



Universidad
de La Laguna

FACULTAD DE FÍSICA

DEPARTAMENTO DE ASTROFÍSICA

TRABAJO DE FIN DE GRADO:

**Aprendizaje automático calibrado en
simulaciones de N-cuerpos para la
producción eficiente de catálogos con halos
de materia oscura de baja masa**

Realizado por:
Guillermo Reyes Peraza

Tutorizado por:
Francisco Shu Kitaura

Contents

1	Formación de estructuras a gran escala en un Universo en expansión	4
2	Función de correlación y espectro de potencias	8
3	Ajuste del espectro de potencias de la simulación SMDPL	12
3.1	Competencias a realizar:	13
3.2	Simulación The Small MultiDark Planck:	14
3.3	FastPM:	16
3.4	Código Patchy:	19
3.5	Modelo de Bias:	21
3.5.1	Bias determinista:	22
3.5.2	Bias estocástico	23
3.6	Aprendizaje automático para obtener los parámetros de Bias galáctico:	24
3.7	Resultados y discusión:	27
4	Conclusión:	39
5	Bibliografía	41
5.1	Referencias	41

Abstract

This work presents an advance in the automatic learning calibrated in simulations of N-bodies that allow to generate efficient catalogs with halos of dark matter of low mass. The fast and precise simulation of dark matter halos catalogs with a small volume and a low number of particles is studied using cutting-edge techniques. Then, once generated, we proceed to calculate the power spectrum of the simulated catalog, in order to verify that the calculated power spectrum is adjusted to the power spectrum of the Small MultiDark Planck (SMDPL) simulation. To this purpose the *Patchy* code is used in its new version that includes a particle mesh module (*FastPM*) to simulate the gravitational evolution and a code to find the parameters of the galactic Bias using Markov chains by minimising the quadratic distance between the reference and the simulated power spectra.

The *FastPM* code is a highly scalable approximated particle mesh (PM) N-body solver, which implements the PM scheme enforcing correct linear displacement (1LPT) evolution via modified kick and drift factors, that solves the Poisson equation of gravity with an enhanced particle mesh. Generating an exact growth of the field of dark matter to a large scale on a mesh, and thus obtaining a gravitational simulation of dark matter.

Making use of the distributions of particles of dark matter obtained with *FastPM*, that is to say after producing the field of dark matter to large scale in a mesh, the *Patchy* code allows us to fill this mesh with dark matter halos by introducing a nonlinear deterministic Bias model, getting this way the transition of the continuous field of dark matter (generated by *FastPM*) to the discrete distribution of halos. Thus this code is capable of generating catalogs of halos of dark matter. However, the production of approximate simulated catalogs with *Patchy* requires a reference catalog constructed from observations or based on a precise N-body simulation, which in our case is represented by the simulation The Small MultiDark Planck.

Once we have generated the dark matter halo catalogs making use of the codes *FastPM* + *Patchy*, we find the Bias parameters with which fit the reference power spectrum through a robust MCMC method relying on the *emcee* code. In this way we get constraints for the various Bias parameter contributions, such as threshold biasing, nonlinear power law Bias, and stochastic Bias.

We also investigated a novel parameter which models the environmental dependence of the stochastic Bias.

Our work demonstrates that our technique can model the power spectrum, and therefore the two-point correlation function, in the reference simulation to great accuracy. Specifically, we show that our method permits us to reach an accuracy within 5% in the power spectrum up to $k \sim 0.4hMpc^{-1}$. However, using our methodology, the histogram of halos in cells does not fit precisely. Therefore we must make an effort in the future to improve the models explored

in this study and so adjust the halo PDF and ultimately the correlation functions up to four points.

Therefore, we may speculate that we lack some ingredient in our Bias model, for example the non-local Bias component. This may be investigated in the future with these techniques. There is still a long way to go to extract all the necessary information from structure formation and be able to apply it efficiently for the next generation of galactic surveys. We hope that this work has contributed to a deeper understanding of the problem and helps to advance the field.

1 Formación de estructuras a gran escala en un Universo en expansión

Abstract

In this chapter, we will make a brief introduction about the formation of large-scale structures in an expanding universe, with the purpose of explaining the important role of the existence of primordial fluctuations of the density of matter.

The Universe, as it is now known from observations of microwave background radiation, began in a hot, dense, nearly uniform state about 13.8 billion years ago. However, if we look at the sky, you can see the structures at all scales, from stars and planets to galaxies, and at much greater scales, one can observe galactic clusters and enormous gaps between galaxies. Then, how have all these structures been formed?

According to current models, once matter (particularly cold dark matter), dominates in the gravitational collapse of the Universe, the small fluctuations of density caused by cosmic inflation begin to amplify causing that the dark matter groups more and more. As long as the density fluctuations remain small, the dark matter can be treated as a fluid without pressure and therefore the perturbations can be descriptions of a linear order and this way, it can be obtained a very good approximation.

Then, when the perturbations have grown sufficiently, a small region can become considerably denser than the average density of the Universe. Thus, as the regions become denser, the linear approximation that describes density heterogeneities begins to fail. At this point, dark matter begins to collapse into a full net of halos before ordinary matter. Then, baryonic matter is expected to follow the evolution of dark matter. This is when structures, such as galactic clusters and galactic halos, begin to form.

Los diferentes objetos astrofísicos que somos capaces de observar, son el resultado de la evolución de las fluctuaciones primordiales. Dichas fluctuaciones de densidad desempeñan el papel de

inicio de formación de estructuras. En un Universo primitivo perfectamente homogéneo no tendría lugar la formación y evolución de estructura alguna, debido a que en estas circunstancias no se daría una inestabilidad gravitacional, y por ende no existirían zonas con mayor y menor densidad, es decir que no se producirían agrupaciones de materia. Sin embargo, gracias a estas fluctuaciones, la concentración de materia era ligeramente superior a la media en algunas regiones del Universo, de manera que debido a la actuación de la gravedad, las zonas con una densidad mayor van a comenzar a atraer cada vez más materia de las regiones con una menor densidad. Por lo tanto con el tiempo, la diferencias de densidad de materia se acentúan cada vez más y finalmente dará origen a un Universo como el nuestro, en el cual la masa está concentrada en galaxias, cúmulos y supercúmulos, estando el resto casi vacío.

El Universo en la actualidad resulta ser bastante "grumoso" en escalas pequeñas (inferiores a 10 Mpc), siendo la sobredensidad típica en las galaxias del orden de:

$$\frac{\delta\rho}{\rho} \approx 10^5$$

Y para las escalas de Cluster es típicamente del orden:

$$\frac{\delta\rho}{\rho} \approx 10^{2-3}$$

Sin embargo se sabe que en momento en el que tuvo lugar el desacoplo de la radiación de la materia ($z \sim 1000$), es decir cuando los fotones del Fondo de Radiación de Microondas (CMB) se desacoplaron y comenzaron a viajar libremente por el espacio-tiempo, periodo en el que además el factor de escala era aproximadamente 10^3 veces menor que el factor de escala actual, el Universo por aquel entonces (antes de la recombinación) era como un medio continuo (fluido) y se piensa que dicho fluido, conocido como plasma primordial, era una especie de medio elástico, "un tira y afloja" entre luz y la materia bariónica, de manera que dichas compresiones y dilataciones de la materia bariónica producían unas vibraciones que se propagaban por todo el plasma, siendo la suma de todas estas oscilaciones las llamadas oscilaciones acústicas de bariones. Y por otro lado mientras la materia ordinaria se encontraba vibrando (debido a la existencia de fuerzas de presión), la materia oscura ya iba agrupándose cada vez más por efecto de la fuerza gravitatoria. Cabe destacar que la materia oscura juega un papel fundamental en la formación de estructuras, ya que únicamente siente la fuerza de la gravedad, permitiendo que la inestabilidad gravitacional pueda formar estructuras compactas al no oponerse ninguna fuerza, como la presión de radiación. Como

resultado, la materia oscura empieza a colapsarse en una red compleja de halos antes que la materia ordinaria. Esta red cósmica se ve afectada por la energía oscura, responsable de la expansión acelerada del Universo. Dependiendo de la componente de energía oscura que es dominante en la actualidad (un 70% aproximadamente, frente a un 30% de materia, la mayoría compuesta por materia oscura), los vacíos cósmicos son más o menos grandes, como ejemplo.

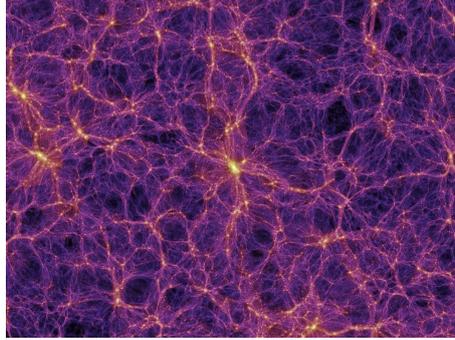


Figure 1: Simulación de la distribución de materia oscura en un fragmento del Universo. Las galaxias se agruparían en las zonas más densas, representadas como más brillantes.

La física de la formación de estructuras en el Universo temprano, es particularmente simple, ya que las perturbaciones de materia oscura con diferentes longitudes de onda evolucionan independientemente. Debido a la expansión del Universo el radio de Hubble crece, incluyendo perturbaciones cada vez mayores. Durante la época de dominación de materia, todas las perturbaciones de materia oscura crecen a través del colapso gravitacional. Sin embargo, las perturbaciones de longitud de onda corta que fueron generadas durante la dominación de radiación, retardaron su crecimiento hasta la dominación de materia. En esta etapa, la materia bariónica luminosa se espera que simplemente siga la evolución de la materia oscura y sus distribuciones deberían parecerse unas a otras.

Mediante el análisis de las inhomogeneidades o fluctuaciones de la temperatura del CMB se puede observar que son muy pequeñas, del orden de 10^{-5} . Dicho mapa de temperaturas está relacionado con el mapa de densidades de aquel plasma primitivo en el momento de la Recombinación, que apenas había evolucionado gravitacionalmente.

Así que la estructura que se observa en la actualidad debe haber crecido desde el tiempo del desacoplamiento de los fotones y la materia, para así poder generar los objetos de materia colapsados que podemos observar hoy en día. Además la pequeña amplitud de dichas fluctuaciones

primordiales resulta ser bastante útil, ya que gracias a ello se pueden analizar como perturbaciones a orden lineal y de esta forma poder obtener una muy buena aproximación.

Luego, cuando las perturbaciones han crecido lo suficiente, una pequeña región puede llegar a ser considerablemente más densa que la densidad media del Universo. En este momento, la física involucrada llega a ser substancialmente más complicada. Cuando las desviaciones de la homogeneidad son pequeñas, la materia oscura puede tratarse como un fluido sin presión y evoluciona según ecuaciones muy simples, debido a que estamos en el régimen lineal ($|\delta(k)| \ll 1$) donde la perturbación empieza a crecer y sigue acoplada con la expansión del Universo. Sin embargo, en regiones que son significativamente más densas que el fondo, dichas perturbación han entrado dentro del régimen no-lineal ($|\delta(k)| > 1$) y se desacoplan de la expansión del Universo, dejando de expandirse y convirtiéndose en un sistema gravitatoriamente ligado. Un signo de que las aproximaciones lineales y de fluido no son válidas, es que la materia oscura empieza a formar caústicas en las que las trayectorias de partículas adyacentes se cruzan, o las partículas empiezan a formar órbitas.

De manera que una vez que la perturbación sea no lineal, los diferentes modos \vec{k} de dichas fluctuaciones se acoplan, por lo que el tratamiento lineal se rompe y llega a ser mucho más difícil de tratar analíticamente, por lo que esto se hace principalmente con simulaciones numéricas.

Más abajo describimos brevemente como solucionamos la gravedad para un sistema de n-cuerpos de partículas de materia oscura y de esta manera modelamos la formación de estructuras en el régimen no lineal.

2 Función de correlación y espectro de potencias

Abstract

Having presented several aspects of the theory of structure formation, we will now introduce the statistical treatment to describe the perturbations statistically. For this, in cosmology, are used statistical quantities such as the two-point correlation function, which is defined as the excess probability, in respect of a random distribution, to find a pair of objects separated by a distance r . Also, the power spectrum turns out to be another fundamental statistical tool for the description of fields in density, which has been defined as the average in the set, of the amplitude of the k mode of the density field in the Fourier space.

In this chapter the mathematical expressions of both functions statistics are developed and obtained, obtaining as a result that the power spectrum is the Fourier transform of the correlation function. Therefore, both functions contain the same information, being suitable the use of one or another function according to utility and context.

Tanto por las condiciones iniciales supuestamente generadas por fluctuaciones cuánticas aleatorias, como por la subsecuente evolución para la cual nunca disponemos de toda la información, el universo observado se puede describir como una instanciación de una teoría probabilística. Por lo que para describir las perturbaciones estadísticamente, en cosmología se emplean cantidades como la función de correlación ξ entre puntos separados por un vector \vec{r} , es decir, las denominadas funciones de correlación de n -puntos, aunque en este trabajo nos centraremos en estudiar la función de correlación de dos puntos (Peebles 1980). Para ello, Consideremos una densidad homogénea de objetos. De manera que la probabilidad de encontrar un objeto en un elemento de volumen dV , es por definición $dP = \rho dV$. No obstante, si se desea calcular la probabilidad conjunta de encontrar un par de objetos a una distancia r , en volúmenes dV_1 y dV_2 , por homogeneidad, se

obtiene que dicha probabilidad es igual al producto de las probabilidades independientes, es decir $dP = \rho^2 dV_1 dV_2$. Sin embargo, en el caso de una distribución de materia, la gravedad actuará como una fuerza de atracción entre pares de partículas, produciendo de esta manera una distribución inhomogénea de materia. Por lo que en este caso la probabilidad bipuntual es:

$$dP = \rho^2(1 + \xi(r_{12}))dV_1 dV_2 \quad (1)$$

Por lo tanto la función correlación de dos puntos se define como el exceso de probabilidad, respecto de una distribución aleatoria, de encontrar un par de objetos en los elementos de volumen dV_1 y dV_2 separados por una distancia r_{12} .

Luego, si definimos el operador $\langle \dots \rangle$ y el valor medio se calcula sobre todo el volumen para un valor fijo de la distancia entre los dos puntos, se puede expresar la correlación de la siguiente manera:

$$\xi(r) = \langle \delta(\vec{x} + \vec{r})\delta(\vec{x}) \rangle, \quad (2)$$

donde $\delta(x)$ representa el contraste de densidad:

$$\delta(\vec{x}) = \frac{\delta\rho(\vec{x})}{\bar{\rho}} = \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}} \quad (3)$$

No obstante, se puede reescribir la expresión anterior considerando que el campo de densidades $\delta(\vec{x})$ se puede también expresar como una suma de modos. De manera que en límite continuo se tiene que el campo de densidades viene dado por:

$$\delta(\vec{x}) = \frac{L^3}{(2\pi)^3} \int \delta(\vec{k}) e^{i\vec{k}\vec{x}} d^3k \quad (4)$$

Por lo tanto, utilizando la ecuación (4) se llega a:

$$\xi(r) = \langle \delta(\vec{x})\delta(\vec{x} + \vec{r}) \rangle = \left\langle \frac{V^2}{(2\pi)^6} \int \int \delta(\vec{k})\delta^*(\vec{k}') e^{-i(\vec{k}' - \vec{k})\vec{x}} e^{-i\vec{k}'\vec{r}} d^3\vec{k} d^3\vec{k}' \right\rangle \quad (5)$$

de forma que haciendo uso de que los modos son variables reales en el espacio de coordenadas (para reproducir correctamente la realidad), de la linealidad del operador $\langle \dots \rangle$, y la transformada de Fourier de la delta de Dirac se obtiene:

$$\xi(r) = \langle \delta(\vec{x})\delta(\vec{x} + \vec{r}) \rangle = \frac{V}{(2\pi)^3} \int \langle |\delta(\vec{k})|^2 \rangle e^{-i\vec{k}\vec{r}} d^3\vec{k} \quad (6)$$

Por otra parte, si consideramos el principio cosmológico, de manera que en escalas espaciales suficientemente grandes, el Universo es homogéneo e isótropo, se esperaría que las fluctuaciones o campos de fluctuaciones sean homogéneos también; esto nos indica que la distribución de δ debe tener un valor central distinto de cero. Por consiguiente, las perturbaciones se describen como campos gaussianos aleatorios en el espacio de Fourier (perturbaciones no correlacionadas), es decir que cada δ se corresponde con una distribución gaussiana independiente.

Ahora, considerando las propiedades estadísticas de una distribución, podemos fijarnos en los momentos de esta, los que serán, para una distribución gaussiana, iguales a cero menos el de orden uno. Siendo dicho modo lo que se conoce como espectro de potencias.

El espectro de potencias resulta ser otra herramienta estadística fundamental para la descripción de campos en densidad, abreviado $P(k)$. El cual, es definido como el promedio en el conjunto, de la amplitud del modo k del campo de densidad en espacio de Fourier. Es decir:

$$P(k) = \langle |\delta^2(\vec{k})| \rangle \quad (7)$$

Luego, introduciendo dicha definición de espectro de potencias en la ecuación (6), la función de correlación queda expresada de la siguiente manera:

$$\xi(r) \equiv \langle \delta(\vec{x})\delta(\vec{x} + \vec{r}) \rangle = \frac{V}{(2\pi)^3} \int P(k) e^{-i\vec{k}\vec{r}} d^3\vec{k} \quad (8)$$

Se puede contemplar de la ecuación anterior (8), que el espectro de potencias (P(k)) resulta ser la transformada de Fourier (Joseph Fourier) de la función de correlación. Por lo tanto, ambas funciones contienen la misma información, siendo adecuado el uso de una u otra función según la utilidad y el contexto.

Además, como asumimos un Universo isotrópico, el espectro de potencia también lo será, por lo tanto el espectro de potencias solo dependerá del módulo del vector $\vec{k}(|\vec{k}|)$. De forma que expandiendo $d^3\vec{k} = k^2 dk \sin\Theta_k d\theta_k d\phi_k$, y haciendo rotar el eje polar hacia la dirección del vector \vec{k} , y además sabiendo que ξ es real, podemos integrar la dependencia angular y reescribir la ecuación anterior de la siguiente forma:

$$\xi(r) = \frac{V}{(2\pi)^3} \int P(k) \frac{\sin(kr)}{kr} 4\pi k^2 dk \quad (9)$$

Por otro lado, usualmente se define el espectro de potencias adimensional $\Delta^2(k)$ como:

$$\Delta^2(k) = \frac{1}{2\pi^2} k^3 P(k) \quad (10)$$

Dicha ecuación nos da información de la varianza por intervalo logaritmico k del campo. Siendo además una forma más intuitiva ya que cuando $\Delta^2(k) \text{ es } \sim 1$ significa que dicho modo no es lineal y por ende su amplitud no es pequeña.

3 Ajuste del espectro de potencias de la simulación SMDPL

Abstract

The goal of the work is to learn and test state-of-the-art techniques in the generation of catalogs of low-mass dark matter halos, and once generated, it will proceed to calculate the power spectrum of said simulated catalog. In order to verify that the calculated power spectrum is adjusted to the power spectrum of the Small MultiDark Planck (SMDPL) simulation. The details of the reference simulation will also be displayed.

Our method relies on low-resolution approximate gravity solvers to simulate the large scale dark matter field, which we then populate with halos according to a flexible nonlinear and stochastic Bias model. In particular, we extend the *Patchy* code with an efficient particle mesh algorithm to simulate the dark matter field (the *FastPM* code), and with a robust MCMC method relying on the emcee code for constraining the parameters of the Bias model. Using the halos in the The Small MultiDark Planck highresolution N-body simulation as a reference catalog.

In this part of the paper, besides describing the different codes that they allow us to achieve the proposed goals, it is also presented a brief introduction on the N-body simulations and the Bias model used by the *Patchy* code. In this way, it is easier to understand the role played by these codes

Finally, we discuss the data obtained and we demonstrate that our technique can model the power spectrum, and therefore the two-point correlation function, in the reference simulation. Specifically, we show that our method permit us to reach percentage accuracy in the power spectrum up to $k \sim 0.4hMpc^{-1}$ (within 5%). However, the histogram of halos in cells does not fit precisely. Therefore we must make an effort in the future to improve the models explored in this study and so adjust the halo PDF and ultimately the correlation function up to four points.

3.1 Competencias a realizar:

En el primer apartado del trabajo, se presenta una breve introducción de la teoría que describe las fluctuaciones de densidad de materia del universo a orden lineal y no lineal, con el fin de explicar el importante papel que tiene la existencia de fluctuaciones primordiales de densidad de materia en la formación de la estructura a gran escala del universo.

Además se han introducido conceptos como el espectro de potencias y la función de correlación, que resultan ser herramientas estadísticas importantes a la hora de describir la estadística de campos en densidad.

La formación de la estructura no lineal no se ha descrito con mucho detalle pero vamos a tratarla en lo que sigue ya que forma parte fundamental de este trabajo.

Por lo que a continuación se procederá a explicar las distintas tareas y competencias realizadas en el IAC para llevar a cabo dicho trabajo.

El objetivo del trabajo consistirá en aprender y poner a prueba las técnicas de estado de arte en la generación de catálogos de halos de materia oscura para surveys de galaxias, y una vez generados, se procederá a obtener el espectro de potencias de dicho catálogo simulado, con el fin de comprobar que el espectro de potencias calculado se ajusta (hasta un cierto valor de escala) a el espectro de potencias de la simulación The Small MultiDark Planck (SMDPL), la cual pertenece a la serie de simulaciones MultiDark con cosmología Planck. Por lo que para ello, en concreto, se utilizará el código *Patchy* en su nueva versión que incluye un módulo de particle mesh (fastPM, Feng et al 2016) para simular la evolución gravitacional y un código para encontrar los parámetros del Bias galáctico mediante cadenas de Markov (emcee, Foreman-Mackey et al 2013 y Vakili et al 2017). Con el fin de estudiar la precisión de dicha metodología e intentar reproducir correctamente la estadística de dos puntos de catálogos de halos de materia oscura poco masivos, simulaciones que presentan un volumen pequeño y realizadas con un bajo número de partículas de materia oscura.

3.2 Simulación The Small MultiDark Planck:

La simulación The Small MultiDark Planck, como ya se había mencionado anteriormente, pertenece a la serie de simulaciones MultiDark con cosmología Planck, proyecto español que se dedica a los esfuerzos de identificación y detección de la materia oscura. Dicha simulación presenta un volumen cúbico (Una caja) que tiene una longitud lateral de sólo $400 \frac{Mpc}{h}$, menos de la mitad del tamaño que la simulación MDPL. El fuerte de esta simulación reside en el gran número de partículas, 3840^3 partículas de materia oscura dentro de un volumen relativamente pequeño. Esto es fundamental para el estudio del Bias de objetos pocos masivos que queremos realizar aquí. También una resolución de masa de $108 \frac{M_{sun}}{h}$ por partícula de materia oscura. Algunos detalles más para esta simulación se describen en Klypin et al. 2014.(figura2).

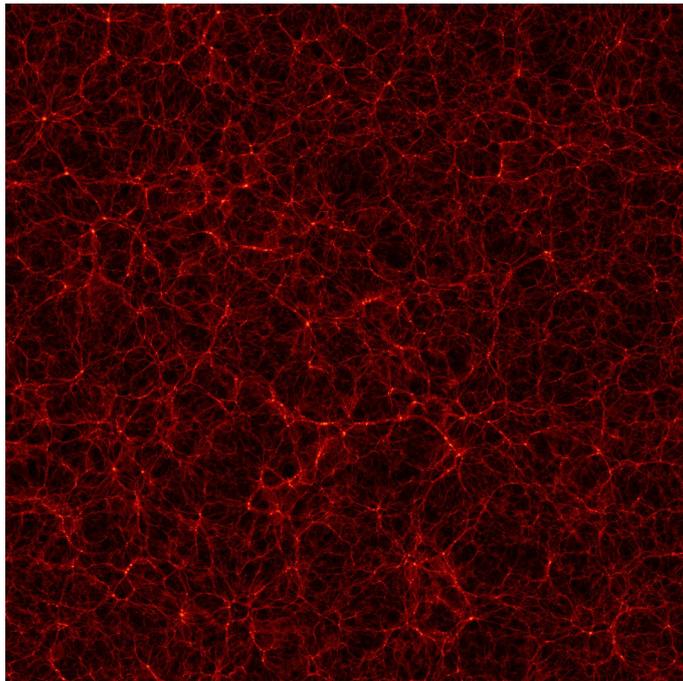


Figure 2: En esta imagen se puede observar una parte de la simulación SMDPL, cuyo espectro de potencias se quiere ajustar. Simulación que presenta un volumen de $400 \frac{Mpc^3}{h^3}$ y con 3840^3 partículas de materia oscura.

A continuación se presentaran en la siguiente tabla (figura3) los diferentes detalles de dicha simulación, entre los que cabe destacar la cosmología empleada y ciertas constantes físicas, que serán necesarias ahora de usar los códigos de *FastPM* y *Patchy*.

<i>Tamaño de la caja</i>	$400 \text{ Mpc}/h$	<i>Longitud lateral del cubo cosmológico</i>
<i>Número de partículas</i>	3840^3	<i>Número total de partículas de materia oscura</i>
<i>Resolución de masa</i>	$9.63 \cdot 10^7 M_{\odot}/h$	<i>Masa de una partícula de materia oscura</i>
<i>Resolución de fuerza</i>	$1.5 \text{ kpc}/h$	<i>Resolución de la fuerza física</i>
<i>Redshift inicial</i>	120	<i>Redshift en el que se inició la simulación</i>
<u>COSMOLOGÍA</u>		
<i>h</i>	0.6777	<i>Parámetro Hubble</i>
Ω_{Λ}	0.692885	<i>Parámetro de densidad para la energía oscura</i>
Ω_m	0.307115	<i>Densidad de la materia (materia oscura + bariones)</i>
Ω_b	0.048206	<i>Parámetro de densidad para la materia bariónica</i>
<i>n</i>	0.96	<i>Normalización del espectro de potencia</i>
σ_8	0.8228	<i>Amplitud de la fluctuación de la densidad de masa en esfera de $8 \text{ Mpc}/h$ (a redshift $z = 0$)</i>
<u>OTRAS CONSTANTES</u>		
<i>G</i>	$6.67428 \cdot 10^{-8} \text{ cm}^3 \text{ g}^{-1} \text{ s}^{-2}$	<i>Constante de la gravedad (Rev. Mod. Phys. 80 (2008) 633-730)</i>
<i>1 Mpc</i>	$30.85677 \cdot 10^{23} \text{ cm}$	<i>1 Mpc en cm</i>
M_{Sun}	$1.98892 \pm 0.00025 \cdot 10^{33} \text{ g}$	<i>La masa solar (Physics Reports 281 (1997) 309)</i>

Figure 3: Tabla en la que se muestran los diferentes variables empleadas en la simulación SMDPL.

Para llevar a cabo los objetivos del trabajo fue necesario el aprendizaje y la utilización de tres códigos: *FastPM*, *Patchy* y *mcmcwrapper* basado en *emcee*.

Por otro lado, cabe destacar que los catálogos de halos de materia oscura simulados haciendo uso de los códigos *FastPM* + *Patchy*, a pesar de presentar el mismo volumen que la simulación SMDPL ($400 \frac{Mpc^3}{h}$), solo fueron necesarias 160^3 partículas de materia oscura, siendo 13824 veces menor que el número de partículas de la simulación SMDPL. Además, es de señalar, que la computación de

dichas simulaciones requieren de un coste y un tiempo de realización bastante menor, resultando ser muy eficientes.

3.3 FastPM:

Antes de proceder, vamos a introducir el concepto de simulaciones de N cuerpos, para así comprender la forma con la que solucionamos la gravedad para un sistema de n-cuerpos de partículas de materia oscura:

En las simulaciones de N cuerpos, el campo de densidad de materia oscura se desarrolla resolviendo las ecuaciones de Boltzmann y Poisson (L. G. Gouy en 1910 y completada por Chapman en 1913) considerando un fondo en expansión. De manera que los solucionadores de gravedad por partículas (PM) son una clase de solucionadores de N cuerpos en los que la evolución del campo de densidad se rige por la dinámica de las partículas de materia oscura. Estas partículas de materia oscura se desarrollan mediante un número finito de pasos. Además, en los solucionadores de PM, el campo de densidad se realiza en una malla con un esquema de interpolación (por ejemplo CIC) que asigna cada partículas con los puntos de dicha rejilla. Entonces, el potencial gravitatorio (y posteriormente la fuerza gravitatoria) se calcula resolviendo la ecuación de Poisson. Esta estimación basada en la red de las fuerzas gravitacionales da lugar a una compensación de la precisión en escalas comparables a la separación entre puntos de la red.

Por otro lado, los pasos de tiempo en una simulación de PM pueden ser lineales o espaciados logarítmicamente en el factor de escala. El espaciado logarítmico de dichos pasos tiene la desventaja de perder precisión en términos de potencia a pequeña escala. Mientras que la elección de pasos de tiempo que son lineales en el factor de escala, da lugar a una agrupación más precisa en pequeñas escalas. Es importante señalar que el número finito de pasos de tiempo en las simulaciones de PM da como resultado un crecimiento a gran escala inexacto.

Sin embargo, en el código *FastPM*, las correcciones se aplican a las ecuaciones que rigen los movimientos de las partículas, de tal manera que se exige un crecimiento a gran escala de forma exacta. Por lo tanto, el error en el crecimiento a gran escala se corrige mediante el uso de las ecuaciones Zeldovic del movimiento para modificar el funcionamiento de los operadores kick y drift en un algoritmo PM puro. Es de destacar que en una simulación PM, el operador Drift cambia la posición de cada partícula manteniendo su momento fijo. Mientras que por otro lado, el operador

Kick cambia el momento de cada partícula, dejando su posición sin cambios. Dichos operadores modificados en el código de *FastPM* se derivan de la integración de las ecuaciones de movimiento Zeldovic. De manera, que se consigue imponer en el código de *FastPM* un crecimiento a gran escala exacto y realizado en un límite de tiempos infinitesimales, al converger los operadores Kick y Drift con los de un algoritmo estándar de PM (Approximation Zeldovich approximation (Chuang et al. 2015a)).

Por lo tanto, el primer programa en utilizarse y aprender su correcto funcionamiento fue *FastPM* (the FastPM code, see Feng et al. 2016). Dicho código, como se ha comentado, es un solucionador de N-cuerpos de malla de partículas aproximado altamente escalable, el cual resuelve la ecuación de Poisson (Siméon-Denis Poisson, que la publicó en 1812) de la gravedad con una malla de partículas (particle mesh) potenciada. Dando como resultado un crecimiento exacto del campo de materia oscura a gran escala sobre una malla, y obteniendo así una simulación gravitacional de la materia oscura. Por otro lado, dicho código presenta un input, en el cual es necesario introducir las siguientes variables para poder generar la simulación gravitacional de la materia oscura:

- Las condiciones iniciales con la que fue hecha la simulación SMDPL.
- El número de partículas de materia oscura necesarias para obtener la simulación gravitacional ($N = 160^3$).
- La masa de cada partícula.
- El desplazamiento al rojo (redshift) de la caja de la simulación SMDPL elegida.
- El volumen de la caja a simular.
- La cosmología usada en la simulación SMDPL y el valor de la constante de Hubble empleado (figura2).
- El factor de escala asociado a dicho desplazamiento al rojo (redshift).

Haciendo uso de la siguiente expresión se pudo obtener el valor del factor de escala correspondiente a dicho desplazamiento al rojo (redshift).

$$z + 1 = \frac{a(t_r)}{a(t_e)} \tag{11}$$

Ya que en cosmología una vez conocido el valor del desplazamiento al rojo z (redshift), se puede obtener información del factor de escala $a(t_e)$ en el tiempo en el que se emitió el fotón, siempre y cuando se conozca el factor de escala $a(t_r)$ en el tiempo en el que fue recibido. Por lo que el desplazamiento al rojo z resulta ser de gran importancia, ya que permite hablar de tiempo y de distancia en términos del factor de escala.

Es habitual etiquetar $a(t_r) = 1$ para la época actual, de tal manera que en un tiempo pasado la misma escala de distancia (como por ejemplo la distancia entre dos galaxias lejanas dadas) disminuye en un factor dado por $a(t_e)$. En otras palabras, si el desplazamiento al rojo medido en una galaxia lejana fuera $z = 1$ significaría $a(t_e = \frac{1}{2})$ y por tanto que dicha galaxia y la nuestra se encontraban, en el momento en que fue emitida la luz, a la mitad de distancia de lo que están hoy en día. De manera que haciendo uso de $a(t_r) = 1$ en la expresión anterior, se puede llegar a la siguiente ecuación:

$$a(t_e) = \frac{1}{z + 1} \quad (12)$$

Por otro lado, la caja de la simulación SMDPL elegida, fue la que presentaba un desplazamiento al rojo $z = 0.5618$ y haciendo uso de la ecuación anterior (12), dicho redshift se corresponde con un factor de escala $a(t_e) = 0.6403$. Cabe destacar, como se ha mencionado anteriormente, que los valores del factor de escala, el redshift y la cosmología introducidos en el input del código de *FastPM* deben coincidir con los de la caja de la simulación SMDPL elegida (caja con $z = 0.5618$) y además si se incluyen las condiciones iniciales de dicha simulación, esto nos permitirá poder hacer una comparación entre la simulación obtenida mediante el uso de los códigos de *FastPM + Patchy* con la simulación de referencia (SMDPL), ya que ambas deberían simular la misma formación de estructura. Para luego, finalmente equiparar los espectros de potencias de ambos catálogos de halos de materia oscura y así estudiar la eficiencia de nuestra metodología a la hora de intentar reproducir la estadística de dos puntos de dicha simulación de referencia.

De manera que, una vez introducido las correspondientes variables en el input y haciendo correr el código de *FastPM* mediante el uso de la terminal, el cual se ejecutaba bastante rápido (en torno a 30 segundos), este genera la simulación gravitacional de las partículas de materia oscura correspondientes a dicho input. No obstante, se prepararon dos input distintos, de forma que en ambos se introdujeron los mismos valores que se han comentado con anterioridad para las

diferentes variables, pero sin embargo, en uno de ellos, en lugar de usar las condiciones iniciales de la simulación SMDPL, se emplearon condiciones iniciales aleatorias para también su posterior análisis. Por lo tanto, haciendo correr dicho programa para cada uno de los inputs, obtuvimos dos simulaciones gravitacionales generadas con condiciones iniciales diferentes.

3.4 Código Patchy:

A continuación, se procedió en segundo lugar a utilizar el código *Patchy* (Kitaura et al 2014), con el cual haciendo uso de las distribuciones de las partículas de materia oscura anteriormente obtenidas con *FastPM*, es decir una vez producido el campo de materia oscura a gran escala en una malla, el código *Patchy* nos permite rellenar dicha malla con halos de materia oscura introduciendo un modelo de Bias determinista no lineal, consiguiendo de esta forma la transición del campo continuo de materia oscura (generado con *FastPM*) a la distribución discreta de halos. De manera dicho código es capaz de generar catálogos de halos de materia oscura (objetos colapsados) permitiendo reconstruir campos de densidad, de velocidad y posición. Sin embargo, la producción de catálogos simulados aproximados con *Patchy* requieren de un catálogo de referencia construido a partir de las observaciones o basado en una simulación exacta del N-cuerpo, que en nuestro caso usaremos la simulación de N-cuerpos SMDPL. Nuestro objetivo consiste en restringir los parámetros que describen el sesgo determinista ($\delta_{th}, \alpha, \rho_\epsilon, \epsilon$) y el parámetro que rige la estocasticidad de la población de halos β . Los parámetros de sesgo se estiman de tal manera que los resúmenes estadísticos de los halos en las simulaciones generadas con *Patchy* coinciden con los resúmenes estadísticos de los halos en el catálogo de referencia.

Para poner en funcionamiento dicho código, al igual que *FastPM*, dispone de un input, en el cual es necesario introducir las siguientes variables:

- El número de mallas a utilizar, que debe coincidir con el número de partículas de materia oscura empleado en el código *FastPM*. En nuestro caso $N = 160$ por cada lado del cubo.
- El volumen de la caja a simular, el cual debe ser equivalente al usado con *FastPM*. Estableciendo $L=400 \frac{Mpc}{h}$, longitud lateral de dicha caja.
- El desplazamiento al rojo (redshift) de la caja de la simulación SMDPL elegida ($z = 0.5618$).
- La cosmología usada en la simulación SMDPL y el valor de la constante de Hubble empleado (figura2).

- El correspondiente valor de la densidad media de objetos colapsados de la simulación SMDPL.
- El directorio en el que se encuentran los archivos generados por *FastPM*, es decir las simulaciones gravitacionales ya comentadas.

Por otra parte, además el código *Patchy* dispone de tres métodos diferentes, modelos que permiten simular la formación de la estructura. Dichos modelos se basan en teoría de perturbaciones Lagrangiana (Lagrangian perturbation theory (LPT))

- A primer orden: Zeldovich approximation.
- A segundo orden incluyendo fuerzas de marea.
- Y una versión que le añade al segundo orden correcciones a escalas pequeñas basadas en el modelo de colapso esférico: Augmented Lagrangian perturbation theory: ALPT.

Sin embargo en nuestro caso, en lugar de usar uno de estos métodos, se hizo uso del código *FastPM*. Por lo que en el input de *Patchy* había que elegir el modelo *FastPM*, para que así dicho código pueda leer los archivos generados por este último (las simulaciones gravitacionales de la materia oscura). Por lo tanto, fue necesario preparar dos input, uno con el directorio de la simulación gravitacional realizada con las condiciones iniciales de la SMDPL y el otro con el directorio del archivo generado con condiciones aleatorias, permanciendo las demás variables con los valores ya comentados.

De manera que una vez ejecutado el código *Patchy*, este nos proporciona para cada uno de los inputs, un espectro de potencias, una PDF, de la cual se hablará más adelante, un catálogo de halos de materia oscura y un campo de posiciones, velocidades y densidades de dichos objetos colapsados. Luego además, mediante el uso de un programa en python, conseguimos obtener las representaciones gráficas de los campos o mapas de densidades.

En la siguiente figura (Figura4) se puede observar una representación bidimensional del campo de densidades del catálogo de halos de materia oscura, el cual se obtuvo haciendo uso de los códigos *FastPM + Patchy* en el que se ha tenido en cuenta las condiciones iniciales de la simulación SMDPL.

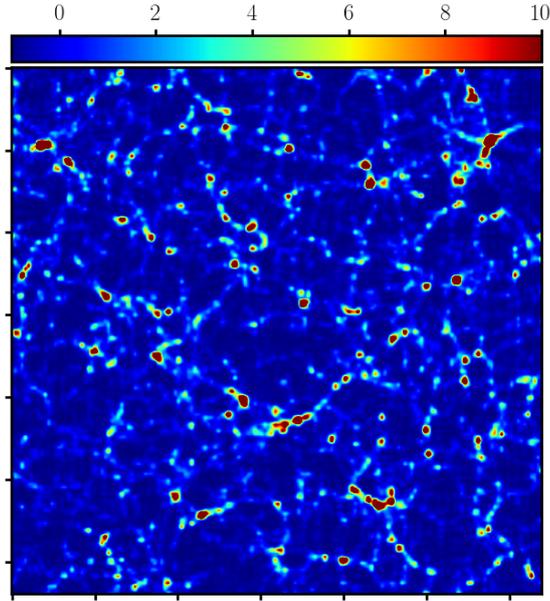


Figure 4: Representación gráfica del campo de densidades de la simulación de halos de materia oscura obtenidos con los códigos *FastPM* + *Patchy* y con las condiciones iniciales de la SMDPL. Simulación que presenta un volumen de $400 \frac{Mpc^3}{h^3}$ y con 160^3 partículas de materia oscura.

3.5 Modelo de Bias:

En esta sección, describimos el modelo de sesgo estadístico usado por el código *Patchy*. Dicho modelo consiste en:

- Un sesgo determinista.
- Un sesgo estocástico.
- Un paso adicional para aplicar distorsiones de desplazamiento al rojo (RSDs) a los catálogos.

Sin embargo, solo se describirán los dos primeros puntos.

3.5.1 Bias determinista:

El número esperado de halos ρ_h en un elemento diferencial de volumen dado (de la celda cósmica) se puede describir en general a partir de una relación de polarización determinista $B(\rho_h|\rho_m)$:

$$\langle \rho_h \rangle_{dV} = f_h B(\rho_h|\rho_m), \quad (13)$$

donde ρ_m es el campo de densidad de materia. El prefactor f_h es un factor de normalización global que puede determinarse exigiendo que el campo de densidad de halos tenga la densidad numérica de la muestra de referencia n_h , es decir, $n_h = \langle \langle \rho_h \rangle_{dV} \rangle_V$. De manera que, formalmente esto puede escribirse de la siguiente forma:

$$f_h = \frac{n_h}{\langle B(\rho_h|\rho_m) \rangle_V} \quad (14)$$

donde $\langle . \rangle_V$ es un promedio de volumen sobre un conjunto. En particular, adoptaremos el siguiente modelo de sesgo determinista compacto:

$$B(\rho_h|\rho_m) = \underbrace{\rho_m^\alpha}_{\text{Bias no lineal}} \times \underbrace{\theta(\rho_m - \rho_{th})}_{\text{Umbral Bias}} \times \underbrace{\exp(-(\frac{\rho_m}{\rho_\epsilon})^\epsilon)}_{\text{corte exponencial}}$$

donde ρ_{th} es el umbral de densidad que suprime la formación de halos en regiones de bajo densidad, y α es un parámetro de sesgo no lineal.

El sesgo de umbral (Kaiser 1984, Bardeen et al., 1986, Sheth et al., 2001, Mo & White 2002) es modelado por una función escalón $\theta(\rho_m - \rho_{th})$ (Kitaura y otros, 2014) y el corte exponencial $\exp(-(\frac{\rho_m}{\rho_\epsilon})^\epsilon)$ (Neyrinck et al., 2014). Por lo tanto, para este modelo de sesgo particular tenemos la siguiente normalización:

$$f_h = \frac{n_h}{\langle \theta(\rho_m - \rho_{th}) \rho_m^\alpha \exp(-(\frac{\rho_m}{\rho_\epsilon})^\epsilon) \rangle_V} \quad (15)$$

La ventaja de este tipo de modelo de sesgo es que, como se había comentado anteriormente, es

flexible y es capaz de incorporar términos adicionales, de manera que cada uno de los términos tiene una interpretación física.

El sesgo de la ley de potencia representa uno de los modelos de sesgo no lineal más sencillo posible: un sesgo Lagrangiano lineal en un marco conmovedor, que puede derivarse de la aproximación log-normal (véase Kitaura et al., 2014) y reanuda en un único parámetro de polarización el desarrollo en serie de Teylor (Brook Taylor,1715) del campo de densidad de materia oscura (Cen & Ostriker 1993, Fry & Gaztanaga 1993, de la Torre y Peacock, 2013).

El sesgo umbral y el corte exponencial describen el hecho de que los halos (o galaxias) sólo pueden residir en regiones que contienen una masa mínima. También representan la pérdida de información con respecto a todo el campo de densidad cósmica a la hora de seleccionar sólo objetos gravitacionalmente colapsados.

3.5.2 Bias estocástico

El número de halos en cada célula del mallado se obtiene de una distribución binomial negativa (NB, Blaise Pascal(1623-1662)), la cual puede caracterizarse por el número esperado de halos en cada celda $\lambda_h = \langle \rho_h \rangle_{dV}$, y un parámetro β que cuantifica la estocasticidad (desviación de la distribución Poissoniana) en la distribución del halo. De manera que, de acuerdo con este modelo, la probabilidad de tener objetos N_h en un elemento de volumen viene dada por:

$$P(N_h|\lambda_h, \beta) = \underbrace{\frac{\lambda_h^{N_h}}{N_h!} e^{-\lambda_h}}_{\text{Distribucin de Poisson}} \times \underbrace{\frac{\Gamma(\beta + N_h)}{\Gamma(\beta)\Gamma(\beta + N_h)^{N_h}}}_{\text{Desviacin Poissoniana}} \times \frac{e^{\lambda_h}}{(1 + \lambda_h/\beta)^\beta}$$

Cuando $\beta \rightarrow \text{inf}$ se puede demostrar que los términos de la segunda fila en la ecuación anterior tienden a uno.

Dado un campo de densidad de materia oscura ρ_m , el campo de densidad de halos puede con-

struirse dibujando muestras del campo de densidad esperado ρ_h mediante el uso de la distribución Binomial Negativo (Nb) (Eq. anterior). Esto está inspirado en el hecho de que la excesiva probabilidad de encontrar halos en regiones de alta densidad genera sobredispersión (Somerville et al., 2001, Casas-Miranda et al., 2002). Esta sobredispersión se modela mediante una distribución NB (Kitaura et al., 2014, Neyrinck et al., 2014).

El sesgo estocástico representa el shot noise o Poisson noise de la transición del campo continuo de materia oscura a la distribución discreta de halo (o galaxia). Como predicho por Peebles (1980), produce una dispersión mayor que Poisson, siempre y cuando la función de correlación de dos puntos permanezca positiva por debajo de la escala del tamaño de la celda.

3.6 Aprendizaje automático para obtener los parámetros de Bias galáctico:

Una vez aprendido a generar catálogos de halos de materia oscura haciendo uso de los códigos *FastPM* + *Patchy*, se procedió a realizar la siguiente actividad: Conseguir encontrar los parámetros de Bias con los que se logra un mejor ajuste del espectro de potencias de dichos catálogos simulados con el espectro de potencias de la simulación SMDPL. Para ello, se hizo uso de un tercer código, *mcmcwrapper*, el cual mediante cadenas de Markov (Andréi Márkov, 1907), haciendo correr el código *Patchy* repetidamente en cada paso, consigue los mejores parámetros del Bias galáctico de forma automática.

Sin embargo, en primer lugar, fue necesario preparar dicho código para su correcto funcionamiento. Para ello, este disponía de un programa realizado en python, el cual tiene implementado el input del código *Patchy*. Por lo tanto, ya una vez aprendido a usar el código *Patchy*, se tuvo que volver a establecer los mismos valores para las siguientes variables:

- El número de mallas a utilizar, que debe coincidir con el número de partículas de materia oscura empleado en el código *FastPM*. En nuestro caso $N = 160$ por cada lado del cubo.
- El volumen de la caja a simular, el cual debe ser equivalente al usado con *FastPM*. Estableciendo $L=400 \frac{Mpc}{h}$, longitud lateral de dicha caja.
- El desplazamiento al rojo (redshift) de la caja de la simulación SMDPL elegida ($z=0.5618$).

- La cosmología usada en la simulación SMDPL y el valor de la constante de Hubble empleado(figura2).
- El correspondiente valor de la densidad media de objetos colapsados de la simulación SMDPL.
- El directorio en el que se encuentran los archivos generados por *FastPM*, es decir las simulaciones gravitacionales ya comentadas.

Por otro lado, además de dichas variables, que ya se habían mencioando anteriormente, fue necesario colocar en el input el mínimo valor entre dos puntos (del espacio de fourier) en los que se va a calcular el espectro de potencias, es decir el valor más pequeño que puede haber entre dos valores de número de ondas consecutivos (la diferencia entre k y $k+1$, siendo $k+1$ el segundo valor de número de ondas usado), ya que una vez utilizado correctamente el código *Patchy*, se genera un fichero, el cual presenta una columna con valores de número de ondas discretos, y para cada valor de k su correspondinte valor $P(K)$ en una segunda columna. De manera que el valor dk elegido, nos permite controlar el número de modos para cada valor de k y obtener un espectro de potencias más suavizado. Por lo tanto, para generar un espectro con un aspecto más suave, se utilizó un valor de $dk=0.015$. Además, se estableció el valor de k más pequeño como $k_{min} = 0.001$.

De manera que, una vez introducidos los valores de las difentes variables en el input de *Patchy*, se procedió a introducir los correspondientes directorios para poder ejecutar correctamente el programa. Para ello, se estableció la ubicación del ejecutable del código *Patchy* y los directorios de los archivos correspondientes a los espectros de potencias de nuestra simulaciones, los cuales se reescribirán en cada paso al leer dicho programa el input anteriormente definido y el espectro de potencias de la simualción SMDPL, la cual tomará como referencia durante cada ejecución.

Sin embargo, para que el código funcione y pueda encontrar los parámetros de Bias que mejor se ajustan, antes es necesario obtener el espectro de potencias de la simulación de N-cuerpos de referencia (SMDPL). Para generar dicha función, se utilizó por separado una nueva versión del código *Patchy*, ya que como se había mencionado en el apartado anterior, este también está diseñado para calcular espectros de potencias, pero fue necesario añadirle una serie de correcciones con el fin de obtener un espectro que no presentara los diferentes problemas que surgen a la hora de calcularlo y evitando así que los valores obtenidos se vean afectados por dichos errores. Por lo tanto, para calcular el espectro de potencias de la simulación de referencia, se preparó un nuevo input, en el cual en lugar de poner el directorio del fichero generado por *FastPM*, se introducía un fichero con los datos de las tres coordenadas espaciales, las tres coordenadas de las velocidades y la

masa de cada halo, obteniendo dichos datos del catálogo de la simulación SMDPL. Además para que el código *Patchy* fuera capaz de leer dicho fichero, cada variable tenía que estar agrupada en una columna, de manera que cada fila se corresponda con las magnitudes de cada halo. De manera que haciendo uso de este fichero y de las demás variables definidas anteriormente, obtuvimos el espectro de potencias correspondiente a la simulación SMDPL (figura5) y también su histograma de número de celdas con un número determinado de halos.

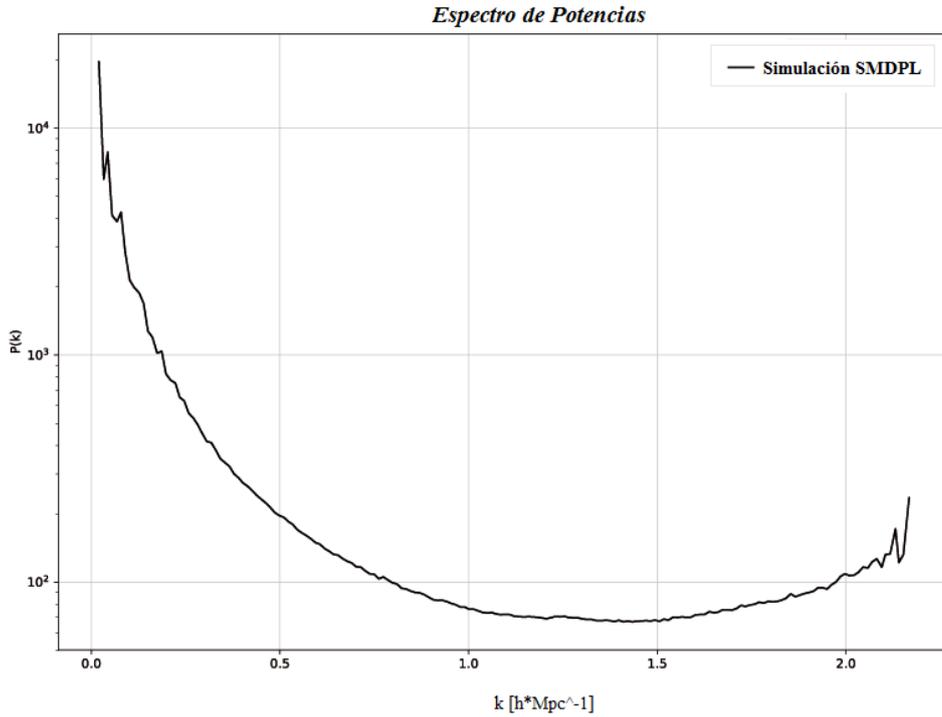


Figure 5: Espectro de potencias de la simulación SMDPL que se desea ajustar.

Además, cabe destacar, que para obtener el espectro de potencias que mejor se ajusta al de la simulación de referencia, se programó en dicho programa un fichero en el cual se iban guardado los cinco parámetros de Bias obtenidos en cada paso y el valor al cuadrado de una sexta cantidad, denominada chi-cuadrado, que denota la distancia cuadrática Euclídea entre el espectro de referencia de la simulación SMDPL y el simulado con *Patchy*.

$$\chi^2 = \sum_k \frac{(P_{ref}(k) - P_{sim}(k))^2}{\sigma_k^2} \quad (16)$$

Ya que como dichos espectros de potencias están realizados sobre el mismo mallado y con el mismo valor de dk , ambas funciones se calcularán sobre los mismos valores discretos de k . Por lo tanto, una vez puesta a correr la cadena, dicho fichero se irá llenando hasta que se realice el último paso establecido (Número de pasos, $n=10000$), de manera que una vez obtenido todos datos, realizando un sencillo programa en python, podemos extraer que parámetros de Bias presentan el menor valor de χ^2 y como dicha variable está relacionada con la distancia que separa a ambos espectros, hemos encontrado los parámetros de Bias que minimizan la distancia entre ambas funciones. También es de resaltar, que en código `mcmcwrapper`, había que añadir los valores máximos y mínimos de los cinco parámetros de Bias entre los que se espera encontrar los valores óptimos, de manera que la cadena irá buscando en dichos intervalos los parámetros que hacen que el valor de χ^2 sea mínimo.

A continuación, una vez preparado, se procedió a correr dicho código para cada uno de las simulaciones gravitacionales de materia oscura obtenidas (una para el caso en el que se utilizó condiciones iniciales y también otra haciendo uso de las condiciones aleatorias). Además es de destacar que antes de dejar corriendo las diferentes cadenas, se hacía uso de un comando en python que nos permitía detener el código una vez se ejecutaba el primer paso. Con el objetivo de facilitar al programa la búsqueda de los parámetros de Bias, ya que obligando a detener dicho código, podíamos intentar ajustar visualmente mediante la representación gráfica de ambos espectros de potencias (es decir, el obtenido en el primer paso y el de referencia) al ir modificando los valores iniciales de dichos parámetros (elegiendo valores dentro de los intervalos asignados para cada uno) con los que el código empieza a probar e iniciar las respectivas cadenas. De forma que después de lograr un ajuste manual, se dejaban corriendo las cadenas aproximadamente durante cuatro días para cada una.

3.7 Resultados y discusión:

A continuación se procederá a analizar las estadísticas obtenidas para cada una de las cadenas mediante la obtención de dos representaciones. La primera de ellas nos proporciona bastante información sobre como el código `mcmcwrapper` va probando diferentes valores para cada uno de los cinco parámetro de Bias, hasta que encuentra unos en los que el programa se encuentra conforme respecto al ajuste, pasando a permanecer dichos valores prácticamente constantes.

En cambio, el segundo fichero nos proporciona los correspondientes valores estimados de los

cinco parámetros obtenidos mediante la representación de la distribución de probabilidad de los mismos en forma de dominios (contornos), en los cuales se nos muestra el valor medio que toma cada uno de los parámetros durante toda la cadena y con sus correspondientes errores.

Después de ejecutar las cadenas MCMC para la simulación gravitacional de materia oscura realizada haciendo uso de las condiciones iniciales de la SMDPL, encontramos las siguientes restricciones sobre los parámetros de polarización. Dichas limitaciones las podemos contemplar en las figuras 6 y 7.

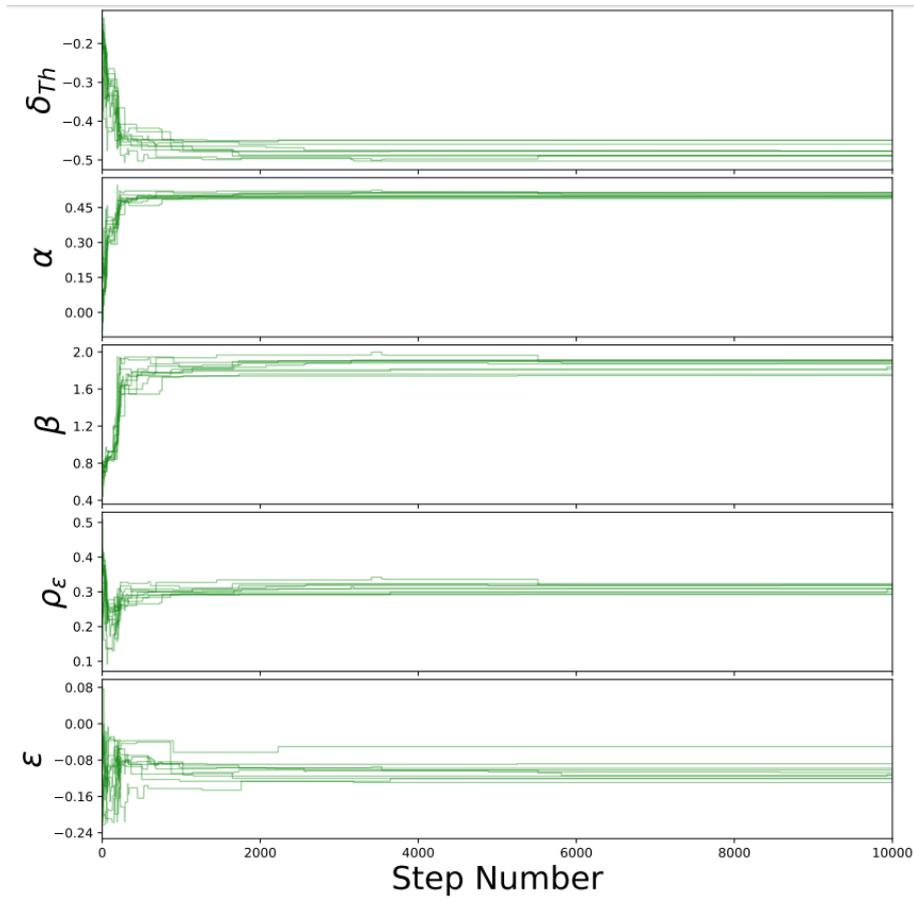


Figure 6: Representación gráfica que describe la búsqueda de los valores de los parámetros de Bias en cada paso $(\delta_{th}, \alpha, \beta, \rho_{\epsilon}, \epsilon)$.

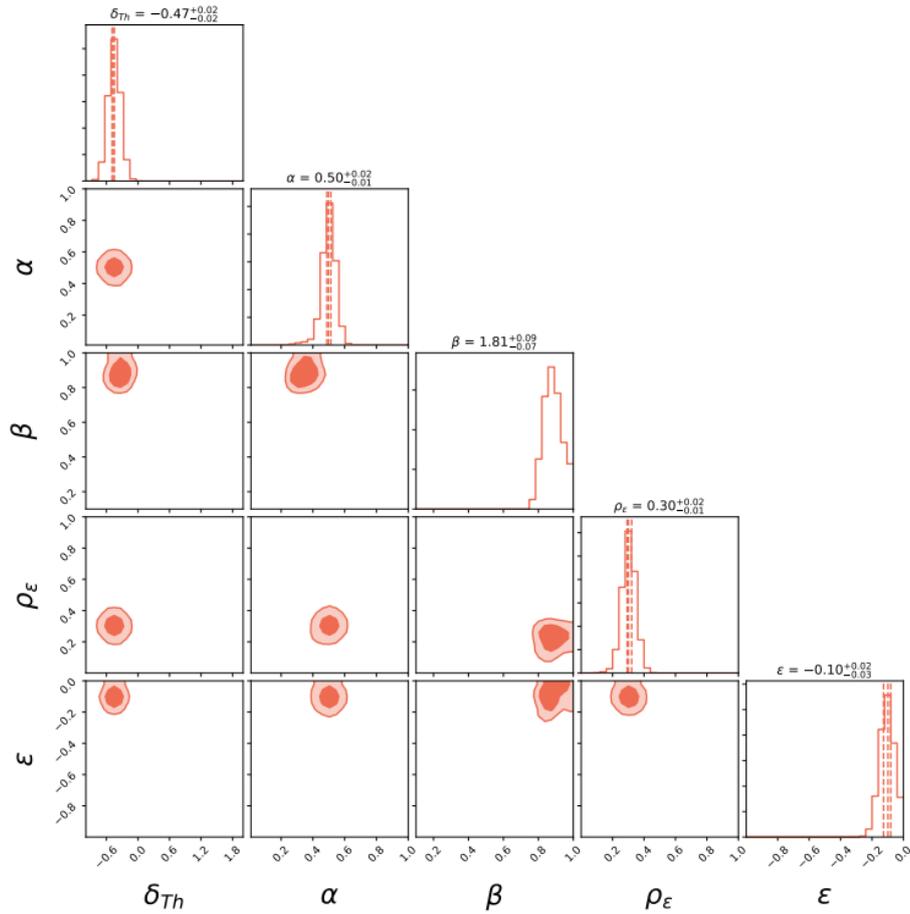


Figure 7: Distribución de probabilidad de los parámetros parciales de polarización ($\delta_{th}, \alpha, \beta, \rho_\epsilon, \epsilon$) con sus correspondientes valores estimados y errores

Se puede contemplar, que el parámetro de sesgo de umbral δ_{th} se encuentra en -0.47, valor que no se había contemplado en los estudios realizados, ya que en nuestro caso se está trabajando con halos de muestreo de masas menores, que tienen una probabilidad no despreciable de residir en regiones de baja densidad en comparación con los estudios llevados sobre campos de desidades que precisan de halos de muestreo en las regiones de alta sobredensidad de materia. Obteniendo un valor de δ_{th} mucho más pequeño y de signo contrario que los obtenidos con alpt (Kitaura et al., 2014).

Por otro lado, nuestro valor estimado del parámetro de polarización no lineal α es de 0.5, siendo cualitativamente más alto que los obtenidos con perturbación Lagrangiana (ALPT) (Kitaura et al., 2014: $\delta_{th} \sim 1.2$ y $\alpha \sim 0.12$) y con fastPM (Vakili et al 2017: $\delta_{th} \sim 1.07$ y $\alpha \sim 0.2$).

Por tanto hemos encontrado un régimen de parámetros de Bias nuevo que se debe al catálogo SMDPL incluyendo halos con un corte de masa considerablemente más bajo que el de los estudios previos que usaban como referencia la BigMDPL de lado de volúmen $2500 h^{-1} \text{ Mpc}$ y mismo número de partículas 3840^3 , por tanto con una resolución de masa de partícula de m_p que es $2.359 \times 10^{10} h^{-1} M_{\odot}$.

El parámetro que gobierna la desviación Poissoniana se encuentra en 1.81. Siendo este valor significativamente mayor que el encontrado con la BigMDPL (ALPT: 0.6 fastPM: 0.74), es decir, indicando que la desviación Poissoniana es menos pronunciada, a diferencia de como se ha encontrado previamente.

Esto es en una primera instancia inesperado. El catálogo de halos de la SMDPL contiene muchos más objetos que se agrupan en escalas pequeñas que la BigMDPL, por tanto generando una correlación a escalas pequeñas. Esta correlación debería contribuir a un exceso de probabilidad de encontrar objetos en elementos de volumen cósmicos y desviarse de la distribución idéntica e independiente que caracteriza a la distribución Poissoniana.

Sin embargo el parámetro de desviación Poissoniana que estamos manejando es global para toda la red cósmica y en consecuencia representa una parámetro estocástico efectivo que incluye zonas de baja densidad poco pobladas y zonas de alta densidad muy pobladas.

Además, los dos últimos parámetros ρ_e, ϵ se estiman en un valor $\sim 0.30, -0.10$ respectivamente. No obstante las limitaciones de ambos parámetros son practicamente consistentes con cero. Su presencia, aunque pequeña, sigue siendo esencial para lograr un buen ajuste del espectro de potencias.

Sin embargo en lugar de emplear dichos valores estimados para calcular el espectro de potencias y comparar su ajuste con el espectro de la simulación de referencia, se utilizaron los parámetros de Bias que presentan un menor valor de χ^2 como se ha comentando en el apartado 3.6. De manera que introduciendo dichos parámetros (los cuales no se desvian mucho de los valores estimados) en el código *Patchy*, se obtuvo el espectro de potencias para nuestra simulación. Cabe destacar que dicho tratamiento se repitió también para la cadena realizada haciendo uso de condiciones iniciales aleatorias.

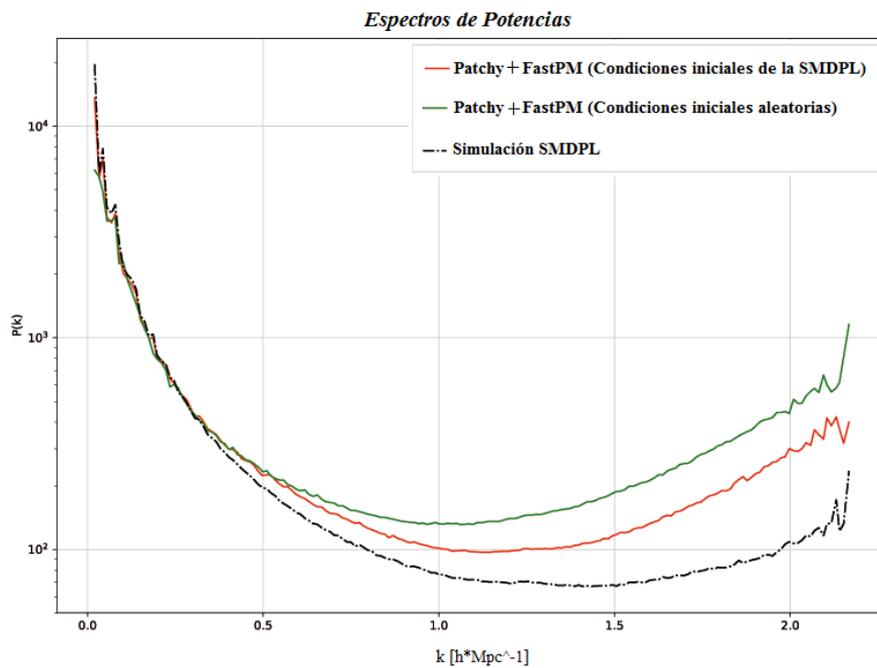
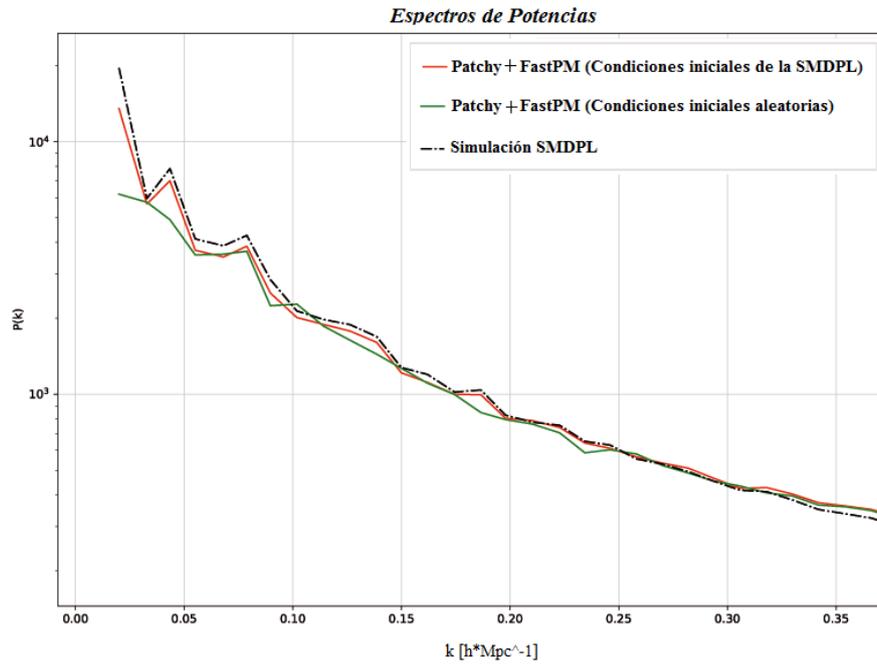


Figure 8: Comparación de los espectros de potencias de la simulaciones obtenidas haciendo uso de los códigos *FastPM* + *Patchy* mediante un método MCMC con el espectro de potencias de la simulación de referencia SMDPL

De las figuras anteriores (figura 8) se puede observar que se ha obtenido un ajuste bastante bueno del espectro de potencias de la simulación de referencia (curva roja) donde demostramos que nuestra técnica puede modelar correctamente dicho espectro. Permittiendonos alcanzar una precisión porcentual en el espectro de potencia dentro del 5% hasta un valor de $k \sim 0.35hMpc^{-1}$. Luego resulta bastante lógico que el espectro de potencias obtenido de la simulación con condiciones iniciales aleatorias (curva verde) se ajusta peor, de manera que los picos no coinciden con la de referencia, ya que se han empleado condiciones diferentes por lo que ambas simulaciones tienen una solución gravitacional distinta, pero aún así se observa que ambas se aproximan bastante hasta un valor de $k \sim 0.35hMpc^{-1}$.

A continuación se volvió a preparar una nueva cadena repitiendo los procedimientos comentados con anterioridad (preparandolo solo para la simulación realizada con las condiciones iniciales de la simulación de referencia), pero en este caso en lugar de utilizar el páremetro de Bias δ_{th} , mi tutor Francisco Shu Kitaura, implementó un nuevo parámetro de sesgo en el código *Patchy*, denominado gamma γ , con el cual sustituimos a δ_{th} y se obtuvieron los siguientes resultados una vez finalizada la cadena (figura 9 y figura 10).

El parámetro gamma describe la dependencia de la desviación de Poissonidad con la densidad local.

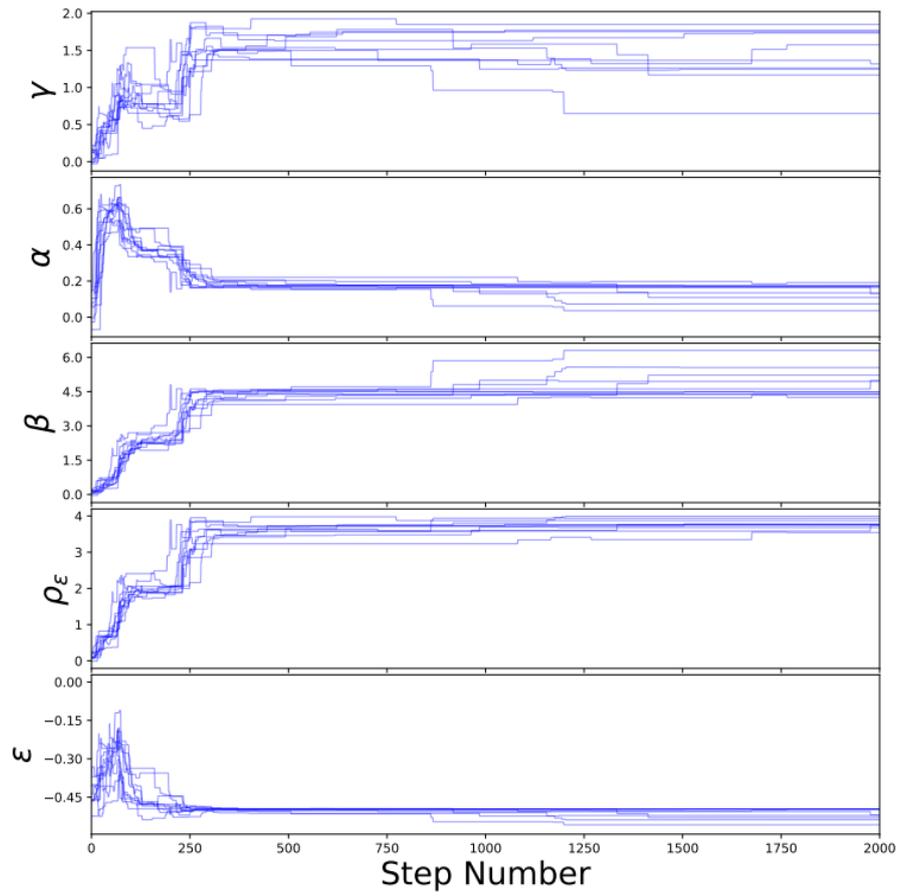


Figure 9: Representación gráfica que describe la búsqueda de los valores de los parámetros de Bias en cada paso $(\gamma, \alpha, \beta, \rho_\epsilon, \epsilon)$.

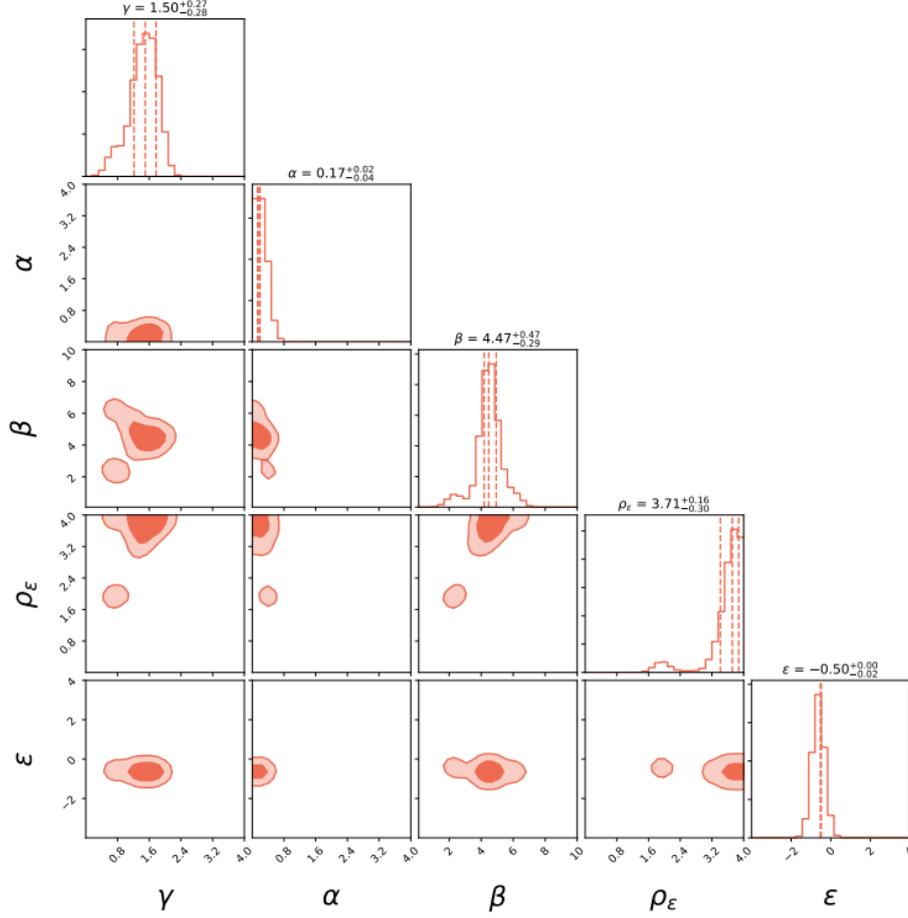


Figure 10: Distribución de probabilidad de los parámetros de Bias ($\gamma, \alpha, \beta, \rho_\epsilon, \epsilon$) con sus correspondientes valores estimados y errores.

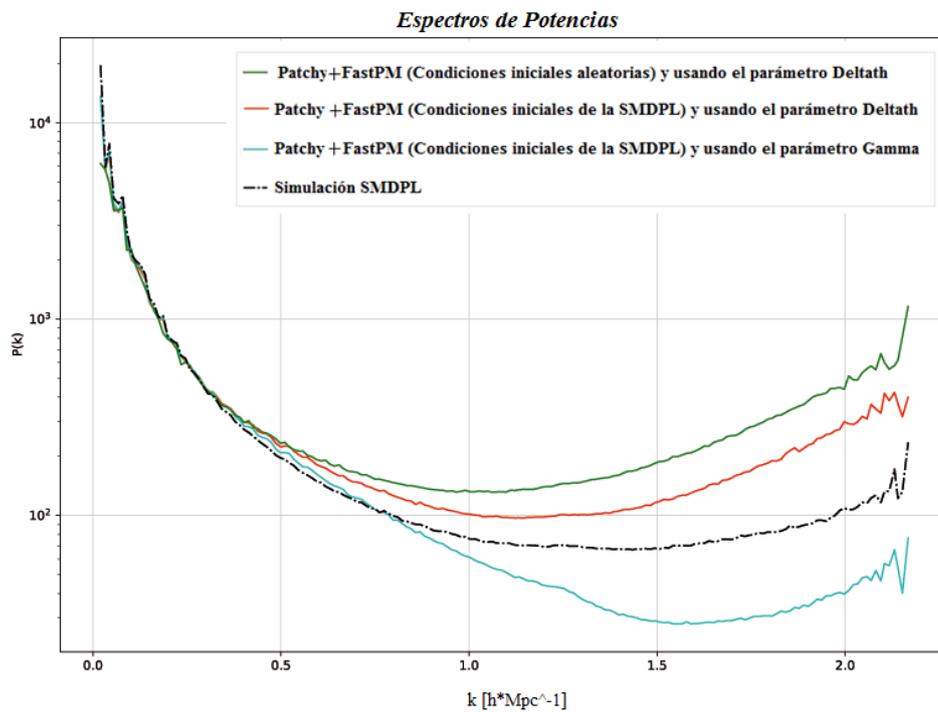
En este caso podemos observar como los valores estimados de los parámetros ($\beta, \rho_\epsilon, \epsilon$) han cambiado significativamente, aumentando su valor, mientras que el parámetro de polarización lial α se ha reducido a 0.17, siendo más pequeño que en caso anterior y ahora comparable con el valor obtenido con alpt (Kitaura et al., 2014)(parámetro con el alpt $\alpha \sim 0.12$).

En cambio el parámetro β se encuentra en 4.47. Siendo ahora dicho valor significativamente mayor que el encontrado en el caso anterior y del obtenido con alpt (aproximadamente 0.6), indicando también que la desviación Poissoniana es bastante menos pronunciada que las que se han encontrado previamente.

Además, los valores estimados de los parámetros ρ_ϵ, ϵ son ahora $\sim 3.71, -0.50$ respectivamente. No obstante, las limitaciones de ambos parámetros ya no son consistentes con cero, su presencia

presenta ahora una mayor importancia, ya que se cree que dichos parámetros serían de crucial importancia cuando se consideran halos de masas menores, que tienen una probabilidad no despreciable de residir en regiones de baja densidad (Neyrinck et al., 2014).

Luego, repitiendo el mismo procedimiento que en el caso anterior, encontramos los parámetros de Bias con los que logra el mejor ajuste del espectro de potencias. Dando como resultado el siguiente ajuste:



Espectros de Potencias

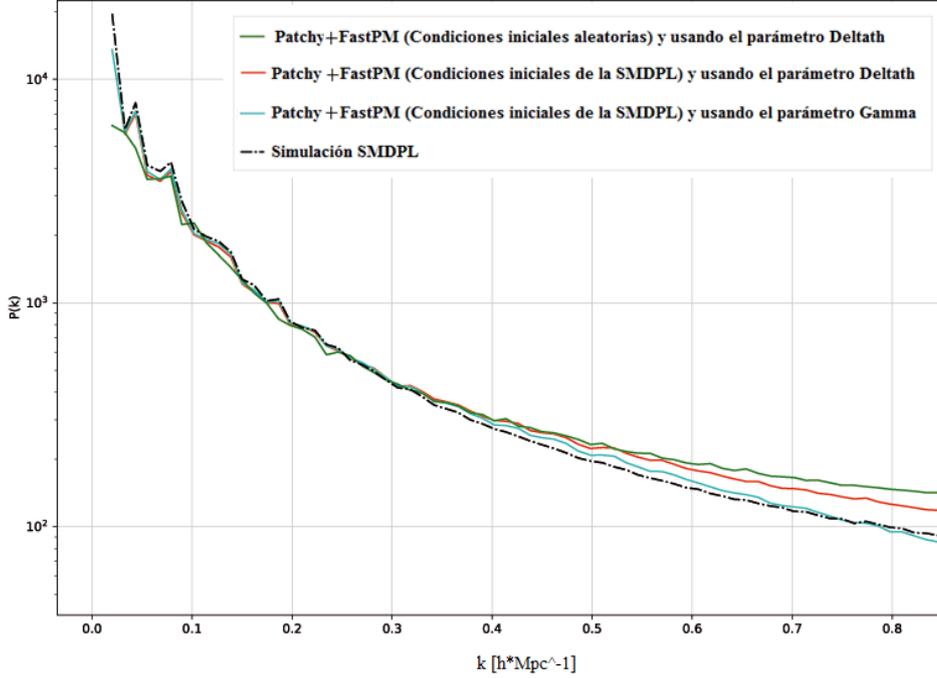


Figure 11: Comparación de los espectros de potencias de las simulaciones obtenidas haciendo uso de los códigos *FastPM* + *Patchy* mediante un método MCMC, utilizando el párametro δ_{th} (Primer caso tratado) y también utilizando el sesgo γ en lugar δ_{th} (segundo caso a tratar) con el espectro de potencias de la simulación de referencia SMDPL.

En las figuras anteriores (figura 11) se puede contemplar que el ajuste obtenido del espectro de potencias haciendo uso del parámetro de Bias γ (curva azul) se ajusta mejor que los resultados obtenidos en los casos anteriores, llegando a modelar correctamente el espectro de potencias de la simulación SMDPL para escalas de superiores a $k \sim 0.35hMpc^{-1}$. Permittiendonos alcanzar una precisión porcentual en el espectro de potencia dentro del 5% hasta un valor de $k \sim 0.4hMpc^{-1}$. Sin embargo dicho espectro se mantiene prácticamente ajustado hasta un valor de $k \sim 0.8hMpc^{-1}$ pero dentro de un porcentaje de error mayor que un 5%. De forma que podemos concluir, que el espectro de potencias obtenido haciendo uso del párametro γ , en lugar del término δ_{th} , resulta ser el que más se ajusta al de la simulación de N-cuerpos de referencia.

Debemos considerar que el parámetro δ_{th} describe el proceso físico según el cual solo se producen halos (o galaxias) en picos de densidad. Este modelo sin embargo es poco preciso para el catálogo de la SMDPL dado que halos de baja masa llegan a poblar zonas de baja densidad en

contraste con el catálogo de la BigMDPL que se usó en estudios anteriores.

De esta manera el término exponencial que reduce de forma progresiva la aparición de objetos en zonas de baja densidad es más apropiado en este estudio. Este trabajo deja clara la necesidad de describir de forma precisa la dependencia del Bias escolástico con el entorno de densidad.

Por otra parte, ya una vez cumplidos los objetivos del trabajo, se procedió a comprobar si con los correspondientes parámetros de Bias obtenidos, con los cuales se obtiene un buen ajuste del espectro de potencias de la simulación SMDPL, se consigue ajustar también las PDFs de nuestras simulaciones, las cuales las calculamos, como se había comentado en el apartado 3.4, haciendo correr el código *Patchy* como se ha explicado e introduciendo los valores encontrados para dichos parámetros. Cabe destacar que dicho ajuste nos interesa, ya que si tenemos que las funciones de correlación de dos puntos se ajustan correctamente, como es en nuestro caso (debido a que el espectro de potencias se corresponde con la transformada de Fourier de la función de correlación, de manera que están relacionados) y también lo están las Pdfs (que son funciones de correlación de un punto), entonces las funciones de correlación de ordenes superiores (como la función de correlación a tres puntos) también estarán correctamente ajustadas, consiguiendo de esta forma reproducir correctamente la estadística de la simulación de referencia. Además resulta importante aclarar que la PDF se corresponde con un histograma que nos da información sobre cuántos objetos (halos) hay en cada celda del mallado.

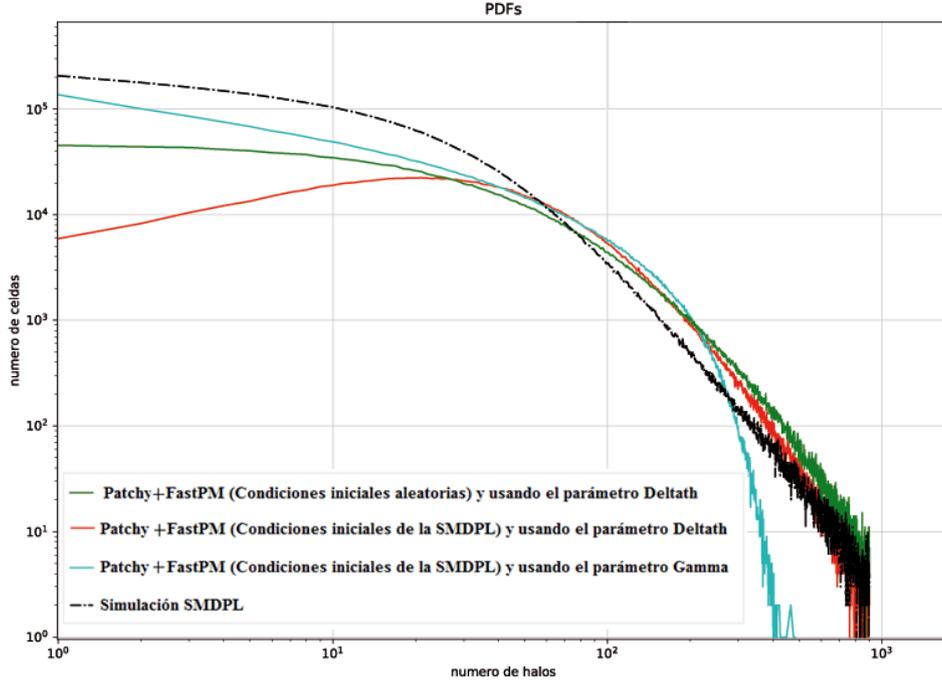


Figure 12: Comparación de las PDFs de las simulaciones obtenidas haciendo uso de los códigos *FastPM* + *Patchy* mediante un método MCMC, utilizando el parámetro δ_{th} (Primer caso tratado) y también utilizando el sesgo γ en lugar δ_{th} (segundo caso) con la PDF de la simulación de referencia SMDPL.

Sin embargo, observando la figura 12 podemos contemplar que para ninguna de las dos cadenas obtenidas se consigue ajustar correctamente la PDF de la simulación de referencia SMDPL, siendo dicho ajuste de bastante interés para estudios futuros, dado que este ajuste determina las funciones de correlación de órdenes mayores. La función de correlación de tres puntos se usa en conjunto con el espectro de potencias para romper la degeneración del Bias con la tasa de crecimiento cósmico (efecto de velocidades peculiares). Por otro lado la función de correlación de cuatro puntos determina las matrices de covarianza que se usan para el analizar los parámetros cosmológicos de datos observacionales. Por tanto debemos hacer un esfuerzo en el futuro para mejorar los modelos explorados en este estudio para ajustar también la PDF de los halos y en definitiva las funciones de correlación hasta cuatro puntos.

4 Conclusión:

Abstract

This work presents an advance towards the fast and accurate generation of simulated dark matter halos catalogs that have a small volume and a low number of particles.

Using our method, *FastPM + Patchy*, we show that our method can achieve a certain accuracy in the power spectrum within 5 % up to a value of $k \sim 0.4hMpc^{-1}$. In addition the catalogs of simulated dark matter halos using the codes *FastPM + Patchy*, only required 160^3 particles of dark matter, being 13824 times smaller than the number of particles in the SMDPL simulation.

It is also important to point out that our technique does not attempt to find halos in a low resolution simulation, but uses sub-grid models extracted from high-resolution simulations to accurately populate such a cosmic network.

The polarization parameters of the halos are to extracted in a calibrated statistical process through the use the reference simulations. Therefore our methodology does not replace the detailed numerical simulations of structure formation, but compliments them.

However, we have verified that for these objects we have not controlled all the statistics. Since we have obtained a very reasonable adjustment of the power spectrum, that is, second-order statistics, but the histogram of halos in cells does not fit precisely. This indicates that the statistics of higher order are not going to reproduce correctly with these models.

Este trabajo presenta un avance hacia la generación rápida y precisa de catálogos de halos de materia oscura simulados que disponen de un volumen pequeño y un bajo número de partículas, extendiéndose en particular, el código *Patchy*. Nuestro modelo consiste en primer lugar, en simular a gran escala el campo de materia oscura, para ello se ha introducido en el código *Patchy* un modelo de formación de estructura de malla de partículas (el código *FastPM*, véase Feng et al., 2016), de forma que luego dicho campo se poblará con halos de acuerdo con un modelo de sesgo flexible,

no lineal y estocástico. De manera que hemos usado una técnica MCMC eficiente para obtener automáticamente los parámetros de Bias que relacionan la población halos con el campo subyacente de materia oscura a gran escala basado en un catálogo de referencia.

Esta técnica es flexible, ya que admite la incorporación de diferentes modelos de sesgo y del número de parámetros de Bias. De forma que nos permite evaluar las distintas degeneraciones y regiones de confianza de los diferentes parámetros.

Por lo tanto utilizando el catálogo halos de la simulación The Small MultiDark de alta resolución como un catálogo de referencia, demostramos que nuestra técnica es capaz de modelar dicho espectro y por consiguiente de reproducir las estadísticas de dos puntos de halos de dicho catálogo referente.

Específicamente, mostramos que nuestro método nos permiten alcanzar una precisión porcentual en el espectro de potencias dentro del 5% hasta un valor de $k \sim 0.4hMpc^{-1}$. Cabe destacar que los catálogos de halos de materia oscura simulados haciendo uso de los códigos *FastPM* + *Patchy*, a pesar de presentar el mismo volumen que la simulación SMDPL ($400 \frac{Mpc^3}{h}$), solo fueron necesarias 160^3 partículas de materia oscura, siendo 13824 veces menor que el número de partículas de la simulación SMDPL.

Esto es posible dado que la resolución del mallado de la materia oscura es suficientemente alto ($2.5 \frac{Mpc}{h}$) para describir la red cósmica donde se encuentran los halos, y además nuestra técnica no pretende encontrar halos en una simulación de baja resolución, sino que usa modelos por debajo de la resolución de la malla (sub-grid models) extraídos de simulaciones de alta resolución para poblar de forma precisa dicha red cósmica. Los parámetros de Bias de los halos hay que extraerlos en un proceso estadístico calibrado con simulaciones de referencia. Por tanto nuestra metodología no viene a sustituir las simulaciones numéricas detalladas de formación de estructuras, sino a complementarlas, aprendiendo de ellas para poder emularlas con gran precisión y eficiencia.

Hemos comprobado que para estos objetos poco masivos aún no tenemos controlada toda la estadística. Si bien obtenemos ajustes del espectro de potencias, es decir, de la estadística de segundo orden, muy razonables, el histograma de recuento de halos en celdas no se ajusta de forma precisa. Esto nos indica que la estadística de mayor orden no la vamos a reproducir correctamente con estos modelos.

Podemos especular que nos falta algún ingrediente en nuestro modelo de Bias, tal como la componente de Bias no local. Esto podrá ser investigado en el futuro con estas técnicas. Queda por tanto aún un largo camino por recorrer para extraer toda la información necesaria de la formación de estructuras y poder aplicarla de forma eficiente para la siguiente generación de surveys galácticos. Confiamos en que este trabajo haya aportado un entendimiento mayor y ayude a progresar el campo.

5 Bibliografía

5.1 Referencias

Capítulo 1:

Liddle, Andrew; David Lyth (2000). Cambridge, ed. *Inflación Cosmológica y Estructuras a Gran Escala*. ISBN 0-521-57598-2.

Padmanabhan, T. (1993). Cambridge University Press, ed. *Formación de estructuras en el Universo*. ISBN 0-521-42486-0.

P. J. E. Peebles y J. T. Yu, "Perturbaciones adiabáticas primigenias en un Universo en expansión," *Astrophysical Journal* 162 (1970), 815.

R. A. Sunyaev, "Fluctuaciones de la radiación del fondo de microondas," en *Estructura a Gran Escala del Universo* ed. M. S. Longair y J. Einasto, 393. Dordrecht: Reidel 1978.

U. Seljak y M. Zaldarriaga (1996). «Una aproximación de integración de la línea de vista de las anisotropías del fondo cósmico de microondas». *Astrophysics J.* 469: 437-;444.

Wayne Hu y Scott Dodelson (2002). «Anisotropías del fondo cósmico de microondas». *Ann. Rev. Astron. Astrophys.* 40: 171-216.

Capítulo 2:

JavierPaz ,Dante (Marzo de 2010). Universidad Nacional de Córdoba,«Halos de materia y estructura en gran escala». http://www2.famaf.unc.edu.ar/publicaciones/documents/serie_d/DAst59.pdf [Accessed 10 Sep 2017]

Corvillón, Andrea(2009).«Función Correlación». <http://www.astro.puc.cl/npadilla/fia3009/trabajos/Andrea.pdf> [Accessed 10 Sep 2017]

IAC. «The perturbed Universe». [Apuntes].

Capítulo 3:

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, *ApJ*, 304, 15
 Casas-Miranda, R., Mo, H. J., Sheth, R. K., & Boerner, G. 2002, *MNRAS*, 333, 730
 Cen, R., & Ostriker, J. P. 1993, *ApJ*, 417, 415
 Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, *AJ*, 145, 10
 Chuang, C.-H., Kitaura, F.-S., Prada, F., Zhao, C., & Yepes, G. 2015, *MNRAS*, 446, 2621
 Chuang, C.-H., Zhao, C., Prada, F., et al. 2015, *MNRAS*, 452, 686 Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, 151, 44
 de la Torre, S., & Peacock, J. A. 2013, *MNRAS*, 435, 743
 Dodelson, Scott (2003). Academic Press, ed. *Cosmología Moderna*
 Feng, Y., Chu, M.-Y., Seljak, U., & McDonald, P. 2016, *MNRAS*, 463, 2273 Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
 Fry, J. N., & Gaztanaga, E. 1993, *ApJ*, 413, 447
 Somerville, R. S., Lemson, G., Sigad, Y., et al. 2001, *MNRAS*, 320, 289
 Kaiser, N. 1984, *ApJ*, 284, L9
 Kitaura, F.-S., & Heß, S. 2013, *MNRAS*, 435, L78 Kitaura, F.-S., Yepes, G., & Prada, F. 2014, *MNRAS*, 439, L21
 Kitaura, F.-S., Gil-Marín, H., Scoccola, C. G., et al. 2015, *MNRAS*, 450, 1836
 Kitaura, F.-S., Rodríguez-Torres, S., Chuang, C.-H., et al. 2016, *MNRAS*, 456, 4156
 Klypin, A., & Holtzman, J. 1997, [arXiv:astro-ph/9712217](https://arxiv.org/abs/astro-ph/9712217)
 Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, *MNRAS*, 457, 4340
 Mo, H. J., & White, S. D. M. 2002, *MNRAS*, 336, 112
 M. Vakili, F.-S. Kitaura, Y. Feng, G. Yepes, C. Zhao, C.-H. Chuang et al., Accurate halo-galaxy mocks from automatic bias estimation and particle mesh gravity solvers, *ArXiv e-prints* (Jan., 2017) , [1701.03765].
 Neyrinck, M. C., Aragón-Calvo, M. A., Jeong, D., & Wang, X. 2014, *MNRAS*, 441, 646
 Peebles, P. J. E. 1980, Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.,
 Ya. B. Zel'dovich, "Una hipótesis, unificando la estructura y la entropía del Universo," *Noticias Mensuales de la Royal Astronomical Society* 160 (1972).
 Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323, 1
<https://www.cosmosim.org/cms/simulations/smdpl/>[Accessed 10 Sep 2017]
 Vakili 2017 <http://adsabs.harvard.edu/abs/2017arXiv170103765V>[Accessed 10 Sep 2017]
 Feng 2016 <http://adsabs.harvard.edu/abs/2016MNRAS.463.2273F> [Accessed 10 Sep 2017]
 emcee Foreman-Mackey <http://adsabs.harvard.edu/abs/2013PASP..125..306F>[Accessed 10 Sep 2017]

<https://arxiv.org/pdf/1701.03765.pdf>: [Accessed 10 Sep 2017]

Procedencia de las imágenes utilizadas

Figura 1: https://www.google.es/search?q=red+cosmica&rlz=1C1ASRM_enES586ES586

[&source=lnms&tbn=isch&sa=X](https://www.google.es/search?q=red+cosmica&rlz=1C1ASRM_enES586ES586&source=lnms&tbn=isch&sa=X)

[&ved=0ahUKEwj5wNS29prWAhWoJMAKHd7](https://www.google.es/search?q=red+cosmica&rlz=1C1ASRM_enES586ES586&ved=0ahUKEwj5wNS29prWAhWoJMAKHd7)

[_C2YQ_AUICigB&biw=1280&bih=894](https://www.google.es/search?q=red+cosmica&rlz=1C1ASRM_enES586ES586&ved=0ahUKEwj5wNS29prWAhWoJMAKHd7_C2YQ_AUICigB&biw=1280&bih=894)

[#imgdii=_P6b6bRR1RUHUM:&imgrc=jmX-Tu9PH_GB1M:](https://www.google.es/search?q=red+cosmica&rlz=1C1ASRM_enES586ES586&ved=0ahUKEwj5wNS29prWAhWoJMAKHd7_C2YQ_AUICigB&biw=1280&bih=894)

Figura 2: https://www.cosmosim.org/resources/simus/smdpl_slice_0042_xy_3000_3.jpg