tanto sólo es necesario guardar el identificador de la galaxia descendiente en cada caso. Este valor será igual al identificador actual de la galaxia cuando esta no se fusione con una central. Por otro lado, la cantidad de progenitores puede tomar cualquier valor partiendo desde cero. En este caso, este número debe ser igualmente registrado, y los progenitores deben ser almacenados mediante una lista ligada entre las galaxias de la instantánea anterior. Cabe destacar que con esta nomenclatura, el indicador del primer progenitor siempre corresponderá al mismo identificador que posee la galaxia, por lo que podría ser omitido. No obstante, para mantener consistencia con otros catálogos similares, el primer progenitor es registrado para todas las galaxias sin importar el número de progenitores que tenga cada una.

Con esta información es posible asignar nuevos tipos de identificadores de las galaxias (como los comúnmente utilizados identificadores tipo *depth-first*), tarea que quedará a

criterio futuro dependiendo de los usos mayoritarios que se le dé a la base de datos.

4.7.2. Inyección en la base de datos

Si bien hemos generado una colección de archivos que almacena un catálogo de galaxias simuladas totalmente utilizable para los propósitos planteados en esta Tesis, la base de datos pública tiene ciertos estándares que no son decididos por los modeladores. Por ello, para concretar la inyección de los datos que hemos generado, debe realizarse una serie de conversiones de unidades y nueva asignación de nombres.

En la primera etapa de publicación, sólo se agrega un conjunto menor de propiedades a la base de datos, las cuales constituyen solamente las características principales de las galaxias. La lista de estas, así como el nombre público, su unidad y la conversión que hay que realizar durante la inyección son mostrados en la Tabla 4.5, en donde se han omitido dos propiedades que no se agregarán para nuestro modelo debido a su ausencia en los archivos creados. Estas corresponden a la tasa de formación estelar del disco de las galaxias, que se puede obtener sustrayendo la del bulbo de las galaxias sobre la total, y el parámetro de espín de los halos de materia oscura que deberá ser extraído del catálogo de propiedades de los halos de ser necesario. Todas las galaxias obtenidas por el modelo serán cargadas en una única tabla, de modo que el valor del corrimiento al rojo será la única forma de discriminar entre la poblaciones de galaxias pertenecientes a diferentes instantáneas. A su vez, no quedará referencia alguna a los fragmentos de la simulación originalmente utilizados para realizar el cálculo.

Esta elección de variables, sus nombres y unidades será el estándar de la base de datos para todos los modelos semianalíticos que procesen las simulaciones disponibles de la colaboración. Una vez cargados los datos, junto con el envío de publicaciones de presentación respectivas, todos quedarán disponibles en el sitio web oficial <https://www.cosmosim.org> donde ya hay habilitada una interfaz de acceso personalizado a todas las tablas mediante SQL, uno de los lenguajes estándar utilizado para acceder a bases de datos en VO.

4.8. Lectura de salidas del código SAG

Durante el presente capítulo hemos descrito todo el proceso de realización de catálogos utilizando el modelo semianalítico SAG sobre la simulación MDPL2. Si bien estos datos estarán disponibles en una base de datos abierta, la cantidad de propiedades incluidas en ella es menor a la almacenada en los archivos originales del modelo. Por ello, estos archivos estarán igualmente disponibles tanto para uso interno como para quienes lo soliciten para concretar proyectos particulares.

No obstante, también hemos visto que la información almacenada está fragmentada en una cantidad importante de archivos HDF5 lo cual complica su lectura de forma sencilla. Si bien unificar total o parcialmente esta colección podría sugerirse como una solución a esta problemática, esta estrategia debería aplicarse sobre cada catálogo creado, restando además transportabilidad ante copias remotas y complicando lecturas parciales de la población de galaxias.

En este contexto, hemos desarrollado un módulo de lectura de salidas genéricas del SAG que resuelve el problema planteado, con soporte no sólo del formato de salidas reducidas utilizado en esta ocasión sino también al formato original del código que tiene arreglos con diferentes dimensiones y unidades arbitrarias. Este paquete está escrito en PYTHON

Nombre en BD	Detalles y unidades en BD	Nombre en SAG y conversión
redshift z	–	Redshift
HostHaloID	–	GalaxyHostID
MainHaloID	–	MainHaloID
HaloID	–	HaloID
ForestID	–	ForestID
GalaxyType	–	Galaxy_Type
HaloMass	M_{\odot}/h	Halo/M200c
Vmax	km/s	Halo/Vmax
Vpeak	km/s	Halo/Vpeak
NFW concentration	–	Halo/cNFW
X	comóvil, Mpc/h	$X \times 10^{-3}$
Y	comóvil, Mpc/h	$Y \times 10^{-3}$
Z	comóvil, Mpc/h	$Z \times 10^{-3}$
Vx	peculiar, km/s	Vx
Vy	peculiar, km/s	Vy
Vz	peculiar, km/s	Vz
MstarSpheroid	M_{\odot}/h	M_star_bulge
MstarDisk	M_{\odot}/h	M_star_disk
McoldSpheroid	M_{\odot}/h	M_gas_bulge
McoldDisk	M_{\odot}/h	M_gas_disk
Mhot	M_{\odot}/h	M_hot
Mbh	M_{\odot}/h	Mbh
SFRspheroid	$M_{\odot}/h/\text{Gyr}$	SFR_bulge $\times 10^9$
SFR	$M_{\odot}/h/\text{Gyr}$	SFR $\times 10^9$
MagStarSDSSu	en reposo, con polvo	Magnitudes/Mag_uS_dust
MagStarSDSSg	en reposo, con polvo	Magnitudes/Mag_gS_dust
MagStarSDSSr	en reposo, con polvo	Magnitudes/Mag_rS_dust
MagStarSDSSi	en reposo, con polvo	Magnitudes/Mag_iS_dust
MagStarSDSSz	en reposo, con polvo	Magnitudes/Mag_zS_dust
MZgasSpheroid	M_{\odot}/h	MZ_gas_bulge
MZgasDisk	M_{\odot}/h	MZ_gas_disk
MZhotHalo	M_{\odot}/h	MZ_hot_halo
MZstarSpheroid	M_{\odot}/h	MZ_stars_bulge
MZstarDisk	M_{\odot}/h	MZ_stars_disk
MeanAgeStars	Gyr	T_stars

Tabla 4.5: Conversión de arreglos entre salidas del SAG para generar tablas que se publicarán en la base de datos (BD).

y ha de ser utilizado desde ese mismo lenguaje. El código completo se distribuye bajo la licencia GNU GPL, versión 3, y está disponible de forma pública en el repositorio web oficial⁶ desde el cual se puede descargar a través de MERCURIAL:

```
hg clone https://bitbucket.org/cnvega/SAGreader
```

El módulo incluye rutinas para leer en conjunto todos los archivos de salidas del SAG localizados en un directorio, los cuales son automáticamente clasificados según la instantánea a que pertenezcan. Mediante instrucciones únicas y simples, permite leer propiedades específicas de todas las galaxias, ya sea de la población completa, una instantánea, o un rango de corrimientos al rojo a elección; así como también permite leer el conjunto de propiedades asociado a una misma galaxia a diferentes épocas. Además, es capaz de interpretar y entregar todos los factores de conversión de unidades para cada propiedad almacenada de una forma explícita y automática.

Por otro lado, el paquete incluye además ejemplos de uso, documentación, y una serie de rutinas adicionales que realizan gráficos característicos de la población de galaxias cargada. En particular, todas las gráficas incluidas en este capítulo en las secciones de cálculo de catálogos (4.5, 4.6 y 4.7) fueron obtenidos con este paquete, y están disponibles para ser recreados ya sea con estos catálogos u otros obtenidos con el modelo SAG. El paquete ha sido creado pensando en datos de gran envergadura, por lo que trata de optimizar tanto el uso de memoria como la cantidad de accesos a disco duro en su ejecución.

Los archivos que componen el paquete en su versión actual son:

- El módulo de lectura `SAGreader.py`, donde se definen todas las clases utilizadas en el proceso de carga y lectura de archivos. Éste puede ser usado independientemente del resto del paquete.
- Dos archivos de código `SAGplots*.py` con rutinas que realizan gráficos de ciertas propiedades de las galaxias utilizando el lector anterior, comparando cada una con datos observacionales.
- Un directorio llamado `Data` que contiene las curvas de los datos observacionales utilizados por las rutinas anteriores.
- Un archivo de código llamado `simu.py` que contiene ejemplos de uso de todo el paquete, y que fue utilizado por el autor para producción científica.
- Un par de archivos `doxygen` para generar la documentación extendida, útil para desarrolladores.

El módulo de lectura utiliza el paradigma de programación orientada a objetos para simplificar su uso. Cada clase, método o rutina tanto del módulo como de los archivos con rutinas para hacer gráficos incluyen su correspondiente documentación. Esta puede ser accedida tanto desde los comandos de ayuda estándar de PYTHON, como del manual de desarrollo que se puede construir utilizando el programa de documentación DOXYGEN.

⁶<https://bitbucket.org/cnvega/SAGreader>

```
1 doxygen config.dox
```

Para más información de la descripción o uso del paquete, existe además un manual breve (por ahora en español únicamente) que se puede solicitar al autor, o directamente seguir cualquiera de las rutinas de ejemplo incluidas en el archivo `simu.py` que contienen las instrucciones usuales de utilización del paquete en su totalidad.

Capítulo 5

Conclusiones

Esta Tesis se ha enfocado en el análisis, modificación y uso del modelo semianalítico de formación de galaxias SAG, con el propósito de estudiar los procesos físicos que conducen la formación y evolución de las galaxias en diferentes escalas espaciales. Como producto final del trabajo, se han generado catálogos de galaxias simuladas con una consistencia predictiva suficiente que permite publicarlos abiertamente en un observatorio virtual para usos generales.

En la primera parte, presentada en el Capítulo 2, hemos detallado toda la intervención necesaria que tuvo que realizarse sobre modelo semianalítico para alcanzar el objetivo planteado y generar el producto deseado. Las modificaciones abarcaron mejoras en rendimiento incluyendo paralelizaciones con multihilos, adaptaciones a nuevos tipos de simulaciones a través de los catálogos de halos en estándar SUSSING, automatización de estrategias de cálculo distribuido y definiciones de nuevos tipos de salida del código aptas para cálculos de gran envergadura. Muchas de estas mejoras permitieron que el modelo SAG pudiese participar del proyecto de comparación nITFy (Knebe et ál., 2015) aún en curso, y en las continuaciones del mismo (Knebe et al. en preparación). Finalmente, se estableció una estrategia para calibrar los parámetros libres del modelo cuando es utilizado sobre simulaciones cuyos catálogos de halos estén almacenados en el formato SUSSING.

En la segunda parte, presentada en el Capítulo 3, hemos detallado todos los cambios realizados al tratamiento físico de las galaxias satélites que son modeladas por el código semianalítico. Se describió un análisis detallado de la tasa de fusión de galaxias obtenida según el comportamiento de los halos de una simulación, y cómo esta sirve para ajustar el modelo analítico utilizado para integrar la evolución orbital de las galaxias satélites huérfanas que emergen al modelar sistemas. Debido a que la integración de órbitas se realiza mediante un código anexo al modelo SAG, todas las adaptaciones a formatos nuevos y rendimiento efectuadas en este último debieron ser igualmente replicadas en el integrador. A su vez, se diseñó e implementó una estrategia para estimar evoluciones orbitales consistentes cuando los halos de materia oscura se ven afectados por las condiciones de borde periódicas de las simulaciones de N -cuerpos. Todas estas mejoras serán publicadas en un trabajo futuro (Vega-Martínez et ál. (a), en preparación). Por otro lado, se estudió el comportamiento de la presión de barrido sobre el gas de las galaxias satélites, y se propuso un nuevo ajuste analítico para modelar la presión ejercida sobre estas galaxias, que depende tanto de la distancia al centro de los halos anfitriones como de la masa de los mismos y del corrimiento al rojo (Vega-Martínez et ál. (b), en preparación).

Finalmente, en el Capítulo 4 hemos descrito el proyecto de colaboración que hemos

iniciado con el grupo MultiDark, consistente en poblar un conjunto de simulaciones cosmológicas de materia oscura con galaxias utilizando el código SAG. En este contexto, hemos descrito el análisis y procesamiento que hemos realizado sobre la primera de las simulaciones consideradas en el proyecto, llamada MDPL2, que abarca un volumen cosmológico de $1 h^{-1}$ Gpc de lado. Hemos detallado la creación y análisis de tres catálogos de galaxias generados con el modelo sobre esta simulación, cada uno realizado en diferentes etapas de desarrollo de los demás temas tratados durante la Tesis. Finalmente, se creó un catálogo de galaxias obtenido con un modelo calibrado, y que es capaz de cumplir con suficiente precisión todas las restricciones observacionales requeridas para su publicación. Este será liberado a mediados del primer cuatrimestre del año 2017, siendo ofrecido libremente mediante la misma base de datos que aloja a las simulaciones del grupo MultiDark (Cora, Vega-Martínez et ál. en preparación; Knebe et ál. en preparación).

Como productos anexos en el desarrollo de la Tesis, se han creado además dos códigos abiertos útiles para procesar la información manipulada durante el desarrollo de este proyecto. El primero consiste en una colección de códigos que permiten leer, particionar y convertir catálogos de halos almacenados en el formato SUSSING, útil tanto en la adaptación de otros códigos semianalíticos a este formato como para el manejo general de este tipo de simulaciones. El segundo corresponde a un paquete de lectura genérico de las salidas del código SAG, diseñado con el fin de simplificar al máximo los procesos de carga y lectura de los productos obtenidos con este modelo para favorecer el análisis y estudio de los mismos. Ambos paquetes de códigos se encuentran abiertamente disponibles en el repositorio oficial del autor¹, ofrecidos mediante licencia GNU GPL.

Proyectos derivados y trabajo a futuro

La cantidad de tópicos contemplados en esta Tesis, a su vez, ha abierto la posibilidad de iniciar varios proyectos de investigación derivados o a modo de continuación del trabajo realizado.

Gracias a las adaptaciones efectuadas al código SAG, este ahora puede aplicarse a un vasto conjunto de simulaciones nuevas que antes no eran procesables por el código. Por ejemplo, hemos iniciado una colaboración con la *Dark Energy Science Collaboration* del relevamiento LSST, para modelar la población de galaxias de la simulación MassiveBlack-II en su versión de materia oscura solamente (Khandai et ál., 2015), de modo que los catálogos generados con ella puedan ser utilizados en la planificación del telescopio. A su vez, mediremos el impacto en la población de galaxias al utilizar simulaciones que consideran materia oscura interactuante con bariones (Schewtschenko et ál., 2015).

El desarrollo del paquete de códigos que permite manipular fácilmente catálogos en formato SUSSING nos ha permitido, entre otros, extraer todos los cúmulos y grupos de la simulación MDPL2 que sean análogos al Grupo Local. Con ello, realizaremos análisis específicos y detallados tanto de la formación de galaxias tipo Vía Láctea como de sus propiedades (Gargiulo et ál., en preparación).

Al estudiar los tratamientos aplicados a las galaxias satélites hemos definido un criterio de fusiones de galaxias que permite reproducir de forma consistente la evolución de la tasa de fusiones histórica. No obstante, aún no está determinado si el valor utilizado en el corte por proximidad que constituye este criterio deba mantenerse constante para corrimientos al rojo altos. Para ello, debemos realizar un estudio de la evolución de los tamaños característicos de galaxias centrales que poseen fusiones, proponiendo para ello utilizar

¹<https://bitbucket.org/cnvega>

simulaciones de alta resolución en masa junto al modelo semianalítico SAG. Por otro lado, el criterio de pérdida de masa por fuerzas de marea utilizado en el integrador debe ser estudiado con más detalle. La evolución de los parámetros que caracterizan los perfiles de densidad de los halos de materia oscura satélites durante su evolución orbital pueden ser estudiados directamente en simulaciones de materia oscura, lo cual permitiría conocer con precisión el mecanismo de pérdida de masa y así medir si existe algún posible sesgo metodológico en los valores determinados por los algoritmos de búsqueda de subhalos.

Durante la creación de los diversos catálogos de galaxias con el modelo SAG utilizando la simulación MDPL2, descubrimos una incompatibilidad entre los tratamientos de efectos de ambiente en galaxias satélites y la evolución del tamaño de los discos de las galaxias utilizando los cambios de momento angular por acreción de material. Deben realizarse análisis más detallados sobre el origen de estas discrepancias, dado que ambos tratamientos permiten reproducir características específicas de las propiedades resultantes de la población de galaxias.

Finalmente, durante el año 2017 se dará inicio a la segunda etapa de la colaboración con el grupo MultiDark que consiste en procesar la simulación SMDPL, muy alta en resolución de masa. Esta simulación permitirá refinar aún más los tratamientos utilizados en el modelo, así como estudiar en más detalle sistemas con galaxias de menores masas en un contexto cosmológico.

Bibliografía

- Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 3, 2046
- Awan H., Gawiser E., Kurczynski P., Jones R. L., Zhan H., Padilla N. D., Arancibia A. M. M., Orsi A., Cora S. A., Yoachim P., 2016, *The Astrophysical Journal*, 829, 1, 50
- Baldry I. K., Driver S. P., Loveday J., Taylor E. N., Kelvin L. S., Liske J., Norberg P., Robotham A. S. G., Brough S., Hopkins A. M., Bamford S. P., Peacock J. A., Bland-Hawthorn J., Conselice C. J., Croom S. M., Jones D. H., Parkinson H. R., et ál., 2012, *Monthly Notices of the Royal Astronomical Society*, 421, 621
- Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *The Astrophysical Journal*, 600, 2, 681
- Baldry I. K., Glazebrook K., Driver S. P., 2008, *Monthly Notices of the Royal Astronomical Society*, 388, 3, 945
- Barnes J., Efstathiou G., 1987, *The Astrophysical Journal*, 319, 575
- Baugh C. M., 2006, *Reports on Progress in Physics*, 69, 12, 3101
- Behroozi P. S., Wechsler R. H., Conroy C., 2013a, *The Astrophysical Journal*, 770, 1, 57
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013b, *The Astrophysical Journal*, 762, 2, 109
- Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013, *The Astrophysical Journal*, 763, 18
- Bell E. F., Phleps S., Somerville R. S., Wolf C., Borch A., Meisenheimer K., 2006, *The Astrophysical Journal*, 652, 1, 270
- Benson A. J., 2010, *Physics Reports*, 495, 2-3, 33
- Benson A. J., 2012, *New Astronomy*, 17, 2, 175
- Bernardi M., Meert A., Sheth R. K., Fischer J.-L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2017, *Monthly Notices of the Royal Astronomical Society*, 467, 2, 2217
- Bertone S., De Lucia G., Thomas P. A., 2007, *Monthly Notices of the Royal Astronomical Society*, 379, 1143
- Binney J., Tremaine S., 1987, *Galactic dynamics*, Princeton, N.J.: Princeton University Press

- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *The Astrophysical Journal*, 379, 440
- Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014, *Astronomy & Astrophysics*, 564, A66
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *Monthly Notices of the Royal Astronomical Society*, 370, 2, 645
- Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 1, 93
- Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *Monthly Notices of the Royal Astronomical Society*, 398, 3, 1150
- Bruzual G., Charlot S., 2003, *Monthly Notices of the Royal Astronomical Society*, 344, 4, 1000
- Bryan G. L., Norman M. L., 1998, *The Astrophysical Journal*, 495, 1, 80
- Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001a, *The Astrophysical Journal*, 555, 1, 240
- Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001b, *Monthly Notices of the Royal Astronomical Society*, 321, 3, 559
- Carretero J., Castander F. J., Gaztanaga E., Crocce M., Fosalba P., 2014, *Monthly Notices of the Royal Astronomical Society*, 447, 1, 646
- Chandrasekhar S., 1943, *The Astrophysical Journal*, 97, 255
- Coelho P., Bruzual G., Charlot S., Weiss A., Barbuy B., Ferguson J. W., 2007, *Monthly Notices of the Royal Astronomical Society*, 382, 2, 498
- Conselice C. J., 2006, *Monthly Notices of the Royal Astronomical Society*, 373, 4, 1389
- Cora S. A., 2006, *Monthly Notices of the Royal Astronomical Society*, 368, 4, 1540
- Cora S. A., Tornatore L., Tozzi P., Dolag K., 2008, *Monthly Notices of the Royal Astronomical Society*, 386, 1, 96
- Croton D. J., Springel V., White S. D. M., De Lucia G., Frenk C. S., Gao L., Jenkins A., Kauffmann G., Navarro J. F., Yoshida N., 2006, *Monthly Notices of the Royal Astronomical Society*, 365, 1, 11
- Croton D. J., Stevens A. R. H., Tonini C., Garel T., Bernyk M., Bibiano A., Hodkinson L., Mutch S. J., Poole G. B., Shattow G. M., 2016, *The Astrophysical Journal Supplement Series*, 222, 2, 22
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *The Astrophysical Journal*, 292, 371
- Davis M., Peebles P. J. E., 1983, *The Astrophysical Journal*, 267, 465

- De Lucia G., Blaizot J., 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 1, 2
- De Lucia G., Kauffmann G., White S. D. M., 2004, *Monthly Notices of the Royal Astronomical Society*, 349, 3, 1101
- De Propris R., Driver S. P., Colless M., Drinkwater M. J., Loveday J., Ross N. P., Bland-Hawthorn J., York D. G., Pimblet K., 2010, *The Astronomical Journal*, 139, 2, 794
- Dolag K., Borgani S., Murante G., Springel V., 2009, *Monthly Notices of the Royal Astronomical Society*, 399, 2, 497
- Dolag K., Vazza F., Brunetti G., Tormen G., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 3, 753
- Domínguez Sánchez H., Pozzi F., Gruppioni C., Cimatti A., Ilbert O., Pozzetti L., McCracken H., Capak P., Le Floch E., Salvato M., Zamorani G., Carollo C. M., Contini T., Kneib J.-P., Le Fèvre O., Lilly S. J., Mainieri V., et ál., 2011, *Monthly Notices of the Royal Astronomical Society*, 417, 2, 900
- Dopita M. A., Kewley L. J., Heisler C. A., Sutherland R. S., 2000, *The Astrophysical Journal*, 542, 224
- Efstathiou G., 2000, *Monthly Notices of the Royal Astronomical Society*, 317, 3, 697
- Efstathiou G., Lake G., Negroponte J., 1982, *Monthly Notices of the Royal Astronomical Society*, 199, 4, 1069
- Einasto J., 1965, *Trudy Astrofizicheskogo Instituta Alma-Ata*, 5, 87
- Evrard A. E., 1989, *The Astrophysical Journal*, 341, L71
- Font A. S., Bower R. G., McCarthy I. G., Benson A. J., Frenk C. S., Helly J. C., Lacey C. G., Baugh C. M., Cole S., 2008, *Monthly Notices of the Royal Astronomical Society*, 389, 4, 1619
- Foster A. R., Ji L., Smith R. K., Brickhouse N. S., 2012, *The Astrophysical Journal*, 756, 2, 128
- Frenk C. S., White S. D. M., Efstathiou G., Davis M., 1990, *The Astrophysical Journal*, 351, 10
- Fujii M., Funato Y., Makino J., 2006, *Publications of the Astronomical Society of Japan*, 58, 4, 743
- Gan J.-L., Kang X., Hou J.-L., Chang R.-X., 2010, *Research in Astronomy and Astrophysics*, 10, 12, 1242
- Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, *Monthly Notices of the Royal Astronomical Society*, 387, 536
- Gargiulo I. D., Cora S. A., Padilla N. D., Munoz Arancibia A. M., Ruiz A. N., Orsi A. A., Tecce T. E., Weidner C., Bruzual G., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 4, 3820

- Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 1, 264
- Gruppioni C., Calura F., Pozzi F., Delvecchio I., Berta S., De Lucia G., Fontanot F., Franceschini A., Marchetti L., Menci N., Monaco P., Vaccari M., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 4, 3419
- Gunn J. E., Gott J. Richard I., 1972, *The Astrophysical Journal*, 176, 1
- Guo Q., Gonzalez-Perez V., Guo Q., Schaller M., Furlong M., Bower R. G., Cole S., Crain R. A., Frenk C. S., Helly J. C., Lacey C. G., Lagos C. d. P., Mitchell P., Schaye J., Theuns T., 2016, *Monthly Notices of the Royal Astronomical Society*, 461, 4, 3457
- Guo Q., White S., Boylan-Kolchin M., De Lucia G., Kauffmann G., Lemson G., Li C., Springel V., Weinmann S., 2011, *Monthly Notices of the Royal Astronomical Society*, 413, 1, 101
- Hashimoto Y., Funato Y., Makino J., 2003, *The Astrophysical Journal*, 582, 1, 196
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 3, 2663
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, *Monthly Notices of the Royal Astronomical Society*, 431, 4, 3373
- Hernquist L., 1990, *The Astrophysical Journal*, 356, 359
- Hirschmann M., De Lucia G., Fontanot F., 2016, *Monthly Notices of the Royal Astronomical Society*, 461, 2, 1760
- Hopkins P. F., Bundy K., Croton D., Hernquist L., Keres D., Khochfar S., Stewart K., Wetzel A., Younger J. D., 2010a, *The Astrophysical Journal*, 715, 1, 202
- Hopkins P. F., Croton D., Bundy K., Khochfar S., van den Bosch F., Somerville R. S., Wetzel A., Keres D., Hernquist L., Stewart K., Younger J. D., Genel S., Ma C.-P., 2010b, *The Astrophysical Journal*, 724, 2, 915
- Huchra J. P., Geller M. J., 1982, *The Astrophysical Journal*, 257, 423
- Ilbert O., McCracken H. J., Le Fèvre O., Capak P., Dunlop J., Karim A., Renzini M. A., Caputi K., Boissier S., Arnouts S., Aussel H., Comparat J., Guo Q., Hudelot P., Kartaltepe J., Kneib J. P., Krogager J. K., et ál., 2013, *Astronomy & Astrophysics*, 556, A55
- Ivezic Z., Tyson J. a., Abel B., Acosta E., Allsman R., AlSayyad Y., Anderson S. F., Andrew J., Angel R., Angeli G., Ansari R., Antilogus P., Arndt K. T., Astier P., Aubourg E., Axelrod T., Bard D. J., et ál., 2008, eprint arXiv:0805.2366
- Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, *The Astrophysical Journal*, 675, 2, 1095

- Jiang T., Hogg D. W., Blanton M. R., 2012, *The Astrophysical Journal*, 759, 2, 140
- Kauffmann G., Heckman T. M., White S. D. M., Charlot S., Tremonti C., Brinchmann J., Bruzual G., Peng E. W., Seibert M., Bernardi M., Blanton M., Brinkmann J., Castander F., Csábai I., Fukugita M., Ivezić Z., Munn J. A., et ál., 2003, *Monthly Notices of the Royal Astronomical Society*, 341, 1, 33
- Khandai N., Di Matteo T., Croft R., Wilkins S., Feng Y., Tucker E., DeGraf C., Liu M.-S., 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 2, 1349
- Kimm T., Yi S. K., Khochfar S., 2011, *The Astrophysical Journal*, 729, 1, 11
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *Monthly Notices of the Royal Astronomical Society*, 457, 4, 4340
- Knebe A., Knollmann S. R., Muldrew S. I., Pearce F. R., Aragon-Calvo M. A., Ascasibar Y., Behroozi P. S., Ceverino D., Colombi S., Diemand J., Dolag K., Falck B. L., Fasel P., Gardner J., Gottlöber S., Hsu C.-H., Iannuzzi F., et ál., 2011, *Monthly Notices of the Royal Astronomical Society*, 415, 3, 2293
- Knebe A., Pearce F. R., Thomas P. A., Benson A., Blaizot J., Bower R., Carretero J., Castander F. J., Cattaneo A., Cora S. A., Croton D. J., Cui W., Cunnama D., De Lucia G., Devriendt J. E., Elahi P. J., Font A., et ál., 2015, *Monthly Notices of the Royal Astronomical Society*, 451, 4, 4029
- Knollmann S. R., Knebe A., 2009, *The Astrophysical Journal Supplement Series*, 182, 2, 608
- Komatsu E., Smith K. M., Dunkley J., Bennett C. L., Gold B., Hinshaw G., Jarosik N., Larson D., Nolte M. R., Page L., Spergel D. N., Halpern M., Hill R. S., Kogut A., Limon M., Meyer S. S., Odegard N., et ál., 2011, *The Astrophysical Journal Supplement Series*, 192, 2, 18
- Kormendy J., Ho L. C., 2013, *Annual Review of Astronomy and Astrophysics*, 51, 1, 511
- Lacey C., Cole S., 1993, *Monthly Notices of the Royal Astronomical Society*, 262, 3, 627
- Lagos C. D. P., Cora S. A., Padilla N. D., 2008, *Monthly Notices of the Royal Astronomical Society*, 388, 2, 587
- Laureijs R., Amiaux J., Arduini S., Auguères J. L., Brinchmann J., Cole R., Cropper M., Dabin C., Duvet L., Ealet A., Garilli B., Gondoin P., Guzzo L., Hoar J., Hoekstra H., Holmes R., Kitching T., et ál., 2011, eprint arXiv:1110.3193
- Lee J., Yi S. K., 2013, *The Astrophysical Journal*, 766, 1, 38
- Lee J., Yi S. K., Elahi P. J., Thomas P. A., Pearce F. R., Behroozi P., Han J., Helly J., Jung I., Knebe A., Mao Y.-Y., Onions J., Rodriguez-Gomez V., Schneider A., Srisawat C., Tweed D., 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 4, 4197
- Levesque E. M., Kewley L. J., Larson K. L., 2010, *The Astronomical Journal*, 139, 712
- Li C., White S. D. M., 2009, *Monthly Notices of the Royal Astronomical Society*, 398, 4, 2177

- LSST Science Collaboration: Abell P. A., Allison J., Anderson S. F., Andrew J. R., Angel J. R. P., Armus L., Arnett D., Asztalos S. J., Axelrod T. S., Bailey S., Ballantyne D. R., Bankert J. R., Barkhouse W. A., Barr J. D., Barrientos L. F., Barth A. J., Bartlett J. G., et ál., 2009, eprint arXiv:0912.0201
- Madau P., Dickinson M., 2014, *Annual Review of Astronomy and Astrophysics*, 52, 1, 415
- Maraston C., 2005, *Monthly Notices of the Royal Astronomical Society*, 362, 3, 799
- McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 2, 593
- McCavana T., Micic M., Lewis G. F., Sinha M., Sharma S., Holley-Bockelmann K., Bland-Hawthorn J., 2012, *Monthly Notices of the Royal Astronomical Society*, 424, 1, 361
- McConnell N. J., Ma C.-P., 2013, *The Astrophysical Journal*, 764, 2, 184
- McKee C. F., Ostriker J. P., 1977, *The Astrophysical Journal*, 218, 148
- Minniti D., Lucas P. W., Emerson J. P., Saito R. K., Hempel M., Pietrukowicz P., Ahumada A. V., Alonso M. V., Alonso-Garcia J., Arias J. I., Bandyopadhyay R. M., Barbá R. H., Barbuy B., Bedin L. R., Bica E., Borissova J., Bronfman L., et ál., 2010, *New Astronomy*, 15, 433
- Mo H., Bosch F., White S., 2010, *Galaxy Formation and Evolution*, Cambridge, UK: Cambridge University Press.
- Mo H. J., Mao S., White S. D. M., 1998, *Monthly Notices of the Royal Astronomical Society*, 295, 2, 319
- Monaco P., Fontanot F., Taffoni G., 2007, *Monthly Notices of the Royal Astronomical Society*, 375, 4, 1189
- More S., Diemer B., Kravtsov A. V., 2015, *The Astrophysical Journal*, 810, 1, 36
- Moster B. P., Somerville R. S., Maubetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, *The Astrophysical Journal*, 710, 2, 903
- Muñoz Arancibia A. M., Navarrete F. P., Padilla N. D., Cora S. A., Gawiser E., Kurczynski P., Ruiz A. N., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 3, 2291
- Muratov A. L., Kereš D., Faucher-Giguère C.-A., Hopkins P. F., Quataert E., Murray N., 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 3, 2691
- Muzzin A., Marchesini D., Stefanon M., Franx M., McCracken H. J., Milvang-Jensen B., Dunlop J. S., Fynbo J. P. U., Brammer G., Labbé I., van Dokkum P. G., 2013, *The Astrophysical Journal*, 777, 1, 18
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *The Astrophysical Journal*, 462, 563
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *The Astrophysical Journal*, 490, 2, 493

- Neyman J., Scott E. L., 1952, *The Astrophysical Journal*, 116, 144
- Ogiya G., Burkert A., 2016, *Monthly Notices of the Royal Astronomical Society*, 457, 2, 2164
- Onions J., Knebe A., Pearce F. R., Muldrew S. I., Lux H., Knollmann S. R., Ascasibar Y., Behroozi P., Elahi P., Han J., Maciejewski M., Merchán M. E., Neyrinck M., Ruiz A. N., Sgró M. A., Springel V., Tweed D., 2012, *Monthly Notices of the Royal Astronomical Society*, 423, 2, 1200
- Orsi A., Padilla N., Groves B., Cora S., Tecce T., Gargiulo I., Ruiz A., 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 1, 799
- Padilla N. D., Salazar-Albornoz S., Contreras S., Cora S. A., Ruiz A. N., 2014, *Monthly Notices of the Royal Astronomical Society*, 443, 3, 2801
- Partridge R. B., 1995, *3K: The Cosmic Microwave Background Radiation*
- Patton D. R., Atfield J. E., 2008, *The Astrophysical Journal*, 685, 1, 235
- Peacock J., 1999, *Cosmological Physics*, Cambridge, UK: Cambridge University Press.
- Peebles P. J. E., 1969, *The Astrophysical Journal*, 155, 393
- Petts J. A., Gualandris A., Read J. I., 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 4, 3778
- Planck Collaboration: Ade P. A. R., Aghanim N., Armitage-Caplan C., Arnaud M., Ashdown M., Atrio-Barandela F., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., Battaner E., Benabed K., Benoît A., Benoit-Lévy A., Bernard J.-P., Bersanelli M., et ál., 2014, *Astronomy & Astrophysics*, 571, A16
- Planck Collaboration: Ade P. A. R., Aghanim N., Arnaud M., Ashdown M., Aumont J., Baccigalupi C., Banday A. J., Barreiro R. B., Bartlett J. G., Bartolo N., Battaner E., Battye R., Benabed K., Benoit A., Benoit-Lévy A., Bernard J. P., Bersanelli M., et ál., 2016, *Astronomy & Astrophysics*, 594, A13
- Press W. H., Schechter P., 1974, *The Astrophysical Journal*, 187, 425
- Riess A. G., Macri L. M., Hoffmann S. L., Scolnic D., Casertano S., Filippenko A. V., Tucker B. E., Reid M. J., Jones D. O., Silverman J. M., Chornock R., Challis P., Yuan W., Brown P. J., Foley R. J., 2016, *The Astrophysical Journal*, 826, 1, 56
- Rollinde E., Petitjean P., Pichon C., Colombi S., Aracil B., D'Odorico V., Haehnelt M. G., 2003, *Monthly Notices of the Royal Astronomical Society*, 341, 1279
- Ruiz A. N., Cora S. A., Padilla N. D., Domínguez M. J., Vega-Martínez C. A., Tecce T. E., Orsi Á., Yaryura Y., Lambas D. G., Gargiulo I. D., Arancibia A. M. M., 2015, *The Astrophysical Journal*, 801, 2, 139
- Schaye J., Crain R. A., Bower R. G., Furlong M., Schaller M., Theuns T., Dalla Vecchia C., Frenk C. S., McCarthy I. G., Helly J. C., Jenkins A., Rosas-Guevara Y. M., White S. D. M., Baes M., Booth C. M., Camps P., Navarro J. F., et ál., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 1, 521

- Schewtschenko J. A., Wilkinson R. J., Baugh C. M., Boehm C., Pascoli S., 2015, *Monthly Notices of the Royal Astronomical Society*, 449, 4, 3587
- SDSS Collaboration: Albareti F. D., Prieto C. A., Almeida A., Anders F., Anderson S., Andrews B. H., Aragon-Salamanca A., Argudo-Fernandez M., Armengaud E., Aubourg E., Avila-Reese V., Badenes C., Bailey S., Barbuy B., Barger K., Barrera-Ballesteros J., Bartosz C., et ál., 2016, eprint arXiv:1608.02013
- Shen S., Mo H. J., White S. D. M., Blanton M. R., Kauffmann G., Voges W., Brinkmann J., Csabai I., 2003, *Monthly Notices of the Royal Astronomical Society*, 343, 3, 978
- Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 1, 575
- Skibba R. A., Sheth R. K., 2009, *Monthly Notices of the Royal Astronomical Society*, 392, 3, 1080
- Soltan A., 1982, *Monthly Notices of the Royal Astronomical Society*, 200, 115
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, *Monthly Notices of the Royal Astronomical Society*, 391, 2, 481
- Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 4, 1105
- Springel V., 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 2, 791
- Springel V., Wang J., Vogelsberger M., Ludlow A., Jenkins A., Helmi A., Navarro J. F., Frenk C. S., White S. D. M., 2008, *Monthly Notices of the Royal Astronomical Society*, 391, 4, 1685
- Springel V., White S. D. M., Jenkins A., Frenk C. S., Yoshida N., Gao L., Navarro J., Thacker R., Croton D., Helly J., Peacock J. A., Cole S., Thomas P., Couchman H., Evrard A., Colberg J., Pearce F., 2005, *Nature*, 435, 7042, 629
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *Monthly Notices of the Royal Astronomical Society*, 328, 3, 726
- Srisawat C., Knebe A., Pearce F. R., Schneider A., Thomas P. A., Behroozi P., Dolag K., Elahi P. J., Han J., Helly J., Jing Y., Jung I., Lee J., Mao Y.-Y., Onions J., Rodriguez-Gomez V., Tweed D., Yi S. K., 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 1, 150
- Taylor J. E., Babul A., 2001, *The Astrophysical Journal*, 559, 2, 716
- Tecce T. E., Cora S. A., Tissera P. B., 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 4, 3170
- Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. d. P., 2010, *Monthly Notices of the Royal Astronomical Society*, 408, 4, 2008
- The Dark Energy Survey Collaboration, 2005, arXiv e-prints, astro-ph/0510346

- Thomas P. A., Onions J., Tweed D., Benson A. J., Croton D., Elahi P., Henriques B., Iliev I. T., Knebe A., Lux H., Mao Y.-Y., Neyrinck M., Pearce F. R., Rodriguez-Gomez V., Schneider A., Srisawat C., 2015, eprint arXiv:1508.05388
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *The Astrophysical Journal*, 688, 2, 709
- Tomczak A. R., Quadri R. F., Tran K.-V. H., Labbé I., Straatman C. M. S., Papovich C., Glazebrook K., Allen R., Brammer G. B., Kacprzak G. G., Kavinwanichakij L., Kelson D. D., McCarthy P. J., Mehrtens N., Monson A. J., Persson S. E., Spitler L. R., Tilvi V., van Dokkum P., 2014, *The Astrophysical Journal*, 783, 2, 85
- Tormen G., 1997, *Monthly Notices of the Royal Astronomical Society*, 290, 411
- Vale A., Ostriker J. P., 2004, *Monthly Notices of the Royal Astronomical Society*, 353, 1, 189
- Wen Z. L., Liu F. S., Han J. L., 2009, *The Astrophysical Journal*, 692, 1, 511
- White S. D. M., Efstathiou G., Frenk C. S., 1993, *Monthly Notices of the Royal Astronomical Society*, 262, 1023
- White S. D. M., Frenk C. S., 1991, *The Astrophysical Journal*, 379, 52
- White S. D. M., Rees M. J., 1978, *Monthly Notices of the Royal Astronomical Society*, 183, 3, 341
- Zehavi I., Zheng Z., Weinberg D. H., Frieman J. A., Berlind A. A., Blanton M. R., Scocimarro R., Sheth R. K., Strauss M. A., Kayo I., Suto Y., Fukugita M., Nakamura O., Bahcall N. A., Brinkmann J., Gunn J. E., Hennessy G. S., et ál., 2005, *The Astrophysical Journal*, 630, 1
- Zentner A. R., Bullock J. S., 2003, *The Astrophysical Journal*, 598, 1, 49
- Zhang C., Yu Q., Lu Y., 2016, *The Astrophysical Journal*, 820, 2, 85
- Zhao D. H., Jing Y. P., Mo H. J., Börner G., 2009, *The Astrophysical Journal*, 707, 1, 354

Apéndice A

Propiedades de la salida reducida

Nombre	Descripción
Identificadores	
ForestID	Identificador del bosque de pertenencia.
HaloID	Identificador del halo de pertenencia.
MainHaloID	Identificador del halo principal.
GalaxyHostID	ID del último halo al cual perteneció.
GalaxyStaticID	ID estática (SID) de las galaxias.
GalaxyType	Tipo de la galaxia.
Propiedades de las galaxias	
X	Posición x comóvil [kpc/ h].
Y	Posición y comóvil [kpc/ h].
Z	Posición z comóvil [kpc/ h].
Vx	Velocidad x peculiar física [km/s].
Vy	Velocidad y peculiar física [km/s].
Vz	Velocidad z peculiar física [km/s].
M_star_disk	Masa estelar del disco [M_{\odot}/h].
M_star_bulge	Masa estelar del bulbo [M_{\odot}/h].
M_ICstars	Masa estelar del halo [M_{\odot}/h].
M_hot	Masa del gas caliente [M_{\odot}/h].
M_gas_disk	Masa del gas frío [M_{\odot}/h].
M_gas_bulge	Masa del gas del bulbo [M_{\odot}/h].
Mbh	Masa del agujero negro central [M_{\odot}/h].
Rhalf_disk	Radio del disco (mitad de masa) [kpc/ h].
Rhalf_bulge	Radio del bulbo (mitad de masa, 3D) [kpc/ h].
SFR	Tasa de formación estelar [$M_{\odot}/h/yr$].
SFR_bulge	Tasa de formación estelar (bulbo) [$M_{\odot}/h/yr$].
SFR_bulge_DI	Tasa de formación estelar por inestabilidades de disco [$M_{\odot}/h/yr$].
InstantSfrQuies	Tasa de formación estelar instantánea, suave [$M_{\odot}/h/yr$].
InstantSfrBulge	Tasa de formación estelar instantánea, brotes [$M_{\odot}/h/yr$].
T_stars	Edad media de la población estelar [Gyr].
Metalicidades	
MZ_stars_disk	Masa de metales del disco, estrellas [M_{\odot}/h].
MZ_stars_bulge	Masa de metales del bulbo, estrellas [M_{\odot}/h].
MZ_ICstars	Masa de metales del halo, estrellas [M_{\odot}/h].

MZ_gas_disk	Masa de metales del disco, gas [M_{\odot}/h].
MZ_gas_bulge	Masa de metales del bulbo, gas [M_{\odot}/h].
MZ_hot_halo	Masa de metales del halo, gas [M_{\odot}/h].
OH_gas_disk_bulge	Tasa de masa O/H en el gas frío.
FeH_stars	Tasa de masa Fe/H estelar (disco + bulbo).
FeH_hot_halo	Tasa de masa Fe/H en el halo.
<hr/>	
Magnitudes absolutas	
Magnitudes/Mag_B_dust	Filtro <i>B</i> con corrección por polvo.
Magnitudes/Mag_I_dust	Filtro <i>I</i> con corrección por polvo.
Magnitudes/Mag_K_dust	Filtro <i>K</i> con corrección por polvo.
Magnitudes/Mag_R_dust	Filtro <i>R</i> con corrección por polvo.
Magnitudes/Mag_U_dust	Filtro <i>U</i> con corrección por polvo.
Magnitudes/Mag_V_dust	Filtro <i>V</i> con corrección por polvo.
Magnitudes/Mag_uS	Filtro <i>u</i> SDSS.
Magnitudes/Mag_gS	Filtro <i>g</i> SDSS.
Magnitudes/Mag_rS	Filtro <i>r</i> SDSS.
Magnitudes/Mag_iS	Filtro <i>i</i> SDSS.
Magnitudes/Mag_zS	Filtro <i>z</i> SDSS.
Magnitudes/Mag_uS_dust	Filtro <i>u</i> SDSS con corrección por polvo.
Magnitudes/Mag_gS_dust	Filtro <i>g</i> SDSS con corrección por polvo.
Magnitudes/Mag_rS_dust	Filtro <i>r</i> SDSS con corrección por polvo.
Magnitudes/Mag_iS_dust	Filtro <i>i</i> SDSS con corrección por polvo.
Magnitudes/Mag_zS_dust	Filtro <i>z</i> SDSS con corrección por polvo.
Magnitudes/Bulge/Mag_uSb	Filtro <i>u</i> SDSS del bulbo.
Magnitudes/Bulge/Mag_gSb	Filtro <i>g</i> SDSS del bulbo.
Magnitudes/Bulge/Mag_rSb	Filtro <i>r</i> SDSS del bulbo.
Magnitudes/Bulge/Mag_iSb	Filtro <i>i</i> SDSS del bulbo.
Magnitudes/Bulge/Mag_zSb	Filtro <i>z</i> SDSS del bulbo.
Magnitudes/Halo/Mag_gSh	Filtro <i>g</i> SDSS del halo.
Magnitudes/Halo/Mag_rSh	Filtro <i>r</i> SDSS del halo.
<hr/>	
Propiedades de los halos	
Halo/M200c	Masa del halo, se mantiene fija en huérfanas [M_{\odot}/h].
Halo/R200c	Radio del halo, se mantiene fijo en huérfanas [kpc/ h].
Halo/Vmax	Velocidad circular del halo (coord. físicas) [km/s].
Halo/Vpeak	Velocidad circular del halo al momento de acreción (coord. físicas) [km/s].
Halo/cNFW	Parámetro de concentración de NFW, se mantiene fijo en huérfanas.
<hr/>	
Árbol de fusión de galaxias	
GalDescendant	SID de la galaxia descendiente.
GalNumProgenitors	Número de progenitores.
GalFirstProgenitor	SID del primer progenitor.
GalNextProgenitor	Lista ligada de progenitores.

Tabla A.1: Lista de propiedades incluidas en la salida reducida del código en formato HDF5, como la que se utilizó para inyectar en la base datos pública las galaxias obtenidas con la simulación MDPL2.