

---

# Historias de formación estelar de galaxias MultiDark

---

Olivia Vidal Velázquez

Tutora: Violeta González Pérez

*Universidad Autónoma de Madrid, Facultad de Ciencias, Departamento de Física Teórica*

Mayo 2021

**Abstract.** Para entender la historia evolutiva del universo, es necesario entender primero cómo se han creado las estrellas, siendo el estudio de la formación estelar en las galaxias imprescindible. En este proyecto se estudia este aspecto utilizando las galaxias creadas computacionalmente por modelos semi-analíticos de formación y evolución de galaxias (GALACTICUS, SAG, SAGE), utilizando la simulación de materia oscura MultiDark Planck 2, que tiene el volumen de un cubo de lado  $1\text{Gpc } h^{-1}$  con más de  $10^{10}$  partículas de masa  $1.5 \times 10^{-9} M_{\odot} h^{-1}$ , es decir, un tamaño comparable al tamaño del Universo observable. Se comparan las galaxias teóricas con los datos observacionales que se tienen en la actualidad. Se estudian en detalle los resultados del modelo semi-analítico SAGE en un amplio rango de tiempos cósmicos, desde el presente,  $z = 0$ , hasta el primer giga año de vida del Universo,  $z = 5$ . Encontrando que, a partir de aproximadamente la mitad de la edad del Universo ( $z > 0.6$ ), el modelo no produce un número suficiente de galaxias con alta SFR en comparación con las observaciones. El modelo reproduce bien la relación creciente que existe entre la tasa de formación estelar y la masa de las galaxias, pero no reproduce las galaxias masivas ( $M_* > 10^{11} M_{\odot}$ ) sin tasa de formación estelar, que existen según las observaciones, a partir de aproximadamente el segundo año de vida del Universo ( $z > 3.61$ ).

<b>1. INTRODUCCIÓN</b>	<b>2</b>
<b>2. MÉTODOS</b>	<b>3</b>
2.1. La edad del universo . . . . .	3
2.2. Simulación de materia oscura con técnica de N-cuerpos. . . . .	3
2.3. Modelos de galaxias. . . . .	5
2.4. Descripción de la simulación MultiDark y los modelos semi-analíticos utilizados. . . .	7
<b>3. RESULTADOS Y DISCUSIÓN</b>	<b>9</b>
3.1. Evolución de la función SFR . . . . .	9
3.2. Evolución de la SFR en función de la masa de las galaxias . . . . .	12
3.3. El 10 % de galaxias modelo más masivas a altos <i>redshifts</i> . . . . .	14
3.4. Galaxias extremadamente masivas a altos <i>redshifts</i> . . . . .	15
<b>4. RESUMEN Y CONCLUSIONES</b>	<b>17</b>
<b>A. Peticiones SQL</b>	<b>20</b>
<b>B. Código de análisis</b>	<b>21</b>

## 1. INTRODUCCIÓN

El modelo cosmológico estándar actual describe la evolución del Universo  $\Lambda$ CDM, el cual está dominado por energía oscura representada por una constante cosmológica  $\Lambda$  y por materia oscura fría (CDM, sus siglas en inglés, *Cold Dark Matter*). En este escenario, las grandes estructuras se forman de manera jerárquica por agregación, es decir, las estructuras más pequeñas se van fusionando entre ellas hasta formar estructuras más grandes. Esta construcción jerárquica de la estructura de las galaxias se debe al crecimiento lineal de fluctuaciones de materia oscura debido a inestabilidades gravitacionales. Las fluctuaciones iniciales siguen una distribución aleatoria gaussiana. Cuando llegan a un valor crítico de densidad de materia, terminan por colapsar localmente formando lo que se denomina halos de materia oscura (e.g. Vega-Martínez, 2017).

Sin embargo, la información extragaláctica que se puede medir proviene de la componente visible del Universo, por lo que se necesita una relación entre esta componente bariónica (estrellas, gas y polvo) y los halos de materia oscura para poder entender, modelar, predecir y comparar con las observaciones los procesos físicos que hay detrás de la formación y evolución de las galaxias. Es decir, además de los halos de materia oscura, hay que considerar otros elementos importantes para la construcción de galaxias de forma jerárquica, en muchas ocasiones, dominados por procesos astrofísicos violentos. Como la física detrás de estos fenómenos no se entiende en profundidad y se trata de sistemas complejos, es necesario considerar las simulaciones numéricas como herramienta teórica para comprender la evolución de las galaxias (e.g. Knebe et al., 2017).

En las últimas décadas se han desarrollado dos estrategias para abarcar estos procesos físicos complejos: las simulaciones numéricas de N-cuerpos y los modelos de formación y evolución de galaxias, dos métodos complementarios para conseguir una comprensión de los procesos astrofísicos detrás de la formación y evolución de galaxias. En este trabajo se utilizarán modelos semi-analíticos, que abordan la formación y evolución de las galaxias a través de un conjunto de ecuaciones diferenciales.

El objetivo de este informe es estudiar de manera teórica la formación y evolución de galaxias

ya que es algo necesario a la hora de entender la historia evolutiva del Universo. Una galaxia es un conjunto de estrellas, nubes de gas, planetas, polvo cósmico, materia oscura, energía oscura, nebulosas y otros cuerpos celestes que permanecen unidos entre sí debido a la interacción gravitacional (AstroMia, 2021). La masa visible de las galaxias proviene de las estrellas, y por esta razón, la variable elegida para el estudio ha sido la tasa de formación estelar. En astrofísica, tanto el estudio de galaxias con alta formación estelar en distintas épocas, como el estudio de galaxias que ya no estén formando estrellas es un tema de investigación clave (e.g. Sherman et al., 2021). La tasa de formación estelar, o *Star Formation Rate (SFR)* en inglés, es la masa total de estrellas formadas por año, por lo que sus unidades habituales son  $M_{\odot}yr^{-1}$ , y se trata de uno de los parámetros fundamentales para estudiar la formación de estrellas en las galaxias<sup>1</sup>. No todas las galaxias forman estrellas al mismo ritmo, la tasa de formación estelar depende principalmente de la masa estelar de la galaxia.

En este proyecto se estudia la tasa de formación estelar de las galaxias a través los modelos semi-analíticos que utiliza la simulación MDPL2 de MultiDark como condiciones iniciales donde comenzar a crear galaxias. El objetivo principal del proyecto es comparar los tres modelos semi-analíticos que existen dentro de esta simulación, SAGE, SAG y GALACTICUS con los datos observacionales, para entender mejor cómo evoluciona la formación estelar galáctica. Existen algunos estudios (Guarnieri et al., 2019; Alcalde Pampliega et al., 2019) que aseguran la existencia de galaxias muy masivas a tiempos cósmicos lejanos que no están formando estrellas. Al tratarse de un número pequeño de galaxias, se convierten en objetos raros y esto hace que su estudio mediante modelos semi-analíticos sea muy interesante, por lo que, se propone estudiar también cómo se comporta el modelo semi-analítico a la hora de buscar estos objetos raros y por tanto, ver si es posible modelizarlos para poder entender sus historias de formación.

Para ello, se va a dividir el informe en tres bloques principales. En la sección de métodos (Sec. 2), tras una pequeña introducción sobre qué es el *redshift* y su relación con la edad del Universo, se explica en detalle que son las simulaciones de N-cuerpos y cómo son utilizadas por los

<sup>1</sup>En este informe se expresará en  $M_{\odot}yr^{-1}h^{-1}$  ya que son las unidades proporcionadas por la base de datos utilizada, para que todo sea consistente. La constante de Hubble se expresa en unidades de  $h$  como  $H = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ , es decir,  $h$  es una expresión de nuestra ignorancia cosmológica

modelos semi-analíticos como base para la creación de sus galaxias. En esta sección también son detallados la simulación (MDPL2) y los modelos semi-analíticos (SAGE, SAG y GALACTICUS) utilizados en este proyecto. A continuación, hay una sección de resultados (Sec. 3), donde primero se realiza una comparativa de la tasa de formación estelar, a distintas épocas cósmicas, dada por los modelos semi-analíticos con los datos observacionales de Gruppioni et al. (2015). De esta manera se va a poder comprobar la afinidad de los datos teóricos con los observados para ésta cantidad. Después, se realiza un estudio más en profundidad de cómo funciona el modelo a tiempos cósmicos lejanos con galaxias muy masivas, (es decir, con un corte de la masa por encima de  $10^{11} M_{\odot} h^{-1}$  (Guarnieri et al., 2019)). Por último, en la Sección 4, se pueden encontrar la discusión y las conclusiones a las que se han llegado.

## 2. MÉTODOS

Las simulaciones numéricas de N-cuerpos, solucionan numéricamente de manera directa ecuaciones no-lineales, es decir, son capaces de describir procesos físicos propios de la formación de galaxias, pero a una escala pequeña, de manera que consideran las galaxias como estructuras complejas. Por otro lado, los modelos semi-analíticos, en particular los que se van a tratar en este proyecto, consisten en la construcción de un conjunto de aproximaciones analíticas con la finalidad de describir la misma física, pero a una escala espacial mayor, es decir, permiten estudiar un amplio rango de parámetros en volúmenes comparables al Universo observable. Se utiliza el problema de N-cuerpos para simular la “red cósmica” de materia oscura y los modelos semi-analíticos introducen gas en estos halos de materia oscura a alto *redshift* y lo dejan evolucionar siguiendo ecuaciones analíticas para poder simular también qué ocurre con la componente bariónica del Universo (Vega-Martínez, 2017).

En esta sección se explican estos dos métodos computacionales tras hacer una acotación sobre la edad del Universo y su relación con el *redshift*. Al final se explican las características de los modelos utilizados para el estudio de este proyecto.

### 2.1 La edad del universo

Al observar espectros de galaxias lejanas, las líneas espectrales están desplazadas con respecto

a las observadas, esto se conoce como el desplazamiento Doppler. Es un fenómeno donde para un observador, la frecuencia, y por tanto, la longitud de onda de una fuente de ondas que se mueve con una velocidad relativa respecto a él cambia, a pesar de que la frecuencia de la fuente no se modifica (Gray & Dunning-Davies, 2008). Si la fuente de ondas se está alejando se observa un desplazamiento al rojo, *redshift* en inglés, simbolizado por  $z$ , es decir, un desplazamiento hacia frecuencias más bajas (longitudes de onda más altas). En cambio si la fuente de ondas se aproxima al observador existe un desplazamiento al azul, es decir a frecuencias más altas (longitudes de onda más bajas) (Gray & Dunning-Davies, 2008).

El *redshift* está directamente relacionado con el factor de expansión del universo, a través de una relación entre este, y la longitud de onda emitida ( $\lambda_e$ ) y observada ( $\lambda_0$ ) del fotón:  $\lambda_0 = \lambda_e(1 + z)$ .

Hubble descubrió que la relación que existe entre el *redshift*, y la velocidad  $v$  de la fuente con respecto al observador viene dada por  $z = v/c$  donde  $c$  es la velocidad de la luz en el vacío, lo que quiere decir que el *redshift* de una luz que proviene de un objeto, como puede ser una galaxia, es proporcional a la distancia ( $d$ ) del mismo con respecto al observador:  $z = \frac{H_0 d}{c}$  (Theuns, 2016), donde  $H_0$  es la constante de Hubble. Esto significa que tanto el *redshift*, como el tiempo que lleva el Universo expandiéndose, como la distancia a la que se encuentran los objetos en expansión están proporcionalmente relacionados; es decir, cuando se habla de *redshift* se está hablando de lo lejos que está el objeto de estudio, tanto espacial como temporalmente.

### 2.2 Simulación de materia oscura con técnica de N-cuerpos.

El problema de los N-cuerpos es un método computacional que consiste en un conjunto de ecuaciones diferenciales ordinarias no lineales difíciles de resolver a pesar de tener un planteamiento sencillo, por lo que es útil en varias disciplinas entre las que se encuentran los sistemas dinámicos y teoría egódica, geometría, topología, cálculo de variaciones y teoría de Morse, análisis numérico y ciencias de la computación (Gil Bor, 2014). El problema de los N-cuerpos pretende estudiar los movimientos individuales de un grupo de corpúsculos de masa dada que interactúan mutuamente según las leyes de gravitación universal de Newton. En astrofísica, se denominan partículas a las identidades de más baja masa conside-

radas en una simulación de N-cuerpos, que suele estar varios órdenes de magnitud por encima de la masa del Sol. Por tanto, estas partículas tienen una masa mínima que proporciona la resolución de la simulación, forman una red similar a las redes cristalinas con defectos y solo interactúan entre ellas gravitacionalmente. Las simulaciones de N-cuerpos son independientes de la masa que se esté estudiando ya que consideran que todas las partículas son iguales. A cualquier perturbación, se van a sentir atraídas o repelidas unas con otras de manera que, se van a juntar más en los lugares donde haya más densidad y van a alejarse más de las zonas más vacías. Debido a que a grandes escalas, la distribución de materia en el Universo es homogénea e isótropa, tratarla como un problema de N-cuerpos es una buena aproximación.

Para las configuraciones con más de dos cuerpos, los problemas de N-cuerpos deben ser simulados utilizando métodos numéricos. La materia oscura fría es el componente de masa dominante en el Universo y la pérdida de energía a través de interacciones electromagnéticas que generen radiación no tiene cabida, ya que las únicas fuerzas válidas para el modelo de materia oscura fría son las gravitacionales y las débiles. Por tanto se toma como referencia para formar las partículas elementales que interactuarán solo gravitacionalmente. De esta manera, el fluido acolisional de materia oscura se puede representar como un grupo de partículas puntuales discretas, todas con la misma masa. Esta representación como un sistema de N-cuerpos es una aproximación que mejorará según se aumente el número de partículas en la simulación. Las simulaciones de N-cuerpos de los universos CDM tienen el tamaño y la calidad suficientes como para que sea posible que la modelización de la formación de galaxias en un volumen coincida bastante bien con los estudios actuales, por lo que proporcionan una buena comprensión de la formación de estructuras (Vega-Martínez, 2017). De esta manera, las simulaciones de materia oscura definen la llamada “red cósmica”, determinando la formación de halos y la estructura a gran escala del Universo. Es decir, esta “red cósmica” describe la topología del Universo  $\Lambda$ CDM.

Las simulaciones de materia oscura son insuficientes para estudiar la formación de galaxias, ya que hace falta incluir las componentes bariónicas y sus procesos. Para ello, se utilizan técnicas semi-analíticas, pero la aplicación de modelos que utilizan estas técnicas requieren de la información de los halos de materia oscura que normalmente

es obtenida de una simulación. En otras palabras, las simulaciones de materia oscura requieren el empleo de modelos semi-analíticos capaces de conectar las propiedades observadas con la distribución de materia oscura.

Para las simulaciones con muchas partículas se requieren cálculos a gran escala en tiempos de ejecución lentos, por lo que se han desarrollado métodos aproximados que reducen la complejidad de los cálculos, como pueden ser los métodos variacionales y la teoría de perturbaciones, gracias a los cuales se pueden producir trayectorias analíticas aproximadas y de esta manera corregirse, además, la integración numérica.

Las simulaciones numéricas son, por tanto, la principal herramienta teórica para estudiar el crecimiento gravitatorio no lineal de la estructura del universo, y para relacionar las condiciones iniciales de materia oscura fría (CDM) con las observaciones de las galaxias hoy en día. La construcción jerárquica de la estructura del universo con sus tres dinámicas no es resoluble analíticamente, además, los rangos dinámicos en cuanto a volumen y masas son enormes, de muchos órdenes de magnitud. Esto hace que sin las simulaciones numéricas directas, esta construcción jerárquica de la estructura del universo sea muy complicada. Combinando estas simulaciones con técnicas que existen para estudiar la formación y evolución de galaxias, se pueden predecir las posiciones, velocidades y propiedades intrínsecas de todas las galaxias, lo que permite establecer vínculos de evolución entre diferentes objetos observados en distintas épocas.

La manera de almacenar la información de los halos de materia oscura de las simulaciones, para después ser utilizada por los modelos semi-analíticos, es a través de una serie de instantáneas (*snapshot* en inglés) de la evolución del Universo en tiempo cósmico, cada cual proporciona distribuciones de partículas caracterizadas por su masa, posición y velocidad, obtenidas como resultado directo de estas simulaciones. En las instantáneas se realiza una búsqueda sobre las densidades para encontrar cada halo de materia oscura que existe a *redshift* 0, en el momento actual, y que excede un cierto límite de masa y se establece un criterio de ligadura energética para construir un sistema. Un halo descendiente (halo más reciente) es aquel que comparte más del 50 % de las partículas con su progenitor (halo más antiguo) (Vega-Martínez, 2017).

Se generan los llamados “árboles de fusiones” de halos utilizados por los modelos semi-analíti-

cos como condiciones iniciales para comenzar a simular la evolución de las galaxias. Estos árboles contienen nodos cada uno correspondiente a un sub-halo de materia oscura y a sus galaxias asociadas. Como se ilustra en la Fig. 1, una vez el árbol esta construido, se recorren sus nodos de acuerdo a la primera búsqueda profunda y se le asocia a cada nodo un rango el cual sirve para la identificación de la raíz del árbol, lo que permite asociar una única identificación para cada halo, y permitiendo así el acceso rápido al árbol completo de cada halo.

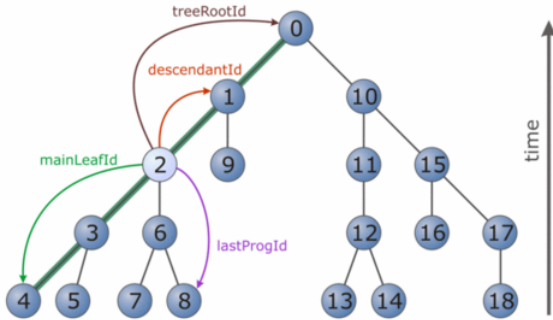


Figura 1: Árbol de fusiones: El nodo más alto (la raíz), representa un halo a  $z = 0$ . Desde él salen ramas que van hacia atrás en el tiempo hacia sus progenitoras, es decir, la línea de tiempo en el esquema va de abajo a arriba. Los números en cada nodo indican el orden de la búsqueda en profundidad, con los progenitores más masivos en el lado más a la izquierda de cada sub-árbol. Estos forman la rama principal (por ejemplo, la línea gruesa verde que parte de la raíz del árbol (0)) del nodo correspondiente. (Jiang & van den Bosch, 2014).

### 2.3 Modelos de galaxias.

Los modelos semi-analíticos han probado ser una herramienta muy potente para el avance de la teoría de la formación de galaxias a pesar de que sigue sin entenderse mucho la física detallada de los procesos de formación estelar y su regulación por retroalimentación. Tratan los procesos físicos mediante técnicas analíticas aproximadas, simulan la evolución del contenido bariónico excluyendo la caracterización y la evolución de los halos de materia oscura, ya que esta información suele estar proporcionada por las simulaciones de N-cuerpos y utilizada como condición inicial.

La información presentada en esta subsección es un resumen del artículo de revisión Baugh (2006) y tiene el fin de explicar algunos de los parámetros que se introducen como condiciones en los modelos semi-analíticos para entender mejor su funcionamiento. Además, se compara como son introducidos estos parámetros en los tres mo-

delos semi-analíticos con los que se trabaja en este proyecto, SAGE, SAG y GALACTICUS.

Para monitorizar la formación de galaxias en la simulación cosmológica, el modelo semi-analítico traza las historias de fusión almacenada en forma de “árboles de fusión”, hace que las galaxias pueblen estos halos. Esta metodología permite probar, durante el post-procesado, diferentes enfoques del enfriamiento del gas, formación estelar, crecimiento de los núcleos activos de las galaxias, retroalimentación, enriquecimiento químico, etc. Esta técnica, aporta posiciones precisas y velocidades particulares para todas las galaxias modelizadas. También permite seguir la historia de evolución de objetos individuales y por consiguiente investigar la relación entre las poblaciones en diferentes épocas. El hecho de que los modelos semi-analíticos establezcan este tipo de conexiones evolutivas hace que este tipo de modelización sea muy potente para interpretar datos observacionales.

El término “semi-analítico” es debido a que los modelos se forman mediante ecuaciones analíticas con parámetros libres que se establecen empíricamente encontrando los mejores ajustes a ciertos datos observacionales. Los modelos semi-analíticos son, por tanto, interpretados como simulaciones simplificadas de procesos de formación de galaxias, donde la formación de estrellas y su regulación mediante procesos de retroalimentación es parametrizado en términos de modelos analíticos físicos simples. Estos modelos relacionan las propiedades de las galaxias que pueblan cada árbol jerárquico de fusiones mediante la resolución de **ecuaciones diferenciales** durante el tiempo de evolución de las mismas. Estas ecuaciones describen el enfriamiento radiactivo del gas, la formación estelar, el crecimiento de agujeros negros supermasivos, procesos de retroalimentación mediante supernovas y AGN, etc (procesos que se detallan al final de esta sección), y vienen dadas por estudios previos, ya sea de observaciones o de otros modelos. Por tanto, modelizar mediante modelos semi-analíticos también sirve para poner a prueba hipótesis sobre los procesos físicos que proporcionan las propiedades de la formación de galaxias, lo que hace que estos modelos puedan ser considerados como laboratorios, es decir, como una ciencia experimental.

Los modelos semi-analíticos modelan muchos procesos físicos, por tanto, contienen una cantidad de parámetros libres muy grande, que para muchos procesos no están bien acotados observacionalmente. La mayoría de estos parámetros

son físicos y se introducen para poder ajustar un amplio rango de observaciones.

Las simulaciones cosmológicas siempre se han esforzado en aumentar el tamaño (y por tanto, la resolución) de la computación de los N-cuerpos, aprovechando todos los algoritmos numéricos y *hardware* avanzados. Los modelos semi-analíticos tienen una demanda computacional mucho más baja que las simulaciones hidrodinámicas, lo que ha permitido la construcción de catálogos de galaxias más grandes y una exploración más rápida del espacio de parámetros libres.

Ya se ha discutido que los halos de materia oscura son la cuna de la formación de galaxias, por lo que los modelos de formación de galaxias de forma jerárquica requieren saber la abundancia de halos de diferentes masas, la historia de formación de cada halo, es decir, el “árbol de fusiones”, y la forma de la estructura interna del halo, en términos de su densidad radial y de su momento angular.

A continuación se detallan las componentes y los procesos físicos bariónicos básicos que se introducen en los modelos semi-analíticos de manera que sean capaces de reproducir la formación y la evolución de las galaxias.

### 2.3.1 El enfriamiento del gas

El enfriamiento del gas está en el centro del proceso de la formación de galaxias ya que establece el ritmo al cual puede comenzar la formación de estrellas. El gas puede enfriarse mediante muchos mecanismos, normalmente siguiendo las etapas mostradas en la Fig. 2. Los más importantes para la formación de galaxias son la emisión de fotones tras transiciones entre niveles de energía y la radiación de frenado (radiación libre-libre) debido a que los electrones son acelerados en un plasma ionizado. Dos de los modelos utilizados en este proyecto, GALACTICUS y SAG, tratan el enfriamiento del gas de manera parecida, la diferencia se encuentra en el código que utilizan a la hora de modelar la dependencia del enfriamiento con la metalicidad (Knebe et al., 2017). Por su parte, la forma de modelizar el enfriamiento del gas que tiene el tercer modelo utilizado, SAGE, tiene alguna modificación más en la que no se entra en detalle debido a que para este estudio no es relevante.

Existen algunos fenómenos, que dependieron del enfriamiento del gas en el pasado, que podrían servir para modificar el modelo estándar de enfriamiento de numerosas maneras, lo que podría afectar al ritmo al que el gas se enfría tanto en

halos de alta masa como en halos de baja masa. Estos fenómenos pueden considerarse como procesos de retroalimentación que regulan el ritmo al cual puede ocurrir la formación de estrellas.

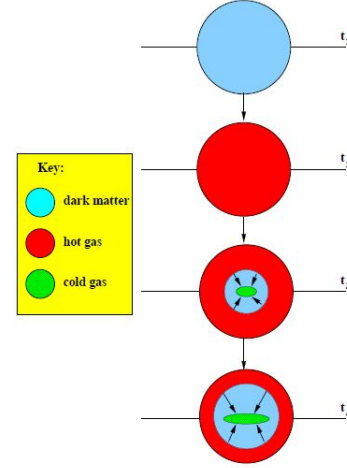


Figura 2: Etapas de enfriamiento del gas:  $t_1$ : El gas tiene la misma distribución espacial que la materia oscura.  $t_2$ : Se crea un halo de gas caliente soportado por la presión del mismo.  $t_3$ : según el gas se enfría, la presión descende y entra en caída libre cayendo al centro del halo.  $t_4$ : El gas que ha tenido el tiempo suficiente para enfriarse desde la formación del halo se encuentra en el radio de enfriamiento. Este radio va creciendo hasta que el halo de gas caliente se enfría del todo o hasta que la fusión con otro halo de lugar a la formación de un nuevo halo. (Baugh, 2006)

### 2.3.2 Formación estelar

La manera de programar la formación estelar de los tres modelos utilizados en este proyecto es completamente distinta (Knebe et al., 2017). GALACTICUS asume que el gas frío de cada galaxia está distribuido según una distribución exponencial radial, y el tamaño de esta distribución se calcula a partir del momento angular del disco resolviendo para el radio de equilibrio del potencial gravitacional del sistema disco + bulbo + materia oscura. En SAG se empiezan a formar estrellas cuando la masa del gas frío del disco supera un límite crítico. Y por último, SAGE calcula la masa del gas frío que está por encima de una densidad superficial crítica para la formación estelar y las estrellas se forman a partir de este gas (Knebe et al., 2017).

### 2.3.3 Procesos de retroalimentación

Los procesos de retroalimentación son uno de los fenómenos más difíciles de modelizar, además de ser los que mayor impacto tienen en la forma de las predicciones teóricas de las propiedades

de las galaxias. Una sobreproducción de galaxias brillantes es un problema que ha persistido en los modelos de formación de galaxias de manera jerárquica durante mucho tiempo (Baugh, 2006).

En el caso de este proyecto los tres modelos incluyen retroalimentación del núcleo activo de las galaxias causada por la acreción de gas en el agujero negro central. GALACTICUS y SAG son modelos con retroalimentación radial causada por la acreción de gas en procesos de enfriamiento del halo caliente al agujero negro central. SAGE además presenta vientos de quásares como forma de retroalimentación.

#### 2.3.4 Evolución química

La formación de estrellas elimina gas frío pero además también cambia el contenido metálico del medio interestelar de una galaxia. Según las estrellas evolucionan, devuelven material con una metalicidad mayor al medio interestelar. Este mecanismo de devolución se produce mediante vientos estelares o explosiones de supernova, lo que produce a su vez una pérdida de masa de la estrella (Baugh, 2006).

Para programar el tratamiento metálico, GALACTICUS utiliza la aproximación de reciclaje instantáneo, con una fracción de reciclaje de 0.46 y una cesión de 0.035 (Knebe et al., 2017). GALACTICUS asume que los metales están completamente mezclados en todas las fases de manera que puede rastrear todos los flujos de masas entre fases. En el modelo químico introducido en SAG, las estrellas con diferentes rangos de masas pueden contaminar tanto el gas frío como el gas caliente debido a la pérdida de masa durante la evolución estelar y la eyección de metales al final de sus vidas. En cambio, SAGE, utiliza un tratamiento metálico simple donde se producen metales en cada evento de formación de estrellas y se reciclan los metales de las estrellas de vida corta de forma instantánea devolviéndose al gas frío.

#### 2.3.5 Fusión de galaxias

El impacto de una fusión de galaxias se cuantifica a partir de la proporción de masa de la galaxia satélite acretada con respecto a la masa de la galaxia central (Baugh, 2006). Los modeladores semi-analíticos tratan como parámetro en sus simulaciones numéricas la fracción de masa que establece el umbral a partir del cual se puede considerar que una fusión sea violenta o importante (Baugh, 2006), como es el caso de los modelos tres

modelos utilizados en este proyecto. SAG y SAGE establecen el umbral para separar las fusiones grandes de las pequeñas en un ratio de masas de 0.3, mientras que GALACTICUS lo establece en 0.25 (Knebe et al., 2017).

#### 2.3.6 Tamaño de las galaxias

Otro parámetro importante en los modelos de formación estelar es la estimación del tamaño del disco galáctico o del bulbo, ya que, una escala del tamaño del disco es necesaria para calcular la profundidad óptica y así poder calcular la extinción por polvo de la luz de las estrellas (Baugh, 2006) lo que condiciona la interpretación de la formación estelar de las galaxias. Esto convierte a este parámetro en uno de los más importantes para el estudio de este proyecto. Una comparación de las predicciones de los modelos de los tamaños de los discos y bulbos de las galaxias con las observaciones puede limitar en los modelos los procesos de enfriamiento y fusión del gas y por tanto, los procesos de formación de estrellas. Por estas razones, es importante que los modelos utilizados en este proyecto establezcan una buena estimación del tamaño de las galaxias, principalmente porque condiciona los procesos de formación estelar, pero también porque en el estudio se realiza un corte de masas, lo que hace necesario que los modelos sean precisos. La masa de los halos de materia oscura de la simulación utilizada en este proyecto, no está bien definida cuantitativamente por lo que existen varias definiciones. Los modelos presentados aplican dos definiciones de masas para definir los halos de materia oscura que forman sus árboles de fusiones (Knebe et al., 2017).

### 2.4 Descripción de la simulación MultiDark y los modelos semi-analíticos utilizados.

Para el estudio de este proyecto, se van a estudiar tres modelos semi-analíticos: SAGE, SAG y GALACTICUS que pueblan con galaxias las simulaciones de materia oscura del proyecto español MultiDark disponibles de manera pública a través de la plataforma de la base de datos CosmoSim. Dentro de todas las simulaciones dentro de MultiDark, SAG, SAGE y GALACTICUS, utilizan la simulación MDPL2 (siglas de *MultiDark PLank*, versión 2), que pertenece a las series de las simulaciones de Multidark con la cosmología de Plank (Klypin et al. (2016) y Riebe et al. (2013)). Esta simulación utiliza  $3840^3$  partículas, con una masa de  $1.5 \times 10^9 M_\odot h^{-1}$ , y empezó a *redshift*  $z = 120$ .



La cosmología utilizada es la de un universo  $\Lambda$ CDM plano son los parámetros cosmológicos de Planck:  $\Omega_m = \Omega_{dm} + \Omega_b = 0.31$ ,  $\Omega_b = 0.048$ ,  $\Omega_\Lambda = 0.69$ ,  $h = 0.6777$ ,  $n_s = 0.96$  y  $\sigma_8 = 0.82$ . Donde  $\Omega_m$  es el parámetro de densidad total de materia, la suma de la materia oscura ( $\Omega_{dm}$ ) y de la densidad de bariones ( $\Omega_b$ ) en la actualidad.  $\Omega_\Lambda$  denota la densidad de energía oscura en la actualidad. Por otro lado, el espectro de potencias utilizado para generar las condiciones iniciales está descrito por el valor de  $\sigma_8$  y con un índice espectral, la normalización del espectro de potencias,  $n_s$ . El volumen de la simulación es un cubo de lado  $1\text{Gpc } h^{-1}$ , el cual es suficientemente grande como para incluir objetos raros interesantes.

Por otro lado, las partículas de materia oscura que dan la resolución de la simulación, tienen una masa de  $1.5 \times 10^9 M_\odot h^{-1}$ . Esta es la masa mínima que va a ser encontrada en la simulación. Existe una regla empírica que considera que la masa mínima tomada generalmente para el halo de materia oscura de una galaxia es de aproximadamente 20 veces el valor mínimo de resolución, ya que es la masa mínima requerida que es considerada para poder estudiar un objeto. La masa mínima de materia oscura considerada en este informe será por tanto  $m = 20 \times 1.51 \times 10^9 = 3.02 \times 10^{10} M_\odot h^{-1}$ . Teniendo en cuenta esto, y tomando como referencia el artículo Knebe et al. (2017), y en particular la Fig. 8 de este, donde se representa el ratio de la masa estelar frente a la masa del halo de materia oscura, en este informe se realizará el corte de masa estelar de las galaxias en  $10^{8.5} M_\odot h^{-1}$ .

La resolución de la fuerza gravitatoria varía desde  $13\text{kpc } h^{-1}$  a altos *redshifts* hasta  $5\text{kpc } h^{-1}$  a bajos *redshifts*. Los catálogos están divididos en 126 *snapshots* entre los *redshifts*  $z = 17$  y  $z = 0$ . Estos *snapshots*, son el conjunto de instantáneas generado tras la simulación, cada instantánea recoge la distribución de partículas y sus velocidades en las diferentes épocas. Sin embargo, para modelar la evolución de las galaxias, es necesaria la información de los halos de materia oscura y de los correspondientes árboles de fusiones que hay en cada instantánea. Los halos y los sub-halos son identificados con ROCKSTAR y los árboles de fusiones construidos con CONSISTENT TREES (Knebe et al., 2017). Todos los modelos siguen y buscan el origen de las subestructuras a partir de la simulación de los N-cuerpos proporcionada por MultiDark, y se ha demostrado que los catálogos resultantes pueden ser utilizados de manera coherente por los modelos semi-analíticos ya que

cumplen los estándares mínimos necesarios (Knebe et al., 2017).

Por tanto, como la mayoría de los modelos semi-analíticos, los tres modelos utilizados, operan con árboles de fusión y halos de materia oscura, las galaxias se forman potencialmente dentro de cada rama de cada árbol de fusión y se definen por un conjunto de propiedades. Algunas de estas propiedades se determinan por métodos directos por la simulación de N-cuerpos, como puede ser la posición del halo o su velocidad, la mayoría de las restantes se establecen a partir de un conjunto de ecuaciones diferenciales, pero esta evolución diferencial a veces se ve interrumpida por eventos fortuitos como puede ser la fusión de dos galaxias.

Como conclusión a estos datos, un catálogo de galaxias para una simulación entera, que contiene típicamente  $\sim 2 \times 10^6$  galaxias juntas con todas sus historias a  $z = 0$ , puede ser construido por cualquier modelo semi-analítico que se desee en unas pocas horas trabajando con ordenadores de alta calidad.

Como ya se ha mencionado al inicio de la sección, dentro de la simulación MDPL2 se encuentran los tres modelos semi-analíticos diferentes: SAG, SAGE y GALACTICUS. A pesar de que el objetivo de cada modelo semi-analítico es el de modelizar la formación de galaxias, es importante tener en cuenta que los modelos han podido ser calibrados para atender a diferentes propósitos, por tanto, existen, naturalmente, variaciones en los catálogos de galaxias entre modelos, debido a las diferentes aproximaciones de calibración y al diseño e implementación de los fenómenos físicos actuales. De hecho, es importante tener varios modelos de formación de galaxias disponibles explorando diferentes aproximaciones de la física de la formación de galaxias. En este caso, SAG, es un modelo potente aportando fracciones de gas razonables y relaciones de metalicidad; SAGE se ajusta a múltiples observaciones simultáneamente, la primera y la más importante, la función de masa estelar y la relación masa estelar - masa del halo, y el punto fuerte de GALACTICUS se encuentra en la función de tasa de formación estelar y su evolución (Knebe et al., 2017).

Todos los catálogos de galaxias están disponibles públicamente en la base de datos de CosmoSim, que proporciona acceso a los datos cosmológicos a través de la *interfaz* de la misma vía el lenguaje de consulta estructurada SQL y pueden ser descargados de manera directa. Sin embargo, en la realización del proyecto, hubo problemas por parte de la base de datos a la hora de proporcio-



nar todos los datos demandados de SAG y de GALACTICUS, la mayoría de ellos no fueron posible de adquirir. Las peticiones utilizadas para realizar este proyecto y los problemas procedentes de la base de datos están explicados con más detalle en el Apéndice A.

### 3. RESULTADOS Y DISCUSIÓN

En esta sección se muestran y se comparan los datos que proporcionan los modelos semi-analíticos con los datos observados. Fundamentalmente se va a hablar del modelo semi-analítico SAGE debido a los problemas que tuvo la base de datos a la hora de proporcionar los datos de los otros modelos. No obstante, los pocos datos que se han obtenido de GALACTICUS serán también mostrados y comparados.

#### 3.1 Evolución de la función SFR

En esta subsección se estudia la función de tasa de formación estelar (SFR), es decir, la densidad numérica por unidad de volumen de la tasa de formación estelar y su evolución.

En la Fig. 3 se muestra la función de la tasa de formación estelar (SFR) del modelo semi-analítico SAGE a distintos *redshifts* en el rango de  $z = 0.142$  a  $z = 5.017$ , es decir, desde aproximadamente el presente hasta el primer giga año de vida del Universo. Mirando la gráfica, lo primero a destacar es la que la forma de la función de SFR es muy parecida en todos los casos, pudiendo describirse con dos rectas con pendientes distintas. Al tratarse de una gráfica donde se representa logaritmo frente a logaritmo, estas dos rectas son en realidad, dos leyes de potencias.

Analizando en primer lugar las funciones SFR generadas con los tres *redshifts* más bajos:  $z = 0.142$ ,  $z = 0.523$  y  $z = 1.077$ , esta claro que se trata de rectas paralelas cada una por encima de la anterior en tiempo cósmico, lo que quiere decir que según el *redshift* aumenta, se encuentra mayor número de galaxias formando estrellas. En los tres casos se puede apreciar cómo, después del punto en el que la pendiente de la recta cambia  $\text{SFR} \simeq 10^{1.3} M_{\odot} h^{-1} \text{yr}^{-1}$  para  $z = 0.142$ ,  $\text{SFR} \simeq 10^{1.4} M_{\odot} h^{-1} \text{yr}^{-1}$  para  $z = 0.523$ ,  $\text{SFR} \simeq 10^{1.6} M_{\odot} h^{-1} \text{yr}^{-1}$  para  $z = 1.077$ , hay una caída mucho más grande, es decir, a partir de esos valores de tasa de formación estelar es raro encontrar galaxias formando tantas estrellas.

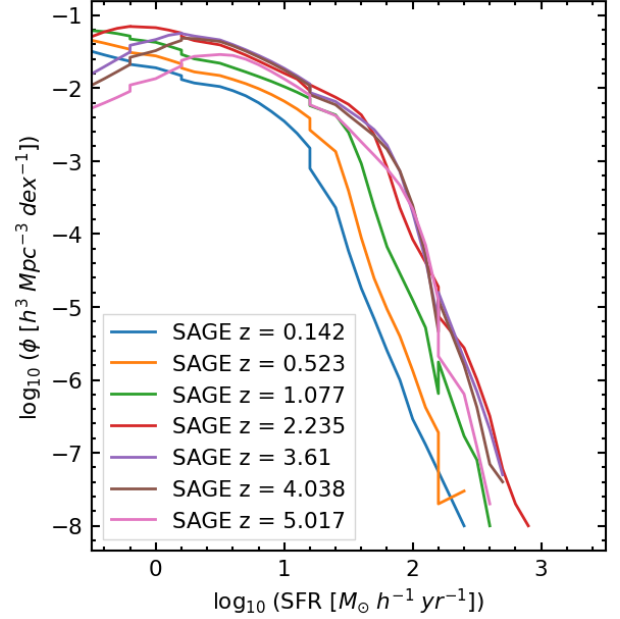


Figura 3: Función de la tasa de formación estelar ( $M_{\odot} h^{-1} \text{yr}^{-1}$ ) del modelo semi-analítico SAGE a distintos *redshifts*. Se pueden distinguir a todos los *redshifts* que la función toma la forma de dos rectas con distintas pendientes, lo que indica un descenso brusco del número de galaxias con alta SFR a partir de un punto determinado.

A *redshifts* más altos ( $z = 2.235$ ,  $z = 3.61$ ,  $z = 4.038$  y  $z = 5.017$ ), a partir de aproximadamente el tercer giga año de vida del Universo, hasta el primero, se puede decir que es más habitual encontrar galaxias con alta formación estelar con este modelo semi-analítico. Sin embargo, se puede ver también que la función de tasa de formación estelar que aumenta con el *redshift* se estabiliza en unos valores determinados, para los tres últimos *redshifts* ( $z = 3.61$ ,  $z = 4.038$  y  $z = 5.017$ ) las curvas prácticamente se superponen a partir de  $\text{SFR} > 10^{0.6} (M_{\odot} h^{-1} \text{yr}^{-1})$ . Es decir, la densidad numérica por unidad de volumen de la SFR tiende a una envolvente a *redshifts* altos. Para estos cuatro valores del *redshift* también existe un cambio en la pendiente de las rectas a partir del cual es más difícil encontrar un número alto de galaxias formando estrellas. Esta vez, todas las curvas, tienen más o menos el mismo punto de inflexión en  $\text{SFR} \simeq 10^{1.7} M_{\odot} h^{-1} \text{yr}^{-1}$  debido a esta estabilización de la densidad numérica por unidad de volumen de la SFR según aumenta el *redshift*.

Por último, cabe destacar una caída a SFR bajas, por debajo de  $\text{SFR} = 10^{0.6} (M_{\odot} h^{-1} \text{yr}^{-1})$ , para los tres *redshifts* más altos estudiados ( $z = 3.61$ ,  $z = 4.038$  y  $z = 5.017$ ). Esto quiere decir que hay menos galaxias formando pocas estrellas, es decir, que hay menor cantidad de galaxias pasivas

los dos primeros giga años del Universo. Estadísticamente, los intervalos del principio son más robustos. Esto puede ser explicado ya que en primera aproximación podría decirse que el número de galaxias totales es constante con el tiempo. A altos *redshifts* hay un mayor número de galaxias formando estrellas con una tasa de formación más alta, por lo que el número de galaxias que hay formando estrellas a baja tasa de formación estelar disminuye.

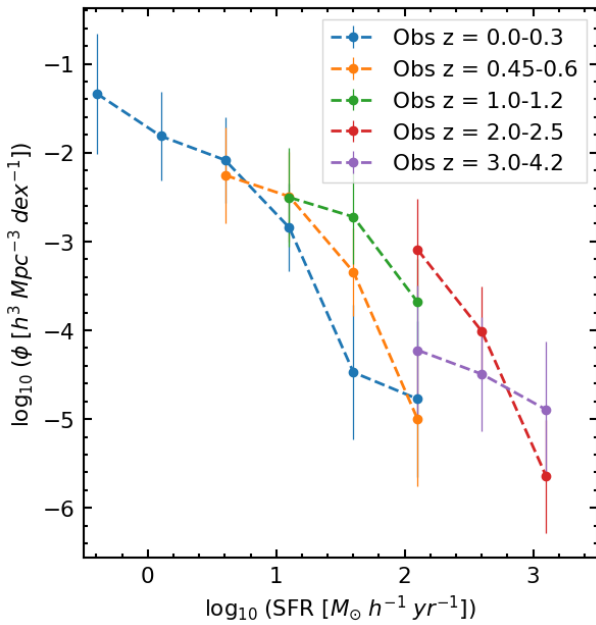


Figura 4: SFR de observaciones a distintos *redshifts*. Se han añadido líneas discontinuas que unen los datos como guía visual para facilitar la lectura y el análisis de la gráfica. No hay suficientes datos para deducir tendencias cuantitativas que puedan ser comparadas con los datos proporcionados por el modelo semi-analítico.

En la Fig. 4 se representa la función de SFR para datos obtenidos observacionalmente a diferentes *redshifts* por Gruppioni et al. (2015) como en el caso del modelo semi-analítico, desde el presente ( $z = 0 - 0.3$ ) hasta el primer giga año y medio del Universo ( $z = 3 - 4.2$ ). Se han añadido unas líneas discontinuas que unen los datos como guía visual para facilitar la lectura y el análisis de la gráfica. Lo primero a destacar en comparación con el modelo es que estos datos son más ruidosos, debido a que estas medidas son complicadas de realizar, sobre todo a la hora de medir bajas tasas de formación estelar. Por esta misma razón no hay suficientes puntos para declarar, como anteriormente, que las funciones se comporten como dos leyes de potencias. Sobre todo a altos *redshifts*, se observa que todos los datos se encuentran solo a altas SFR. Esto quiere decir que

existe la misma tendencia cualitativa en comparación con los datos obtenidos con el modelo semi-analítico SAGE, sin embargo, observacionalmente no hay suficientes datos para deducir tendencias cuantitativas.

En la última parte de esta subsección, se muestra un análisis comparativo, *redshift* por *redshift*, de los datos obtenidos observacionalmente frente a los obtenidos mediante el modelo semi-analítico SAGE a los *redshifts* de los cuales se tienen datos observacionales. También se representan algunos datos obtenidos con el modelo semi-analítico GALACTICUS, pero debido a problemas con la base de datos no pudieron obtenerse más. Se detallan estos problemas en el Apéndice A.

De las Figs. 5(a), 5(b) y 5(c), se puede decir que el modelo semi-analítico se ajusta bien a los datos observados cuando se estudian SFR bajas. Sin embargo, tiene una caída de galaxias formando estrellas con alta SFR (a partir de  $SFR = 10^{1.6} (M_{\odot} h^{-1} yr^{-1})$ ) demasiado rápida, a bajos *redshifts* hay más galaxias formando estrellas con alta SFR, que las predichas por el modelo, difiriendo en los tres casos en torno a un orden de magnitud en los puntos por encima de  $SFR > 10^{1.6} M_{\odot} h^{-1} yr^{-1}$ , lo que lleva a la conclusión de que al modelo semi-analítico SAGE, no tiene suficiente número de galaxias con una alta SFR. Esto podría indicar que este modelo no tiene suficiente gas frío y denso, necesario para formar estrellas, o que su evolución no es adecuada.

Las Figs. 5(d) y 5(e) confirman la existencia de este problema con el modelo semi-analítico SAGE de manera más pronunciada, es decir, a altos *redshifts* ( $z = 2.235$  y  $z = 3.61$ ), el modelo predice menos galaxias formando estrellas con altas SFR que las observadas, esta vez con una diferencia entre los datos observados y los modelados de hasta aproximadamente 2 órdenes de magnitud a partir de SFR por encima de  $10^2 M_{\odot} h^{-1} yr^{-1}$ . Incluso, el último dato observado a *redshift* 3.61, con tasa de formación estelar de  $10^3 M_{\odot} h^{-1} yr^{-1}$ , SAGE no lo modeliza, es decir, no encuentra galaxias a tan alto *redshift* ( $z = 3.61$ ) con tasa de formación estelar tan alta ( $SFR = 10^3 M_{\odot} h^{-1} yr^{-1}$ ). En este rango de *redshifts* no se puede comparar si las galaxias con poca SFR se ajustan bien ya que observacionalmente es muy complicado medir galaxias que están formando pocas estrellas debido a que su luminosidad es menor y por tanto, no se tienen datos.

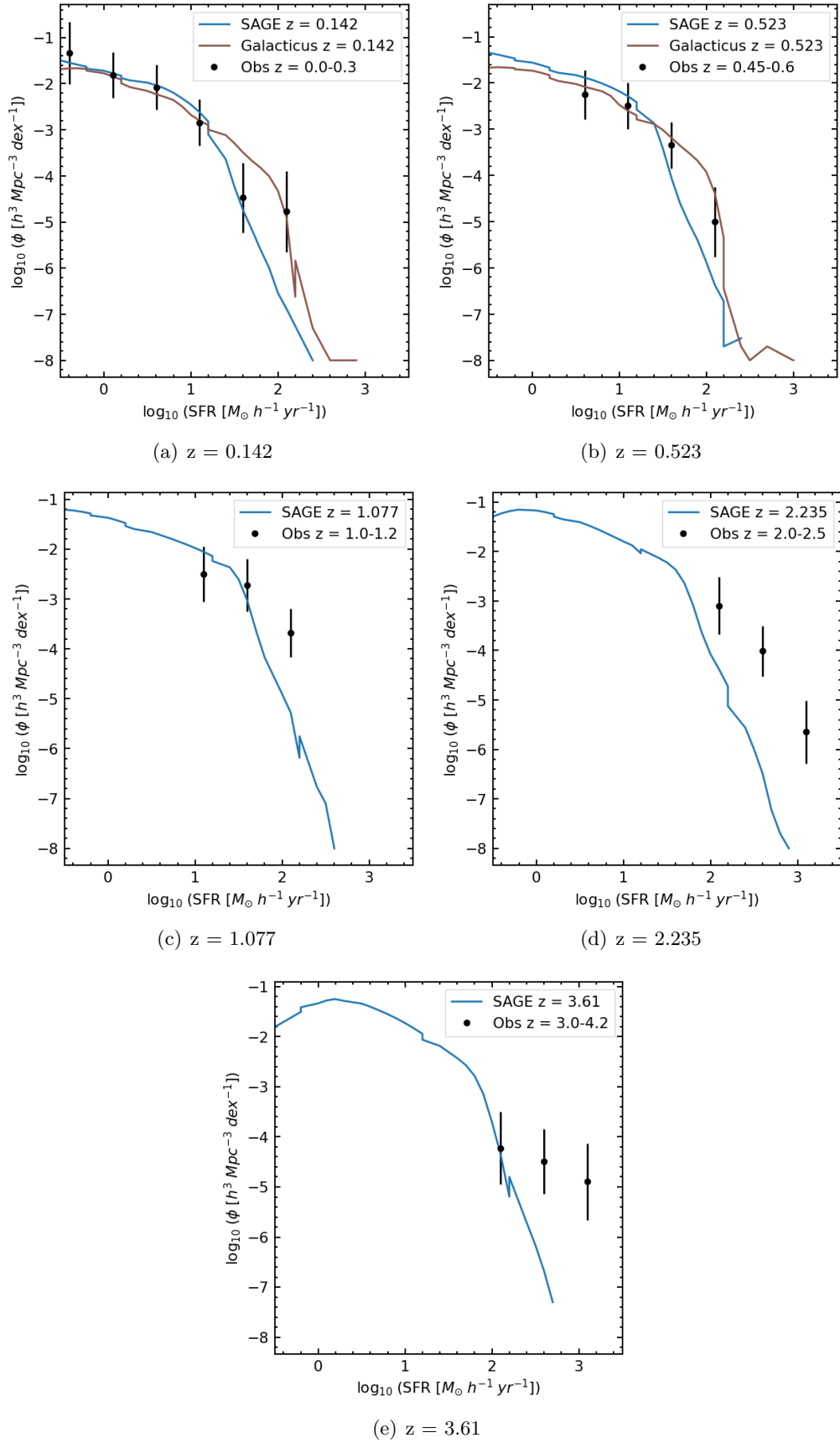


Figura 5: Función de SFR del modelo semi-analítico SAGE (líneas azules) frente a la Función de SFR de las observaciones (puntos negros) representando los redshift por separado. En 5(a) y 5(b) se representa también la función SFR del modelo semi-analítico GALACTICUS (líneas marrones). A bajos *redshifts* los datos obtenidos con SAGE se ajustan bien a los datos observados, pero según aumenta el *redshift*, deja de producir suficientes galaxias con alta SFR. En cambio, se puede ver en 5(a) y 5(b) que GALACTICUS se ajusta mejor a los datos observados que SAGE.

Por otro lado, teniendo en cuenta los datos obtenidos a bajos *redshifts* con el modelo semi-analítico GALACTICUS, en las Figs. 5(a) y 5(b) esta claro que este modelo se ajusta mejor que SAGE para tasas de formación estelar altas (a partir de  $\text{SFR} = 1.6 \log_{10}(M_{\odot} h^{-1} \text{yr}^{-1})$ ). Esta diferencia se debe a que ambos modelos tratan el gas de manera distinta, como ya fue mencionado en la sección de métodos (Sec. 2).

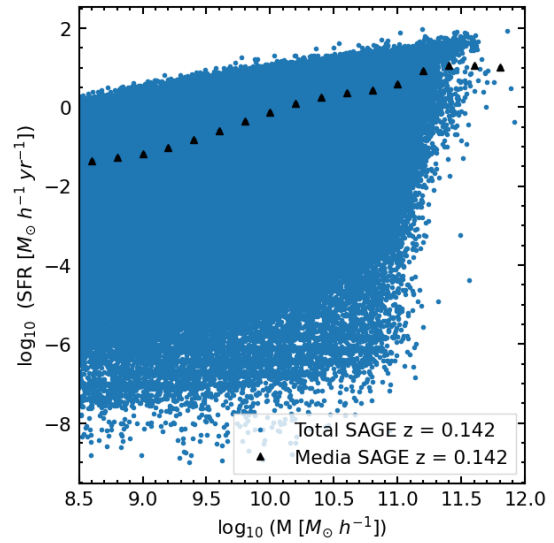
Debido a los problemas con la base de datos CosmoSim por los cuales no pudieron obtenerse datos de los modelos semi-analíticos SAG y GALACTICUS (salvo los mostrados previamente de este último) (Apéndice A), a continuación se discuten los datos obtenidos con estos dos modelos a partir del artículo Knebe et al. (2017). Según este informe, GALACTICUS y SAG tienen más galaxias con alta SFR comparándolos con SAGE. Esto ya ha sido mostrado en las Figs. 5(a) y 5(b) dónde se representan los datos obtenidos con GALACTICUS. Ambos se ajustan a los datos observacionales y por tanto forman estrellas al ritmo adecuado (es decir, el observado), al menos a *redshift* entre 0 y 0.3, que es el rango de *redshift* en el que han sido estudiadas las medidas. Esta similitud entre SAG y GALACTICUS puede ser debida a que estos dos modelos tratan el gas de manera más parecida.

### 3.2 Evolución de la SFR en función de la masa de las galaxias

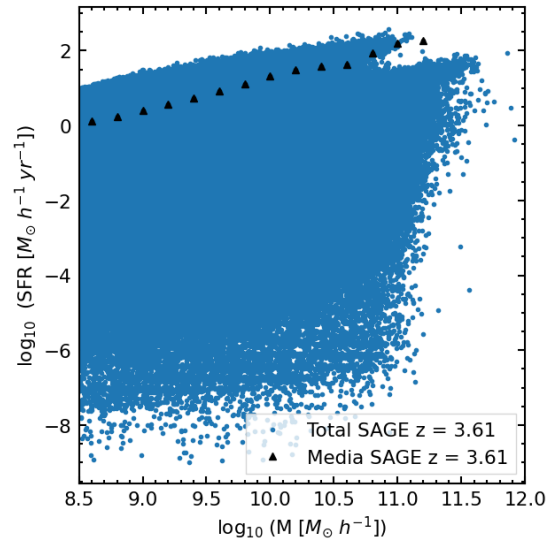
En esta subsección se estudia la SFR con respecto a la masa estelar de las galaxias. Se realiza el corte de las masas en  $10^{8.5}$ , no se muestran masas más pequeñas ya que se acercan mucho al límite de resolución, como ya se ha explicado en la sección 2.

Se representa, mediante círculos azules, en las gráficas de la Fig. 6 la SFR de las galaxias frente a su masa estelar al mínimo y al máximo *redshift* estudiados hasta ahora, que han sido comparados con los datos observacionales, es decir,  $z = 0.142$  y  $z = 3.61$ . A pesar de que se puede intuir una pequeña tendencia creciente, no se puede decir que exista una relación clara entre ambas variables. La tendencia lineal creciente se ve de manera más clara mediante la media de la de tasa de formación estelar frente a los intervalos de masa correspondiente, representada en la Fig. 6 con triángulos negros. Poniendo énfasis a la forma de la función, a la derecha de la gráfica, a masas muy altas (por encima de  $10^{11} M_{\odot} h^{-1}$ ), se puede ver como la dispersión de la función de puntos azules decrece, pasando de tener más de 8 órdenes de magnitud por debajo de  $10^{11} M_{\odot} h^{-1}$  a pasar

a tener aproximadamente 2 por encima de ésta cantidad. Estas gráficas muestran que según aumenta la masa de las galaxias, aumenta también su tasa de formación estelar, cuanto más masiva sea una galaxia más estrellas estará formando a un determinado *redshift*, tanto bajo como alto. Es decir, el modelo semi-analítico SAGE asocia una alta tasa de formación estelar a galaxias muy masivas, lo que está de acuerdo con las observaciones descritas en Popesso et al. (2019), donde se puede ver la misma tendencia creciente en la Fig. 7.



(a)  $z = 0.142$



(b)  $z = 3.61$

Figura 6: Los círculos azules representan la tasa de formación estelar frente a la masa de todas las galaxias. Los triángulos negros representan la media de tasa de formación estelar frente a la masa de las galaxias a *redshifts*  $z = 0.142$  y  $z = 3.61$ . Se aprecia una tendencia creciente de la media, a mayor masa mayor tasa de formación estelar.

Cabe destacar que la media de la tasa de formación estelar frente a los intervalos de masa (los triángulos negros de la Fig. 6) parece indicar que la SFR es proporcional a la masa. Para comprobar esto, en la Fig. 7 se representa con triángulos negros la media de la SFR entre la masa frente a los intervalos de masas de todas las galaxias, esta vez a *redshifts* altos, a partir de  $z = 3.61$  ya que las próximas secciones se centran en la tasa de formación estelar de las galaxias entre el primer y el segundo giga año del Universo (entre  $z = 3.61$  y  $z = 5.017$ ). Como era de esperar, la media es prácticamente constante, pero con una pequeña tendencia creciente, lo que indica que efectivamente, la tasa de formación estelar es prácticamente proporcional a la masa de las galaxias, pero que la SFR aumenta un poco más en proporción a la masa según se va aumentando la masa. En esta figura (Fig. 7) se representa además de la media con triángulos negros, la SFR dividida entre la masa frente a la masa del 10 % de las galaxias más masivas con puntos azules para aportar claridad. Mirando los puntos que representan el 10 % de las galaxias, efectivamente la función se aplanan con respecto a la dibujada en la Fig. 6, pero su forma en general sigue siendo la misma. En la Fig. 7 se ve más claramente y de forma más precisa el cambio en la dispersión de los puntos azules. Ahora el cambio en la dispersión se ve a partir de masas por encima de  $10^{10.8} M_{\odot} h^{-1}$ , donde la dispersión es de menos de 1 orden de magnitud con respecto a los más de 6 órdenes de magnitud que pueden verse para masas por debajo de este límite. Este cambio en la dispersión concuerda con el cambio de pendiente de las dos rectas estudiadas en la Fig. 3 discutidas en la subsección 3.1, en ambos casos quiere decir que a partir de estos valores, SFR por encima de en torno a  $10^{1.5} M_{\odot} h^{-1} yr^{-1}$ , es más difícil encontrar galaxias formando tantas estrellas.

La mayoría de los puntos a masas por encima de  $10^{10.8} M_{\odot}$ , incluyendo las galaxias extremadamente masivas representadas mediante estrellas rojas (Fig. 7(a) y 7(b)), se encuentran por encima de la media, lo que quiere decir que la SFR con respecto a la masa es mayor en proporción que a masas más bajas de  $10^{10.8} M_{\odot}$ , donde se puede ver que los puntos están más repartidos por encima y por debajo de la media, habiendo incluso más por debajo de esta.

La conclusión a la que se llega en este apartado es que a una galaxia que sea muy masiva el modelo semi-analítico SAGE va a asociarle una alta SFR.

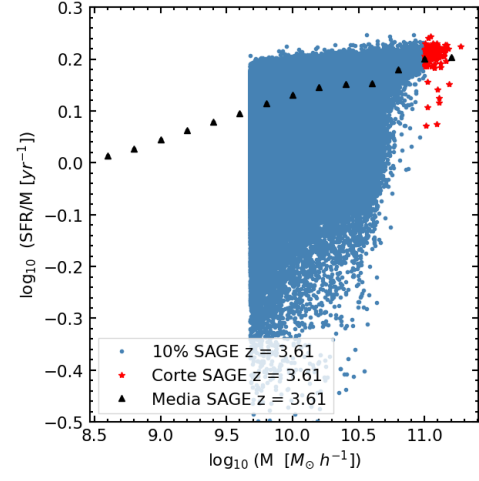
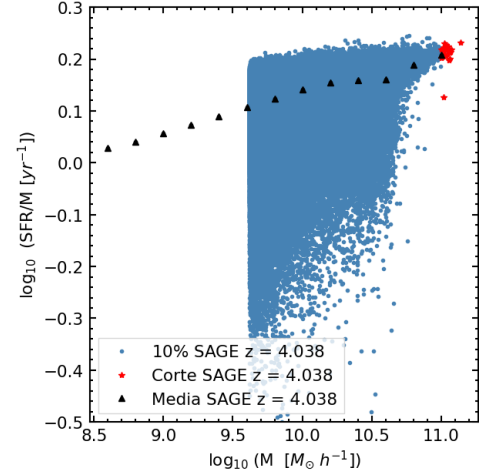
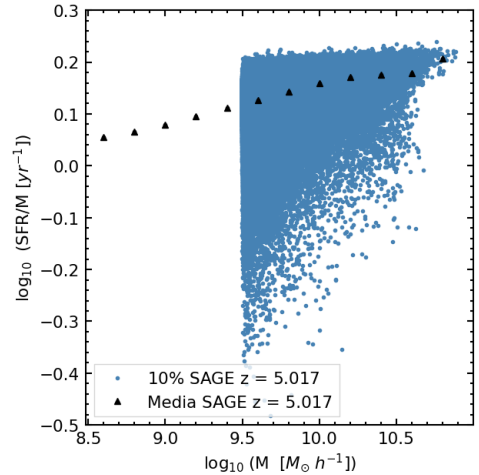
(a)  $z = 3.61$ (b)  $z = 4.038$ (c)  $z = 5.017$ 

Figura 7: Los triángulos negros representan la media de la tasa de formación estelar entre la masa frente a los intervalos de masa de todas las galaxias a distintos *redshifts*. Por claridad se representa también la tasa de formación estelar entre la masa del 10 % de las galaxias más masivas frente a su masa mediante círculos azules. Estas gráficas confirman que el modelo semi-analítico SAGE asocia una mayor SFR a las galaxias según aumentan su masa, a pesar de que la SFR aumente de manera un poco más rápida que las masas.



### 3.3 El 10 % de galaxias modelo más masivas a altos *redshifts*

La SFR es un indicativo de cómo va evolucionando la masa en el modelo jerárquico de formación de estructura, al principio se tienen objetos pequeños que van juntándose y creciendo. Por tanto, en principio se podría suponer que una galaxia, al tener una SFR grande, en general, su masa debería ser pequeña, ya que hay que darle tiempo (*redshift*) para que obtenga mayor masa estelar. Por tanto, a *redshifts* altos, se esperaría encontrar pocas galaxias masivas con alta SFR.

Esta sección se centra en el estudio de las propiedades del 10 % de las galaxias más masivas del modelo a *redshifts* altos, entre el primer y el segundo giga año del Universo (de  $z = 3.61$  a  $z = 5.017$ ).

Se representa en la Fig. 8, una vez más mediante triángulos negros, la media de la SFR frente a los intervalos de masa, para los tres *redshifts* más altos. Con los puntos azules se muestra la tasa de formación estelar frente a la masa de las galaxias del 10 % de las galaxias más masivas. Las estrellas rojas en 8(a) y 8(b) representan las galaxias extremadamente masivas que se discuten en la siguiente subsección (Sec. 3.4).

Analizando los puntos azules, el 10 % de las galaxias más masivas, y poniendo atención a la forma de la función, se destaca en los tres casos el cambio de dispersión mencionado en la sección anterior (Sec. 3.2), de más de 4 órdenes de magnitud a aproximadamente uno, de la función a partir de una masa de aproximadamente  $10^{10.8} M_{\odot} h^{-1}$ . De nuevo, en la zona de masas más altas, esta vez a partir de  $10^{10.5} M_{\odot} h^{-1}$ , se puede ver que según aumenta la masa hay cada vez menos galaxias con baja tasa de formación estelar, hasta al final donde prácticamente solo hay galaxias con alta masa y alta tasa de formación estelar, exceptuando algunos casos puntuales que se pueden contar ya que son puntos distinguibles. Esto sigue indicando que el modelo semi-analítico SAGE asocia una alta tasa de formación estelar a galaxias más masivas.

Después de ver la tendencia creciente de la SFR frente a la masa de las galaxias, se estudia la función de SFR del 10 % de las galaxias más masivas a altos *redshifts*, entre el primer y el segundo giga año del universo, y se compara con la función de SFR de todas las galaxias ya representada en la Fig. 3, con el objetivo de comprobar si el modelo semi-analítico SAGE es capaz de modelar galaxias muy masivas con baja SFR.

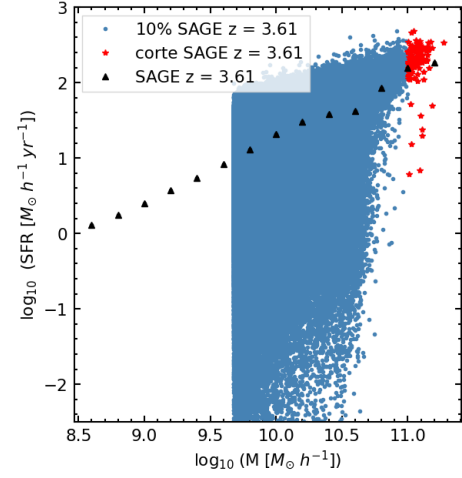
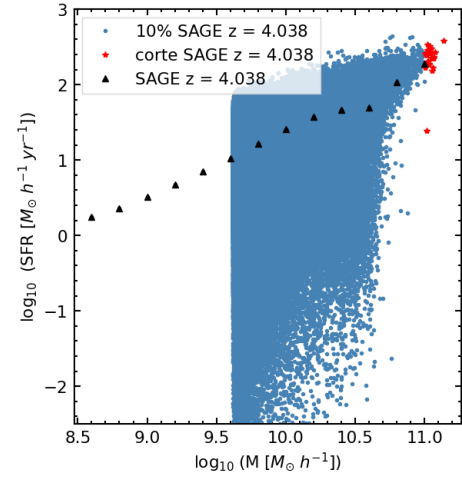
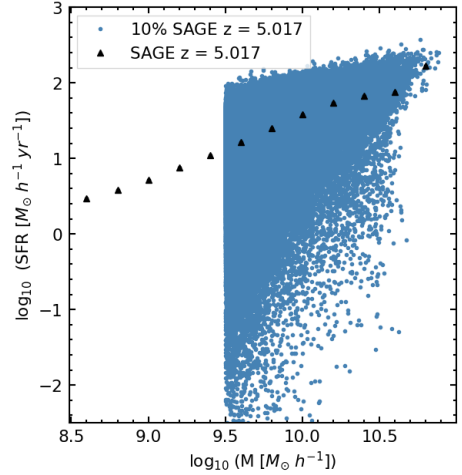
(a)  $z = 3.61$ (b)  $z = 4.038$ (c)  $z = 5.017$ 

Figura 8: Los círculos azules representan la tasa de formación estelar del 10 % de las galaxias más masivas frente a su masa, las estrellas rojas para el caso de las gráficas 8(a) y 8(b) representan la tasa de formación estelar de las galaxias con masa por encima de  $10^{11} M_{\odot} h^{-1}$  y los triángulos negros representan la media de la tasa de formación estelar frente a los intervalos de masa de todas las galaxias a distintos *redshifts*. Las galaxias más masivas son las que tienen mayor SFR.

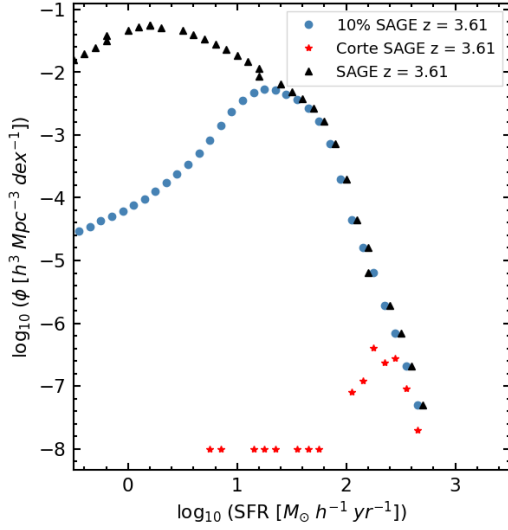
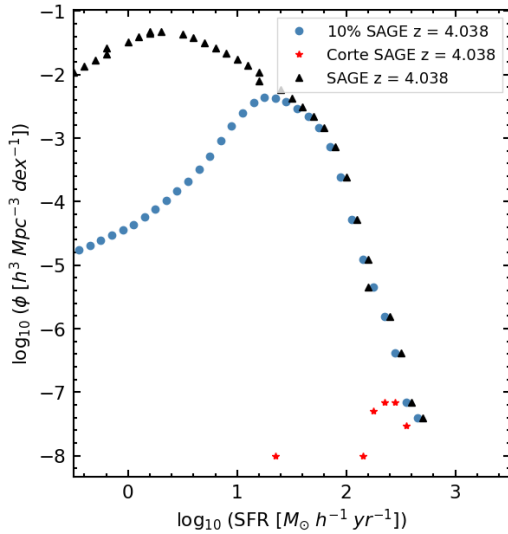
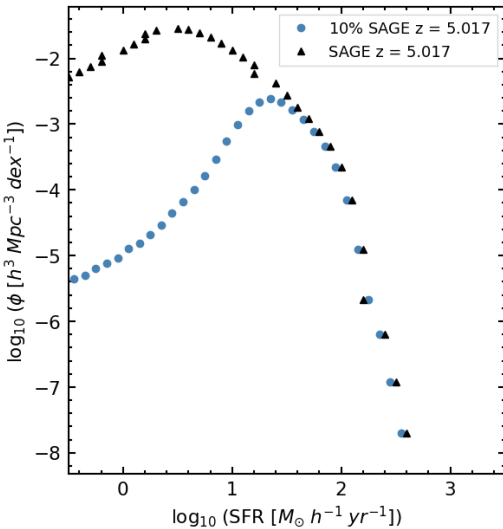
(a)  $z = 3.61$ (b)  $z = 4.038$ (c)  $z = 5.017$ 

Figura 9: Función SFR comparando todas las galaxias (triángulos negros) con el 10 % de las galaxias más masivas encontradas (círculos azules, se discuten en la Sec. 3.3) y con las galaxias con un corte de masa por encima de  $10^{11} M_{\odot} h^{-1}$  (estrellas rojas, se discuten en la Sec. 3.4) a *redshift*  $z = 3.61$ ,  $z = 4.038$  y  $z = 5.017$ . Se confirma que las galaxias con alta SFR son las galaxias más masivas.

Esto se muestra en la Fig. 9, donde se representa la función de SFR de todas las galaxias mediante triángulos negros, la función SFR del 10 % de las galaxias más masivas con puntos azules y las galaxias extremadamente masivas con estrellas rojas, a *reshifts* 3.61, 4.038 y 5.017 respectivamente. Estas últimas se ven en detalle en la siguiente subsección (Sec. 3.4). En estas gráficas la tendencia mencionada antes parece mucho más clara, prácticamente todas las galaxias con alta tasa de formación estelar son las galaxias más masivas.

Mirando solo la curva del 10 % de las galaxias más masivas (puntos azules), se puede ver como vuelve a existir un cambio de pendiente como el estudiado en la función de SFR de todas las galaxias en la primera subsección (Sec. 3.1), y a partir del cual podemos distinguir dos leyes de potencias. Sin embargo, esta vez encontramos una recta con pendiente positiva y la otra negativa, y en los tres casos un punto máximo en la función en  $\log(\text{SFR}) \simeq 1.1 \log(M_{\odot} h^{-1} \text{yr}^{-1})$ . Después de este valor máximo de la curva se encuentran las galaxias con alta formación estelar. A partir de este máximo, es decir, a partir de una SFR de aproximadamente  $10^{1.5} M_{\odot} h^{-1} \text{yr}^{-1}$ , comparando los puntos azules que representan las galaxias más masivas, con los triángulos negros que representan el total de galaxias modelizadas, se puede ver que los puntos azules prácticamente superponen los triángulos negros, es decir, en este lado de la función (a partir de  $\text{SFR} = 10^{1.5} M_{\odot} h^{-1} \text{yr}^{-1}$ ), prácticamente todas las galaxias que tienen esta tasa de formación estelar son las galaxias masivas. Si uno se fija en lo que pasa antes del máximo ( $\text{SFR} < 10^{1.5} M_{\odot} h^{-1} \text{yr}^{-1}$ ), las galaxias más masivas se alejan de la curva azul, lo que significa que según disminuye su masa, más galaxias hay con baja SFR.

### 3.4 Galaxias extremadamente masivas a altos *redshifts*

Existen observaciones que aseguran la existencia de galaxias muy masivas (con masas por encima de  $10^{11} M_{\odot}$ ) a muy alto *redshift* (en los primeros años de formación del Universo) sin formación estelar (Guarnieri et al., 2019; Alcalde Pampliega et al., 2019). No se trata de muchas galaxias, pero son muy características debido al hecho de ser muy masivas y no tener formación estelar tan pronto. Se consideran objetos raros y es por esta razón que son objetos interesantes para poner a prueba los modelos teóricos.



A priori, mirando tan solo las gráficas de las Figs. 6, 7 y 8, las galaxias masivas del modelo semi-analítico SAGE forman más estrellas si su masa es más grande, por lo que no hay indicios de que SAGE contemple la existencia de galaxias masivas que no formen estrellas.

Para ver realmente como interpreta el modelo semi-analítico SAGE la tasa de formación estelar de las galaxias más masivas a altos *redshifts*, lo primero es buscar las galaxias masivas que van a estudiarse modelizadas por SAGE. La galaxia más masiva modelizada a *redshifts* altos se obtiene a partir de la interfaz SQL de Cosmosim (Apéndice A). En particular a  $z = 3.61$ , la galaxia más masiva tiene una masa de  $1.88 \times 10^{11} M_{\odot} h^{-1}$ , a  $z = 4.038$  una masa máxima de  $1.38 \times 10^{11} M_{\odot} h^{-1}$  y a  $z = 5.017$  una masa máxima de  $7.68 \times 10^{10} M_{\odot} h^{-1}$ .

A *redshift* tan alto, entre el primer y el segundo giga año del Universo, se habla de galaxias muy masivas a partir de masas estelares del orden de  $10^{11} M_{\odot}$ , ya que la galaxia más grande encontrada tiene una masa estelar de  $1.7 \cdot 10^{11} M_{\odot}$  a *redshift* 3.717 (Glazebrook et al. (2017)). La detección de una galaxia tan masiva con sus propiedades es algo muy complicado para los modelos semi-analíticos.

Teniendo en cuenta los datos de las masas máximas encontradas, y teniendo en cuenta el corte a partir del cual se considera que se tienen galaxias masivas (Guarnieri et al. (2019)), se realizó un corte en las masas a partir de  $10^{11} M_{\odot} h^{-1}$ . A partir de este corte, el modelo semi-analítico SAGE, encuentra 131 galaxias a *redshift* 3.61 y 24 a *redshift* 4.038. Para *redshift* 5.017, sin embargo, la masa máxima que modeliza SAGE, es más pequeña que  $10^{11} M_{\odot} h^{-1}$ , por lo que, en esta subsección solo se estudian los datos proporcionados por SAGE a  $z = 3.61$  y a  $z = 4.038$ . Este número de galaxias muy masivas que SAGE es capaz de modelizar, conociendo el volumen de la simulación (Sec. 2), significa que a *redshift* 3.61 hay una densidad numérica (número de galaxias por unidad de volumen) de  $4.1 \times 10^{-8} Mpc^{-3}$  y a *redshift* 4.038 una densidad numérica de  $7.5 \times 10^{-9} Mpc^{-3}$ . Se pueden comparar estas densidades con las encontradas para galaxias con masa por encima de  $10^{11} M_{\odot}$  en otros modelos semi-analíticos a *redshift*  $z = 3.9$  y  $z = 4.1$  (Guarnieri et al. (2019)). En el modelo descrito en Gonzalez-Perez et al. (2014) se encuentra una densidad de  $1.6 \times 10^{-5} Mpc^{-3}$ , para el modelo descrito en Henriques et al. (2015)  $5.2 \times 10^{-7} Mpc^{-3}$ ,  $5 \times 10^{-6} Mpc^{-3}$  para el modelo

descrito en Lacey et al. (2016) y  $5 \times 10^{-6} Mpc^{-3}$  en De Lucia et al. (2006). Al comparar estas densidades numéricas, se destaca que varían mucho de unos modelos a otros, pero que todas tienen en común que no son muy grandes, es decir, en general en los modelos semi-analíticos no existen muchas galaxias por unidad de volumen muy masivas.

En la Fig. 8 se representa mediante estrellas rojas (en las gráficas 8(a) y 8(b)) la tasa de formación estelar frente a la masa de las galaxias más masivas encontradas por encima del corte de masa de  $10^{11} M_{\odot} h^{-1}$ . Analizando estos puntos, se puede ver como prácticamente todas las galaxias con masa por encima de  $10^{11} M_{\odot} h^{-1}$  se encuentran en la media o por encima de la media de la SFR. Tan solo hay 8 galaxias muy masivas de 131 para el caso de *redshift* 3.61 que tengan más baja SFR que la media y tan solo 1 galaxia de 24 para el caso de *redshift* 4.038. Lo interesante es ver si estas pocas galaxias con la tasa de formación estelar más baja que la media han dejado de formar estrellas, es decir, ver si esta SFR es nula o prácticamente nula de forma que el modelo semi-analítico modele galaxias de acuerdo con las observaciones de Guarnieri et al. (2019) y Alcalde Pampliega et al. (2019). Para ello, se estudia la función de SFR de las galaxias masivas por encima de  $10^{11} M_{\odot}$  a *redshifts* 3.61 y 4.038 y se compara con la función de SFR de todas las galaxias ya dibujada en la Fig. 3. Esto se muestra representado en la Fig. 9, en las gráficas 9(a) y 9(b) mediante estrellas rojas, donde se puede ver claramente que todas las galaxias representadas tienen una SFR bastante alta, por encima de  $10^{0.8} M_{\odot} h^{-1} yr^{-1}$  para *redshift* 3.61 y por encima de  $10^{1.4} M_{\odot} h^{-1} yr^{-1}$ .

Por otro lado, se destaca también, sobre todo se ve más claro en la gráfica 9(a) a *redshift* 3.61 ya que hay más galaxias, que al final de la curva, a altas SFR a partir de en torno a  $10^{2.2} M_{\odot} h^{-1} yr^{-1}$ , las estrellas rojas, es decir, el número de galaxias muy masivas, están muy poco por debajo de los triángulos negros, que representan la cantidad total de galaxias formando estrellas con esa tasa de formación estelar. Esto indica que, prácticamente todas las galaxias con alta tasa de formación estelar, por encima de  $10^{2.2} M_{\odot} h^{-1} yr^{-1}$  para *redshift* 3.61, son las galaxias masivas. Este comportamiento también se aprecia en la gráfica 9(b) a *redshift* 4.038, solo que al haber menos galaxias con masa por encima de  $10^{11} M_{\odot} h^{-1}$  se ve con menos claridad.

A la conclusión a la que se llega a partir de

aquí es que según el modelo semi-analítico SAGE es prácticamente imposible encontrar galaxias extremadamente masivas con baja o nula SFR, por debajo de  $10^{0.8} M_{\odot} h^{-1} \text{yr}^{-1}$ , a *redshifts* altos, los dos primeros giga años del Universo, lo cual, entra en desacuerdo con los datos observacionales de Guarnieri et al. (2019) y Alcalde Pampliega et al. (2019). El desacuerdo entre el modelo y las observaciones podría deberse a que, o bien el modelo no captura la formación de galaxias de forma realista a alto *redshift*, o a que las estructuras de la simulación de N-cuerpos se forman con demasiada lentitud, es decir, que el problema venga del modelo cosmológico en el que se basan estas simulaciones. Estos objetos estadísticamente raros encontrados en las observaciones son, por tanto, interesantes porque pueden ayudar a mejorar los modelos semi-analíticos.

#### 4. RESUMEN Y CONCLUSIONES

En este trabajo se ha estudiado la historia evolutiva del Universo utilizando modelos teóricos de formación y evolución de galaxias. La información extragaláctica que se puede medir proviene de la componente visible del Universo, es decir, de las estrellas, por esta razón la variable elegida para el estudio ha sido la tasa de formación estelar (SFR), la masa total de estrellas formada por año en una galaxia. Para ello, se ha realizado un estudio en detalle (y se ha comparado con las observaciones de Gruppioni et al. (2015)) de las galaxias creadas computacionalmente por el modelo semi-analítico SAGE de formación y evolución de galaxias utilizando la simulación de materia oscura MultiDark Plank 2 (MDPL2) con un tamaño comparable al tamaño del Universo observable.

Según el modelo de formación jerárquica, las galaxias se van formando por agregación, partiendo de masas más pequeñas hasta formar galaxias con cada vez más masa. Hay observaciones (Guarnieri et al., 2019; Alcalde Pampliega et al., 2019) que aseguran la existencia de galaxias extremadamente masivas, con masa por encima  $10^{11} M_{\odot}$ , sin tasa de formación estelar entre el primer y el segundo giga año del Universo. Las estrellas se empezaron a formar a aproximadamente  $z \sim 10$  (Loeb & Barkana, 2001), es decir, en la mitad del primer giga año del Universo, y según el modelo de formación jerárquica, crear galaxias de  $10^{11} M_{\odot}$  en menos de un giga año es poco probable, lo que convierte a estos objetos en objetos raros y muy interesantes a la hora de modelizar. Es

por esta razón que la última parte de este proyecto (Sec. 3.4) se centra en estudiar estos objetos, para finalmente ver que el modelo semi-analítico utilizado (SAGE) no es capaz de encontrarlos, lo que nos lleva a preguntarnos si el problema está en los modelos de galaxias o en el modelo cosmológico en el cual se basa el modelo, ya que el ritmo de formación de estructuras cambia según el modelo cosmológico elegido.

Como ya se ha mencionado, en este proyecto se muestra cómo funciona el modelo semi-analítico SAGE de la simulación de MultiDark Plank 2, al modelizar la tasa de formación estelar de las galaxias, a bajas y a altas masas estelares. Para ello, lo primero que se ha hecho es un estudio sobre cómo evoluciona con el tiempo cósmico medido a través del *redshift* (ver la definición en la sección 2.1) la tasa de formación estelar y se compara con las observaciones obtenidas de Gruppioni et al. (2015). Encontramos en el modelo, una caída de galaxias formando estrellas a partir de una cierta SFR que variaba según el *redshift* entre los valores  $z = 0.142$  y  $z = 3.61$ . Por lo que a partir de un valor de SFR de en torno a  $10^{1.5} M_{\odot} h^{-1} \text{yr}^{-1}$  encontrar galaxias formando tantas estrellas se vuelve estadísticamente más difícil. No se ha podido comprobar con las observaciones este cambio de comportamiento a partir de un punto, debido a que observacionalmente es muy difícil medir bajas tasas de formación estelar y por tanto la cantidad de puntos es mucho menor.

Para el caso del modelo también se destaca una caída a SFR bajas según el *redshift* comienza a ser alto, lo que significa que a *redshifts* altos (menor edad del Universo), hay menos galaxias formando pocas estrellas. Haciendo la comparación de los datos del modelo con los observados, *redshift* por *redshift*, se llega a la conclusión de que el modelo semi-analítico SAGE tiene una caída de galaxias formando estrellas con alta SFR demasiado rápida comparada con las observaciones. Se observan más galaxias formando estrellas con alta SFR ( $> 10^{1.6} M_{\odot} h^{-1} \text{yr}^{-1}$ ) que las predichas por el modelo, lo que indica que el modelo no tiene el suficiente gas frío y denso necesario para formar estrellas.

A pesar de los problemas encontrados con la base de datos CosmoSim para obtener datos para los modelos SAG y GALACTICUS (ver Apéndice A), se ha podido estudiar el modelo semi-analítico GALACTICUS en los *redshifts*  $z = 0.142$  y  $0.523$ , y se ha visto que este modelo (GALACTICUS) se ajusta mejor a los datos observacionales en esta situación que SAGE, lo que muestra que GALAC-

TICUS tiene mejor tratamiento del gas que SAGE.

Seguidamente se ha realizado un estudio de la relación entre la SFR y la masa estelar de las galaxias y de la evolución de esta relación, a partir del cual se ha visto que para el modelo semi-analítico SAGE la tasa de formación estelar aumenta según aumenta la masa de las galaxias para todos los *redshifts* hasta un valor de en torno a  $SFR = 10^2 M_\odot h^{-1} yr^{-1}$  a partir del cual no hay más galaxias modelo. En la forma de la función existe una dispersión que decrece hasta 6 órdenes de magnitud para masas por encima de  $10^{11} M_\odot h^{-1}$  que indica que a partir de esta masa la probabilidad de encontrar galaxias con baja SFR es prácticamente nula. Además la tendencia creciente de la función se mantiene, aunque más levemente, para el cociente entre la SFR y la masa con respecto a la masa de las galaxias, es decir, son cantidades más o menos proporcionales pero con un leve incremento superior de la SFR con respecto a la masa. Este incremento de la tasa de formación estelar con la masa coincide con las observaciones de Popesso et al. (2019).

Al encontrar esta semejanza entre el modelo semi-analítico SAGE y las observaciones, se decide ir un paso más allá y poner a prueba el modelo con los objetos extremadamente masivos y estadísticamente improbables buscados y observados por Guarnieri et al. (2019) y Alcalde Pampliega et al. (2019), con el fin de poner restricciones a los modelos teóricos. Estas galaxias, además de ser extremadamente masivas ( $M > 10^{11} M_\odot$ ), según las observaciones no están formando estrellas, es decir, no tienen tasa de formación estelar. Esto a priori SAGE no parece contemplarlo.

Por esta razón, en las dos últimas subsecciones, (Secs. 3.3 y 3.4) se estudian las galaxias más masivas a alto *redshift* modeladas por SAGE, para comprobar si existe alguna con SFR baja. Primero, para ir poco a poco con el análisis se realiza un primer corte tomando el 10% de las galaxias más masivas a *redshift* 3.61, 4.038 y 5.017. Tras comprobar que para estas galaxias más masivas la tendencia creciente de la tasa de formación estelar frente a la masa de las galaxias modeladas por SAGE se sigue manteniendo, se ha comparado la función de SFR de todas las galaxias frente a la función SFR de las galaxias por encima del corte y se ha comprobado que prácticamente todas las galaxias con alta tasa de formación estelar son las galaxias masivas, como se puede ver en la 9, los puntos azules que representan el 10% de las galaxias más masivas están superpuestos a los puntos negros que representan el total de las galaxias a

tasas de formación estelar altas, por encima de  $10^{1.5} M_\odot h^{-1} yr^{-1}$ .

Por último, se ha realizado un segundo corte, esta vez centrado en estudiar las galaxias extremadamente masivas, con masas por encima de  $10^{11} M_\odot h^{-1}$ . De estas sólo se han encontrado para *redshifts*  $z = 3.61$  y  $z = 4.038$ . En el estudio de la función de la tasa de formación estelar de estas galaxias extremadamente masivas, la Fig.9 muestra como todas ellas están por encima de una SFR de torno a  $10 M_\odot h^{-1} yr^{-1}$ , es decir, todas tienen una tasa de formación estelar grande. De esta figura se destaca también que a SFR más altas aun, por encima de  $10^{2.2} M_\odot h^{-1} yr^{-1}$ , las estrellas rojas que representan las galaxias más masivas se encuentran muy poco por debajo de los triángulos negros que representan la cantidad total de galaxias formando estrellas con esa tasa de formación estelar, es decir, que prácticamente todas las galaxias formando estrellas a ese ritmo ( $> 10^{2.2} M_\odot h^{-1} yr^{-1}$  son las galaxias más masivas).

Lo que resulta interesante es que este modelo parece fallar a altos *redshifts* tanto por no tener galaxias con baja SFR como con alta. Debido a los problemas que se tuvieron con la base de datos (Apéndice A, no se pudieron hacer estas comparaciones con los otros dos modelos, SAG y GALACTICUS, pero según el informe de Knebe et al. (2017) GALACTICUS es el mejor de los modelos para modelizar la función de tasa de formación estelar y su evolución, por lo que quizá con este modelo podría haberse encontrado lo esperado observacionalmente. Según Knebe et al. (2017), SAG tiene también más galaxias con alta SFR en comparación con SAGE, por lo que quizá este otro modelo también hubiera podido cambiar las conclusiones, teniendo en cuenta además que SAG utiliza la función de tasa de formación estelar como restricción para la calibración de sus parámetros.

Cuando un modelo no concuerda con las observaciones es el momento en el que se pueden hacer nuevos descubrimientos y aspectos a mejorar en el modelo. En el caso del modelo semi-analítico SAGE habría que buscar la manera de cambiar las ecuaciones que dirigen la evolución del gas, de modo que las galaxias tengan más gas frío y denso.

## AGRADECIMIENTOS

“The CosmoSim database used in this paper is a service by the Leibniz-Institute for Astrophysics

sics Potsdam (AIP). The MultiDark database was developed in cooperation with the Spanish MultiDark Consolider Project CSD2009-00064.”

The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. ([www.gauss-centre.eu](http://www.gauss-centre.eu)) and the Partnership for Advanced Supercomputing in Europe (PRACE, [www.prace-ri.eu](http://www.prace-ri.eu)) for funding the MultiDark simulation project by providing computing time on the GCS Supercomputer SuperMUC at Leibniz Supercomputing Centre (LRZ, [www.lrz.de](http://www.lrz.de)).

The Bolshoi simulations have been performed within the Bolshoi project of the University of California High-Performance AstroComputing Center (UC-HiPACC) and were run at the NASA Ames Research Center.

## Referencias

- Alcalde Pampliega B., et al., 2019, *The Astrophysical Journal*, 876, 135
- AstroMia 2021, *Las Galaxias*, <https://www.astromia.com/universo/galaxias.htm>
- Baugh C. M., 2006, *Reports on Progress in Physics*, 69, 3101
- De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, 366, 499
- Gil Bor R. M., 2014, *Miscelánea Matemática de la Sociedad Matemática Mexicana*, 57.1, 83
- Glazebrook K., et al., 2017, *Nature*, 544, 71
- Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, *MNRAS*, 439, 264
- Gray R., Dunning-Davies J., 2008, *Department of Physics, University of Hull*
- Gruppioni C., et al., 2015, *MNRAS*, 451, 3419
- Guarnieri P., et al., 2019, *MNRAS*, 483, 3060
- 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, 2663
- Jiang F., van den Bosch F. C., 2014, *MNRAS*, 440, 193
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Knebe A., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 474, 5206–5231
- Lacey C. G., et al., 2016, *MNRAS*, 462, 3854
- Loeb A., Barkana R., 2001, *ARA&A*, 39, 19
- Popesso P., et al., 2019, *MNRAS*, 490, 5285
- Riebe K., et al., 2013, *Astronomische Nachrichten*, 334, 691
- Sherman S., et al., 2021, *arXiv e-prints*, p. arXiv:2105.03766
- Theuns T., 2016, *Institute for Computational Cosmology, Ogden Centre for Fundamental Physics, Physics Department, Durham University*.
- Vega-Martínez C. A., 2017, *PhD thesis, Universidad Nacional de la plata - Facultad de Ciencias Astronómicas y Geofísicas*

## A. Peticiones SQL

La forma acceder la base de datos de CosmoSim es a través de la *interfaz* de la plataforma mediante formularios de consulta en lenguaje de consulta estructurada (en inglés, *Structured Query Language (SQL)*). En este formulario hay dos opciones para elegir la duración del tiempo que tardan en cargarse los datos pedidos, *short queue*, 40 segundos o *long queue* 1 hora. Los problemas que se han tenido al solicitar los datos con los modelos SAG o GALACTICUS son debidos a este tiempo, el tiempo se acababa antes de que la base de datos proporcionara ningún dato. En alguna ocasión, si la consulta era repetida muchas veces, la base de datos terminaba por proporcionar los datos, y es por esta razón por la que en el informe hay algún dato de GALACTICUS. Sin embargo, no se encontró la razón a esto, y repetir peticiones tan largas, tantas veces terminó por ser inviable. Se intentó contactar con el equipo de mantenimiento de la base de datos para buscar una solución, pero no se obtuvo respuesta. A continuación se muestran algunos de los formularios utilizados para solicitar los datos en la base de datos.

Lo primero que se hizo, fue pedir una tabla de *redshifts* con sus respectivos *snapnum* (número correspondiente a cada *snapshot*), para la base de datos MDPL2, que engloba SAGE, GALACTICUS y SAG. De esta manera se obtiene la correlación entre *redshift* y *snapnum* para poder realizar las futuras peticiones con el *snapnum*, lo que facilita la petición ya que son números enteros y mucho más precisos.

### Código A1:

```
SELECT DISTINCT zred, snapnum, aexp
FROM MDPL2.Redshifts
ORDER BY snapnum DESC
```

Para pedir los datos necesarios para representar la función SFR (Sec. 3.1) se utilizó un código SQL (Código A2) que realiza directamente una división en intervalos de SFR, gracias a la función FLOOR, y cuenta el número de galaxias en cada intervalo, de manera que a partir de estos datos se podía representar directamente la función SFR. El *snapnum* se iba cambiando según el redshift necesario, los códigos que se muestran en este apéndice están todos con *snapnum* = 56 correspondiente a *redshift* = 3.61. Con el código A1, puede verse la correlación entre *redshift* y *snapnum* para poder ir variando esta variable en el código.

### Código A2:

```
SELECT 0.1*(0.5+FLOOR(LOG10(SFR)/0.1)) AS
StarFR,
COUNT(*) AS num
FROM MDPL2.SAGE
WHERE snapnum = 56 AND (MstarDisk+
MstarSpheroid) > 0
GROUP BY FLOOR(LOG10(SFR)/0.1)
ORDER BY 0.1*(0.5+FLOOR(LOG10(SFR)/0.1))
```

Para ver la relación entre la masa y la SFR (Sec. 3.2) se solicitaron todas las masas, con sus respectivas SFR ordenadas según la masa descendiente (Código A3). Este código sirvió después para el estudio de las dos últimas subsecciones de la sección de resultados (3.3 y 3.4). Al tener las masas ordenadas, se podían encontrar de manera sencilla tanto las masas máximas a cada *redshift*, como la cantidad de galaxias totales que hay en el modelo por encima de 20 veces el límite de resolución que se había establecido a  $M > 10^{8.5} M_{\odot} h^{-1}$ . Se contaron en esta lista la cantidad de galaxias por encima del límite de  $10^{11} M_{\odot} h^{-1}$  y se calculó el 10 % del total de galaxias, para así poder poner los límites en el código con la función LIMIT. Como las galaxias están ordenadas de forma descendiente, para coger el 10 % más masivas basta con coger el 10 % primeras.

### Código A3:

```
SELECT LOG10(MstarSpheroid + MstarDisk)
as Mass,
LOG10(SFR) as SFR
FROM MDPL2.SAGE
WHERE snapnum=56
AND LOG10(MstarSpheroid + MstarDisk)
>8.5
AND SFR > 0
ORDER BY Mass DESC
```

El código A4 muestran un ejemplos de la función LIMIT que se ha utilizado para realizar los cortes en masa, las 131 galaxias con masa por encima de  $10^{11} M_{\odot} h^{-1}$  a *redshift* 3.61 (*snapnum* 56) según la lista de masas ordenadas proporcionada por el código A3.

### Código A4:

```
SELECT LOG10(MstarSpheroid + MstarDisk)
as Mass,
LOG10(SFR) as SFR
FROM MDPL2.SAGE
WHERE snapnum=56
AND (MstarSpheroid + MstarDisk)>0
AND SFR > 0
ORDER BY Mass DESC
LIMIT 131
```

## **B. Código de análisis**

Para realizar las gráficas mostradas en este informe he utilizado el lenguaje de programación Python. He creado unos códigos que pueden ser encontrados de forma pública en la plataforma Github a través del enlace: <https://github.com/OliviaV57/TFG-TUT>. El código está documentado con comentarios y un fichero README explicativo.