

UNIVERSIDAD NACIONAL DE COLOMBIA

TESIS DE MAESTRÍA EN ASTRONOMÍA

**FUNCIÓN DE CORRELACIÓN DE
MATERIA DE DOS PUNTOS EN EL
FORMALISMO COMBINADO DE
PERTURBACIONES-HALO**

Autor:

Fernando Alonso VELEZ R.

Director:

Dr. Leonardo CASTAÑEDA C.

*Tesis sometida como requisito parcial
para optar al grado de Magister en Astronomía*

en la

Facultad de Ciencias
Observatorio Astronómico

28 de agosto de 2020

UNIVERSIDAD NACIONAL DE COLOMBIA

Resumen

Facultad de Ciencias
Observatorio Astronómico

Maestría en Astronomía

FUNCIÓN DE CORRELACIÓN DE MATERIA DE DOS PUNTOS EN EL FORMALISMO COMBINADO DE PERTURBACIONES-HALO

por Fernando Alonso VELEZ R.

La formación de estructuras a gran escala en el universo es un área de gran interés científico en cosmología, en cuyo estudio diversas herramientas estadísticas ayudan a evaluar los modelos analíticos que describen el origen y evolución de estructuras. En esta dirección, el presente trabajo estima la función de correlación de dos puntos en el campo de materia sobre los resultados de la simulación BolshoiP a fin de comparar tales estimaciones con lo predicho por la teoría no lineal de perturbaciones aplicada al modelo del halo llamado *formalismo perturbaciones - halo*(FPH), implementado por Mead et. al. [14].

Las estimaciones de la función de correlación de dos puntos que se obtuvieron son consistentes con lo reportado en la literatura en los cuatro estimadores empleados. En contraste con los resultados del modelo FPH se observan puntos donde el espectro de potencias obtenido a partir de la función de correlación es sobrestimado, particularmente a redshift mayores a 1 y bajas escalas.

Palabras Claves: Cosmología, Halo, Perturbaciones, Función de correlación

UNIVERSIDAD NACIONAL DE COLOMBIA

Abstract

Facultad de Ciencias
Observatorio Astronómico

Magister en Astronomía

TWO POINT CORRELATION FUNCTION OF MATTER IN THE FORMALISM HALO - PERTURBATION

by Fernando Alonso VELEZ R.

The formation of large scale structures in the universe is an area of great scientific interest in cosmology, in whose study various statistical tools help to evaluate the analytical models that describe the origin and evolution of structures. In this direction, the present work estimates the correlation function of two points in the field of matter on the results of the BolshoiP simulation in order to compare such estimates with that predicted by the non-linear theory of perturbations applied to the halo model called formalism perturbation - halo (FPH), implemented by Mead et. al. [14].

The estimates of the two-point correlation function that were obtained are consistent with what is reported in the literature in the four estimators used. In contrast to the results of the FPH model, points are observed where the power spectrum obtained from the correlation function is overestimated, particularly at redshifts greater than 1 and low scales.

Keywords: Cosmology, Halo, Perturbations, correlation function

A mi hijo... mi razón de ser y vivir...

Agradecimientos

La consecución de un logro es con frecuencia el resultado de un cúmulo de aportes que directa o indirectamente recibimos de muchas personas y en tal sentido agradezco a mi familia el apoyo y comprensión brindado, al maestro Alonso Medina Cardona por compartir todo su saber y permitirme iniciar como aficionado a la astronomía alimentando mi interés en la disciplina y fundamentalmente este trabajo no hubiera sido posible sin la guía, apoyo permanente y orientación de mi maestro y amigo Leonardo Castañeda Colorado. Así mismo, personas como el Ingeniero Daniel Burbano, Andres Moreno y Juan Felipe Bravo quienes me brindaron un apoyo oportuno para la finalización de este trabajo.

Contenido

Abstract	iii
Abstract	v
Agradecimientos	ix
1. Marco Teórico	1
1.1. Modelo Estándar	1
1.2. El universo inhomogéneo	2
1.2.1. El gauge Poisson	3
Gauge Newtoniano Conforme	4
Dinámica de partículas en un universo en expansión	4
La ecuación de Vlasov	4
Dinámica Euleriana	6
1.3. Teoría lineal de perturbaciones	8
1.4. Teoría no lineal de perturbaciones	10
1.5. Modelo de colapso esférico	11
1.6. Modelo del Halo	14
1.6.1. Función de masa	15
1.6.2. Sesgo	16
1.6.3. Perfiles del halo	17
1.6.4. Función de correlación	18
1.6.5. El espectro de potencias en el modelo del halo	20
1.6.6. Estimadores de la Función de Correlación	21
1.7. Formalismo perturbaciones-halo	23
1.8. Modelo Analítico	24
1.9. Procesos Puntuales	28
1.9.1. Campos Binomiales	28
1.9.2. Campos Poisson	29
1.9.3. Campos de Cox	30
2. Simulación Bolshoi	33
2.1. Simulaciones cosmológicas	33
2.2. Simulación Bolshoi	33
2.2.1. Consultas	38
2.3. Estimación de la función de correlación de dos puntos	38
2.3.1. Código correlator.c	38
Descripción	38
Argumentos	39

Ejecución	40
2.3.2. Resultados	41
3. Resultados	43
3.1. Optimización del código	43
3.2. Selección de Snapshots	44
3.3. Distribución de masa de halos	45
3.4. Estimación de la función de correlación de dos puntos	46
3.5. Espectro de potencias en la simulación BolshoiP y comparación con el modelo analítico	51
4. Conclusiones y Perspectivas	57
4.1. Conclusiones	57
4.1.1. Conclusiones físicas	57
4.1.2. Conclusiones estadísticas	57
4.1.3. Conclusiones del desarrollo computacional	58
4.2. Recomendaciones	58
4.3. Perspectivas	58
A. Apéndices	59
A.1. Campos aleatorios	59
A.2. Partición de los conteos	61
A.2.1. Cantidad de núcleos por conteo	61
A.2.2. Número de puntos por núcleo	61
A.3. Ecuaciones de Einstein en el Gauge Newtoniano Conforme	62
A.4. Otros resultados	66
A.5. Código Fuente	67
A.5.1. Código del programa para la estimación de la función de correlación	67
A.5.2. Código R para la simulación de campos aleatorios	80
A.5.3. Código Python para el cálculo del espectro de potencias	85
Bibliografía	87

Lista de Figuras

1.1. Mapeo del universo cercano realizado por el SDSS	3
1.2. Campo aleatorio (izquierda) $\xi(r_{12}) = 0$ y campo con estructuras (derecha) $\xi(r_{12}) > 0$	19
1.3. Componentes de la función de correlación de dos puntos	19
1.4. Campos puntuales empleados para la estimación de la función de correlación.	22
1.5. Espectro de potencias en el modelo analítico.	25
1.6. Espectro de potencias en el modelo analítico	26
1.7. Espectro de potencias por componente	27
1.8. Componente no lineal del término halo-halo	27
1.9. Ejemplos simulados de campos binomiales.	29
1.10. Ejemplos simulados de campos puntuales Poisson.	30
1.11. Ejemplos simulados de campos de Cox.	31
2.1. Desarrollo histórico de las simulaciones cosmológicas	34
2.2. Cajas (snapshots) de la simulación BolshoiP (A) $z = 10.18$, (B) $z = 7.59$, (C) $z = 0.99$, (D) $z = 0.01$	35
3.1. Tiempos efectivos de procesamiento	44
3.2. Selección de Snapshots	45
3.3. Funciones de masa estimadas	46
3.4. Estimador natural y de Landy-Szalay de la función de correlación de dos puntos a diferentes redshift	47
3.5. Error de estimación a diferentes redshift	48
3.6. Estimadores de la función de correlación de dos puntos	50
3.7.	51
3.8.	52
3.9. La línea continua es la estimación del espectro de potencias a $z = 4.06$ en tanto que las líneas no continuas provienen de la teoría combinada perturbaciones-halo.	53
3.10. La línea continua es la estimación del espectro de potencias a $z = 1.07$ en tanto que las líneas no continuas provienen de la teoría combinada perturbaciones-halo.	53
3.11. La línea continua es la estimación del espectro de potencias a $z = 0.01$ en tanto que las líneas no continuas provienen de la teoría combinada perturbaciones-halo.	54
3.12. Comparación del Espectro de Potencias para el término halo-halo	55
A.1. Estimador de Davis-Peebles y Hamilton de la función de correlación de dos puntos a diferentes redshift	66

Lista de Tablas

2.1. Detalle de la información contenida en la tabla FOF de la simulación BolshoiP	37
3.1. Características de los snapshots empleados	45
3.2. Estimación de parámetros para la sección de tendencia lineal	49

Abreviaturas

FLRW	F riedman- L emaitre- R obertson- W alker
SDSS	S loan D igital S ky S urvey
ART	A daptive R efinement T ree
FOF	F riends O f F riends
MPI	M essage P assing I nterface
WMAP	W ilkinson M icrowave A nisotropy P robe
SQL	S ecquential Q uery L anguage

Notación y Convenciones

Signatura de la métrica $(-, +, +, +)$
Velocidad de la luz $c = 1$

Indices

Griegos μ 0, 1, 2, 3

Latinos i 1, 2, 3

$A_{b'c}$ $\partial_c A_b$

Tensor de Ricci

$$R_{abc}^d = \Gamma_{ab,c}^d - \Gamma_{ac,b}^d + \Gamma_{ab}^e \Gamma_{ec}^d - \Gamma_{ac}^e \Gamma_{eb}^d$$
$$R_{ab} = R_{abc}^c$$

Capítulo 1

Marco Teórico

1.1. Modelo Estándar

La cosmología moderna se fundamenta en la teoría de la relatividad general de Einstein como teoría de gravedad, que junto al principio cosmológico, constituye el marco sobre el cual se construye el modelo estándar de la cosmología. Dicho principio implica considerar que a grandes escalas, la materia en el universo esta uniformemente distribuida a lo cual se le denomina *homogeneidad* y adicionalmente que en cualquier dirección que se observe, el universo tenga las mismas propiedades, es decir que no existan direcciones privilegiadas, esto se conoce como *isotropía*. Las hipótesis de homogeneidad e isotropía conllevan a una manera única de escribir la métrica de la forma

$$ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right], \quad (1.1)$$

donde (t, r, θ, ϕ) son el tiempo cosmológico y las coordenadas comóviles respectivamente. La función $a(t)$ es el factor de escala cosmológico y k es la curvatura tri-espacial la cual es constante. La ecuación 1.1 es esféricamente simétrica como una de las consecuencias de la isotropía.

Las ecuaciones de campo de Einstein relacionan precisamente, la geometría del espacio-tiempo con el contenido de materia en el universo, lo cual se expresa en forma tensorial de la forma:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathcal{R} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \quad (1.2)$$

donde $R_{\mu\nu}$ es el tensor de Ricci, \mathcal{R} es el escalar de Ricci, Λ es la constante cosmológica, G es la constante de gravitación y T_{μ}^{ν} es el tensor de momento-energía. La expresión dada en 1.1 se conoce como la métrica de Friedmann-Robertson-Walker, de tal manera que las ecuaciones de campo de Einstein para esta métrica y con tensor de momento-energía dado por, $T_0^0 = -\rho$ y $T_i^i = p$ con $i = 1, 2, 3$ son:

$$H^2 = \frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} - \frac{k}{a^2} + \frac{\Lambda}{3} \quad (1.3)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}, \quad (1.4)$$

donde H es el parámetro de Hubble y Λ la constante cosmológica, así las expresiones 1.3 y 1.4 se conocen como las *ecuaciones de Friedmann* y en general el modelo cosmológico que implican se denomina el universo de Friedmann-Lemaitre-Robertson-Walker (FLRW).

En este contexto, el contenido del universo se considera como un fluido cosmológico el cual está constituido fundamentalmente por materia, radiación y vacío. Estos componentes se introducen en las ecuaciones de campo como densidades de energía Ω_m , Ω_k y Ω_Λ respectivamente. En este modelo, la componente Ω_Λ es la responsable de la expansión acelerada actual del universo [10].

Adicionalmente, se considera un parámetro de densidad de curvatura denotado por Ω_k de tal manera que la ecuación 1.3 se puede reescribir como:

$$\begin{aligned} 1 &= \frac{8\pi G\rho}{3H^2} - \frac{k}{a^2H^2} + \frac{\Lambda}{3H^2} \\ 1 &= \Omega_m + \Omega_k + \Omega_\Lambda, \end{aligned} \quad (1.5)$$

Las observaciones cosmológicas actuales han permitido estimar el valor de los parámetros cosmológicos con una precisión mayor al 5% (NASA/WMAP Science Team) con $\Omega_{m0} \approx 0.3$, $\Omega_{\Lambda0} \approx 0.7$ y $\Omega_{k0} \approx 0$. La métrica dada en 1.1 da cuenta de la suave distribución del universo a muy grandes escalas, sin embargo para explicar la formación de estructuras observadas se pueden considerar perturbaciones a la métrica, las cuales a través de las ecuaciones de campo implican perturbaciones en el tensor de momento-energía. La teoría de tales perturbaciones se describen en la siguiente sección.

La formación de estructura en el universo es uno de los más importantes y activos campos de investigación tanto en cosmología teórica como en la observacional. La literatura acerca del tema es amplia y rica, sin embargo, algunas aclaraciones importantes acerca del marco conceptual y las aproximaciones usadas en aspectos específicos no están presentes en la literatura.

1.2. El universo inhomogéneo

En cosmología moderna el universo es considerado como homogéneo e isotrópico en un sentido estadístico [6] a escalas de 100 *Mpc* y mayores. Sin embargo a escalas menores se observan estructuras tales como los cúmulos de galaxias los cuales han sido catalogados en varios proyectos, siendo uno de ellos el SDSS como se muestra en la figura¹ 1.1.

Una brillante y excelente referencia en el tema es el libro de P.J.E. Peebles titulado *The large scale structure of the universe*[20] del cual se toma la frase: **La discusión de como las irregularidades en la distribución de materia se comportan en la expansión del universo es considerablemente simplificada en la aproximación límite de la relatividad general, la mecánica newtoniana, la cual aplica en una pequeña región comparada con la longitud de Hubble cH^{-1} (y grande comparada con el radio de Schwarzschild de cualquier objeto colapsado). El resto del universo solo puede afectar la región a través de un campo de marea. Aunque el punto fue claramente expuesto por Lemaitre (1931), no siempre se ha reconocido que la aproximación Newtoniana no es un modelo pero un caso límite válido sin importar lo que suceda en la parte distante del universo.**

¹Tomado de <https://blog.sdss.org/2015/12/29/how-sdss-talked-about-light-for-iy12015>

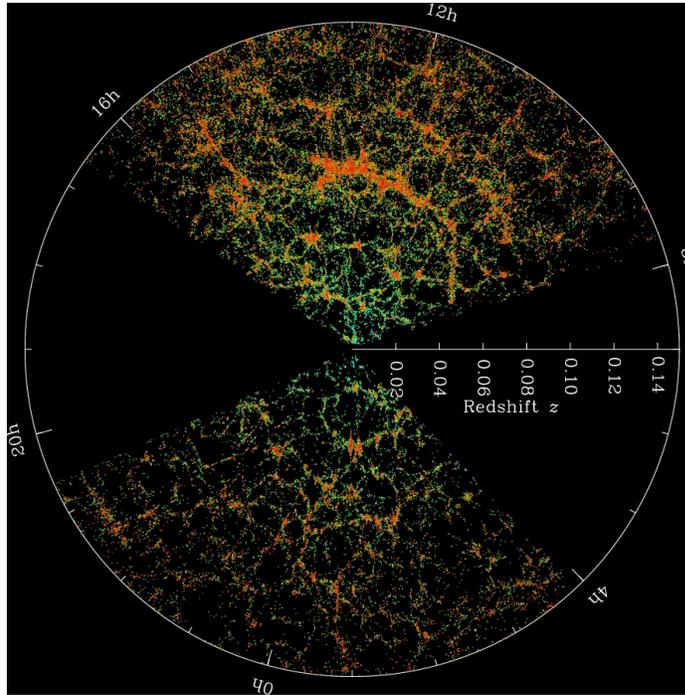


FIGURA 1.1: Mapeo del universo cercano realizado por el SDSS

1.2.1. El gauge Poisson

La representación del universo inhomogéneo y las estructuras que allí se generan requiere establecer una clase de mapeos independientes del sistema de coordenadas llamados Gauges, con los cuales es posible describir la física del universo de forma invariante. Un gauge que permite tal representación es el gauge Poisson.

A escalas menores al radio de Hubble (subhorizonte), se aborda el contenido del universo como un fluido de materia oscura fría (CDM) en una época en la cual es no relativista, lo cual reduce ostensiblemente la complejidad del problema. La dinámica del fluido CDM desde las ecuaciones de campo de Einstein es descrita solo por el potencial del mismo fluido y en ese caso el tensor métrico es:

$$ds^2 = a^2(\tau) \left(-\{1 + 2\Psi(\mathbf{x}, \tau)\} d\tau^2 + 2\omega_i(\mathbf{x}, \tau) d\tau dx^i + [\{1 - 2\Phi(\mathbf{x}, \tau)\} \gamma_{ij} + h_{ij}(\mathbf{x}, \tau)] dx^i dx^j \right), \quad (1.6)$$

donde la contribución por ondas gravitacionales satisface $\gamma^{ij} h_{ij} = 0$. El gauge Poisson tiene la ventaja que la condición $\nabla \cdot \omega = \nabla \cdot \mathbf{h} = 0$ fija el gauge. Desde las ecuaciones de campo, cuando $\sum_{ij} \approx 0$, los potenciales escalares están relacionados por $\Psi = \Phi$. Uno de los más interesantes aspectos de gauge Poisson es que se puede obtener el potencial escalar y vectorial directamente desde la distribución instantánea de esfuerzo-energía sin un tiempo de integración requerido. Una versión restringida del gauge Poisson, se obtiene cuando $\omega_i = h_{ij} = 0$, el cual es conocido en la literatura como *Gauge Newtoniano Conforme*.

Gauge Newtoniano Conforme

Este gauge permite describir un universo perturbado respecto a la versión homogénea y plana que se obtiene al usar la métrica de FLRW. Para ello el elemento de línea está dado por:

$$ds^2 = a^2 \left[-(1 + 2\Psi)d\tau^2 + (1 - 2\Phi)\delta_{ij}dx^i dx^j \right], \quad (1.7)$$

donde los términos Ψ es el potencial Newtoniano y Φ la perturbación de la curvatura espacial [6]. Estos dos términos son considerados pequeños, al punto que cualquier término no lineal es despreciado. Así mismo, se puede ver que en caso de ser nulos ($\Psi = \Phi = 0$) se recupera la métrica del background.

Así, para un campo escalar ψ cualquiera en el universo perturbado se tendrá que

$$\psi(\tau) = \psi_0(\tau) + \delta\psi(\tau)$$

siendo $\psi_0(\tau)$ el valor del background y $\delta\psi(\tau)$ el término asociado a la perturbación del campo [10].

Dinámica de partículas en un universo en expansión

El gauge Poisson ofrece la posibilidad de comprender de una forma directa algunas consecuencias en el movimiento de las partículas en un universo perturbado. Una vez se conoce la métrica, la ecuación de movimiento en la relatividad general esta en las ecuaciones de campo a través de la condición $T_{,b}^{ab} = 0$. Sin embargo, como en el caso del fluido de materia oscura fría, podemos ir más lejos y obtener algún indicio acerca de la naturaleza del movimiento bajo la influencia de la gravedad. Así, la ecuación de movimiento para partículas libres en el espacio-tiempo se denominan geodésicas y esta dada por:

$$\frac{d^2 x^\alpha}{d\tau^2} + \Gamma_{\beta\gamma}^\alpha \frac{dx^\beta}{d\tau} \frac{dx^\gamma}{d\tau} = 0, \quad (1.8)$$

donde τ es el tiempo propio y Γ es el término de contiene la física asociada a la gravedad. Sin embargo, en lugar de resolver (1.8) la dinámica de las partículas no-relativistas en un universo en expansión, puede ser obtenida en el denominado limite Newtoniano de la cosmología.

La ecuación de Vlasov

En el limite no relativista, se puede considerar un conjunto de partículas de masa en reposo m que interaccionan solo gravitacionalmente en un universo en expansión.

$$\frac{d\mathbf{v}}{dt} = Gm \sum_i \frac{\mathbf{r}_i - \mathbf{r}}{|\mathbf{r}_i - \mathbf{r}|^3}, \quad (1.9)$$

donde \mathbf{v} es la velocidad de las partículas. La expansión del universo solo genera una redefinición de la posición y del potencial gravitacional para las partículas. En el límite continuo (1.9) se expresa así:

$$\frac{d\mathbf{v}}{dt} = G \int d^3\mathbf{r} \frac{\rho(\mathbf{r})}{|\mathbf{r}_i - \mathbf{r}|^3} (\mathbf{r}_i - \mathbf{r}), \quad (1.10)$$

En el contexto de la cosmología, la ecuación (1.10) considera un conjunto de coordenadas (t, \mathbf{x}) que permite cubrir la variedad espacio-temporal, de tal manera que en el gauge Poisson esta coinciden con las coordenadas Eulerianas del fluido y adicionalmente se usa el conjunto espacial de coordenadas referidas como *coordenadas comóviles* y las coordenadas físicas están definidas por $\mathbf{r} = a(t)\mathbf{x}$. Para la coordenada temporal se emplea el tiempo cósmico t ó el tiempo conforme τ donde $d\tau = \frac{dt}{a(t)}$. Así, el campo de velocidades en tiempo conforme es:

$$\mathbf{v}(\mathbf{x}, \tau) = \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{d\tau} \frac{d\tau}{dt} = H(\tau)\mathbf{x} + \frac{d\mathbf{x}}{d\tau} = H(\tau)\mathbf{x} + \mathbf{u}(\mathbf{x}, \tau), \quad (1.11)$$

donde el primer término del lado derecho de (1.11) corresponde a la ley de Hubble y el segundo término es el campo de velocidades peculiares. El siguiente paso es calcular:

$$\frac{d\mathbf{v}}{dt} = \frac{d\tau}{dt} \frac{d}{d\tau} (H(\tau)\mathbf{x} + \mathbf{u}(\mathbf{x}, \tau)) = \frac{1}{a} \left(\frac{\partial H}{\partial \tau} \mathbf{x} + H\mathbf{u} + \frac{d\mathbf{u}}{d\tau} \right), \quad (1.12)$$

El primer término de la derecha en la expresión (1.12) es la ecuación de Friedmann, el segundo define la ecuación de movimiento en la aproximación de baja velocidad ($v \ll c$). Desde la ecuación (1.10) se tiene que en principio $\rho(\mathbf{r}, t)$ podría tener contribuciones relativistas provenientes de radiación o energía del vacío, en general, se puede escribir la densidad como:

$$\rho(\mathbf{r}, t) = \bar{\rho}(t) + \delta\rho(\mathbf{r}, t), \quad (1.13)$$

donde la fuente para las velocidades peculiares es el segundo término de (1.13). Usando coordenadas comóviles, la ecuación de movimiento (1.10) y la función:

$$\delta(\mathbf{x}, \tau) = \frac{\rho(\mathbf{x}, \tau) - \bar{\rho}(\tau)}{\bar{\rho}(\tau)}, \quad (1.14)$$

se tiene que el potencial cosmológico queda definido por:

$$\Phi(\mathbf{x}, \tau) = -\frac{1}{2} \frac{\partial H(\tau)}{\partial \tau} x^2 + \phi_p(\mathbf{x}, \tau), \quad (1.15)$$

donde de (1.10)

$$\frac{d\mathbf{v}}{dt} = -\frac{1}{2} \nabla_{\mathbf{x}} \Phi(\mathbf{x}, \tau), \quad (1.16)$$

y definiendo $\mathbf{p} = a(\tau)m\mathbf{u}$ y la ecuación de movimiento para el campo de velocidades peculiares entonces:

$$\frac{d\mathbf{p}}{dt} = -am\nabla_{\mathbf{x}} \Phi(\mathbf{x}, \tau), \quad (1.17)$$

La ecuación de gravedad para el potencial de perturbación es

$$\boxed{\nabla^2 \phi_p(\mathbf{x}, \tau) = \frac{3}{2} \Omega_m(\tau) H^2(\tau) \delta(\mathbf{x}, \tau).} \quad (1.18)$$

Los resultados en la aproximación Newtoniana son idénticos a los obtenidos por la teoría de perturbaciones cosmológicas relativista en el gauge Poisson para fluidos no relativistas. Finalmente, la ecuación diferencial para las partícula de materia oscura en

coordenadas comóviles es:

$$\boxed{\frac{d^2 \mathbf{x}}{d\tau^2} + H(\tau) \frac{d\mathbf{x}}{d\tau} = -\nabla_{\mathbf{x}} \phi_p(\mathbf{x}, \tau)}, \quad (1.19)$$

la cual es usualmente empleada en los códigos de simulación de N-cuerpos.

Para materia oscura, la función densidad número de partículas $f(\mathbf{x}, \mathbf{p}, \tau)$ en el espacio de fase satisface la ecuación de Boltzmann sin colisiones dada por:

$$\boxed{\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{d\mathbf{x}}{d\tau} \cdot \nabla_{\mathbf{x}} f + \frac{d\mathbf{p}}{d\tau} \cdot \nabla_{\mathbf{p}} f = 0} \quad (1.20)$$

y cuando solo actúa la gravedad se denomina ecuación de Vlasov y se expresa de la forma:

$$\frac{df}{d\tau} = \frac{\partial f}{\partial \tau} + \frac{\mathbf{p}}{ma} \cdot \nabla_{\mathbf{x}} f - am \nabla_{\mathbf{x}} \phi_p \cdot \nabla_{\mathbf{p}} f = 0, \quad (1.21)$$

Esta ecuación debe ser resuelta de forma consistente con (1.17) y (1.18) lo cual es en general una tarea compleja y no se aborda en este trabajo, en lugar de ello se emplean los momentos del momentum para describir la evolución espacial.

Dinámica Euleriana

Para iniciar con las ecuaciones de momentos del momentum se define:

$$\begin{aligned} \int d^3 \mathbf{p} f(\mathbf{x}, \mathbf{p}, \tau) &= \rho(\mathbf{x}, \tau) \\ \int d^3 \mathbf{p} \frac{\mathbf{p}}{am} f(\mathbf{x}, \mathbf{p}, \tau) &= \rho(\mathbf{x}, \tau) \mathbf{U}(\mathbf{x}, \tau) \\ \int d^3 \mathbf{p} \frac{p_i p_j}{a^2 m} f(\mathbf{x}, \mathbf{p}, \tau) &= \rho(\mathbf{x}, \tau) u_i(\mathbf{x}, \tau) u_j(\mathbf{x}, \tau) + \sigma_{ij}, \end{aligned} \quad (1.22)$$

Este conjunto ecuaciones define el campo de velocidades $\mathbf{U}(\mathbf{x}, \tau)$ y el tensor de anisotropías σ_{ij} . Tomando el momentum de orden cero, la ecuación (1.21) se expresa como:

$$\frac{\partial}{\partial \tau} \int d^3 \mathbf{p} f + \nabla_{\mathbf{x}} \cdot \int d^3 \mathbf{p} \frac{\mathbf{p} f}{am} - am \nabla_{\mathbf{x}} \phi_p \cdot \int d^3 \mathbf{p} \nabla_{\mathbf{p}} f = 0, \quad (1.23)$$

por el teorema de Gauss el último término se puede reescribir como:

$$\int d^3 \mathbf{p} \nabla_{\mathbf{p}} f = \oint d^2 \mathbf{S} f = 0, \quad (1.24)$$

y usando las definiciones dadas en (1.22) y $\delta(\mathbf{x}, \tau)$ se tiene que la ecuación (1.23) se expresa así:

$$\frac{\partial}{\partial \tau} \delta(\mathbf{x}, \tau) + \nabla_{\mathbf{x}} \cdot ([1 + \delta(\mathbf{x}, \tau)] \mathbf{U}(\mathbf{x}, \tau)) = 0, \quad (1.25)$$

el primer término en (1.25) es nulo, esto es la conservación de la masa en el background cosmológico. Esto implica que

$$\boxed{\frac{\partial}{\partial \tau} \delta(\mathbf{x}, \tau) + \nabla_{\mathbf{x}} \cdot ([1 + \delta(\mathbf{x}, \tau)] \mathbf{U}(\mathbf{x}, \tau)) = 0} \quad (1.26)$$

La primera ecuación de momentum está dada por:

$$\int d^3 \mathbf{p} \frac{\partial f}{\partial \tau} \frac{p_i}{a m} + \nabla_{\mathbf{x}} \cdot \int d^3 \mathbf{p} \frac{\mathbf{p} f p_i}{a^2 m^2} - \nabla_{\mathbf{x}} \phi_p \cdot \int d^3 \mathbf{p} \nabla_{\mathbf{p}}(f) p_i = 0, \quad (1.27)$$

analizando cada término en (1.27) se tiene que el primero

$$\frac{\partial}{\partial \tau} \int d^3 \mathbf{p} \frac{f p_i}{a m} + H(\tau) \int d^3 \mathbf{p} \frac{f p_i}{a m} = \frac{\partial}{\partial \tau} (\rho(\mathbf{x}, \tau) \mathbf{u}_i(\mathbf{x}, \tau)) + H(\tau) \rho(\mathbf{x}, \tau) \mathbf{u}_i(\mathbf{x}, \tau), \quad (1.28)$$

como se puede ver desde (1.22) el segundo término en (1.27) puede ser escrito como:

$$\sum_j \frac{\partial}{\partial x_j} \int d^3 \mathbf{p} \frac{p_i p_j f}{a^2 m^2} = \sum_j \frac{\partial}{\partial x_j} (\rho(\mathbf{x}, \tau) \mathbf{U}_i(\mathbf{x}, \tau) \mathbf{U}_j(\mathbf{x}, \tau) + \sigma_{ij}), \quad (1.29)$$

El último término en (1.27) requiere especial atención. Usando la notación de índices repetidos para operadores diferenciales se obtiene:

$$\sum_j \frac{\partial}{\partial x_j} \phi_p \int d^3 \mathbf{p} \frac{\partial f}{\partial p_j} p_i = \sum_j \frac{\partial}{\partial x_j} \phi_p \int d^3 \mathbf{p} \frac{\partial(f p_i)}{\partial p_j} - \sum_j \frac{\partial}{\partial x_j} \phi_p \int d^3 \mathbf{p} f \delta_{ij}, \quad (1.30)$$

el primer término de la derecha es nulo por el teorema de Gauss, quedando al final

$$\frac{\partial}{\partial x_i} \phi_p \int d^3 \mathbf{p} f = \frac{\partial \phi_p}{\partial x_i} \rho(\mathbf{x}, \tau), \quad (1.31)$$

con el resultado anterior y restando \mathbf{U}_i veces la ecuación de continuidad (1.26) se obtiene la ecuación de Euler

$$\boxed{\frac{\partial \mathbf{U}(\mathbf{x}, \tau)}{\partial \tau} + H(\tau) \mathbf{U}(\mathbf{x}, \tau) + (\mathbf{U}(\mathbf{x}, \tau) \cdot \nabla_{\mathbf{x}}) \mathbf{U}(\mathbf{x}, \tau) = -\nabla_{\mathbf{x}} \phi_p - \frac{1}{\rho} \nabla_{\mathbf{x}} \cdot \sigma} \quad (1.32)$$

El conjunto de ecuaciones (1.26) y (1.32) son los bloques fundamentales de la teoría de perturbaciones Euleriana. Como se puede ver, el problema en el límite no relativista para el fluido de materia oscura es reducido a un problema de fluidos. Es común desde (1.25) asociar el campo $\mathbf{U}(\mathbf{x}, \tau)$

$$\theta(\mathbf{x}, \tau) \equiv \nabla \cdot \mathbf{U}(\mathbf{x}, \tau), \quad (1.33)$$

la cual es la divergencia del campo. La vorticidad es

$$\mathbf{W}(\mathbf{x}, \tau) \equiv \nabla_{\mathbf{x}} \times \mathbf{U}(\mathbf{x}, \tau), \quad (1.34)$$

La ecuación (1.25) para la vorticidad es:

$$\boxed{\frac{\partial \mathbf{w}(\mathbf{x}, \tau)}{\partial \tau} + H(\tau) \mathbf{W}(\mathbf{x}, \tau) - \nabla_{\mathbf{x}} \times (\mathbf{U}(\mathbf{x}, \tau) \times \mathbf{W}(\mathbf{x}, \tau)) = -\nabla_{\mathbf{x}} \times \left(\frac{1}{\rho} \nabla_{\mathbf{x}} \cdot \sigma \right)} \quad (1.35)$$

1.3. Teoría lineal de perturbaciones

Se define un campo de fluctuaciones en la densidad de energía $\rho(\mathbf{x}, \tau)$ como:

$$\delta(\mathbf{x}, \tau) = \frac{\rho(\mathbf{x}, \tau) - \bar{\rho}(\tau)}{\bar{\rho}(\tau)}, \quad (1.36)$$

donde τ es el tiempo conforme definido como $\tau = \int \frac{dt}{a(t)}$ y \mathbf{x} son las coordenadas comóviles.

Las ecuaciones de Einstein para la perturbación $\delta(\mathbf{x}, \tau)$ definen un potencial escalar $\phi(\mathbf{x}, \tau)$ tal que la ecuación de movimiento en coordenadas comóviles y en tiempo conforme es [23]:

$$\nabla^2 \phi(\mathbf{x}, \tau) = \frac{3}{2} \Omega_m(\tau) \mathcal{H}^2(\tau) \delta(\mathbf{x}, \tau), \quad (1.37)$$

donde $\mathcal{H}(\tau)$ es la tasa de expansión conforme y $\Omega_m(\tau)$ es el parámetro de densidad de materia.

Por otra parte, si se considera ahora la ecuación de continuidad $T^0{}_{;\nu} = 0$ expresadas en coordenadas comóviles \mathbf{x} se obtiene:

$$\nabla_{\mathbf{x}} \cdot [(1 + \delta(\mathbf{x}, \tau)) \mathbf{U}(\mathbf{x}, \tau)] + \frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} = 0, \quad (1.38)$$

Ahora, tomando la componente $T^i{}_{;\nu} = 0$ también en coordenadas comóviles:

$$\frac{\partial \mathbf{U}(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H} \mathbf{U}(\mathbf{x}, \tau) + \mathbf{U}(\mathbf{x}, \tau) \cdot \nabla \mathbf{U}(\mathbf{x}, \tau) = -\nabla_{\mathbf{x}} \phi(\mathbf{x}, \tau), \quad (1.39)$$

Las ecuaciones 1.37, 1.38 y 1.39 describen la densidad de contraste y el campo de velocidades peculiares.

La teoría lineal de perturbaciones se obtiene cuando se desprecian los términos $\nabla_{\mathbf{x}} \cdot \delta(\mathbf{x}, \tau) \mathbf{U}(\mathbf{x}, \tau)$ y $\mathbf{U}(\mathbf{x}, \tau) \cdot \nabla \mathbf{U}(\mathbf{x}, \tau)$ en las ecuaciones 1.38, 1.39 respectivamente, de tal manera que se obtiene:

$$\frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} + \nabla_{\mathbf{x}} \cdot \mathbf{U}(\mathbf{x}, \tau) = 0 \quad (1.40)$$

$$\frac{\partial \mathbf{U}(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H} \mathbf{U}(\mathbf{x}, \tau) = -\nabla_{\mathbf{x}} \phi(\mathbf{x}, \tau), \quad (1.41)$$

Dado que el campo de velocidades peculiares, es un campo vectorial, éste puede ser descrito por su rotacional y divergencia los cuales son aplicados a la ecuación 1.41 obteniendo:

$$\frac{\partial \theta(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H} \theta(\mathbf{x}, \tau) + \nabla_{\mathbf{x}}^2 \phi(\mathbf{x}, \tau) = 0 \quad (1.42)$$

$$\frac{\partial W(\mathbf{x}, \tau)}{\partial \tau} + \mathcal{H} W(\mathbf{x}, \tau) = 0, \quad (1.43)$$

donde $\theta(\mathbf{x}, \tau) = \nabla \cdot \mathbf{U}(\mathbf{x}, \tau)$ y $W(\mathbf{x}, \tau) = \nabla \times U(\mathbf{x}, \tau)$. Así, al sustituir la expresión 1.37 y 1.42 en la ecuación 1.43 se obtiene:

$$\boxed{\frac{\partial^2 \delta(\mathbf{x}, \tau)}{\partial \tau^2} + \mathcal{H} \frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} = \frac{3}{2} \Omega_m \mathcal{H}^2 \delta(\mathbf{x}, \tau)} \quad (1.44)$$

Suponiendo que la solución tiene la forma $\delta(\mathbf{x}, \tau) = \delta_1(\mathbf{x})D(\tau)$ entonces la ecuación 1.44 se puede expresar como:

$$\frac{\partial^2 D}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial D}{\partial t} = 4\pi G \rho D, \quad (1.45)$$

donde se ha tenido en cuenta que $\frac{dt}{d\tau} = a(t)$. Por otra parte, $H(t)$ se relaciona con los parámetros cosmológicos como sigue:

$$H^2(t) = H_0^2 \left[\frac{\Omega_m}{a^3} + \frac{1 - (\Omega_m + \Omega_r)}{a^2} + \Omega_\Lambda \right], \quad (1.46)$$

De tal manera que al tomar las dos primeras derivadas de $H(t)$ se puede verificar que satisface la ecuación:

$$\frac{\partial^2 H}{\partial t^2} + 2 \frac{\dot{a}}{a} \frac{\partial H}{\partial t} = 4\pi G \rho H, \quad (1.47)$$

Esto implica que una solución de la ecuación 1.45 es justamente $H(t)$ o más generalmente,

$$\boxed{D_-(t) \propto H(t)} \quad (1.48)$$

El signo menos como subíndice denota que es la solución decreciente de la ecuación puesto que la primera derivada de $H(t)$ es negativa. Sin embargo, como la ecuación diferencial 1.45 es homogénea y lineal entonces se puede construir una segunda solución a través del Wroskiano, obteniéndose:

$$\boxed{D_+(t) \propto H(t) \int_0^t \frac{dl}{a^2(l)H^2(l)}} \quad (1.49)$$

De esto se tiene que la solución general de la ecuación diferencial (1.44) se puede expresar como:

$$\boxed{\delta(\mathbf{x}, t) = A(\mathbf{x})D_+(t) + B(\mathbf{x})D_-(t)} \quad (1.50)$$

La solución $D_+(t)$ es llamada *función lineal de crecimiento* y para la época de dominación de materia $D_+(t) \propto t^{2/3}$, en tanto que $D_-(t) \propto t^{-1}$ con lo cual las perturbaciones de materia crecen como una ley de potencias en el régimen lineal. Las condiciones iniciales para esta solución, vienen dadas por la función de transferencia, la cual modela de forma estocástica las condiciones iniciales tras la era de dominio de radiación [25].

1.4. Teoría no lineal de perturbaciones

Al no considerar la linealización antes mencionada y abordar la solución de las ecuaciones 1.38 y 1.39 se tiene la teoría no lineal de perturbaciones, de tal manera que si tomamos la divergencia estas ecuaciones se obtiene:

$$\frac{\partial \delta(\mathbf{x}, \tau)}{\partial \tau} + \theta(x, \tau) = -\nabla_x \cdot [\delta(\mathbf{x}, \tau) \mathbf{U}(\mathbf{x}, \tau)] \quad (1.51)$$

$$\frac{\partial \theta(x, \tau)}{\partial \tau} + \mathcal{H}(\tau)\theta(x, \tau) + \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\delta(\mathbf{x}, \tau) = \nabla_x \cdot [\mathbf{U}(\mathbf{x}, \tau) \cdot \nabla_x \mathbf{U}(\mathbf{x}, \tau)], \quad (1.52)$$

Así, haciendo una representación en el espacio de Fourier es posible determinar los modos espectrales y estudiar su evolución, con lo cual la ecuación 1.51 y 1.52 se expresan como:

$$\frac{\partial \tilde{\delta}(\mathbf{k}, \tau)}{\partial \tau} + \tilde{\theta}(\mathbf{k}, \tau) = -\int d^3 \mathbf{q}_1 d^3 \mathbf{q}_2 \delta_D(\mathbf{k} - \mathbf{q}_{12}) \alpha(\mathbf{q}_1, \mathbf{q}_2) \tilde{\theta}(\mathbf{q}_1, \tau) \tilde{\delta}(\mathbf{q}_2, \tau), \quad (1.53)$$

$$\begin{aligned} \frac{\partial \tilde{\theta}(\mathbf{k}, \tau)}{\partial \tau} + \mathcal{H}(\tau)\tilde{\theta}(\mathbf{k}, \tau) + \frac{3}{2}\Omega_m(\tau)\mathcal{H}^2(\tau)\tilde{\delta}(\mathbf{k}, \tau) = \\ -\int d^3 \mathbf{q}_1 d^3 \mathbf{q}_2 \delta_D(\mathbf{k} - \mathbf{q}_{12}) \beta(\mathbf{q}_1, \mathbf{q}_2) \tilde{\theta}(\mathbf{q}_1, \tau) \tilde{\delta}(\mathbf{q}_2, \tau), \end{aligned} \quad (1.54)$$

donde δ_D es la función delta de Dirac, $\tilde{\theta}$ y $\tilde{\delta}$ denotan la transformada de Fourier de θ y δ respectivamente, $\mathbf{q}_{12} = \mathbf{q}_1 + \mathbf{q}_2$ y α y β son potenciales que reúnen las no-linealidades y están definidos por

$$\alpha(\mathbf{q}_1, \mathbf{q}_2) = \frac{\mathbf{q}_{12} \cdot \mathbf{q}_1}{q_1^2}, \quad \beta(\mathbf{q}_1, \mathbf{q}_2) = \frac{q_{12}^2 \mathbf{q}_1 \cdot \mathbf{q}_2}{2q_1^2 q_2^2}, \quad (1.55)$$

De esta manera, las soluciones no lineales están dadas por [2]:

$$\delta(\mathbf{k}, \tau) = \sum_{n=0}^{\infty} D_1(\tau)^n \delta_n(\mathbf{k}) \quad (1.56)$$

$$\theta(\mathbf{k}, \tau) = -\mathcal{H}f(\Omega_m, \Omega_\Lambda) \sum_{n=0}^{\infty} D_1(\tau)^n \theta_n(\mathbf{k}), \quad (1.57)$$

con

$$\delta_n(\mathbf{k}) = \int d^3 \mathbf{q}_1 \cdots \int d^3 \mathbf{q}_n \delta_D(\mathbf{k} - \mathbf{q}_{1, \dots, n}) F_n^{(s)}(\mathbf{q}_1, \dots, \mathbf{q}_n) \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n) \quad (1.58)$$

$$\theta_n(\mathbf{k}) = \int d^3 \mathbf{q}_1 \cdots \int d^3 \mathbf{q}_n \delta_D(\mathbf{k} - \mathbf{q}_{1, \dots, n}) G_n^{(s)}(\mathbf{q}_1, \dots, \mathbf{q}_n) \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n), \quad (1.59)$$

donde los kernels F_n y G_n son funciones homogéneas de los vectores $(\mathbf{q}_1, \dots, \mathbf{q}_n)$, los cuales están dados a través de las siguientes formulas de recurrencia:

$$F_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_m)}{(2n+3)(n-1)} [(2n+1)\alpha(\mathbf{k}_1, \mathbf{k}_2)F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) + 2\beta(\mathbf{k}_1, \mathbf{k}_2)G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n)], \quad (1.60)$$

$$G_n(\mathbf{q}_1, \dots, \mathbf{q}_n) = \sum_{m=1}^{n-1} \frac{G_m(\mathbf{q}_1, \dots, \mathbf{q}_m)}{(2n+3)(n-1)} [3\alpha(\mathbf{k}_1, \mathbf{k}_2)F_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n) + 2n\beta(\mathbf{k}_1, \mathbf{k}_2)G_{n-m}(\mathbf{q}_{m+1}, \dots, \mathbf{q}_n)], \quad (1.61)$$

siendo $\mathbf{k}_1 = \mathbf{q}_1 + \dots + \mathbf{q}_m$ y $\mathbf{k}_2 = \mathbf{q}_{m+1} + \dots + \mathbf{q}_n$ y $F_1 = G_1 = 1$. Para el caso de $n = 2$ se tiene :

$$F_2^s(\mathbf{q}_1, \mathbf{q}_2) = \frac{5}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_1}{q_1 q_2} \left(\frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{2}{7} \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2} \quad (1.62)$$

$$G_2^s(\mathbf{q}_1, \mathbf{q}_2) = \frac{3}{7} + \frac{1}{2} \frac{\mathbf{q}_1 \cdot \mathbf{q}_1}{q_1 q_2} \left(\frac{q_1}{q_2} + \frac{q_2}{q_1} \right) + \frac{4}{7} \frac{(\mathbf{q}_1 \cdot \mathbf{q}_2)^2}{q_1^2 q_2^2} \quad (1.63)$$

El superíndice en las soluciones indica que estas deben ser simétricas respecto al cambio de las \mathbf{q}_i [3].

La evolución del campo de fluctuaciones en la densidad de materia $\delta(\mathbf{x}, \tau)$ explica la formación de estructuras a gran escala en el universo desde la teoría de perturbaciones, sin embargo, localmente se requiere una descripción al menos teórica de la manera y condiciones en las cuales se sucede el colapso gravitatorio de una determinada masa en un objeto sobredenso respecto al background, una de cuyas formas se describe en la siguiente sección.

1.5. Modelo de colapso esférico

Una de las aproximaciones más simples y útiles resulta de suponer que los halos de materia oscura son formados a partir del colapso esférico de las perturbaciones de la densidad de materia. Este modelo permite mostrar de manera detallada el procedimiento seguido para encontrar la densidad crítica necesaria para la formación de un halo. Se define una región esférica inicial de radio comovil R_0 con una densidad inicial δ_0 . Se asume que las fluctuaciones son Gaussianas con rms (desviación estándar) en la escala de R_0 y es mucho menor a la unidad $|\delta_0| \ll 1$. Esto significa que la masa M_0 dentro de R_0 es:

$$M_0 = \frac{4}{3} \pi R_0^3 \bar{\rho} (1 + \delta_0) \approx \frac{4}{3} \pi R_0^3 \bar{\rho}_0 \quad (1.64)$$

donde $\bar{\rho}$ denota la densidad de fondo.

Tras la evolución del universo, la región cambia de tamaño y suponiendo que la masa original siga contenida en la región entonces se tendrá:

$$M_0 = \frac{4}{3}\pi R^3 \bar{\rho}(\tau)(1 + \delta) \quad (1.65)$$

donde R es el nuevo radio de la región. Por otra parte, tomando la expresión (1.50) para el campo de densidad y considerando $\Omega_i \approx 1$ se obtiene que [20]:

$$\delta = \frac{3}{5} \left(\frac{t}{t_i} \right)^{2/3} \left(\delta_i - t_i \frac{\nabla \cdot \mathbf{v}_i}{a_i} \right) + \frac{2}{5} \left(\frac{t_i}{t} \right) \left(\delta_i + \frac{3}{2} t_i \frac{\nabla \cdot \mathbf{v}_i}{a_i} \right), \quad (1.66)$$

donde t es el tiempo cosmológico y el subíndice i denota el valor de la magnitud correspondiente evaluada en el tiempo t_i . Si se considera adicionalmente, que el colapso inicia cuando la velocidad es cero y que el interés se centra en la parte creciente de la solución, entonces se denotará $\delta' = \frac{3}{5}\delta_i$.

Por otra parte, el radio de la esfera que encierra la masa M cambia en el tiempo y se denota por $r(t)$ de tal manera que su aceleración en el límite Newtoniano será:

$$\frac{d^2 r}{dt^2} = -\frac{GM}{r^2} \quad (1.67)$$

y asumiendo que la masa se mantiene en la región entonces la primera integral es:

$$\left(\frac{dr}{dt} \right)^2 = \frac{2GM}{r} + C \quad (1.68)$$

donde la constante de integración C indica energía de la región y determina el tipo de solución en las ecuaciones (1.67) y (1.68) así:

$$C > 0 : r = A(\cosh \theta - 1) \quad t = B(\sinh \theta - \theta), \quad (1.69)$$

$$C < 0 : r = A(1 - \cos \theta) \quad t = B(\theta - \sin \theta), \quad (1.70)$$

siendo $A^3 = GMB^2$ y dado que en coordenadas físicas $r = aR$ entonces:

$$r_b = A_b(\cosh \eta - 1) \quad t = B_b(\sinh \eta - \eta) \quad A_b^3 = GMB_b^2 \quad (1.71)$$

Usando expansiones en series y teniendo en cuenta que $\delta_i \ll 1$, $\theta_i \ll 1$ y $\eta_i \ll 1$ entonces:

$$t \approx \frac{B\theta^3}{6} \left(1 + \frac{\theta^2}{20} \right) \quad (1.72)$$

$$r \approx \frac{A\theta^3}{2} \left(1 + \frac{\theta^2}{12} \right), \quad (1.73)$$

de tal manera que de la ecuación de tiempo (1.72) se obtiene:

$$\theta^2 \approx \left(\frac{6t}{B} \right)^{2/3} \left(1 - \frac{1}{20} \left(\frac{6t}{B} \right)^{2/3} \right)^{2/3}, \quad (1.74)$$

y reemplazando está expresión en la ecuación de radio (1.73)

$$r \approx \left(\frac{9}{2}GMt^2\right) \left[1 + \frac{1}{20} \left(\frac{6t}{B}\right)^{2/3}\right] \quad (1.75)$$

Reemplazando estos resultados en la densidad inicial de contraste se tiene:

$$\rho = \bar{\rho}(1 + \delta_i) = \frac{3M_0}{4\pi R_0} \quad (1.76)$$

$$\bar{\rho}(1 + \delta_i) = \frac{1}{6\pi Gt^2} \left[1 - \frac{3}{20} \left(\frac{6t}{B}\right)^{2/3}\right], \quad (1.77)$$

de donde

$$\frac{\rho}{\bar{\rho}} = \left[1 + \delta_c - \frac{3}{20} \left(\frac{6t}{B}\right)^{2/3}\right] = 1 + \delta_i \quad (1.78)$$

donde $\delta_c = 3/5(\Omega_i - 1)$. Así, las constantes A y B se expresan como:

$$A = \frac{3r_i}{10(\delta_c - \delta_i)} \quad B = 6t_i \left[\frac{3}{20(\delta_c - \delta_i)}\right]^{3/2} \quad (1.79)$$

y para el background

$$A_b = \frac{3r_i}{10\delta_c} \quad B_b = 6t_i \left(\frac{3}{20\delta_c}\right)^{3/2} \quad (1.80)$$

Usando las ecuaciones (1.71) para las condiciones iniciales entonces:

$$r_0 = \frac{1}{10\delta_c} r_0 (\cosh \eta_0 - 1), \quad (1.81)$$

Este resultado fija la evolución de la masa relativa al background de la forma

$$\cosh \eta = 2\Omega^{-1} - 1, \quad (1.82)$$

así, igualando los tiempos para el background y para la evolución del halo se lora fijar el parámetro θ

$$\sinh \theta - \theta = \left(1 - \frac{\delta_i}{\delta_c}\right)^{3/2} (\sinh \eta - \eta), \quad (1.83)$$

finalmente para un modelo abierto se puede escribir:

$$\left(\frac{r_0}{r}\right)^3 = \left[\left(1 - \frac{\delta_i}{\delta_c}\right) \frac{\cosh \eta - 1}{\cosh \theta - 1}\right]^3, \quad (1.84)$$

De manera análoga para un universo cerrado y plano, tomando el limite cuando $\Omega \rightarrow 1$, $\eta \rightarrow 0$, se tiene:

$$\frac{R_0}{R} = \frac{6^{2/3}}{2} \frac{(\theta - \sin \theta)^{2/3}}{1 - \cos \theta}, \quad (1.85)$$

Así, de la relación entre η y el tiempo cosmológico se puede obtener [23]:

$$\frac{1}{1+z} = \left(\frac{3}{4}\right)^{2/3} \frac{(\theta - \sin \theta)^{2/3}}{(5/3)|\delta_0|}, \quad (1.86)$$

1.6. Modelo del Halo

El modelo estándar de la cosmología (Λ CDM) considera la existencia de materia oscura como componente importante del universo, la cual solo interactúa de forma gravitacional con la materia visible, de tal manera que la formación de aglomeraciones de materia oscura llamadas *halos* resulta una conclusión natural. La descripción físico-matemática de tales aglomeraciones se conoce como *modelo del halo*, el cual durante los últimos años ha servido para el estudio de las estructuras a gran escala del universo. Este modelo permite estudiar la función de correlación de N puntos tanto en campos de densidad de materia oscura (desde simulaciones cosmológicas de N -cuerpos) como en muestreos observacionales de galaxias pudiendo con ello estudiar aspectos como la distribución de las mismas. El Modelo del Halo hace algunos supuestos importantes como son:

- Toda la materia del universo está dividida en unidades distintas llamadas Halos
- A pequeñas escalas, las estadísticas del campo de densidad de masa son determinadas por la distribución espacial dentro del halo.
- A escalas mayores al tamaño típico de un halo, la estadística importante es la distribución espacial de halos.
- Los halos son espacialmente excluyentes, lo cual significa que un punto solo puede pertenecer a un halo en un momento t .

En general, el Modelo del Halo tiene tres componentes, *la función de masa, el sesgo* (bias) y la distribución espacial de materia oscura dentro del halo también llamada *perfil del halo*.

El objetivo principal es describir el campo de densidad de masa como un conjunto discreto de halos correlacionados, para lo cual considere halos de masa M_i , centrados en \mathbf{x}_i y a un redshift z de forma que la descripción discreta de la densidad es:

$$\rho(\mathbf{x}; z) = \sum_i \rho_h(\mathbf{x} - \mathbf{x}_i; m_i, z), \quad (1.87)$$

y que expresándolo de forma normalizada se tiene:

$$u(\mathbf{x} - \mathbf{x}_i) = \frac{\rho(\mathbf{x} - \mathbf{x}_i)}{\int \rho(\mathbf{x} - \mathbf{x}'_i) d^3x'_i}, \quad (1.88)$$

la masa puede ser expresada como una integral sobre funciones continuas

$$\rho(\mathbf{x}) = \sum_i m_i u(\mathbf{x} - \mathbf{x}_i | m_i) \quad (1.89)$$

$$= \sum_i \int \delta(m - m_i) \delta(\mathbf{x}' - \mathbf{x} | m) dm d^3x', \quad (1.90)$$

la densidad número de halos es

$$n(m) = \left\langle \sum_i \delta(m - m_i) \delta^3(\mathbf{x} - \mathbf{x}_i) \right\rangle, \quad (1.91)$$

donde $\langle \dots \rangle$ denota el ensamble promedio sobre realizaciones del campo cosmológico. De acuerdo con [5] la densidad media del universo es:

$$\bar{\rho} = \langle \rho(\mathbf{x}) \rangle = \left\langle \sum_i m_i u(\mathbf{x} - \mathbf{x}_i | m_i) \right\rangle \quad (1.92)$$

$$= \int mn(m) dm \int u(\mathbf{x} - \mathbf{x}') d^3x' \quad (1.93)$$

$$= \int mn(m) dm, \quad (1.94)$$

donde el ensamble promedio ha sido reemplazado por el promedio sobre la función de masa de los halos y sobre el espacio.

1.6.1. Función de masa

Esta función permite describir la distribución de masa de halos de materia oscura en el universo. Un modelo de función de masa fue propuesto por Press & Schechter [21], el cual permite predecir la masa de las aglomeraciones de materia oscura considerando que estas siguen un proceso jerárquico de formación, esto significa que primero se forman pequeñas aglomeraciones que luego dan origen a otras de mayor tamaño. Este modelo supone que la distribución espacial puede ser descrita de manera Gaussiana, la cual es utilizada para construcción del modelo con el cual se determina la correlación entre cualquier par de objetos.

Adicionalmente, el modelo también supone que las regiones sobredensas $\delta > \delta_{sc}$ colapsan a objetos observables, donde δ_{sc} es la densidad crítica para que ello ocurra. De esta forma, la búsqueda de objetos de masa M se logra utilizando un suavizado de la sobredensidad (campo) inicial mediante un filtro de radio $R = R(M)$, el cual se distribuye de forma Gaussiana. Así, la varianza del campo esta dada por:

$$\sigma^2(R) = \int \frac{k^3 P^{lin}(k) dk}{2\pi^2} \frac{1}{k} |W(kR)|^2 \quad (1.95)$$

donde $W(x)$ es el filtro utilizado, usualmente $W(x) = e^{-\frac{x^2}{2}}$. Dada la relación de la varianza con el espectro de potencia entonces la evolución temporal esta dada por:

$$\sigma^2(\tau) = D^2(\tau) \sigma^2(\tau_0) \quad (1.96)$$

Si $P(\delta, V)$ denota la probabilidad de que un volumen V contenga una masa cuya desviación promedio este entre δ y $\delta + d\delta$, entonces la aproximación Gaussiana es:

$$P(\delta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\delta^2}{2\sigma^2}\right) \quad (1.97)$$

con lo cual, la fracción de objetos con masa mayor que $m = 4/3\bar{\rho}\pi R_f^3$ puede ser calculada como la probabilidad que en un volumen la densidad supere un valor crítico δ_{sc} , esto es:

$$P(\delta > m) = \int_{\delta_{sc}}^{\infty} P(\delta)d\delta \quad (1.98)$$

$$F(m) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\delta_{sc}}^{\infty} \exp\left(-\frac{\delta^2}{2\sigma^2}\right)d\delta \quad (1.99)$$

de tal manera que la fracción de objetos colapsados con masas entre m y $m + dm$ es:

$$f(m)dm = F(m) - F(m + dm) = -\frac{\partial F}{\partial m}dm \quad (1.100)$$

A partir de este resultado, la densidad número de objetos estará dada por el producto del número de parches por la fracción de objetos dividido en el volumen, esto es:

$$n(m)dm = \frac{V(m/\bar{\rho})f(m)dm}{V} = -\rho \frac{\partial F}{\partial m} \frac{dm}{m} \quad (1.101)$$

Entonces haciendo $\nu = \frac{\delta_{sc}^2}{\sigma^2}$ y $x = \frac{\delta^2}{\sigma^2}$ se tiene que la expresión 1.99 se tiene:

$$F(m) = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{\nu}}^{\infty} \exp\left(-\frac{x}{2}\right)dx \quad (1.102)$$

con lo cual la derivada parcial de F es:

$$\frac{\partial F}{\partial m} = \frac{2}{\sqrt{2\pi}} \exp(-\nu/2) \frac{d\sigma}{dm} \quad (1.103)$$

así, la densidad número en el modelo de Press & Shechter es:

$$n_{ps}dm = -2\bar{\rho} \frac{1}{\sqrt{2\pi}} \sqrt{\nu} e^{-\nu/2} \frac{d\sigma}{\sigma} \quad (1.104)$$

1.6.2. Sesgo

A partir del campo de densidad de materia, unas condiciones iniciales dadas en el universo temprano por las funciones de transferencia y la teoría de perturbaciones es posible describir la formación de estructuras tras la evolución del universo, tales como los halos de materia oscura, sin embargo, la formación de nuevas estructuras a partir de dichos halos requiere una descripción mas rigurosa y detallada. En tal sentido, el campo de densidad de halos constituye un elemento central para poder estudiar esa dinámica de formación de estructuras. Una forma de abordar este problema es considerar el campo de densidad de halos como una versión sesgada del campo de densidad de materia con lo cual se da origen a un término de sesgo o bias por su nombre en ingles. [24].

Para empezar, considere el campo de densidad de todos los halos con masas en un rango de m a $m + dm$, suavizado con algún filtro a una escala R . Por otra parte, se asume que la densidad de halos δ^h es un mapeo determinístico \mathcal{F} del campo de densidad de materia

$\delta(\mathbf{x})$, el cual se aplica en todo punto \mathbf{x} del campo así:

$$\delta^h(\mathbf{x}|R, m) = \mathcal{F}_{(m,R)}[\delta(\mathbf{x}|R)] \quad (1.105)$$

donde \mathcal{F} es una función que depende de la masa del halo y un filtro de escala R y la densidad $\delta(\mathbf{x}|R)$

$$\delta(\mathbf{x}|R) = \frac{1}{V} \int \delta(\mathbf{y}) W(|\mathbf{x} - \mathbf{y}|, R) d^3y \quad (1.106)$$

con $W(|\mathbf{x}|, R)$ algún filtro normalizado. La expansión de Taylor de la función $\mathcal{F}_{(m,R)}$ alrededor del punto $\delta = 0$ es:

$$\mathcal{F}_{(m,R)}[\delta(\mathbf{x}|R)] = \sum_{i=0} \frac{b_i(m|R)}{i} [\delta(\mathbf{x}|R)]^i \quad (1.107)$$

de tal manera que δ^h se puede escribir como:

$$\delta^h(\mathbf{x}|R, m) = \sum_{i=0}^{\infty} \frac{b_i(m)}{i} [\delta(\mathbf{x}|R)]^i \quad (1.108)$$

donde los coeficientes b_i están dados por:

$$b_i(m) = \left. \frac{\partial^i \mathcal{F}_{(m)}[\chi]}{\partial \chi^i} \right|_{\chi=0} \quad (1.109)$$

Para el caso en el cual $b_i = 0$ para todo $i > 1$ se tiene un modelo de sesgo lineal dado por:

$$\delta^h(\mathbf{x}|R, m) = b_0(m) + b_1(m)\delta(\mathbf{x}|R) \quad (1.110)$$

1.6.3. Perfiles del halo

El perfil de densidad de masa del halo es uno de los aspectos mejor descritos en relación a los resultados de las simulaciones de alta resolución disponibles. A pesar que los halos no tienen una simetría exactamente esférica, esta geometría ha sido de gran utilidad para la descripción del perfil de densidad del halo. En esencia, el perfil del halo se entiende como una función que describe la distribución espacial de materia oscura dentro del halo. Navarro -et al (1996) [15] obtienen una descripción de tal perfil como una función biparamétrica dada por la expresión:

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2} \quad (1.111)$$

donde r_s es un parámetro que define una escala de radio con una correspondiente densidad ρ_s . En este perfil la densidad $\rho(r)$ se comporta como r^{-1} en el núcleo ($r \ll r_s$) y en la parte más externa ($r \gg r_s$) como r^{-3} . Este como otros perfiles del halo son ajustados a partir de simulaciones cosmológicas.

De esta manera, la masa contenida en el halo se obtiene por integración de [1.111](#)

$$\int_0^r \rho(\mathbf{r}) d\mathbf{r} = 4\pi\rho_0 r_s^3 \left[\ln \left(1 + \frac{r}{r_s} \right) - \frac{r/r_s}{1 + r/r_s} \right] \quad (1.112)$$

Por otra parte, el modelo de colapso esférico predice que las estructuras en equilibrio virial poseen un contraste de densidad $\Delta = \rho_{<r}/\rho_{crit}$ aproximadamente igual a 100 para el modelo de Λ CDM [\[15\]](#) donde $\rho_{<r}$ es la densidad media de una esfera de radio r y $\rho_{crit} = 3H^2/8\pi G$ es la densidad crítica, la cual es necesaria para que la curvatura de universo sea cero, H es la constante de Hubble y G la constante gravitacional universal. El radio donde se alcanza este equilibrio se denomina usualmente radio virial r_{vir} y se lo asocia al tamaño característico del halo simulado. La masa del halo, denominada comúnmente masa virial M_{vir} está relacionada con r_{vir} a través de la expresión [\[7\]](#):

$$M(r)|_{r=r_{vir}} = \frac{4}{3}\pi\Delta\rho_{crit}r_{vir}^3 \quad (1.113)$$

De la ecuación [1.112](#) y [1.113](#) se tiene que el perfil del halo se puede reescribir como:

$$\frac{\rho(r)}{\rho_{crit}} = \frac{\delta_c}{(cr/r_{vir})(1 + cr/r_{vir})^2} \quad (1.114)$$

donde c es el parámetro de concentración definido por $c = r/r_{vir}$ y δ_c es la densidad característica del halo dada por:

$$\delta_c = \frac{\Delta c^3}{3(\ln 1 + c - c/(1 + c))} \quad (1.115)$$

1.6.4. Función de correlación

La función de correlación de dos puntos, se define como el exceso de probabilidad que dos partículas en volúmenes dV_1 y dV_2 respectivamente estén separadas por una distancia r_{12}

$$dP = n^2 dV_1 dV_2 [1 + \xi(r_{12})] \quad (1.116)$$

Una vez el campo de densidad es definido, el siguiente paso es obtener medidas estadísticas tales como la función de correlación de dos, tres ó N puntos. En general, dicha correlación es función de un conjunto de parámetros tales como posición, masa, redshift entre otros, y que bajo el principio cosmológico se reduce a depender de la distancia entre los halos. En términos del campo de densidad la función de correlación de dos puntos se expresa como:

$$\begin{aligned} \xi(\mathbf{r}_1, \mathbf{r}_2) &= \langle \delta(\mathbf{r}_1), \delta(\mathbf{r}_2) \rangle \quad (1.117) \\ &= \left\langle \sum_i \delta(\mathbf{x}_1 - \mathbf{x}_i) \delta(m_1 - m_i), \sum_j \delta(\mathbf{x}_2 - \mathbf{x}_j) \delta(m_2 - m_j) \right\rangle \\ &= n(m_1) \delta(\mathbf{x}_1 - \mathbf{x}_2) \delta(m_2 - m_1) + n(m_1)n(m_2)(1 + \xi_{hh}(m_1, m_2, \mathbf{x}_1, \mathbf{x}_2)), \end{aligned} \quad (1.118)$$

El primer término en [\(1.118\)](#) corresponde a la correlación debida a elementos dentro de un mismo halo la cual domina a pequeñas escalas, en tanto que el segundo término es la

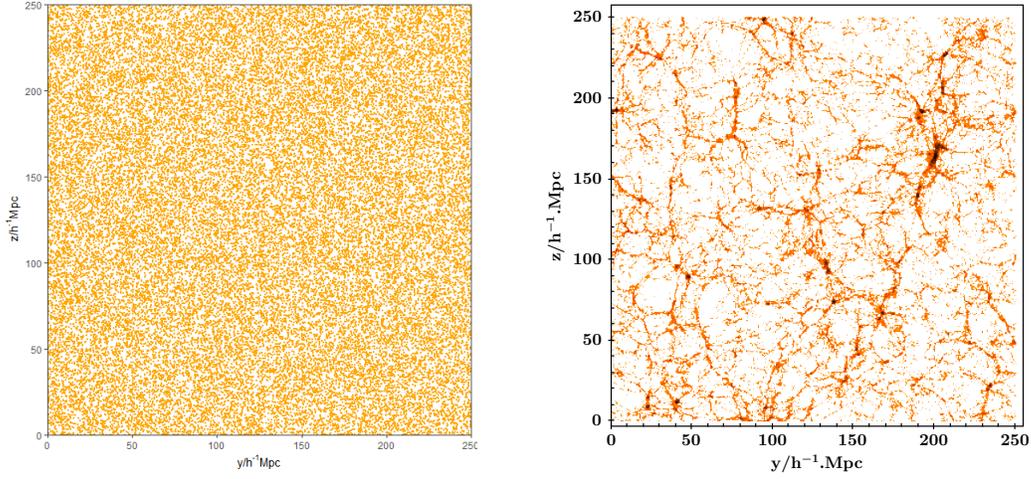


FIGURA 1.2: Campo aleatorio (izquierda) $\xi(r_{12}) = 0$ y campo con estructuras (derecha) $\xi(r_{12}) > 0$

contribución debida a halos diferentes y por tanto su efecto se hará mas evidente a gran escala. En esta segunda contribución ξ_{hh} denota el exceso de probabilidad con respecto a un campo aleatorio uniforme. En tal sentido, la función de correlación de dos puntos se expresa estructuralmente como la suma de dos contribuciones:

$$\xi(\mathbf{r}_1, \mathbf{r}_2) = \xi_{1h}(\mathbf{r}_1, \mathbf{r}_2) + \xi_{2h}(\mathbf{r}_1, \mathbf{r}_2), \quad (1.119)$$

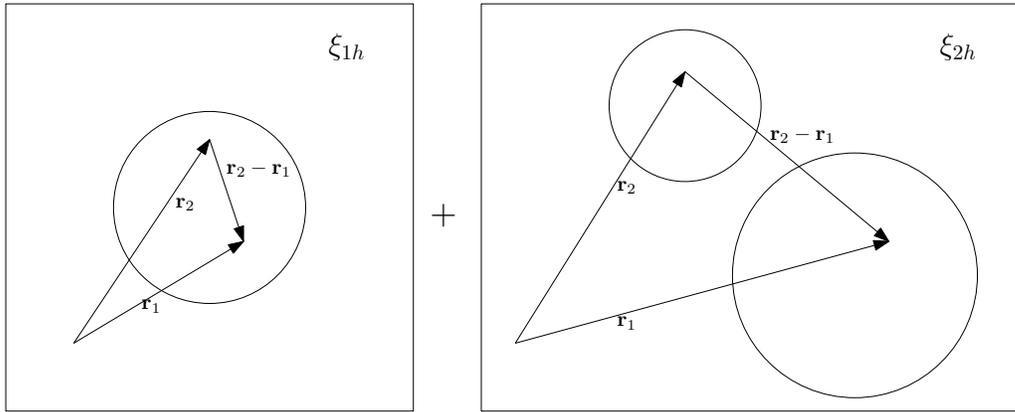


FIGURA 1.3: Componentes de la función de correlación de dos puntos

Así, cada término se puede expresar como:

$$\xi_{1h}(\mathbf{r}_1, \mathbf{r}_2) = \int \frac{m_1^2 n(m_1)}{\bar{\rho}^2} dm_1 \int u_{m_1}(\mathbf{r}_1 - \mathbf{x}_1) u_{m_1}(\mathbf{r}_2 - \mathbf{x}_1) d^3 x_1 \quad (1.120)$$

$$\xi_{2h}(\mathbf{r}_1, \mathbf{r}_2) = \int dm_1 \int n(m_1)n(m_2)dm_2 \int \frac{u_{m_1}(\mathbf{r}_1 - \mathbf{x}_1)}{\bar{\rho}} d^3x_1 \times \int \frac{u_{m_2}(\mathbf{r}_2 - \mathbf{x}_2)}{\bar{\rho}} \xi_{hh}(m_1, m_2, \mathbf{x}_1, \mathbf{x}_2) d^3x_2 \quad (1.121)$$

El término ξ_{hh} es requerido para conocer la correlación halo-halo y puede ser aproximado por:

$$\xi_{hh}(m_1, m_2, r) \approx b(m_1)b(m_2)\xi(r) \quad (1.122)$$

donde $b(m_i)$ es el término de sesgo y $\xi(r)$ es la correlación dependiendo unicamente de la distancia entre los dos elementos de masa.

En general, las funciones de correlación implican en su calculo la convolución de los perfiles de masa, por lo cual es conveniente trabajar con sus representaciones de Fourier.

1.6.5. El espectro de potencias en el modelo del halo

El espectro de potencias es la representación de la función de correlación en el espacio de Fourier. La principal razón para ello radica en que la convolución de funciones en el espacio real, se convierten en multiplicaciones en el espacio de Fourier. Para ello consideremos:

$$u_m(\mathbf{k}) = \int \frac{1}{(2\pi)^3} u_m(\mathbf{x}) e^{-\mathbf{k}\cdot\mathbf{x}} d^3x \quad (1.123)$$

y la representación integral de la función delta de Dirac:

$$\delta_D(\mathbf{k}_{123\dots N}) = \int \frac{1}{(2\pi)^3} e^{-i\mathbf{x}\cdot(\mathbf{k}_1+\mathbf{k}_2+\dots+\mathbf{k}_N)} d^3x \quad (1.124)$$

donde $\mathbf{k}_{123\dots N} = \mathbf{k}_1 + \mathbf{k}_2 + \dots + \mathbf{k}_N$. Un punto importante en el modelo del Halo es que para la descripción del agrupamiento (clustering) no lineal de los halos de materia oscura se tiene en cuenta que ellos mismo están formados siguiendo fluctuaciones del campo de densidad lineal.

Aplicando la transformada de Fourier a la componente ξ_{1h} dada en la expresión (1.120) se tiene:

$$\int \frac{m_1^2 n(m_1)}{\bar{\rho}^2} dm_1 \int u_{m_1}(\mathbf{r}_1 - \mathbf{x}_1) u_{m_1}(\mathbf{r}_2 - \mathbf{x}_1) d^3x_1 = \int d^3\mathbf{k}_1 \int d^3\mathbf{k}_2 \int \frac{m_1^2 n(m_1)}{\bar{\rho}^2} u_m(\mathbf{k}_1) u_m(\mathbf{k}_2) (2\pi)^3 \delta_d(\mathbf{k}_{12}) e^{i(\mathbf{k}_1\cdot\mathbf{r}_1 + \mathbf{k}_2\cdot\mathbf{r}_2)} d^3x_1 \quad (1.125)$$

y dado que $u_m(-\mathbf{k}_1) = u_m(\mathbf{k}_1)^*$ entonces la contribución de un halo a la correlación en el espacio de Fourier es:

$$\xi_{1h}(\mathbf{r}_1, \mathbf{r}_2) = (2\pi)^3 \int d^3\mathbf{k} \int \frac{m^2 n(m)}{\bar{\rho}^2} |u_m(\mathbf{k})|^2 e^{i\mathbf{k}\cdot(\mathbf{r}_1 - \mathbf{r}_2)} dm \quad (1.126)$$

con lo cual el espectro de potencias de un halo es identificado como:

$$P_{1h}(\mathbf{k}) = (2\pi)^3 \int \frac{m^2 n(m)}{\bar{\rho}^2} |u_m(\mathbf{k})|^2 dm \quad (1.127)$$

Para el espectro de potencias de dos halos se tiene que la transformada de Fourier de la contribución correspondiente es:

$$P_{2h}(\mathbf{k}) = (2\pi)^6 \int n(m_1)n(m_2) \frac{m_1 m_2}{\bar{\rho}^2} u_{m_1}(\mathbf{k}_1) u_{m_2}(\mathbf{k}_2) P_{hh}(\mathbf{k}, m_1, m_2) dm_1 dm_2 \quad (1.128)$$

donde P_{hh} es la transformada de Fourier del término ξ_{hh}

$$\xi_{hh}(\mathbf{x}_1, \mathbf{x}_2, m_1, m_2) = \int P_{hh}(\mathbf{k}, m_1, m_2) e^{i\mathbf{k} \cdot (\mathbf{x}_1 - \mathbf{x}_2)} d^3\mathbf{k} \quad (1.129)$$

En general se tiene entonces que el espectro de potencias se puede expresar como:

$$P(k) = P_{1h}(k) + P_{2h}(k) \quad (1.130)$$

Así mismo, la transformada de Fourier para la expresión (1.122) es:

$$P_{hh}(k) \approx b(m_1)b(m_2)P(k) \quad (1.131)$$

donde $P(k)$ el elemento central es como determinar $P(k)$, siendo la opción más común el espectro lineal $P_{lin}(k)$, sin embargo un uso más refinado lo constituye el espectro no lineal.

1.6.6. Estimadores de la Función de Correlación

Para campos de densidad de halos proveniente de simulaciones cosmológicas, la función de correlación de dos puntos se puede estimar directamente realizando los respectivos conteos en relación a un campo aleatorio binomial (Ver sección 1.9.1).

Para ello, se parte de un conjunto de N elementos (halos de materia oscura en nuestro caso) que se representan por puntos distribuidos en una ventana tridimensional W y cuya posición se especifica en algún sistema de coordenadas, particularmente rectangulares para este caso y se denomina el *campo objetivo*. Por otra parte, se genera artificialmente un campo aleatorio binomial con $N_r \geq N$ puntos en la mismas ventana W , el cual sirve como *campo de comparación*.

La estimación de $\xi(r)$ se realiza comparando la distribución de los objetos en el *campo objetivo* con la correspondiente en el *campo de comparación*, realizando para ello el conteo exhaustivo de la cantidad de pares de puntos que se encuentran en un rango de distancia r y $r + dr$ denotados como DD y RR para el campo objetivo y de comparación respectivamente. Adicionalmente si se superponen los campos en la misma ventana W se puede realizar el respectivo conteo de pares entre puntos de los dos campos, lo cual es denotado como DR . La figura 1.4 ilustra estos conteos en un análogo de dos dimensiones, de tal manera que en el caso real de estudio el anillo de las figuras corresponde a un casquete esférico.

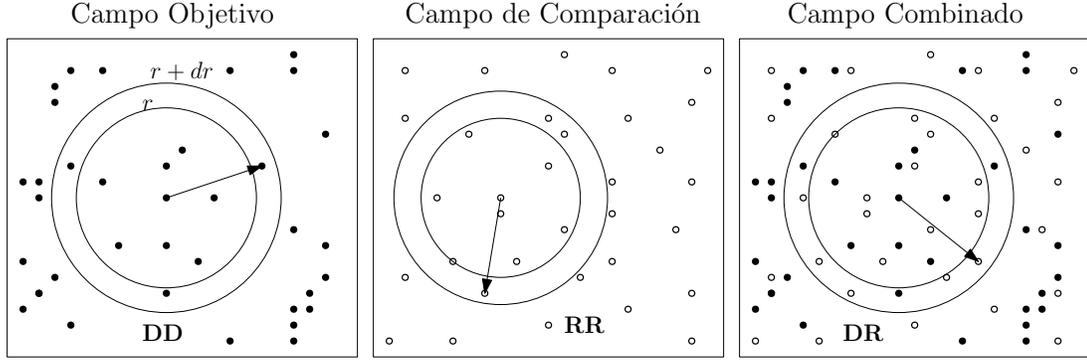


FIGURA 1.4: Campos puntuales empleados para la estimación de la función de correlación.

Existen varios estimadores de la función de correlación de dos puntos entre los cuales tiene:

- **Peebles-Hauser:** Propuesto en 1974, este compara el exceso de pares de puntos en un rango de distancia entre el campo objetivo y el campo de comparación. Se define como:

$$\hat{\xi}_{PH}(r) = \left(\frac{N_r}{N} \right)^2 \frac{DD(r)}{RR(r)} - 1 \quad (1.132)$$

- **Davis-Peebles:** También conocido como estimador estándar, es propuesto en 1983 por los autores que le dan su nombre, presentando menor sesgo y mejorando la corrección el efecto de borde, el cual se genera cuando se cuentan pares de puntos cerca al borde de la ventana pues naturalmente se subestima el resultado. Este estimador se define como:

$$\hat{\xi}_{DP}(r) = \frac{N_r}{N} \frac{DD(r)}{DR(r)} - 1 \quad (1.133)$$

- **Hamilton:** Aunque sigue siendo un estimador sesgado, la magnitud de este es menor al de los estimadores mencionados previamente [26].

$$\hat{\xi}_{HM}(r) = \frac{DD(r)RR(r)}{[DR(r)]^2} - 1 \quad (1.134)$$

- **Landy-Szalay:** Al igual que el estimador de Hamilton, fue propuesto en 1993 y presentando un sesgo menor al de los dos primeros estimadores.

$$\hat{\xi}_{LS}(r) = 1 - \left(\frac{N_r}{N} \right)^2 \frac{DD(r)}{RR(r)} - 2 \frac{DR(r)}{RR(r)} \quad (1.135)$$

Varios estimadores para el error pueden ser empleados, entre los cuales el más sencillo se deriva de considerar una distribución Poisson, en cuyo caso está dado por:

$$\hat{\sigma} = \frac{1 + \xi(r)}{\sqrt{DD(r)}} \quad (1.136)$$

Otros estimadores pueden ser consultados en [26].

En un artículo de año 2000, [11] presentan un estudio comparativo en el cual concluyen que a pequeña y mediana escala ($r < 100 Mpc h^{-1}$) los diferentes estimadores exhiben un comportamiento muy similar, pero a escalas mayores se logran identificar diferencias. Así mismo, destacan el estimador de Hamilton y Landy-Szalay por presentar menor variabilidad pese a no ser insesgados.

1.7. Formalismo perturbaciones-halo

El formalismo perturbaciones-halo presentado por [24] puede ser descrito a partir de tres elementos:

1. Los Halos son trazadores sesgados de la distribución de masa. Para describir ese sesgo se asume que la relación entre el campo de densidad de halos y el campo de densidad de materia oscura es no lineal, local y determinística. Esto permite el uso del formalismo descrito por [8] y previamente mencionado en la sección 1.6.2.
2. El campo de densidad de materia oscura es propagado a régimen no lineal usando la teoría Euleriana de perturbaciones.
3. Se asume que las galaxias solo se forman en halos cuya masa supere un determinado umbral y son insertadas en los mismos usando el modelo del Halo

Partiendo de estos elementos y teniendo la densidad del campo de materia oscura (1.108), se procede a considerar la evolución del mismo a régimen no lineal mediante la teoría de perturbaciones. La idea principal que se emplea de esta teoría es que cada modo de Fourier del campo de densidad puede ser expandido como una serie,

$$\delta(\mathbf{k}, a) = \sum_n D_1^n(a) \delta_n(\mathbf{k}) \quad (1.137)$$

$$\delta_n(\mathbf{k}) = \int \frac{\prod_{i=1}^n [d^3 k_i \delta_1(\mathbf{k}_i)]}{(2\pi)^{3n-3}} [\delta^D(\mathbf{k})]_n F_n^{(s)}(\mathbf{k}_1, \dots, \mathbf{k}_n) \quad (1.138)$$

donde $\delta_n(\mathbf{k})$ es el n -ésimo orden de la perturbación Euleriana y $D_1^n(a)$ es el factor de crecimiento lineal. Así, tomando la transformada de Fourier de la relación de sesgo (1.108), a tercer orden e insertando la expansión dada en (1.137) se obtiene:

$$\begin{aligned} \delta^h(\mathbf{k}|R, m) = & b_1(m) [\delta_1(\mathbf{k}|R) + \delta_2(\mathbf{k}|R) + \delta_3(\mathbf{k}|R)] + \\ & \frac{b_2(m)}{2!} \int \frac{1}{(2\pi)^3} [\delta_1(\mathbf{q}_1|R) \delta_1(\mathbf{k} - \mathbf{q}_1|R) + 2\delta_1(\mathbf{q}_1|R) \delta_2(\mathbf{k} - \mathbf{q}_1|R)] d^3 q_1 + \\ & \frac{b_3(m)}{3!} \int \frac{1}{(2\pi)^6} [\delta_1(\mathbf{q}_1|R) \delta_1(\mathbf{q}_2|R) \delta_1(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2|R)] d^3 q_1 d^3 q_2 \quad (1.139) \end{aligned}$$

donde $\delta_n(\mathbf{q}|R) = W(|\mathbf{q}|R) \delta_i(\mathbf{q})$. El siguiente paso es insertar la expresión para δ_n dada en (1.138) en el campo de densidad de halos $\delta^h(\mathbf{k}|R, m)$ de tal manera que el

campo de densidad de halos a régimen cuasi-lineal puede ser escrito como una expansión en serie perturbada del campo de densidad de materia oscura así:

$$\delta^h(\mathbf{k}, a|m, R) = \sum_n D_1^n(a) [\delta^h(\mathbf{k}|m)]_n; \quad (1.140)$$

$$[\delta^h(\mathbf{k}|m, R)]_n = \int \frac{\prod_{i=1}^n [d^3 q_i \delta_1(\mathbf{q}_i)]}{(2\pi)^{3n-3}} [\delta^D(\mathbf{k})]_n F_n^{(h)}(\mathbf{q}_1, \dots, \mathbf{q}_n|m, R) \quad (1.141)$$

donde $[\delta^h(\mathbf{k}|m, R)]_n$ es la perturbación de orden n del campo de densidad de halos y la abreviación $[\delta^D(\mathbf{k})]_n = \delta^D(\mathbf{k} - \mathbf{q}_1 - \dots, \mathbf{q}_n|R)$ ha sido usada [24]. Las funciones $F_n^{(h)}(\mathbf{q}_1, \dots, \mathbf{q}_n|m, R)$ son los kernels de la teoría no lineal de perturbaciones.

Este formalismo llamado perturbaciones-halo, fue implementado por Mead et.al en el 2015, permitiendo conocer el efecto de la teoría no lineal de perturbaciones sobre la forma de modelar la densidad de halos en relación al uso la teoría lineal, esto se describe en la siguiente sección.

1.8. Modelo Analítico

El uso de las simulaciones cosmológicas ha permitido desarrollar una amplia descripción de la dinámica y evolución de los halos de materia oscura como elementos fundamentales en la evolución del universo, sin embargo dado que la simulaciones asumen que las partículas no colisionan entre si, entonces los procesos físicos diferentes a la gravedad que contribuyen al resultado de la simulación no se logran comprender [14].

Las pequeñas fluctuaciones en el campo de densidad de materia pueden ser descritas mediante la teoría lineal de perturbaciones en el universo temprano, pero al paso del tiempo tales fluctuaciones crecen y se hacen necesario el uso de la teoría no lineal. Un elemento importante en la comprensión de la formación de estructuras en el universo es el espectro de potencias del campo de densidad de materia, el cual puede ser calculado desde las simulaciones antes mencionadas. Alternativamente, los modelos analíticos calibrados desde las simulaciones proveen también aproximaciones para dicho espectro. Un modelo ampliamente usado es el *Halofit* de Smith et al. (2003) y para el 2009 Heitmann et al. proponen el *Emulator Code* en el cual se logra correr puntos específicos del espacio de parámetros y a partir de estos resultados interpolar para puntos no considerados, obteniendo resultados precisos para $k \leq 10 Mpc^{-1}$ y $z < 4$ [14].

Un nuevo enfoque para modelar el espectro de potencias de masa, lo presenta Mead et al. [14] en el cual expone una variante optimizada del modelo del halo permitiendo predecir el espectro para número de onda $k \approx 10 Mpc^{-1}$ a la vez que incorpora el formalismo de perturbaciones-halo al implementarlo computacionalmente en el código *HMcode* desarrollado en FORTRAN.

Particularmente, la función de correlación para el término halo-halo, descrito en la ecuación (1.121), se calcula empleando la aproximación lineal y también la versión no lineal obtenida desde el formalismo de perturbaciones-halo.

Un aspecto importante del modelo propuesto por Mead et al. es que logra cuantificar el impacto de efectos bariónicos no considerados por otros modelos tales como los núcleos

activos de galaxias y la explosión de supernovas. Usando los parámetros cosmológicos de la simulación BolshoiP y que se describen con mayor detalle en la sección 2.2, se obtuvo del código HMcode el espectro de potencias a diferentes redshift de los cuales la figura 1.5 presenta cuatro curvas para cada uno de las cuatro opciones de cálculo que provee el código.[2008-02-26]

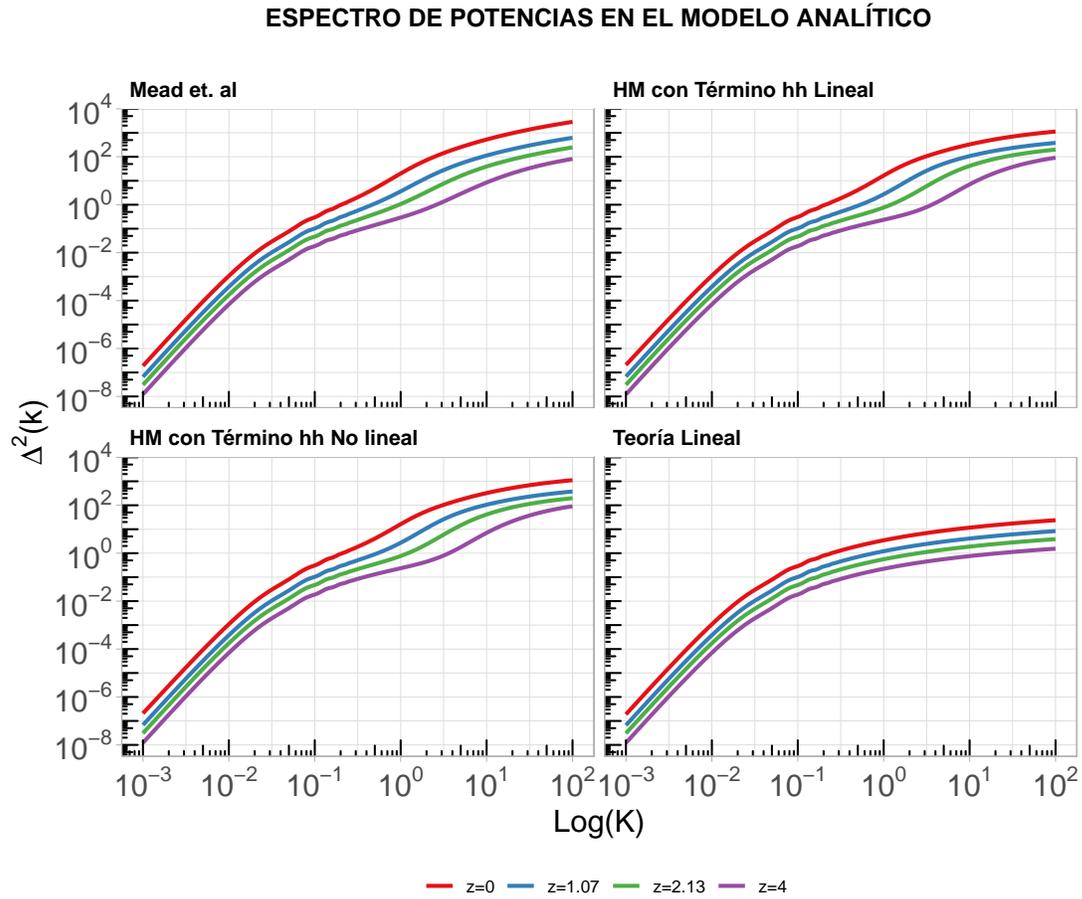


FIGURA 1.5: Espectro de potencias en el modelo analítico.

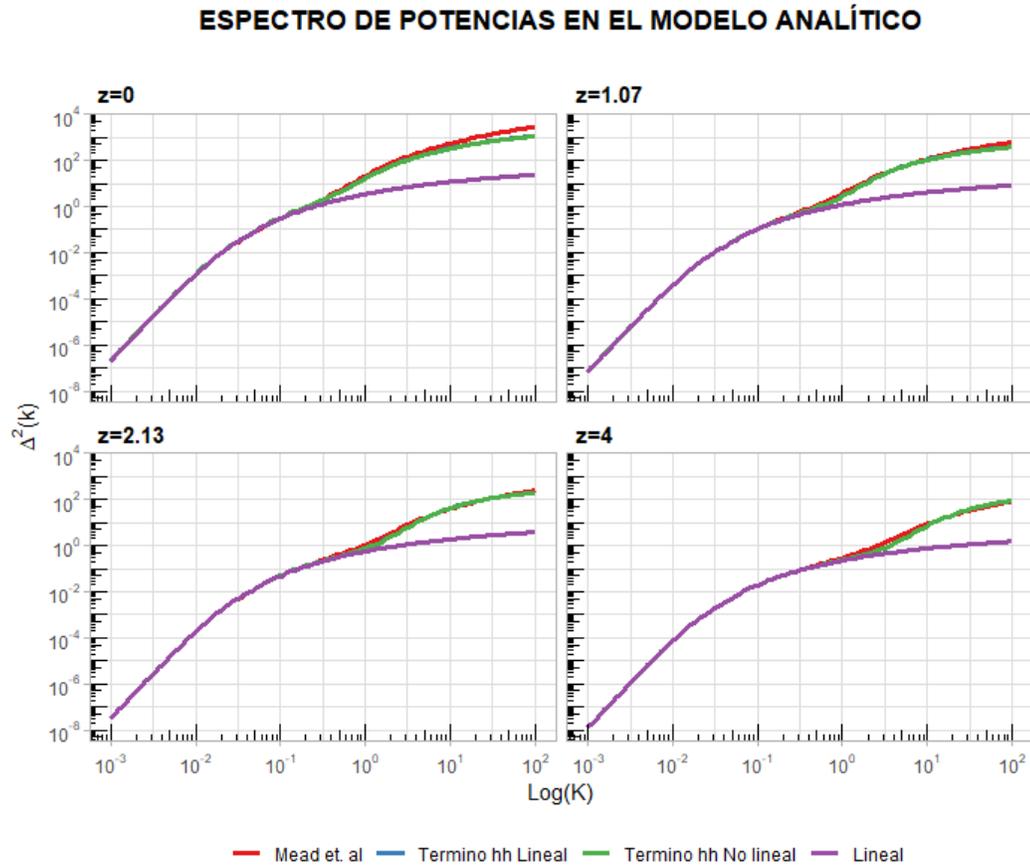


FIGURA 1.6: Espectro de potencias en el modelo analítico

La figura 1.5 presenta para cada cálculo del código, los resultados para cuatro valores de redshift, en tanto que la figura 1.6 permite comparar a cada redshift el espectro de potencias usando los diferentes modelos analíticos para describir la formación de estructura.

Seguindo la ecuación (1.130) se tiene que el espectro de potencias esta compuesto por dos elementos correspondientes al término de un halo y de dos halos, los cuales se presentan en la figura 1.7.

Por otra parte, la figura 1.8 muestra como varia la componente no lineal del término halo-halo respecto al mismo valor usando el espectro lineal. Esto se hace más importante a menor redshift donde se predice un aumento de aproximadamente un 2.3% en la formación de estructura a escalas intermedias.

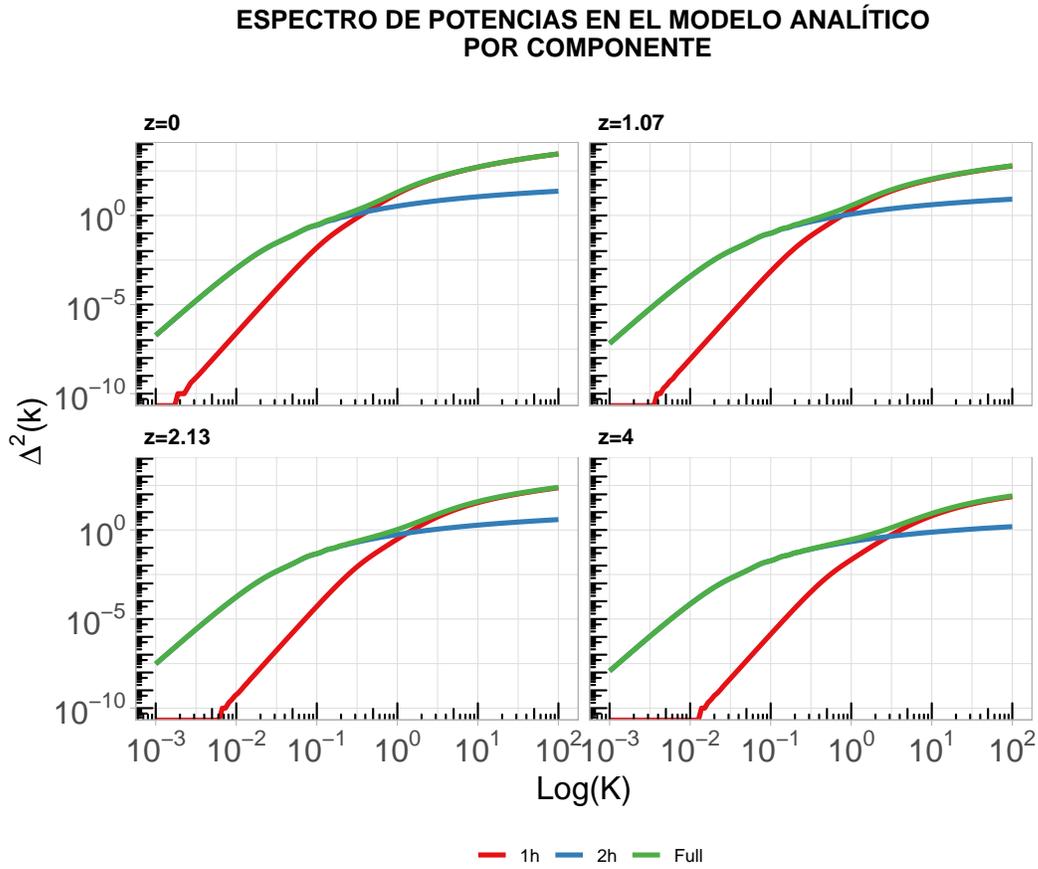


FIGURA 1.7: Espectro de potencias por componente

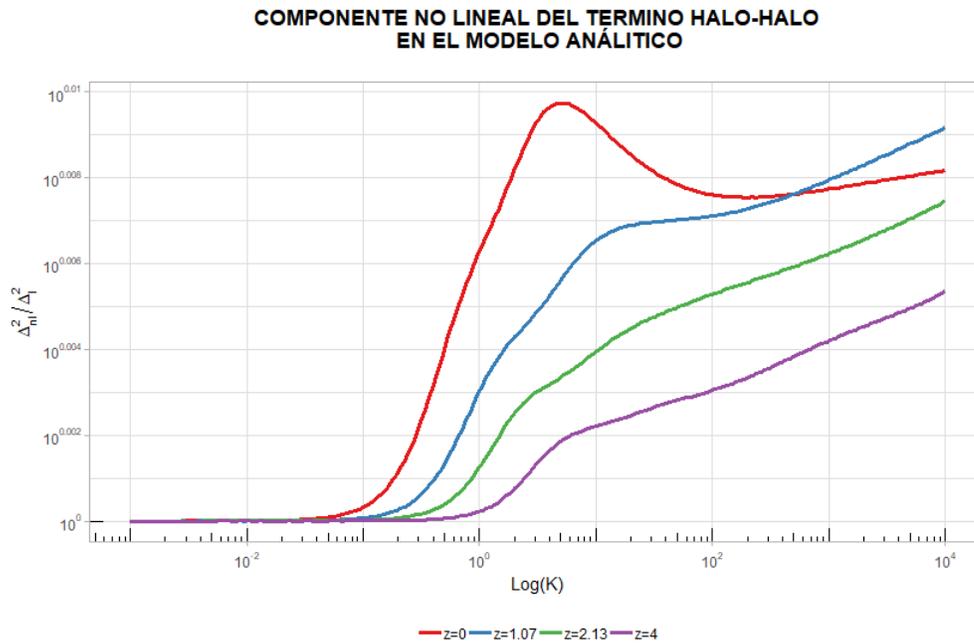


FIGURA 1.8: Componente no lineal del término halo-halo

Por otra parte, contrastar estos modelos requiere el uso de resultados de simulación que por el carácter discreto de sus resultados conduce a modelar los campos de halos como procesos estocásticos espaciales que en su versión discreta se conocen como *Campos Puntuales Espaciales* y cuyo estudio esta en los orígenes del modelo del halo y otros problemas cosmológicos, razón por la cual se presenta a continuación una breve conceptualización al respecto (en el Apéndice A.1 se presentan algunos elementos básicos sobre campos aleatorios).

1.9. Procesos Puntuales

Un proceso ² puntual $\Phi(A)$ $A \subset \mathbb{R}^d$ se define como una variable aleatoria sobre un espacio medible formado por una familia de puntos $\mathbf{x} \in \mathbb{R}^d$, la cual asigna a cada subconjunto A un número positivo correspondiente a la cantidad de puntos que este contiene. Estos puntos deben ser simples y localmente finitos. Lo primero indica que no hay puntos múltiples (esto es que cada punto ocupa un lugar o posición diferente) y la segunda hace referencia a que para cualquier subconjunto A se tiene que $\Phi(A) < \infty$.

Para $d = 3$ se tiene conjuntos en el espacio, por lo cual se denotará el volumen de A mediante $V(A)$ y la probabilidad que el subconjunto contenga k puntos por $P[\Phi(A) = k]$. En general un proceso puntual se caracteriza mediante una *función de intensidad* $\lambda(\mathbf{x})$, para lo cual sea dV un elemento de volumen que contiene el punto \mathbf{x} entonces:

$$\lambda(\mathbf{x}) = \lim_{dV \rightarrow 0} \frac{E[\Phi(A)]}{dV} \quad (1.142)$$

lo cual corresponde al número medio de puntos por unidad de volumen y que para procesos homogéneos e isotrópicos se reduce a $\lambda(\mathbf{x}) = \lambda$. En cosmología este valor se conoce como *densidad número* y con frecuencia se denota por \bar{n} [26].

Existe una amplia variedad de tipos de campos puntuales, dos de los cuales se son los campos Binomiales y los de Neymann-Scott [17].

1.9.1. Campos Binomiales

En este tipo de campo, se tiene que N puntos se distribuye de manera uniforme sobre una región (ventana) $W \subset \mathbb{R}^3$ por lo cual es un proceso homogéneo e isotrópico de donde:

$$\lambda = \frac{N}{V(W)} \quad (1.143)$$

de tal manera que para una región $A \subset W$, el número esperado de puntos es:

$$E[\Phi(A)] = \lambda V(A) \quad (1.144)$$

Así, el nombre del campo viene dado porque la probabilidad de hallar en una región A , k puntos está dada por una distribución binomial, esto es $P[\Phi(A) = k] = \text{Binomial}(N, p_A)$, donde $p_A = \frac{V(A)}{V(W)}$.

²Proceso puntual es equivalente a Campo puntual

La figura 1.9 ilustra dos campos binomiales en el plano simulados con el software libre R [22].

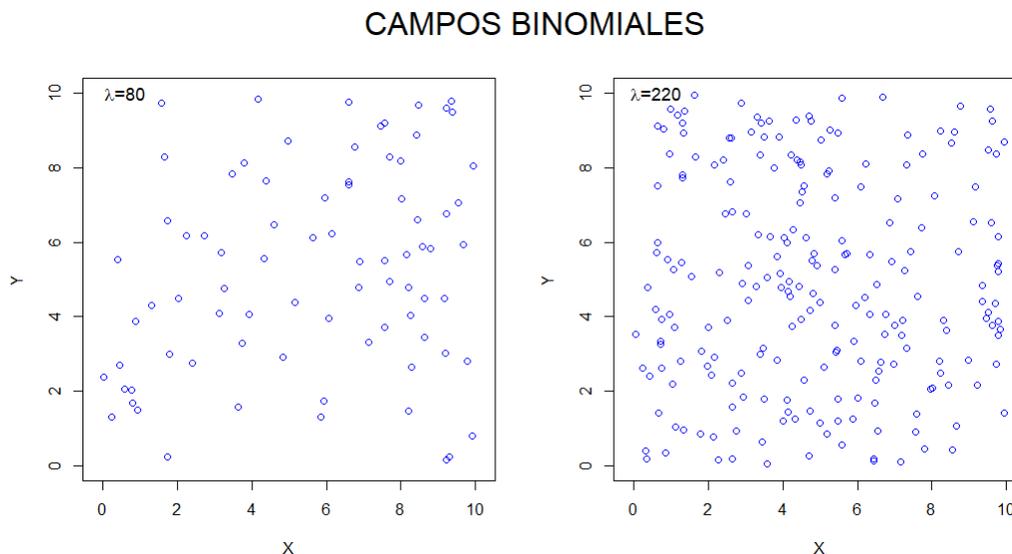


FIGURA 1.9: Ejemplos simulados de campos binomiales.

1.9.2. Campos Poisson

Dentro de los campos puntuales, el caso del campo Poisson es un caso de central importancia, pues este se usa en muchos casos como punto de referencia para el estudio y caracterización de otros tipos de campos. Un campo Poisson $\Phi(A)$ está caracterizado fundamentalmente por dos propiedades citeIP08:

1. $\Phi(A)$ es el número de puntos sobre la región acotada A y sigue una distribución de probabilidad Poisson con parámetro $\lambda(\mathbf{x})V(A)$.
2. Un conjunto de k regiones disyuntas A_i tiene asociadas k variables aleatorias Poisson independientes $\Phi(A_i)$.

El parámetro $\lambda(\mathbf{x})$ se denomina intensidad del campo y puede ser un valor constante $\lambda(\mathbf{x}) = \lambda$, en cuyo caso se tiene un *Campo Poisson Homogéneo* ó puede ser un valor que depende de la posición $\lambda(\mathbf{x})$ definiendo así a un *Campo Poisson No Homogéneo*.

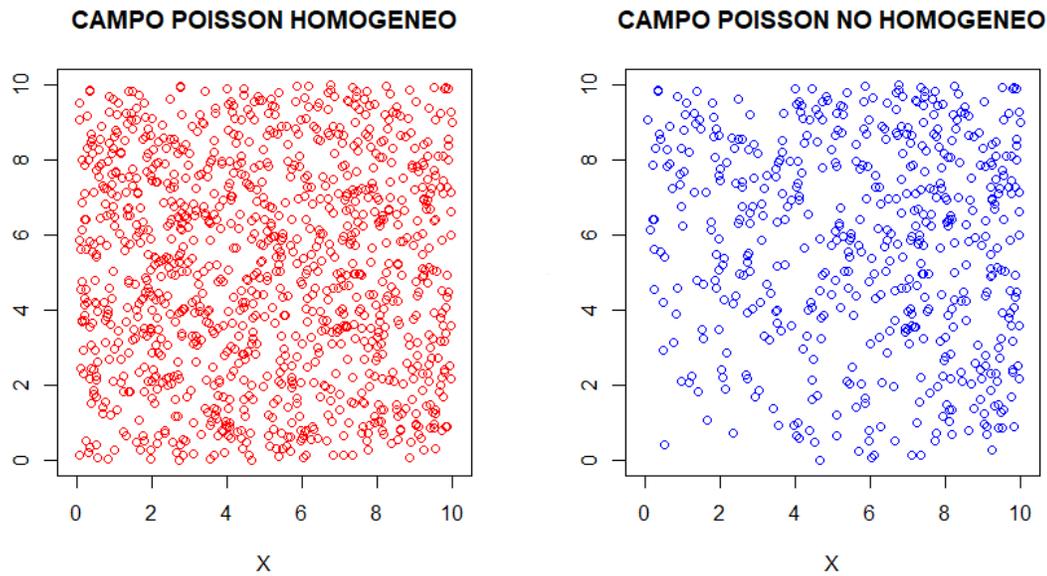


FIGURA 1.10: Ejemplos simulados de campos puntuales Poisson.

1.9.3. Campos de Cox

Los campos de Cox son una generalización de los campos Poisson no homogéneos donde la función de intensidad $\lambda(\mathbf{x})$ es aleatoria. Los campos de Cox son el resultado de dos mecanismos aleatorios, razón por la cual también son llamados campos Poisson doblemente estocásticos.

Dada una función de intensidad $\lambda(\mathbf{x})$ se genera un campo inicial de puntos padre C y condicionado sobre este, se genera un segundo campo X de puntos hijos con la misma función de intensidad $\lambda(\mathbf{x})$. Este tipo de campo admite muchas variaciones, siendo uno de ellos los *Campos de Neyman-Scott* en los cuales ambas etapas del proceso se generan con un proceso Poisson homogéneo. Este tipo de procesos puntuales tienen su origen por 1952 en la teoría de la distribución espacial de las galaxias propuesto por Neyman-Scott [16] la cual daría las bases para el posterior modelo del halo. Para 1958 los mismos autores abordan desde una perspectiva estadística problemas de la cosmología y proponen el campo de Neyman-Scott como elemento para modelar los cúmulos de galaxias [17].

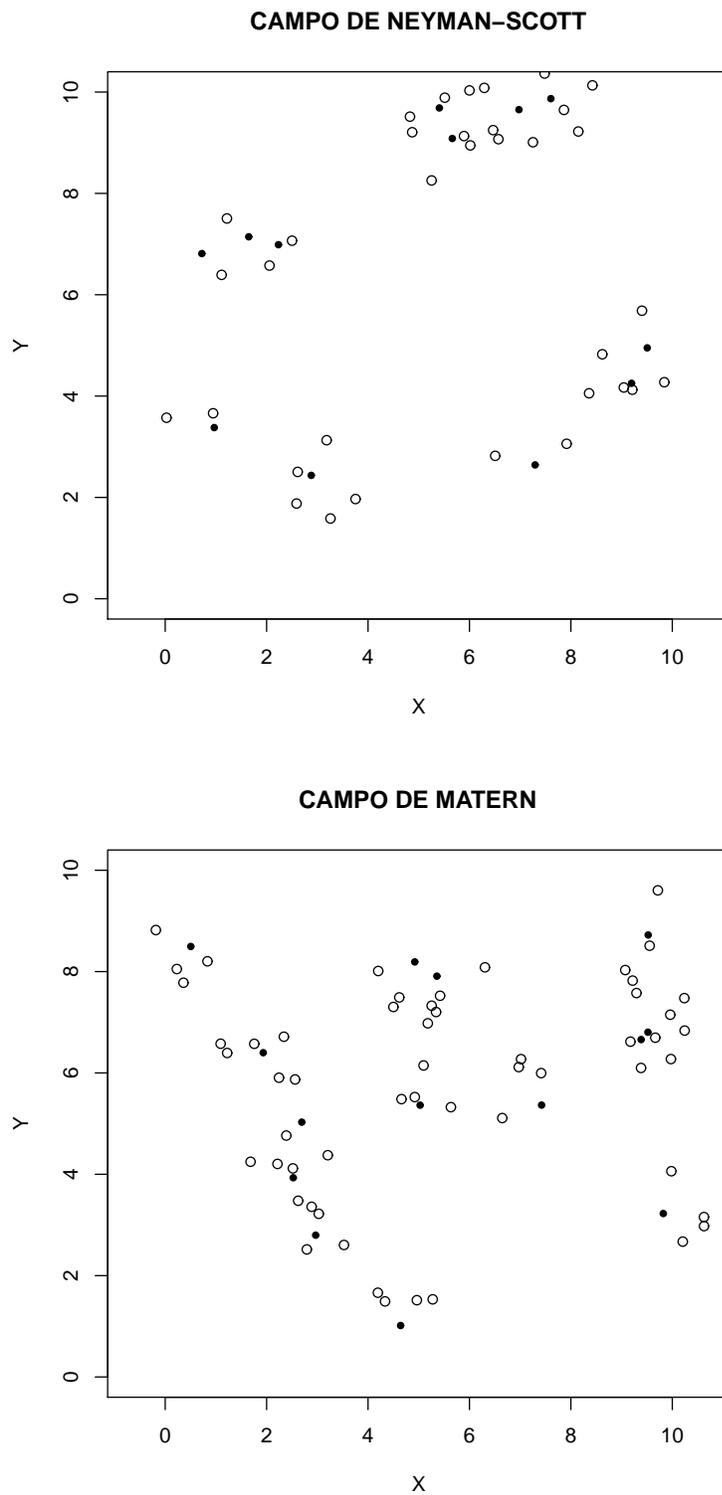


FIGURA 1.11: Ejemplos simulados de campos de Cox.

En la figura 1.11 (arriba) se presenta una simulación de un campo de Neyman-Scott

en la cual los puntos padre están representados con puntos negros (llenos) y los puntos hijos con círculos negros (puntos de fondo blanco). Una característica importante es que el número de puntos alrededor de cada punto padre está dado por una distribución Poisson, en tanto que un campo de Matérn (abajo) se obtiene donde el número de puntos hijo es constante.

Capítulo 2

Simulación Bolshoi

2.1. Simulaciones cosmológicas

De acuerdo con el modelo estándar de la cosmología, en el universo temprano, pequeñas fluctuaciones en los campos de densidad de materia son consideradas como las responsables de la posterior formación de las estructuras que hoy se observan, tales como galaxias, cúmulos y supercúmulos de galaxias. Es por ello que el estudio de la evolución de tales fluctuaciones aporta valiosa información del origen, formación y evolución de las mayores estructuras existentes en el universo. Esto se logra aproximadamente en una simulación cosmológica, al recrear computacionalmente la acción de la gravedad sobre un determinado número de partículas durante periodos de tiempo que emulan la edad del universo.

Desde los años setenta se ha venido desarrollando simulaciones cosmológicas, que inicialmente describían la distribución de partículas de materia oscura debido a la acción de la gravedad. Hace unas dos décadas se ha venido incluyendo materia ordinaria ó visible, nuevos fenómenos físicos como el efecto de las supernovas, los agujeros negros y la formación estelar como se muestra en la figura 2.1¹. Algunas de estas simulaciones son Millennium, Millennium XXL, Bolshoi entre otras.

Las simulaciones cosmológicas posibilitan así mismo, la realización de estimaciones estadísticas, que sobre el universo real pueden ser imposibles o de gran complejidad observacional, constituyéndose por ello en predicciones del modelo teórico en el cual se fundamenta la simulación. La contrastación de tales predicciones junto con los registros observacionales disponibles, (muestreos de galaxias como el SDSS) son un importante elemento de control, pues establecen la capacidad del modelo cosmológico para dar cuenta de las diferentes estructuras detectadas y sus características, tales como masa, luminosidad y distribución espacial entre otras.

2.2. Simulación Bolshoi

Esta simulación cosmológica de N-cuerpos fue desarrollada inicialmente entre el 2008 y 2009 en el supercomputador Pleiades del NASA Ames Research Center, empleando el código ART (Adaptive Refinement Tree) y OpenMP para el proceso de paralelizado. Esta simulación, utilizó los resultados del WMAP5 para especificar la cosmología correspondiente. Recientemente, esta simulación ha sido nuevamente realizada (BolshoiP)

¹Tomada de <http://www.illustris-project.org/about/>

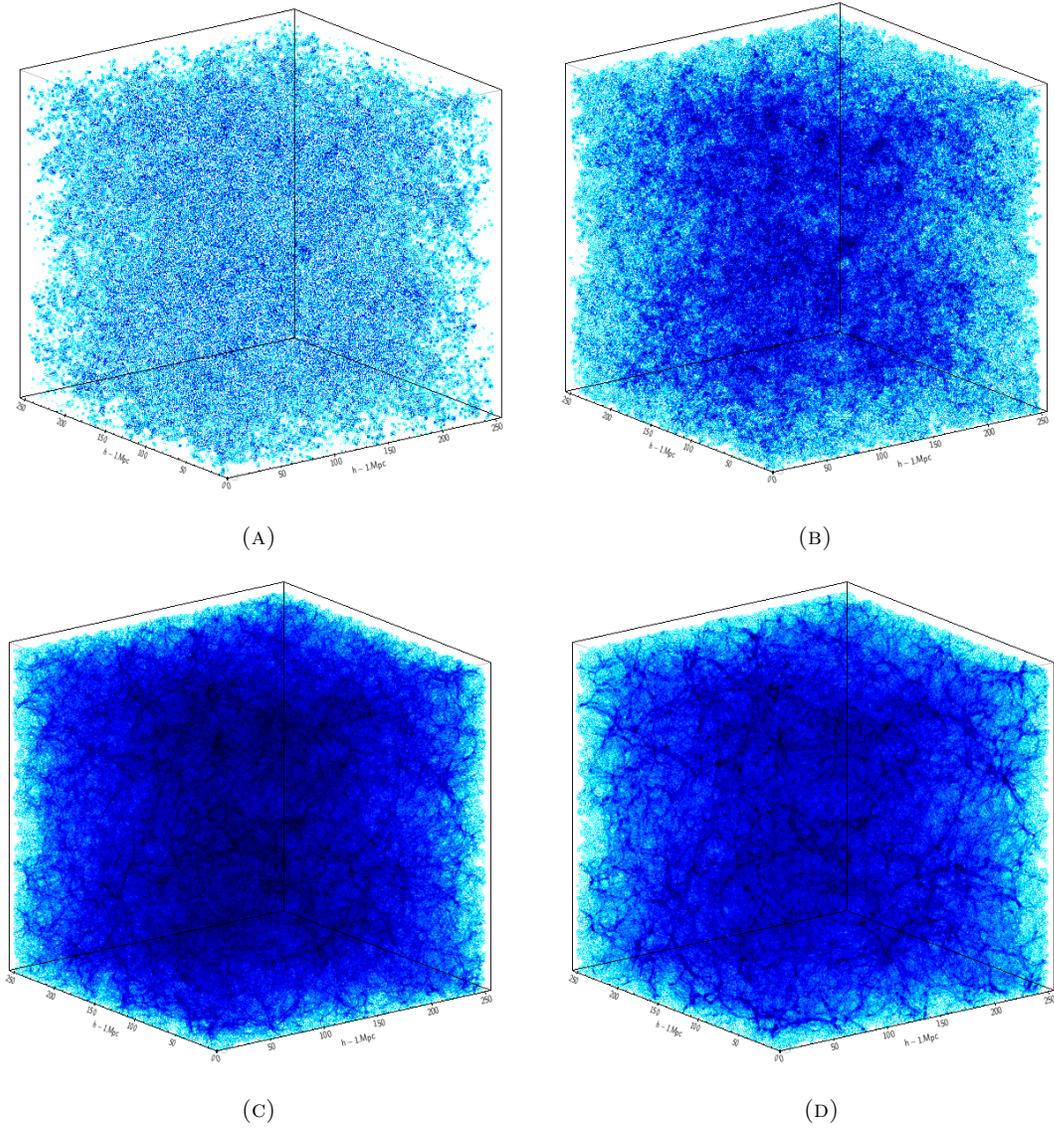


FIGURA 2.2: Cajas (snapshots) de la simulación BolshoiP (A) $z = 10.18$,
(B) $z = 7.59$, (C) $z = 0.99$, (D) $z = 0.01$

La información de la simulación está organizada en una base de datos a la cual se tiene acceso desde la página web www.cosmosim.org. Esta base de datos contiene los resultados de varias simulaciones, una de las cuales corresponde a BolshoiP cuya información está organizada en 10 tablas cuyo nombre y descripción se relacionan a continuación:

- **AvailHalos:** Contiene la lista de los Snapshots² disponibles en las tablas FOF y BDM.
- **BDMV:** Catálogo de halos construido mediante el algoritmo Limite máximo de densidad (BDM) y calculado usando como criterio de sobredensidad $360\rho_{back}$ (densidad de background o de fondo).
- **BDMW:** Catálogo BDM de halos calculado usando como criterio de sobredensidad $200\rho_{back}$
- **FOF:** Catálogo de halos construido mediante el algoritmo Friends of Friends (FOF) con longitud de enlace 0.17 (nivel 0).
- **FOF1-FOF5:** Catálogo de halos FOF de nivel 1 a 5 esto indica longitudes de enlace entre 0.085 a 0.010625.
- **Redshifts:** Lista de los Snapshots disponibles y sus respectivos Redshifts.

El catálogo empleado en el presente trabajo es el construido bajo el algoritmo FOF, cuya información se describe en la tabla 2.1.

²En una simulación un Snapshots corresponde a un registro de posiciones, velocidades y demás variables en un instante de tiempo cosmológico, que da cuenta de la evolución del universo hasta dicho momento.

Grupo	Columna	Unidades	Descripción
Identificación	<i>fofId</i>		Identificador del grupo en la base
	<i>snapnum</i>		
Inf. FOF	<i>level</i>		Información acerca del grupo en el archivo original
	<i>NInFile</i>		
Posición	<i>x</i>	$1/h \text{ Mpc}$	Coordenadas coomoviles
	<i>y</i>		
	<i>z</i>		
Movimiento	<i>vx</i>	km/s	Componentes de velocidad peculiar
	<i>vy</i>		
	<i>vz</i>		
	<i>angMom_x</i>	$1/h^2 M_{\odot}$	Componentes de momento angular
	<i>angMom_y</i>	$Mpc \text{ km/s}$	
	<i>angMom_z</i>		
	<i>np</i>		Número de partículas del grupo
	<i>mass</i>	$1/h M_{\odot}$	Masa del grupo
	<i>size</i>	$1/h \text{ Mpc}$	Radio de la esfera
	<i>delta</i>	ρ_b	Sobredensidad en términos de la densidad de background
	<i>spin</i>		Parámetro de spin
Dispersión	<i>disp</i>	$1/h \text{ Mpc}$	Dispersión espacial
	<i>disp_v</i>	km/s	Dispersión de la velocidad
Forma y orientación	<i>axis1</i>		Semi eje mayor (no normalizado)
	<i>axis2</i>		Semi eje intermedio (no normalizado)
	<i>axis3</i>		Semi eje menor (no normalizado)
	<i>axis1_x</i>		Dirección del eje mayor
	<i>axis1_y</i>		
	<i>axis1_z</i>		
	<i>axis2_x</i>		Dirección del eje intermedio
	<i>axis2_y</i>		
	<i>axis2_z</i>		
	<i>axis3_x</i>		Dirección del eje menor
	<i>axis3_y</i>		
	<i>axis3_z</i>		

TABLA 2.1: Detalle de la información contenida en la tabla FOF de la simulación BolshoiP

2.2.1. Consultas

La información de la simulación BolshoiP es publica y es accesible desde la página <http://www.cosmosim.org/query> mediante una consulta SQL que permita extraer la información deseada. Tras en registro en la plataforma *cosmosim* se accede en el menú de consulta (query) donde se relaciona el código correspondiente para ello, que en términos generales corresponde a especificar:

- a. ¿Que columnas desea consultar?
- b. ¿De que tabla?
- c. ¿Con que criterio de selección?

Aunque se puede especificar más elementos como orden, límite de registros, incluso cálculos básicos como conteos y sumas, estas tres preguntas conforman una consulta básica como la que se realizó en este trabajo.

A continuación se presenta la consulta realizada las obtener las coordenadas cartesianas comoviles de cada halo presente en la tabla FOF para un redshift cero (snapnum 212).

```
SELECT p.x, p.y, p.z
FROM MDR1.FOF AS p
WHERE snapnum = 212
```

Produciendo como resultado un archivo de extensión `.csv` con la siguiente estructura:

```
"row_id", "p.x", "p.y", "p.z"
1,12.678099632263184,2.5998001098632812,3.2825000286102295
2,14.963399887084961,4.0843000411987305,4.260300159454346
3,2.1138999462127686,9.705400466918945,2.308500051498413
4,1.5198999643325806,10.84630012512207,2.388400077819824
```

2.3. Estimación de la función de correlación de dos puntos

2.3.1. Código correlator.c

La función de correlación de dos puntos se estimó usando el código `correlator.c` el cual fue desarrollado por el finlandés Pauli Pihajoki en el 2006 para el observatorio de Tuorla de la universidad de Turku Finlandia y se encuentra disponible desde la página [18] http://www.astro.utu.fi/research/CosmoS/tools/tools_p1.shtml.

Este código permite calcular cuatro estimadores de la función de correlación de dos puntos de un campo como el obtenido en las simulaciones cosmológicas. Para ello, el programa realiza el conteo por pares de puntos que estén en un rango de distancia que el usuario establece. Esto se realiza en dos campos de puntos, uno correspondiente a la caja de simulación que se estudia y otro correspondiente a un campo aleatorio de comparación.

Descripción

Este código desarrollado en lenguaje C calcula la función de correlación de dos puntos mediante los siguientes cuatro estimadores:

1. Estimador de Peebles - Hauser
2. Estimador de Davis - Peebles
3. Estimador de Hamilton
4. Estimador de Landy - Szalay

Así mismo, se estima el error Poisson asociado para la correlación en cada rango de distancia, siguiendo a [26].

Argumentos

Originalmente cuatro argumentos debían ser especificados para ejecutar este código. Sin embargo, con las mejoras introducidas estos aumentaron a seis, como son:

1. **Número de procesadores:** Debiendo ser un múltiplo cuatro, este valor especifica cuantos núcleos se emplearán en el cálculo.
2. **Archivo de salida:** Se especifica un nombre que tendrá el archivo de resultados.

Ejemplo: salida.dat

3. **Archivo de datos:** Este archivo está compuesto por dos partes importantes. La primera corresponde a los *Bins* ó simplemente los rangos de distancia que se desean usar para realizar los conteos por pares, esta opción aunque poco práctica tiene la utilidad que permite usar amplitudes variables para los rangos de distancia si se requiere. Esta parte inicia en la primera fila con el número de rangos que se van a emplear y posteriormente los rangos, especificados mediante dos columnas que corresponden a r y $r + dr$ respectivamente.

La segunda parte, hace referencia las coordenadas rectangulares de los puntos provenientes de la simulación. En las tres primeras líneas de este archivo se especifica en su orden: número de puntos en el archivo, coordenadas del punto inferior de la caja y las coordenadas del punto superior de la caja. Los datos están separados por espacios y se usa punto para la posición decimal.

Ejemplo:

Fila 1: 10

Fila 2: 0.01 0.02

Fila 3: 0.02 0.03

.

.

.

Fila 11: 0.09 0.10

Fila 12: 92

Fila 13: 0 0 0

Fila 14: 500 500 500

Fila 15: 366.620 325.945 280.183

Fila 16: 372.183 361.054 282.921

Fila 17: 383.345 242.158 241.393

.

```

.
.
Fila 106: 104.881 435.226 328.974

```

4. **Campo de comparación:** Este archivo contiene las coordenadas rectangulares de los puntos del campo de comparación, el cual corresponde a un campo aleatorio binomial que puede contar con la misma cantidad de puntos que en el archivo de datos, aunque [26] sugiere usar un campo con por lo menos tres veces el número de puntos que el archivo de datos y se especifican de la misma manera que en el caso anterior, incluyendo en las tres primeras filas número de puntos en el archivo, coordenadas del punto inferior de la caja y las coordenadas del punto superior de la caja, como de muestra a continuación.

Ejemplo:

```

Fila 1: 92
Fila 2: 0 0 0
Fila 3: 500 500 500
Fila 4: 166.620 125.945 480.183
Fila 5: 272.183 61.054 262.921
Fila 6: 283.345 242.158 231.393
.
.
.
Fila 95: 204.881 335.216 28.974

```

5. **Nombre del archivo de control:** Es el archivo control que permite monitorear el desarrollo y avance del cálculo. Se especifica con extensión `.dat`.
6. **Tipo de estimador:** El número del estimador que se desea calcular de acuerdo con la lista que se relaciona en la descripción.

Ejecución

Desde Linux, la ejecución del programa comprende dos pasos:

Compilación El archivo `correlator.c` se compila y genera un archivo ejecutable `correlator.o`. La opción `-lm` se emplea para llamar las librerías matemáticas.

Versión original:

```
gcc correlator.c -o correlator.o -lm
```

Versión optimizada:

```
mpicc correlator.c -o correlator.o -lm
```

Ejecución En la ejecución se indica como se mencionó antes, el archivo de salida, el archivo de datos, el campo de comparación y el tipo de estimador requerido. Se debe tener en cuenta que los archivos deben estar en la misma ruta del código.

Versión original:

```
./correlator.o Salida.dat Datos.dat CMP.dat 4
```

Versión optimizada:

```
mpiexec -n 24 ./correlator.o Salida.dat Datos.dat CMP.dat STATUS.dat 4
```

2.3.2. Resultados

Originalmente, el archivo de resultados esta compuesto por cuatro columnas, siendo las dos primeras los rangos de distancias especificados en el archivo de datos es decir los *Bins*, la tercera columna es la estimación puntual de la correlación y la última es el error asociado a esa estimación. Tras los cambios efectuados y descritos previamente a estructura del archivo de salida es:

Ejemplo:

```
0 0.05 3532 3226 3226 0.0948543 0.0192763
0.05 0.1 567152 23436 23226 23.4008 0.160109
0.1 0.15 1940522 62428 62886 29.8723 0.12311
0.15 0.2 3074056 121928 121726 24.2506 0.0723735
0.2 0.25 4241076 201687 202466 19.9548 0.0465701
0.25 0.3 5461532 302201 299428 17.2214 0.0332993
```

```
.
.
.
```

donde las dos primera columnas siguen siendo los bins, pero las columnas tres a cinco corresponden a los conteos DD, DR y RR respectivamente, en tanto que la sexta y quinta contiene en su orden la estimación de la función de correlación de dos puntos y su error de estimación.

Capítulo 3

Resultados

3.1. Optimización del código

El código descrito en la sección 2.3.1 fue diseñado para funcionar de manera secuencial, desarrollando cada uno de los conteo (DD, DR y RR) con un solo procesador, lo cual para cantidades pequeñas de puntos ($n < 10000$) funciona bien, sin embargo, para el presente trabajo resulto poco eficiente al considerar que el número de puntos superaba los 10 millones, de tal manera que fue necesario implementar una serie de ajustes y modificaciones importantes que elevaron sustancialmente la eficiencia computacional del código, más aun teniendo en cuenta que la capacidad de procesamiento disponible.

. Para ello se implementaron varios ajustes siendo el más importante la paralelización del cálculo, de forma que se pudo emplear toda la capacidad de procesamiento disponible. Los principales cambios realizados fueron:

1. **Paralelización:** Empleando la librería MPI versión 1.4 se logro ejecutar los tres conteos de manera simultanea utilizando para ello varios procesadores en cada uno.
2. **Partición optima:** En virtud de la carga de trabajo descompensada que presentaban los procesadores con una asignación uniforme de puntos a procesar, se desarrollo una regla que logró balancear dicha carga computacional, disminuyendo la diferencia en los tiempos de procesamiento entre los núcleos (Ver Apéndice [A.2](#))
3. **Bins constantes:** Se modifiko los rangos de distancia originales (bins), los cuales podían ser de diferentes tamaños. En su lugar se empleo bins de igual longitud facilitando la asignación de las distancias calculadas entre cualquier par de puntos a un bin en específico.
4. **Reducción de ciclos:** Con el cambio anterior se pudo eliminar un ciclo muy importante con el cual se identificaba el bin correspondiente a una distancia, pero cuyo costo era elevado en tiempo de procesamiento.
5. **Estructura de resultados:** Originalmente, el código obtenía la estimación de la función de correlación dedos puntos y su respectivo error, pero no reportaba como resultado los conteo base con los cuales está era calculada. El archivo de resultados se ajusto para que que incluyera los conteos originales facilitando con ello el calculo de otros estimadores no solicitados en la corrida inicial.
6. **Archivo de STATUS:** En casos con gran número de puntos, donde el tiempo de procesamiento excedía las 24 horas no se podía conocer el estado del calculo, de tal

manera que se incorporó en el código la generación de un archivo de STATUS del proceso, permitiendo monitorear el desarrollo del mismo.

De esta manera, estos cambios permitieron reducir el tiempo de ejecución en dos ordenes de magnitud como se muestra en la figura 3.1.

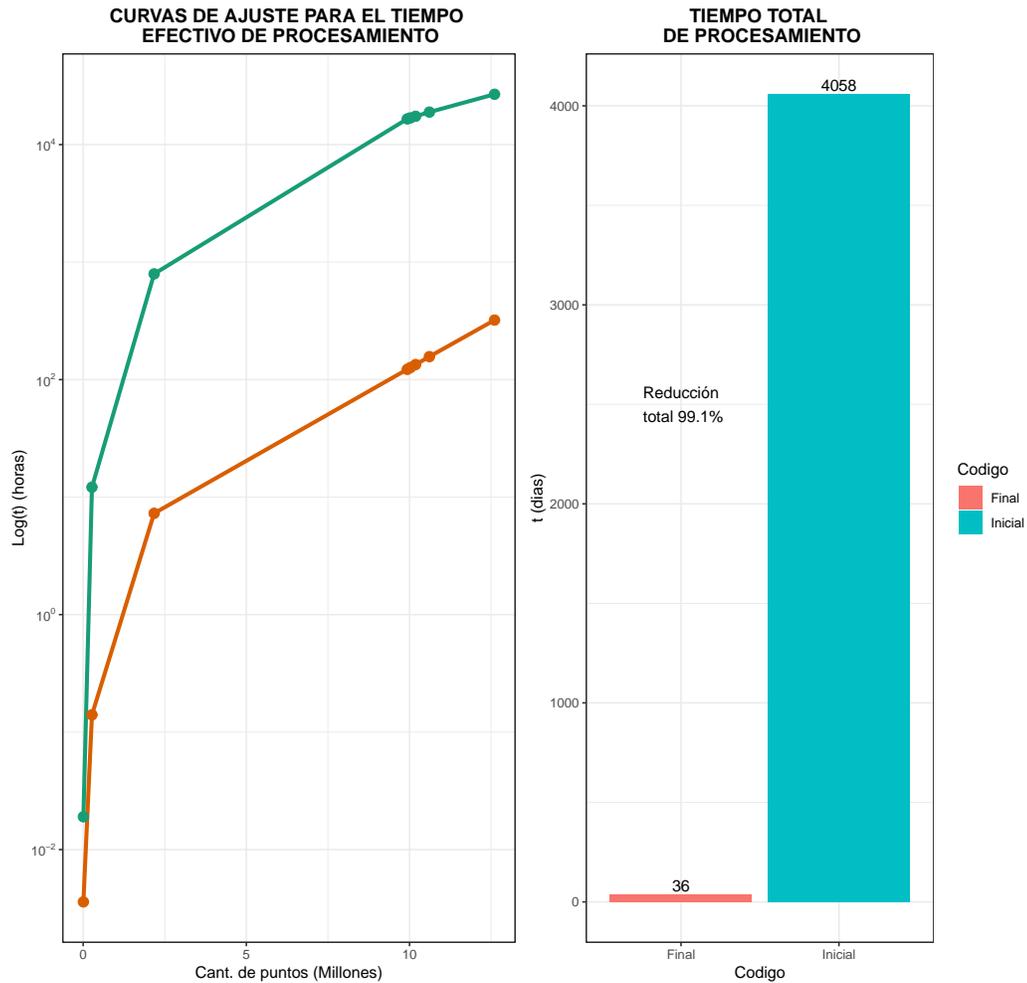


FIGURA 3.1: Tiempos efectivos de procesamiento

3.2. Selección de Snapshots

La información disponible en la base de datos se agrupa en snapshots (cajas) que son registros de todas las variables a un redshift específico. La figura 3.2 muestra la distribución de la cantidad de halos en cada caja.

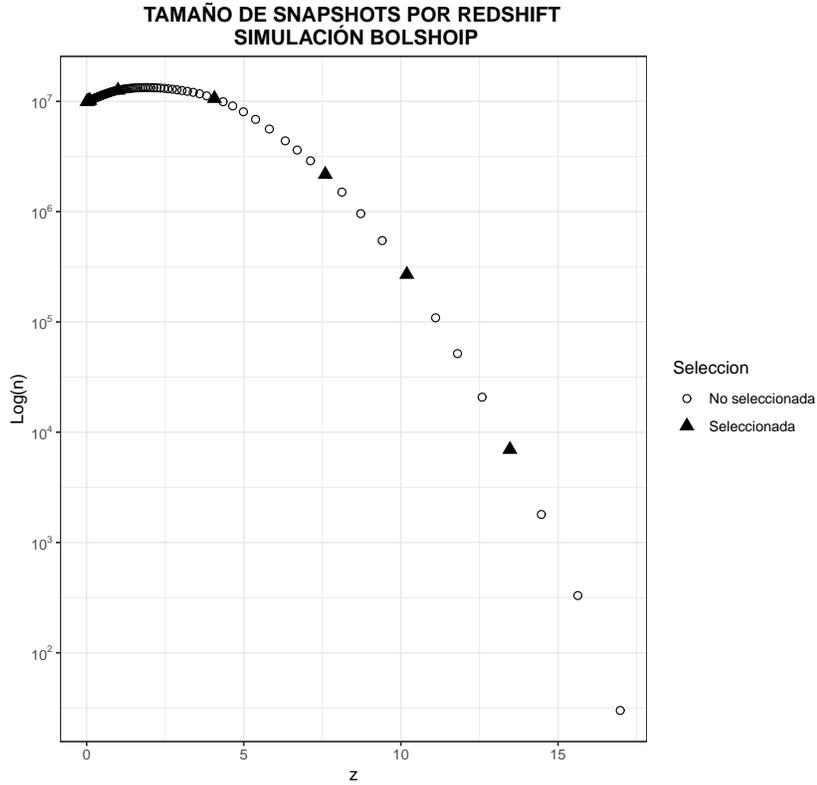


FIGURA 3.2: Selección de Snapshots

A partir de la tabla FOF de la simulación BolshoiP se seleccionaron ocho snapshots a diferentes redshift como se presenta en tabla 3.1.

snapnum	Redshift	Num. de halos
212	0.010	9931775
208	0.050	10029117
202	0.100	10186939
136	0.996	12604837
76	4.067	10609037
58	7.593	2174183
50	10.189	269547
42	13.468	6974

TABLA 3.1: Características de los snapshots empleados

3.3. Distribución de masa de halos

Para cuatro de las ocho cajas seleccionadas, se presenta en la figura 3.3 la densidad número de halos por rango de masa, en la cual se observa que para ordenes de masa menores a 10^{11} el número de halos es muy similar para $z \leq 4$. Para halos con masa hasta 10^{13} , la densidad número evoluciona de forma aproximadamente igual para $z < 1$, mostrando una marcada diferencia con la densidad para $z = 4.067$. También resulta

importante que la curvatura de los perfiles observados decrece conforme evoluciona las estructuras en la simulación.

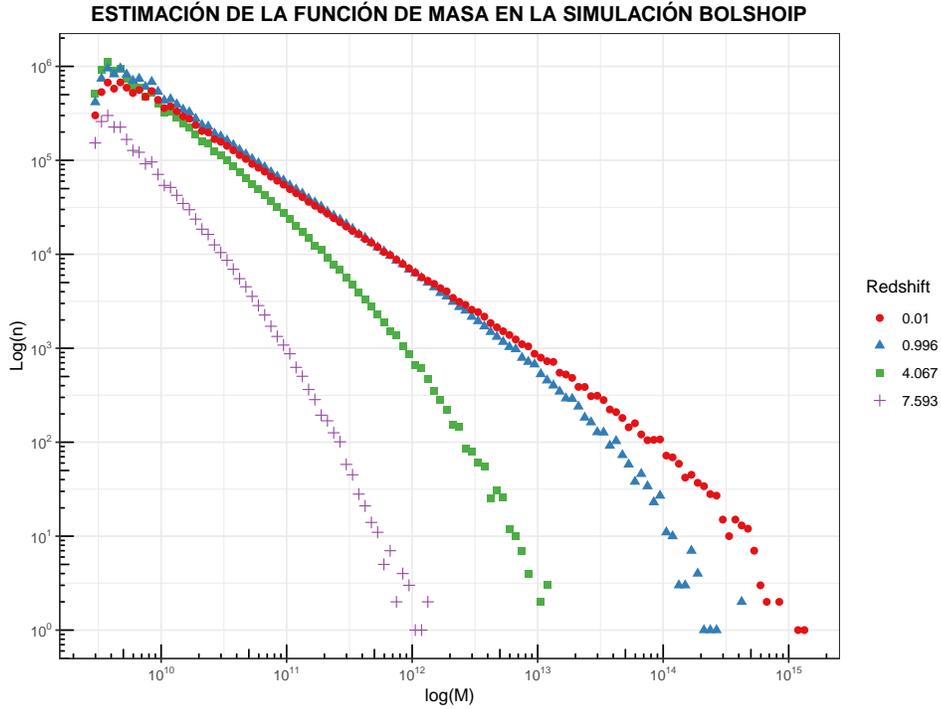


FIGURA 3.3: Funciones de masa estimadas

Las curvas de función de masa presentadas en la figura 3.3 muestran el crecimiento en número de halos con diferentes masas a medida que evoluciona el universo, particularmente para $z = 0.996$ la abundancia de halos es ligeramente mayor en relación a $z = 0.01$ para halos con masas alrededor de $10^{11} M_\odot$, similares en el rango medio de la escala de masa $10^{12} - 10^{13} M_\odot$ y para valores superiores el número de halos a redshift 0.01 aumenta.

3.4. Estimación de la función de correlación de dos puntos

Las figuras 3.4 y A.1 muestran los estimadores obtenidos para la función de correlación de dos puntos en cuatro de los snapshot seleccionados. En ellas, se destaca que a altos redshift la correlación a corta distancia es alta, mostrando un decrecimiento paulatino con la evolución del universo y logrando valores similares a $z = 0.01$ desde un megaparsec en adelante.

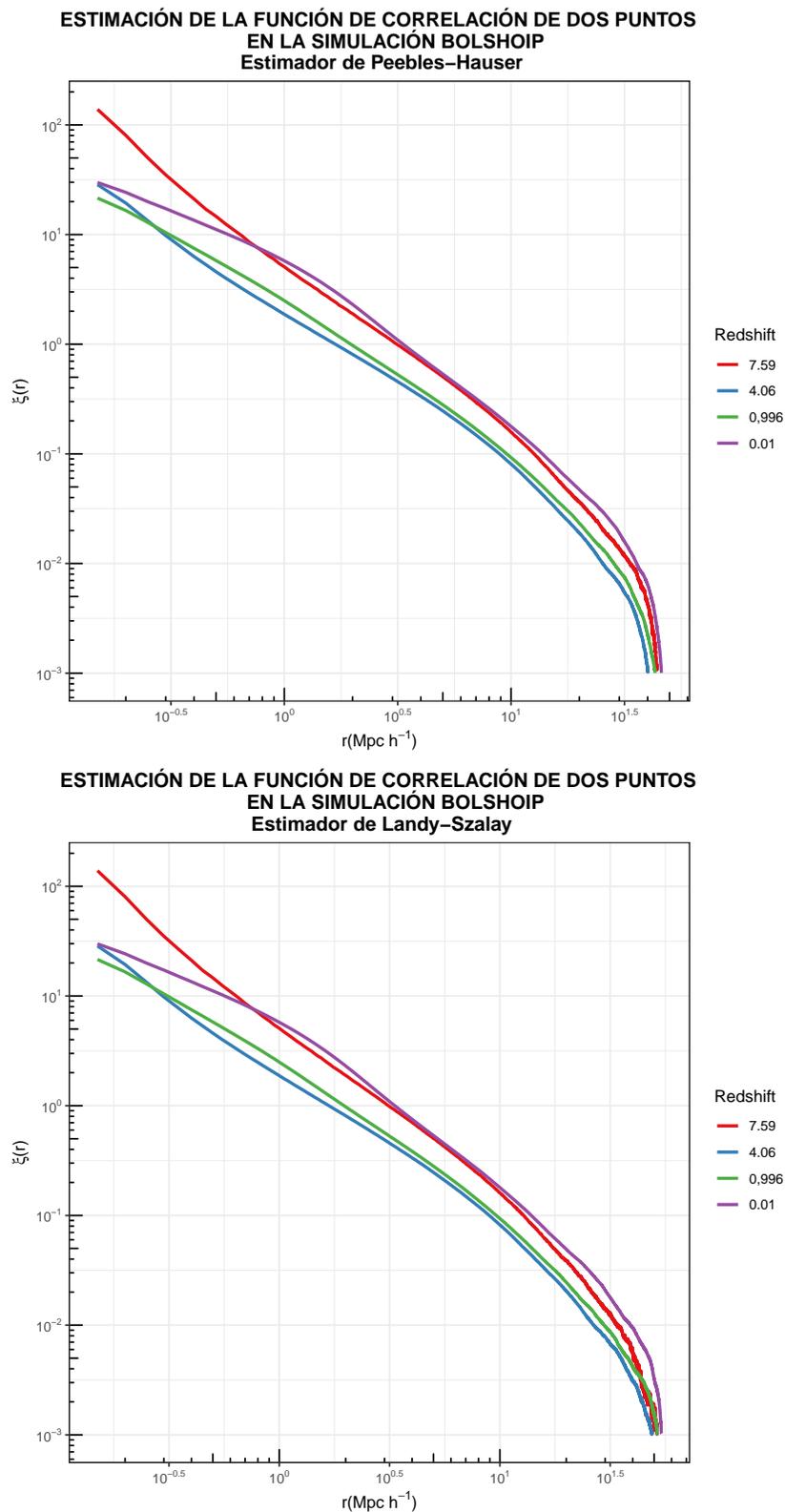


FIGURA 3.4: Estimador natural y de Landy-Szalay de la función de correlación de dos puntos a diferentes redshift

El error descrito en 1.136 es estimado para cada redshift y estimador, siendo presentados en la figura 3.5 solo el correspondiente a Peebles-Hauser y Landy-Szalay pues las curvas para los otros dos estimadores se superponen en gran medida. Se destaca que en los cuatro redshift mostrados que el error de estimación en Landy-Szalay es mayor a bajos valores de r , pero coinciden desde 10 Mpc con el valor correspondiente para Peebles-Hauser.

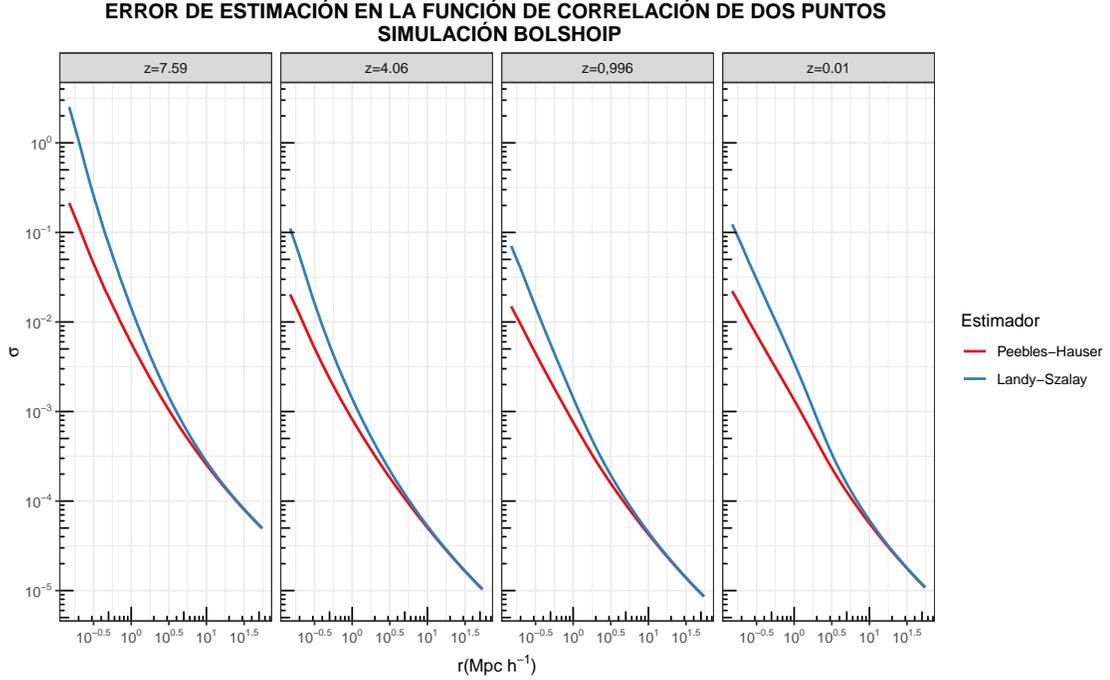


FIGURA 3.5: Error de estimación a diferentes redshift

Por otra parte, para escalas entre $1 - 46 Mpc$ se observa una tendencia aproximadamente lineal, la cual es modelada en escala logarítmica de acuerdo con 3.4 y los resultados para el estimador natural y de Landy-Szalay se resumen en la tabla 3.2. En general los valores obtenidos con el estimador de Landy-Szalay son comparables con valores reportados para cúmulos de galaxias $\gamma \approx 1.8$ [9], en tanto que el estimador natural muestra valores alrededor de 2 para dicho parámetro. Por su parte r_0 exhibe valores comparables.

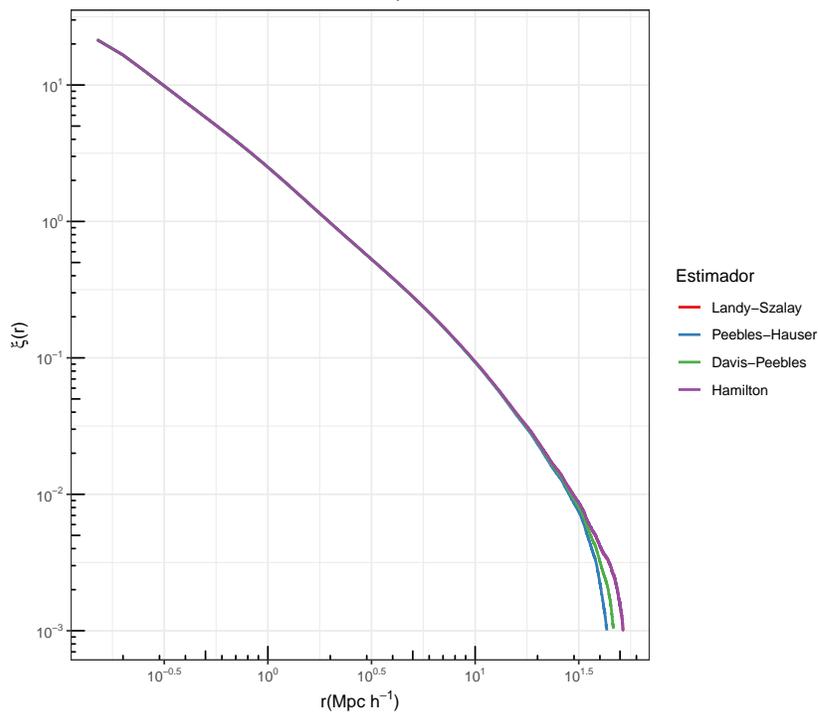
$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma} \quad (3.1)$$

z	Landy-Szalay		Peebles-Hauser	
	γ	$\mathbf{r_0}$	γ	$\mathbf{r_0}$
7.59	1.965 ± 0.012	3.330 ± 0.037	2.110 ± 0.023	3.461 ± 0.0707
4.06	1.853 ± 0.012	2.140 ± 0.022	2.030 ± 0.025	2.299 ± 0.046
0.996	1.799 ± 0.010	2.272 ± 0.020	2.026 ± 0.024	2.516 ± 0.050
0.01	1.777 ± 0.009	3.327 ± 0.031	1.930 ± 0.017	3.467 ± 0.055

TABLA 3.2: Estimación de parámetros para la sección de tendencia lineal

Tal como lo menciona [11] los estimadores para un mismo redshift presentan valores muy similares, lo cual en estos resultados se verifica hasta $32 Mpc$ a partir de lo cual los estimadores de Peebles-Hauser y Davis-Peebles decrecen más rápidamente que los de Hamilton y Szalay los cuales coinciden en casi exactamente como se observa en la figura 3.6.

ESTIMACIÓN DE LA FUNCIÓN DE CORRELACIÓN DE DOS PUNTOS
EN LA SIMULACIÓN BOLSHOIP
Redshift 0,996



ESTIMACIÓN DE LA FUNCIÓN DE CORRELACIÓN DE DOS PUNTOS
EN LA SIMULACIÓN BOLSHOIP
Redshift 0.01

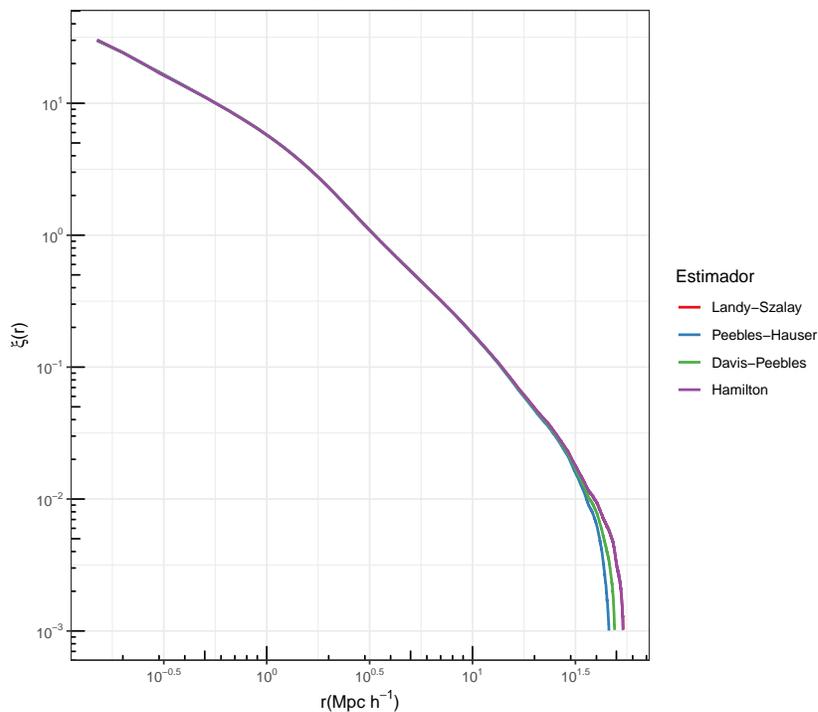


FIGURA 3.6: Estimadores de la función de correlación de dos puntos

3.5. Espectro de potencias en la simulación BolshoiP y comparación con el modelo analítico

A partir de los resultados obtenidos de las correlaciones, se calculó el espectro de potencias [19, Pag. 498] para cada estimador y cada redshift, como se muestra en la figura 3.7. En general, a las escalas consideradas los cuatro estimadores exhiben resultados muy similares, sin embargo a $z = 0.01$ y escalas pequeñas muestran unas leves variaciones.

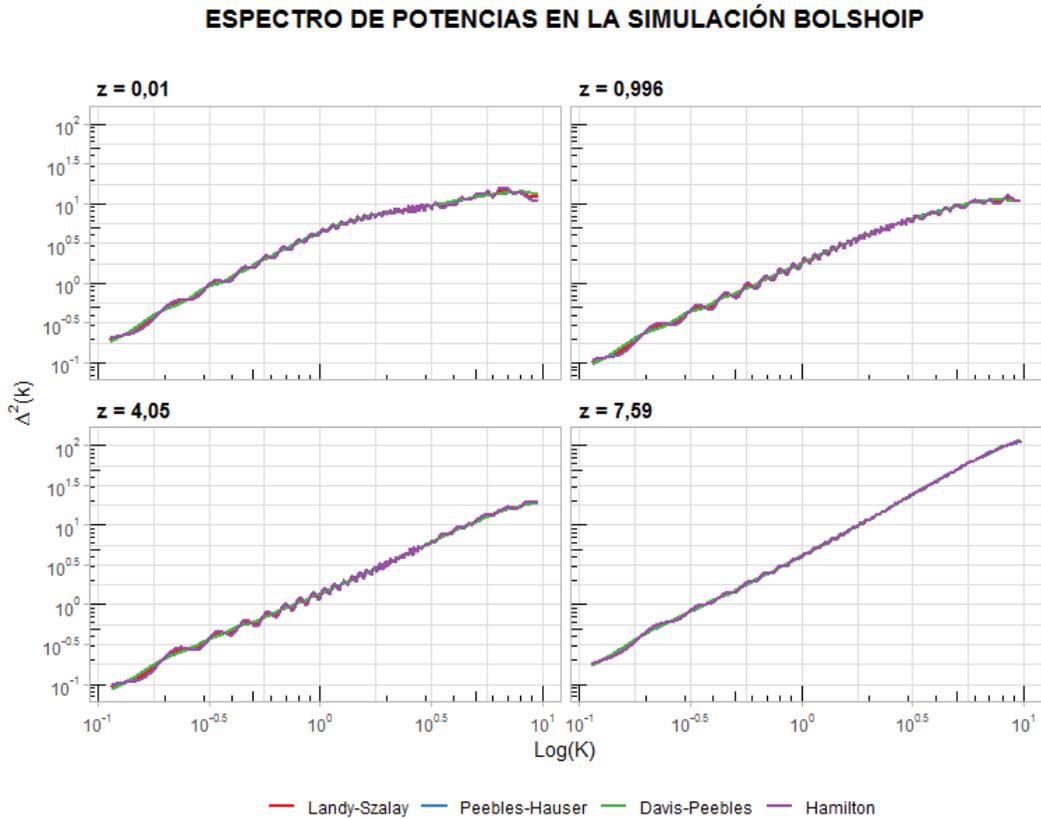


FIGURA 3.7

Dada la proximidad en los resultados entre los estimadores, se presenta en la figura 3.8 el espectro de potencias calculado a partir del estimador de Landy-Szalay en cada uno de los redshift considerados. Destaca que para épocas tempranas ($z = 7.59$) y números de onda mayores (pequeñas escalas) el valor del espectro parece estar sobreestimado, sugiriendo mayor evidencia de clustering en relación a redshift más recientes.

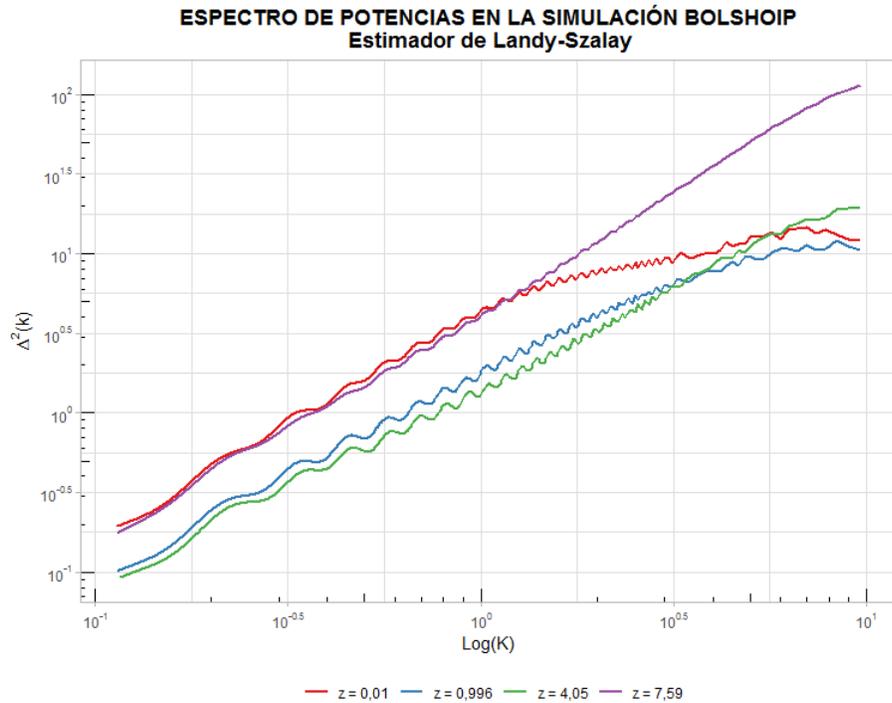


FIGURA 3.8

A un $z = 4$ los valores del espectro de potencia calculados desde la simulación difieren del modelo teórico, sobrestimado la formación de estructura a todas las escalas como se ilustra en la figura 3.9.

Por otra parte, dado que la función de correlación estimada corresponde al término halo-halo, es de esperar que el respectivo espectro de potencias difiera del espectro completo principalmente a escalas pequeñas debido a la ausencia del término de un halo. Este comportamiento se observa en la figura 3.11 al comparar el resultado de la simulación BolshoiP con el correspondiente espectro obtenido con el código HMcode.

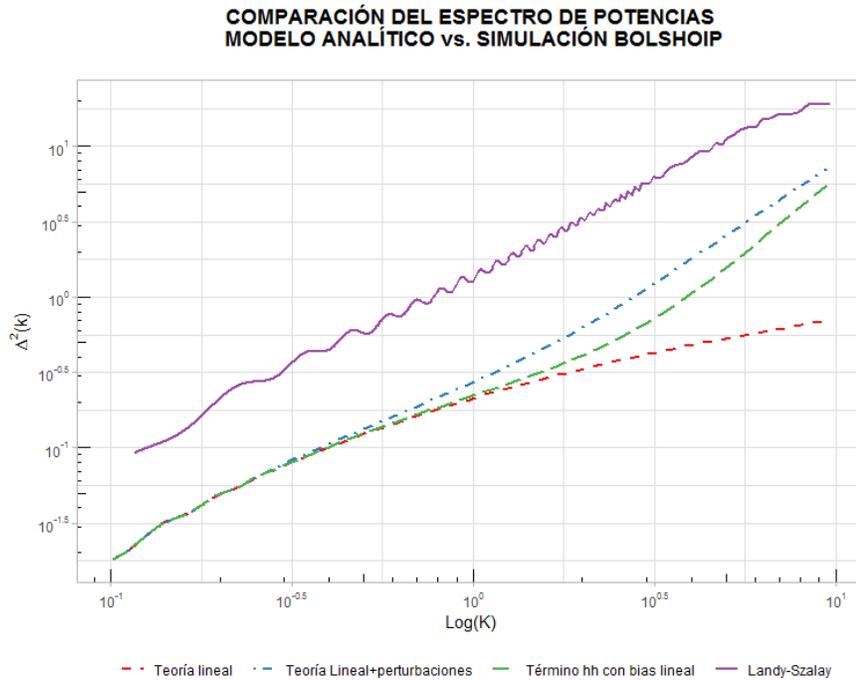


FIGURA 3.9: La línea continua es la estimación del espectro de potencias a $z = 4.06$ en tanto que las líneas no continuas provienen de la teoría combinada perturbaciones-halo.

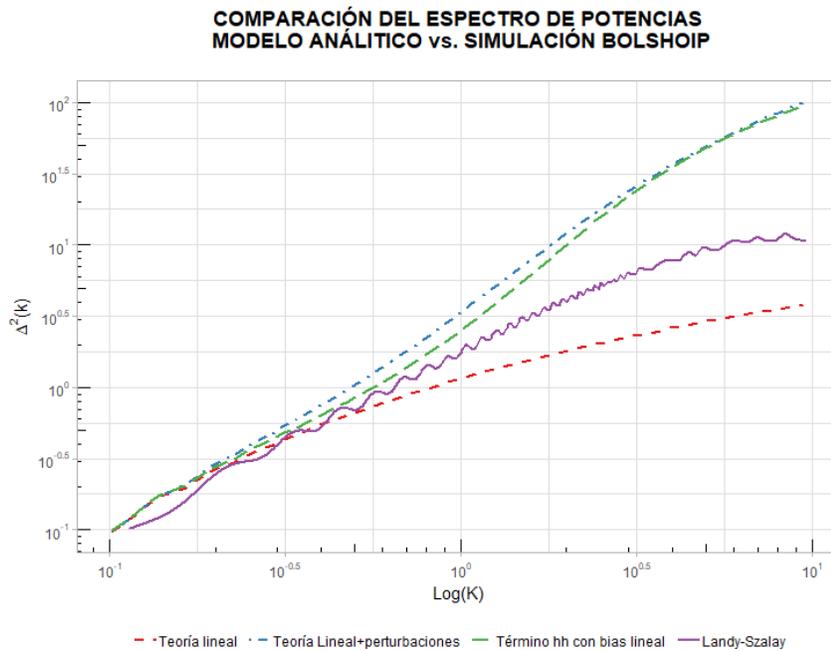


FIGURA 3.10: La línea continua es la estimación del espectro de potencias a $z = 1.07$ en tanto que las líneas no continuas provienen de la teoría combinada perturbaciones-halo.

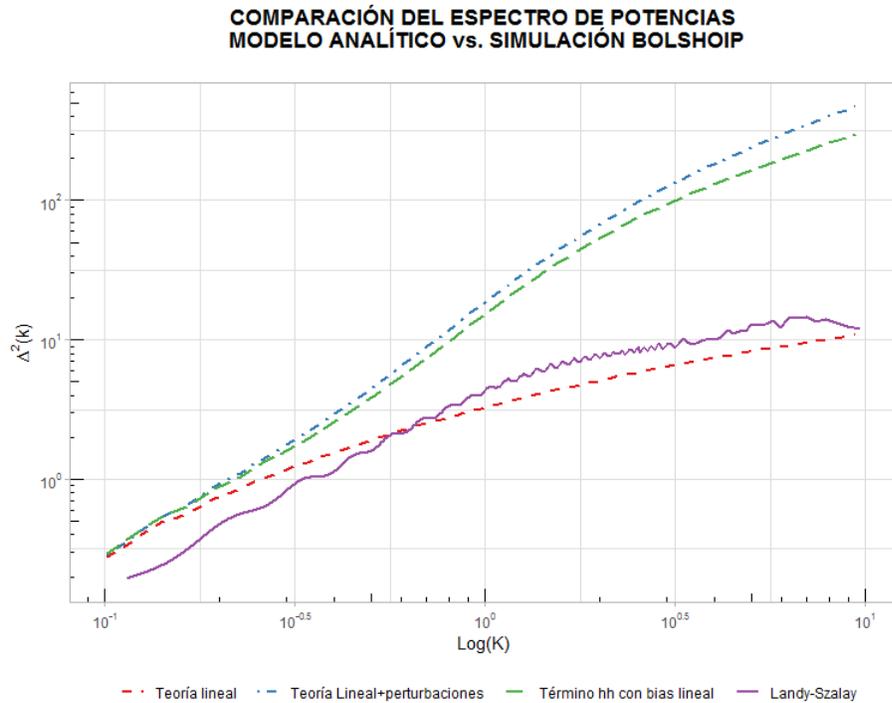


FIGURA 3.11: La línea continua es la estimación del espectro de potencias a $z = 0.01$ en tanto que las líneas no continuas provienen de la teoría combinada perturbaciones-halo.

Tomando el termino halo-halo del modelo analítico y la estimación obtenida desde la simulación, se presenta en la figura 3.12 las curvas correspondientes, siendo importante la proximidad de la estimación al valor analítico descrito por la curva azul, siendo mejor el ajuste para $Z = 0$. Esto significa que a dicho redshift la abundancia de pares de halos detectados a las diferentes distancias estudiadas es muy similar al valor predicho por el formalismo perturbaciones-halo implementado en el código `HMcode`.

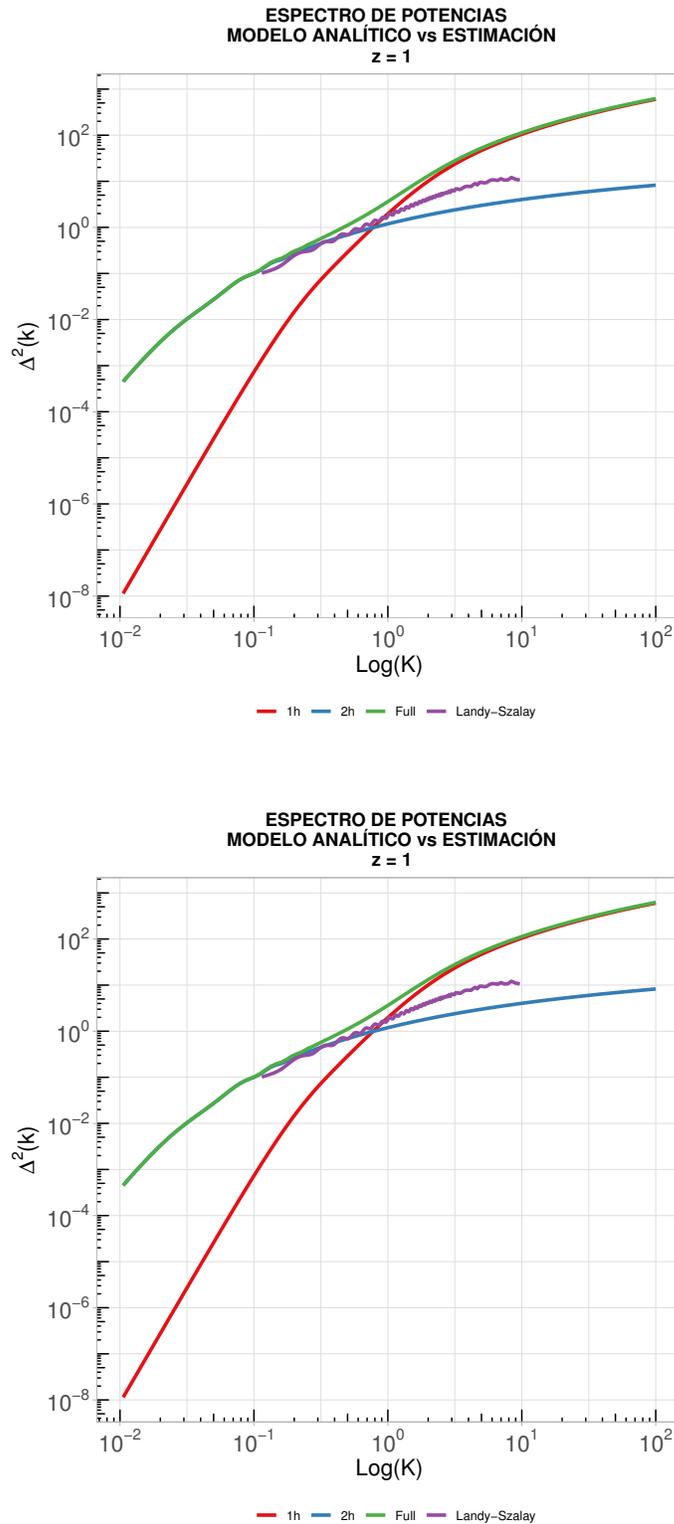


FIGURA 3.12: Comparación del Espectro de Potencias para el término halo-halo

Capítulo 4

Conclusiones y Perspectivas

La naturaleza de este trabajo en un contexto de la cosmología moderna, continua teniendo particular importancia al usar estadísticas como la función de correlación para extraer ya sea de muestreos galáctico o de simulaciones cosmológicas información que conduzca a mejorar los actuales modelos que describen la formación y evolución del universo, entre los cuales la formación de estructuras a gran escala sigue siendo un tema de interés científico de primer nivel. Más aún, en el contexto nacional, el presente trabajo constituye un aporte en términos de desarrollo y mejoramiento de herramientas eficientes, que optimizando la capacidad de computo permitan apoyar líneas de investigación que relacionen simulaciones o datos observacionales.

4.1. Conclusiones

Como resultado del presente trabajo se presentan las siguientes conclusiones:

4.1.1. Conclusiones físicas

1. Las estimaciones logradas de la función de correlación de dos puntos sobre la simulación BolshoiP muestran que a distancias mayores a 40 Mpc la correlación decrece por debajo de 10^{-3} a los diferentes redshift considerados.
2. La teoría no lineal de perturbaciones aplicada al modelamiento de la densidad de halos de materia oscura (formalismo perturbaciones-halo) evidencia que existe elementos relevantes que pueden ser explicados en el espectro de potencias hasta en un 2.3% a escalas intermedias ($K \approx 5$).

4.1.2. Conclusiones estadísticas

1. Los estimadores obtenidos para la función de correlación de dos puntos sobre la simulación BolshoiP, presentan en general buena precisión a $r > 0.5 Mpc$ y mejora a medida que el redshift decrece, particularmente, el caso del estimador de Landy-Szalay se presenta muy similar al de Hamilton.
2. Los resultados obtenidos para la función de correlación de dos puntos con los cuatro estimadores descritos están en concordancia con lo descrito por [11], donde se indica la paridad de dichos estimadores bajas distancias ($r < 32 Mpc$ según estos resultados).

4.1.3. Conclusiones del desarrollo computacional

1. El trabajo de paralelización del código original a través del uso MPI, resultó fundamental para la realización de esta tesis, considerando la disponibilidad del recurso computacional existente, logrando una reducción estimada del 99% en el tiempo total de ejecución. Así mismo, resulta importante mencionar que fue necesario el acompañamiento de un especialista en computación de alto desempeño dada la especificidad del tema.
2. En virtud de la limitación en la capacidad de procesamiento, en este trabajo se tomaron campos aleatorios de comparación con el mismo tamaño de los campos de halos, aún cuando la literatura [27] sugieren usar factores por encima de tres, sin embargo esto excedía por mucho la capacidad de procesamiento disponible luego de la optimización del código sin embargo, los errores en las estimaciones se mantienen en valores aceptables a escalas $r < 10 \text{ Mpc}$. Por debajo de estas distancias los errores en la correlación son menores a 0.1 ($\sigma < 0.1$)

4.2. Recomendaciones

1. Esta y otras líneas de investigación se apoyan parcialmente en recursos computacionales que permiten el uso de diversas fuentes de información disponibles en la actualidad y para el futuro se prevé aumenten de manera importante, de tal manera que un incremento en la capacidad de procesamiento en el programa de maestría sería un aporte de considerable beneficio.
2. Las diferentes herramientas de tipo estadístico también constituyen un elemento presente en estas líneas de investigación, pudiendo ser un excelente complemento a los cursos de formación para la maestría en astronomía de cara a los futuros ciclos de formación que los estudiantes puedan tomar.

4.3. Perspectivas

1. Otros aspectos acerca de la formación de estructura pueden ser abordados tomando este trabajo como punto de partida, tales como considerar la estimación de la función de correlación de dos puntos sobre el campo aleatorio pero bajo un enfoque de puntos marcados, es decir poder incorporar en la función de correlación información física adicional como la masa de cada halo. Por otra parte, la estimación del sesgo (bias), resulta importante a fin de completar la descripción de la formación y evolución de estructuras a gran escala. Finalmente, la descripción de las estructuras obtenidas en las simulaciones puede ser estudiada más ampliamente mediante la detección de las estructuras filamentosarias que visualmente parecen generarse y que se corresponden con resultados de los muestreos galácticos como el SDSS.
2. Es posible evaluar el efecto de usar estrategias de muestreo (por ejemplo bootstrap) sobre las cajas de simulación a fin de replicar estos resultados y hallar las fracciones de muestreo que permitan reducir el esfuerzo computacional conservando un nivel de precisión comparable.

Apéndice A

Apéndices

A.1. Campos aleatorios

En la descripción teórica del origen, formación y evolución del universo se establece un importante vínculo con diferentes observables astrofísicos, tales como velocidades, distribuciones de masa y la radiación cósmica de fondo entre otros, los cuales pueden ser considerados como realizaciones de un proceso estocástico que subyace en la naturaleza del fenómeno mismo, siendo los *Campos Aleatorios ó Estocásticos* un caso particular de ellos. Este enfoque aporta elementos claves para el desarrollo de una descripción estadística de aspectos específicos del universo en concordancia con los modelos físicos que dan cuenta del origen y dinámica de los fenómenos de estudio en astronomía. Particularmente, la formación de estructuras a gran escala del universo se consideran como el resultado del crecimiento de pequeñas fluctuaciones primordiales sobre un universo homogéneo amplificado por la inestabilidad gravitacional [2]. La naturaleza de tales fluctuaciones no son deterministas sino más bien estocásticas debido principalmente que no se tiene acceso observacional al campo de fluctuaciones primordiales lo cual permitiría definir las condiciones iniciales para una ecuación de evolución determinista y por otra parte la escala de evolución cosmológica supera ampliamente cualquier observación astronómica existente, de tal manera que no es posible ver la evolución completa de un solo sistema.

Así, algunos conceptos importantes para la descripción de los campos aleatorios en el contexto de los procesos estocásticos [1] se enuncian:

Definición A.1. *Un Proceso Estocástico es una familia de variables aleatorias $X = \{X_t, t \in T\}$ definidas sobre un espacio de probabilidad común $(\Omega, \mathfrak{F}, P)$ y con valores en un espacio medible (Ω', \mathfrak{F}') llamado espacio de estados. El conjunto de parámetros T se llama dominio de definición del proceso, \mathfrak{F} es una σ -álgebra definida sobre Ω y P es una medida sobre los elementos de la σ -álgebra.*

Definición A.2. *Un proceso estocástico $X = \{X_t, t \in T\}$ cuyo dominio de definición del proceso T sea de dimensión N y todas las v.a. X_t sean vectores de dimensión d se denomina Campo Aleatorio y se denota por $X^{N,d}$.*

Particularmente si cada una de las variables aleatorias del proceso están definidas sobre \mathbb{R}^N y toman valores en \mathbb{R}^d , entonces se tendrá un campo aleatorio real $X^{N,d}$. En general, un campo aleatorio puede ser caracterizado completamente si todas las variables aleatorias de valor vectorial X^d se pueden describir mediante una función de

distribución conjunta ¹ la cual esta dada por:

$$F_{t_1, \dots, t_N}(B) = P\{(X(t_1), \dots, X(t_N)) \in B\} \quad (\text{A.1})$$

Donde B es un elemento de la σ -álgebra \mathfrak{B}^{Nd} y $t_i \in \mathbb{R}^N$. Notese que en el caso de un proceso estocástico real la función de distribución se define sobre \mathbb{R}^N en tanto que en el campo aleatorio esta definida sobre $\mathbb{R}^{N,d}$

Definición A.3. Si \mathbf{X} es un campo aleatorio con función de distribución $F(\mathbf{X})$, entonces el valor esperado del campo esta dado por:

$$E[\mathbf{X}(\mathbf{t})] = \int_{\mathbf{X}(\mathbf{t}) \in \mathbf{B}} \mathbf{X}(\mathbf{t}) dF(\mathbf{X}(\mathbf{t})) \quad (\text{A.2})$$

Definición A.4. Si \mathbf{X} es un campo aleatorio con función de distribución $F(\mathbf{X})$, entonces la función de covarianza del campo esta dada por:

$$\text{Cov}(X(\mathbf{t}_1), X(\mathbf{t}_2)) = E[(\mathbf{X}(\mathbf{t}_1) - E[\mathbf{X}(\mathbf{t}_1)])(\mathbf{X}(\mathbf{t}_2) - E[\mathbf{X}(\mathbf{t}_2)])] \quad (\text{A.3})$$

Definición A.5. Un campo aleatorio $\mathbf{X}^{N,d}$ se dice estacionario en sentido estricto si para un τ arbitrario se satisface que

$$P\{(X(t_1), \dots, X(t_N)) \in B\} = P\{(X(t_1 + \tau), \dots, X(t_N + \tau)) \in B\} \quad (\text{A.4})$$

$$F(\mathbf{X}(\mathbf{t} + \tau)) = F(\mathbf{X}(\mathbf{t})) \quad (\text{A.5})$$

donde $\tau, \mathbf{t} \in \mathbb{R}^N$

Esta característica es importante porque establece la invarianza de la función de distribución conjunta ante traslaciones, lo cual es equivalente a decir que el campo aleatorio es estrictamente homogéneo.

Esta condición de *Estacionariedad Estricta* para muchos casos se puede reducir a una estacionariedad débil (homogeneidad débil) si el campo satisface:

1. $E[X(\mathbf{t})] = \mathbf{m}$ para $\mathbf{m} \in \mathbb{R}^d$ constante.
2. $E[(X(\mathbf{s}) - \mathbf{m})(X(\mathbf{t}) - \mathbf{m})]$ solo sea función de $\mathbf{s} - \mathbf{t}$

En general un campo homogéneo en sentido estricto siempre es homogéneo en sentido débil, pero lo contrario no es cierto en general.

Definición A.6. Un campo aleatorio homogéneo $\mathbf{X}^{n,d}$ se dice isotrópico si:

$$\text{Cov}(X(\mathbf{t}), X(\mathbf{t} + \tau)) = \text{Cov}(\|\tau\|) \quad (\text{A.6})$$

es decir, que la función de covarianza solo depende de la magnitud de τ y no de su dirección.

En resumen un campo aleatorio que sea invariante a traslaciones se denomina *estrictamente homogéneo* en tanto que si es invariante a rotaciones el campo se dice *isotrópico*.

¹También llamada función de distribución finito-dimensional

Esto significa que no hay puntos ni direcciones preferenciales en donde las características del campo sean particularmente diferentes respecto a otro punto o dirección.[1]

A.2. Partición de los conteos

Para el proceso de conteo de puntos en la estimación de la función de correlación de dos puntos se logro inicialmente paralelizar los tres tipos de conteo DD, DR y RR asignando a cada uno la tercera parte de los núcleos disponibles y al interior de cada conteo se distribuyó de manera uniforme el número de puntos que cada procesador debía procesar, sin embargo tal distribución de núcleos para los tres conteos y del número de puntos por núcleo mostró rápidamente que la carga computacional que se genera en cada proceso y núcleo no es la misma, resultado que unos procesadores terminaban mucho antes que otros quedando unos pocos sobrecargados y aumentando el tiempo de procesamiento, de tal manera que hallar una forma optima para tal distribución resulto de interés a fin de lograr mejorar los tiempos de procesamiento. De esta manera, se desarrolla una expresión con la cual se determina el número correcto de núcleos para cada proceso y la cantidad de puntos por núcleo que reduzca al máximo el tiempo de procesamiento.

A.2.1. Cantidad de núcleos por conteo

Consideremos que t sea el tiempo que le toma a un procesador calcular la distancia entre dos puntos y asignarlo a un bin (rango de distancia) y Sea n el número de puntos en una caja, de tal manera que en cada proceso se tiene que el tiempo total de procesamiento es:

$$\text{DD ó RR: } t \sum_{i=1}^{n-1} (n-i) = \frac{tn(n-1)}{2} \quad (\text{A.7})$$

$$\text{DR: } tn^2 \quad (\text{A.8})$$

De tal manera que si consideramos la razón de entre el número de operaciones del conteo DR y DD se tendrá:

$$\lim_{n \rightarrow \infty} \frac{tn^2}{\frac{tn(n-1)}{2}} = \lim_{n \rightarrow \infty} \frac{2n^2}{n^2 - n} = 2 \quad (\text{A.9})$$

Lo cual indica que para cajas con gran cantidad de puntos el conteo DR tardará el doble de que los procesos DD y RR, por tal razón se propone asignar a los conteos DD, RR y DR una cantidad de 1/4, 1/4 y 1/2 de los procesadores disponibles respectivamente.

A.2.2. Número de puntos por núcleo

Sea k el número de núcleos asignados al conteo DD o RR y m el número de puntos para los cuales el primer núcleo debe calcular las distancias con todos los demás puntos

de la caja, entonces la cantidad de distancias que se deberán calcular está dado por:

$$\sum_{i=1}^{n-1} (n-i) = \frac{n(n-1)}{2}$$

Lo que se desea es que cada núcleo tenga aproximadamente la misma cantidad de operaciones (distancias) a calcular para lo cual:

$$\begin{aligned} \sum_{i=1}^m (n-i) &= \frac{n(n-1)}{2k} \\ mn - \frac{m(m+1)}{2} &= \frac{n(n-1)}{2k} \\ m^2 + (1-2n)m + \frac{n(n-1)}{k} &= 0 \end{aligned}$$

Resolviendo la ecuación se tiene:

$$m = \left[\frac{4(n^2 - n) - k(1 - 2n)^2}{4k} \right]^{1/2} + n - \frac{1}{2} \quad (\text{A.10})$$

A.3. Ecuaciones de Einstein en el Gauge Newtoniano Conforme

Considere la métrica dada por:

$$ds^2 = a^2(\tau) \left\{ -(1 + 2\Psi)d\tau^2 + (1 - 2\Phi)\delta_{\mu\nu}dx^\mu dx^\nu \right\} \quad (\text{A.11})$$

donde $a(\tau)$ denota el factor de escala en términos del tiempo conforme, Φ la perturbación de la curvatura espacial y Ψ el potencial Newtoniano. Así, los términos de conexión están dados por:

$$\begin{aligned} \Gamma_{00}^0 &= H + \Psi' & \Gamma_{0\mu}^0 &= \Psi_{,\mu} & \Gamma_{\mu\nu}^0 &= H\delta_{\mu\nu} - [2H(\Psi + \Phi) + \Phi']\delta_{\mu\nu} \\ \Gamma_{00}^\mu &= \Psi_{,\mu} & \Gamma_{0\nu}^\mu &= H\delta_\nu^\mu - \Phi'\delta_\nu^\mu & \Gamma_{\nu\kappa}^\mu &= -(\Phi_{,\kappa}\delta_\nu^\mu + \Phi_{,\nu}\delta_\kappa^\mu) + \Phi_{,\mu}\delta_{\nu\kappa} \end{aligned}$$

y por otra parte, los valores correspondientes en la métrica de background (cuando $\Phi = \Psi = 0$) son:

$$\begin{aligned} \bar{\Gamma}_{00}^0 &= H & \bar{\Gamma}_{0\mu}^0 &= 0 & \bar{\Gamma}_{\mu\nu}^0 &= H\delta_{\mu\nu} \\ \bar{\Gamma}_{00}^\mu &= 0 & \bar{\Gamma}_{0\nu}^\mu &= H\delta_\nu^\mu & \bar{\Gamma}_{\nu\kappa}^\mu &= 0 \end{aligned}$$

de donde los términos en la métrica perturbada tiene la forma $\Gamma_{bc}^a = \bar{\Gamma}_{bc}^a + \delta\Gamma_{bc}^a$ y por tanto los términos $\delta\Gamma_{bc}^a$ asociados a las perturbaciones son:

$$\begin{aligned} \delta\Gamma_{00}^0 &= \Psi' & \delta\Gamma_{0\mu}^0 &= \Psi_{,\mu} & \delta\Gamma_{\mu\nu}^0 &= -[2H(\Psi + \Phi) + \Phi']\delta_{\mu\nu} \\ \delta\Gamma_{00}^\mu &= \Psi_{,\mu} & \delta\Gamma_{0\nu}^\mu &= -\Phi'\delta_{\nu}^\mu & \delta\Gamma_{\nu\kappa}^\mu &= -(\Phi_{,\kappa}\delta_{\nu}^\mu + \Phi_{,\nu}\delta_{\kappa}^\mu) + \Phi_{,\mu}\delta_{\nu\kappa} \end{aligned}$$

De igual manera, ocurre si se considera el tensor de Ricci en la métrica perturbada, $R_{ab} = \bar{R}_{ab} + \delta R_{ab}$ y dado que:

$$R_{ab} = \Gamma_{ab,c}^c - \Gamma_{ac,b}^c + \Gamma_{cd}^c\Gamma_{ab}^d - \Gamma_{db}^c\Gamma_{ac}^d$$

entonces las componentes son:

$$\begin{aligned} R_{00} &= -3H' + 3\Phi'' + \nabla^2\Psi + 3H(\Psi' + \Phi') \\ R_{0\mu} &= 2(\Psi' + H\Phi)_{,\mu} \\ R_{\mu\nu} &= (H' + 2H^2)\delta_{\mu\nu} \\ &\quad + [-\Psi'' + \nabla^2\Phi - H(\Psi' + 5\Phi') - (2H' + 4H^2)(\Psi + \Phi)]\delta_{\mu\nu} + (\Phi - \Psi)_{,\mu\nu} \end{aligned}$$

Subiendo el primer índice se obtiene:

$$R_b^a = g^{ac}R_{cb} = (\bar{g}^{ac} + \delta g^{ac})(\bar{R}_{cb} + \delta R_{cb}) = \bar{R}_b^a + \delta g^{ac}\bar{R}_{cb} + \bar{g}^{ac}\delta R_{cb}$$

de donde las componentes están dadas por:

$$\begin{aligned} R_0^0 &= \frac{3}{a^2} + \frac{1}{a^2}[-3\Phi'' - \nabla^2\Psi - 3H(\Psi' + \Phi') - 6H'\Psi] \\ R_\mu^0 &= -\frac{2}{a^2}(\Phi' + H\Psi)_{,\mu} \\ R_\nu^\mu &= \frac{1}{a^2}(H' + 2H^2)\delta_{\mu\nu} \\ &\quad + \frac{1}{a^2}[-\Phi'' + \nabla^2\Phi - H(\Psi' + 5\Phi') - (2H' + 4H^2)(\Psi + \Phi)]\delta_{\mu\nu} + \frac{1}{a^2}(\Phi - \Psi)_{,\mu\nu} \end{aligned}$$

Así el escalar de Ricci será:

$$\begin{aligned} R &= R_0^0 + R_\mu^\mu \\ &= \frac{6}{a^2}(H' + H^2) + \frac{1}{a^2}[-6\Phi'' + 2\nabla^2(2\Phi - \Psi) - 6H(\Psi' + 3\Phi') - 12(H' + H^2)\Psi] \end{aligned}$$

Con lo anterior se puede obtener el tensor de Einstein $G_b^a = R_b^a - \frac{1}{2}\delta_b^a R$ así:

$$\begin{aligned} G_0^0 &= -\frac{3H^2}{a^2} + \frac{1}{a^2}[-2\nabla^2\Phi + 6H\Phi' + 6H^2\Psi] \\ G_\mu^0 &= R_\mu^0 \\ G_\nu^\mu &= -\frac{2}{a^2}(H' + H^2)\delta_{\mu\nu} + \frac{1}{a^2}[2\Phi'' + \nabla^2(\Psi - \Phi) + H(2\Psi' + 4\Phi') + (4H' + 2H^2)\Psi]\delta_{\mu\nu} \\ &\quad + \frac{1}{a^2}(\Phi - \Psi)_{,\mu\nu} \end{aligned}$$

Estas componentes tienen la forma $G_b^a = \bar{G}_b^a + \delta G_b^a$ donde δG_b^a está compuesto por los términos que relacionan a Φ y Ψ . Una descripción más amplia es presentada por [13, 4], donde también se relaciona el tensor de momento-energía perturbado, obteniéndose con ello las ecuaciones de campo de Einstein bajo la teoría de perturbaciones, las cuales se reducen a la ecuación de Poisson en la aproximación sub-horizonte ($d \ll d_H^2$).

² d_H es la distancia de Hubble

A.4. Otros resultados

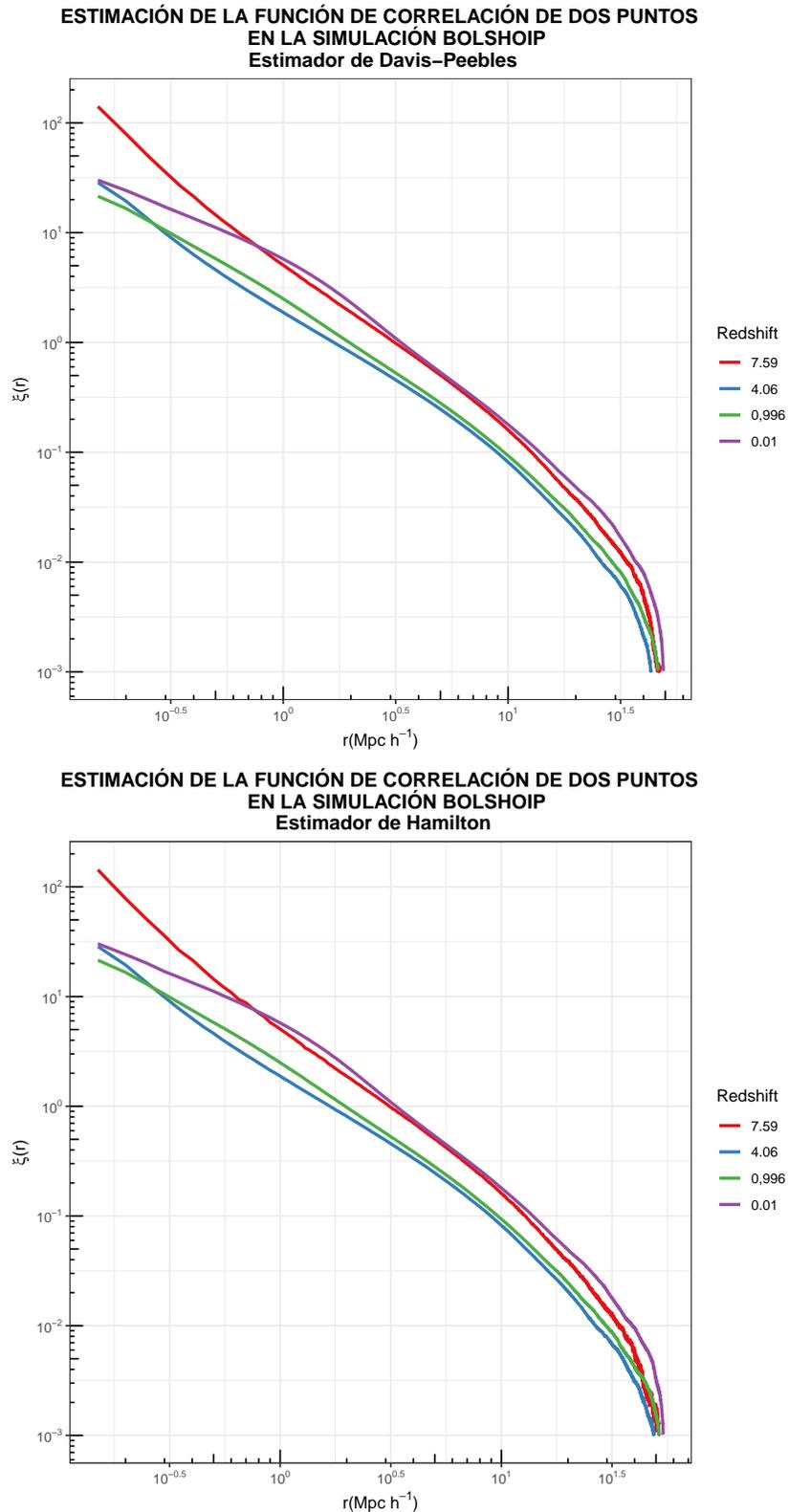


FIGURA A.1: Estimador de Davis-Peebles y Hamilton de la función de correlación de dos puntos a diferentes redshift

A.5. Código Fuente

A.5.1. Código del programa para la estimación de la función de correlación

```
#include <stdlib.h>
#include <stdio.h>
#include <string.h>
#include <math.h>
#include <time.h>
#include "mpi.h"

/* DEFINE TIPOS VARIABLES ESPECIFICAS */
typedef unsigned long ULONG;
typedef double LDOUBLE;

/* DEFINE LA ESTRUCTURA _p3 Struct defs and inlines using them. */
typedef struct _p3 {
LDOUBLE x, y, z;
} p3;

/* DEFINE PRODUCTO PUNTO Y NORMA DE UN VECTOR */
#define IPROD(a, b) ((a).x*(b).x + (a).y*(b).y + (a).z*(b).z)
#define NORM(a) (sqrt(IPROD((a), (a))))

/* DEFINE A CONSTANT TO DIVIDE THE PROBLEM IN SUBSET OF PROCESS TO
SOLVE THE PROBLEM */
#define SUBP 2 // Define dos grupos de procesos 1:DR y 2: DD, RR

typedef struct _field {
ULONG num_pts;
p3 low_bound, high_bound;
p3 *pts;
} field;

typedef struct _dataset {
int num_bins;
LDOUBLE **bins;
field data;
} dataset;

/* DEFINICIÓN DE FUNCIONES, ARGUMENTOS Y PUNTEROS */
void read_bins(FILE *f, dataset *storage);
void read_data(FILE *f, field *storage);
void correlate(int correlation_function,
              dataset *data, field *cmp, FILE *o, FILE *s);

/*Crear arreglo de particion de segmentos */
```

```

ULONG * partition (ULONG, int, int);
ULONG * MakeArray (int, int);
ULONG * SumArray (ULONG *, ULONG *, int);

void printArray (ULONG *, int, char *);

/* A prototype for a correlation function.
 * Parameters: ratio of N_r/N., DD, DR, RR. */
typedef LDOUBLE (*correlation_function)(LDOUBLE, ULONG, ULONG, ULONG);

/* Supported correlation methods. */
#define PEEBLES_HAUZER 1
#define DAVIS_PEEBLES 2
#define HAMILTON      3
#define LANDY_SZALAY 4
#define NUM_METHODS   4

/* Corresponding functions. */
inline LDOUBLE ph_corr(LDOUBLE rat, ULONG dd, ULONG dr, ULONG rr) {
return rat*rat*dd/rr - 1;}
inline LDOUBLE dp_corr(LDOUBLE rat, ULONG dd, ULONG dr, ULONG rr) {
return rat*dd/dr - 1;}
inline LDOUBLE ham_corr(LDOUBLE rat, ULONG dd, ULONG dr, ULONG rr) {
return (LDOUBLE)dd*rr/(dr*dr) - 1;}
inline LDOUBLE ls_corr(LDOUBLE rat, ULONG dd, ULONG dr, ULONG rr) {
return 1 + rat*rat*dd/rr - 2*rat*dr/rr;}

/** DECLARATION OF VECTORS **/
long *VR;

/** DECLARATION OF MPI VARIABLES **/

static int v, sv;

int p;          // Number of processes.
int whoami;    // Rank of the processes, it is like ID.
int subp;      // Number of processes in the sub set.
int subset1;   // Id grupos principales de procesos: 1:DR  0: DD, RR
int subset2;   // Id subprocesos DD y RR asi: 1:RR  0:DD
int subrank1; // Id procesos dentro del conteo DR subrank1=0,1,2,...,11
int subrank2; // Id procesos dentro del conteo DD y RR subrank2=0,1,2,...,5

int namel;          // name length
MPI_Status status;
//Create a new communicator por the gropus of DD, DR, RR.
MPI_Comm MPI_SUB_COMM_PS, MPI_SUB_COMM_DD, MPI_SUB_COMM_DR, MPI_SUB_COMM_RR;

```

```

/*****
** M A I N   P R O G R A M **
*****/
int main(int argc, char **argv) {

/** DECLARATION OF GENERAL VARIABLES **/

dataset data;
field cmp;
int i, corr_func;
int a,b;

/** DECLARATION OF TIME VARIABLES **/

clock_t t_ini, t_fin;
double secs;
t_ini = clock();

/** DECLARATION OF FILE VARIABLES **/

FILE *IN, *OUT, *CMP, *STATE;
//Start parallel program
MPI_Init (&argc, &argv);
//Get version of MPI
MPI_Get_version (&v,&sv);
//Error Handler for MPI
MPI_Errhandler_set (MPI_COMM_WORLD, MPI_ERRORS_RETURN);
MPI_Comm_size (MPI_COMM_WORLD,&p); //Get the number of processes
MPI_Comm_rank (MPI_COMM_WORLD,&whoami); //Get the ID of each process

printf ("Version de MPI %d\n",v); //Print MPI version
printf ("Numero de procesos: %d\n",p); //Print size of processes
printf ("ID proceso: %d\n",whoami); //Print ID of each process

if (argc != 5 && argc != 6) {
printf( "correlator: arguments: outputfile datafile "
"comparison sample file [correlation method]\n"
"Supported correlation methods:\n"
"1 => Peebles-Hauser\n"
"2 => Davis-Peebles\n"
"3 => Hamilton\n"
"4 => Landy-Szalay (default)\n");
exit(EXIT_FAILURE);}

/* Verify the correlation selection. */
if (argc == 6) {
corr_func = atoi(argv[5]);

```

```
if (corr_func <= 0 || corr_func > NUM_METHODS) {
    fprintf(stderr, "Invalid correlation method %d\n",
        corr_func);
    exit(EXIT_FAILURE);
}
else
    /* Prefer Landy-Szalay correlation function by default. */
    corr_func = 4;

/* Open the file handles, exit on errors.*/
/*printf("paso por aqui");*/
OUT = fopen(argv[1], "w");
if (OUT == NULL) {
    perror("fopen");
    exit(EXIT_FAILURE);
}

IN = fopen(argv[2], "r");
if (IN == NULL) {
    perror("fopen");
    exit(EXIT_FAILURE);}

CMP = fopen(argv[3], "r");
if (CMP == NULL) {
    perror("fopen");
    exit(EXIT_FAILURE);}

STATE = fopen (argv[4], "w");
if (STATE == NULL) {
    perror("fopen");
    exit(EXIT_FAILURE);
}

/* Read in the data. */
read_bins(IN, &data);
read_data(IN, &data.data);
read_data(CMP, &cmp);

/* Close input files. */
fclose(IN);
fclose(CMP);

/* Calculate and store the correlation. */
correlate(corr_func, &data, &cmp, OUT, STATE);

/* Final cleanup. */
fclose(OUT);
```

```

fclose (STATE);
free(cmp.pts);
free(data.data.pts);
for (i=0; i < data.num_bins; i++) {
free(data.bins[i]);
}
free(data.bins);

t_fin = clock();
secs = (double)(t_fin - t_ini) / CLOCKS_PER_SEC;
printf("Tiempo de ejecución %.16g segundos\n", secs);

MPI_Finalize ();
exit (EXIT_SUCCESS);
//return 0;
}

/*****
* CONSTRUCTION OF FUNCTIONS *
*****/
/*****/
/*****/
inline void check_malloc(void *p, int line) {
if (p == NULL) {
fprintf(stderr, "malloc returned null at line %d", line);
exit(EXIT_FAILURE);
}}
/*****/

void read_bins(FILE *f, dataset *d) {
int i;

/* First the number of bins. */
fscanf(f, "%d", &d->num_bins);

if (d->num_bins <= 0) {
fprintf(stderr, "Invalid number of bins: %d\n",
d->num_bins);
exit(EXIT_FAILURE);}

/* Allocate space. */
printf("Allocating space for %d bins...\n", d->num_bins);
d->bins = (LDOUBLE**)malloc(d->num_bins * sizeof(LDOUBLE*));
check_malloc(d->bins, __LINE__);

for (i=0; i < d->num_bins; i++) {
d->bins[i] = (LDOUBLE*)malloc(2 * sizeof(LDOUBLE));
check_malloc(d->bins[i], __LINE__); }

```

```

/* Read the bins. */
for (i=0; i < d->num_bins; i++) {
/*printf("Reading in bin %d... ", i);*/
fscanf(f, "%lg %lg", &d->bins[i][0], &d->bins[i][1]);
/*printf("%lg %lg\n", d->bins[i][0], d->bins[i][1]);*/
}}
/*=====*/

void read_data(FILE *f, field *d) {
ULONG i;

/* First, get the number of points. */
fscanf(f, "%lu", &d->num_pts);

if (d->num_pts <= 0) {
fprintf(stderr, "Invalid number of data points: %lu\n",
d->num_pts);
exit(EXIT_FAILURE);
}

/* Allocate space. */
printf("Allocating space for %lu points...\n", d->num_pts);
d->pts = (p3*)malloc((int)d->num_pts * sizeof(p3));
check_malloc(d->pts, __LINE__);

/* Low and high bounds. */
printf("Reading high and low bounds...\n");
fscanf(f, "%lg %lg %lg",
&d->low_bound.x, &d->low_bound.y, &d->low_bound.z);
fscanf(f, "%lg %lg %lg",
&d->high_bound.x, &d->high_bound.y, &d->high_bound.z);
printf("low: %lg %lg %lg high: %lg %lg %lg\n",
d->low_bound.x, d->low_bound.y, d->low_bound.z,
d->high_bound.x, d->high_bound.y, d->high_bound.z);

/* Then get the data. */
for (i=0; i < d->num_pts; i++) {
fscanf(f, "%lg %lg %lg",
&d->pts[i].x, &d->pts[i].y, &d->pts[i].z);}}
/*=====*/
void correlate(int corr_func, dataset *d, field *c, FILE *out, FILE *state) {
ULONG *dd, *ddt;
ULONG *dr, *drt;
ULONG *rr, *rrt;
ULONG i, j,df;
ULONG lif, lsp; //li = Limite inferior, ls = Limite superior

```

```

int k, ps;
p3 sep;
LDOUBLE l;
LDOUBLE delta;
ULONG *vSegmentosP1;
ULONG *vSegmentosP2;
ULONG temp;

correlation_function cfunc;

/* Get the right function. */
switch (corr_func) {
case PEBBLES_HAUSER:
cfunc = &ph_corr;
break;
case DAVIS_PEBBLES:
cfunc = &dp_corr;
break;
case HAMILTON:
cfunc = &ham_corr;
break;
case LANDY_SZALAY:
default:
cfunc = &ls_corr;
break;}

subset1=whoami % SUBP;
subrank1 = whoami / SUBP;
subset2=subrank1 % SUBP;
subrank2 = subrank1 / SUBP;

//Gets a vector with segments' numbers to divide the mount of constelations
    by p processes that balance the load between processes//

// Particion para el conteo DR
if (subset1==1){vSegmentosP1=partition(d->data.num_pts,p/SUBP, subset1);}
// Particion para los conteos DD y RR
else{vSegmentosP2=partition(d->data.num_pts,p/4, subset1);}

//if(whoami<3){printArray (vSegmentosP,p/SUBP,"Resultado Vector:");}

if (subset1 == 0) MPI_Comm_split (MPI_COMM_WORLD, 0, whoami, &MPI_SUB_COMM_PS);
if (subset1 == 1) MPI_Comm_split (MPI_COMM_WORLD, 1, whoami, &MPI_SUB_COMM_DR);

if (subset1 == 0 && subset2 == 0) MPI_Comm_split (MPI_SUB_COMM_PS, 0,
                                                subrank1, &MPI_SUB_COMM_DD);
if (subset1 == 0 && subset2 == 1) MPI_Comm_split (MPI_SUB_COMM_PS, 1,

```

```

subrank1, &MPI_SUB_COMM_RR);

/* Allocate. */

if (whoami == 0) {
ddt = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(ddt, __LINE__);
memset(ddt, 0, d->num_bins * sizeof(ULONG)); //Zero the storages
}
if (whoami == 1) {
drt = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(drt, __LINE__);
memset(drt, 0, d->num_bins * sizeof(ULONG)); //Zero the storages
}
if (whoami == 2) {
rrt = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(rrt, __LINE__);
memset(rrt, 0, d->num_bins * sizeof(ULONG)); //Zero the storages
}

if (subset1 == 0 && subset2 == 0){
dd = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(dd, __LINE__);
memset(dd, 0, d->num_bins * sizeof(ULONG)); //Zero the storages
}
if (subset1 == 1){
dr = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(dr, __LINE__);
memset(dr, 0, d->num_bins * sizeof(ULONG)); //Zero the storages
}
if (subset1 == 0 && subset2 == 1){
dd = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(dd, __LINE__);
memset(dd, 0, d->num_bins * sizeof(ULONG)); //Zero the storages

dr = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(dr, __LINE__);
memset(dr, 0, d->num_bins * sizeof(ULONG)); //Zero the storages

rr = (ULONG*)malloc(d->num_bins * sizeof(ULONG));
check_malloc(rr, __LINE__);
memset(rr, 0, d->num_bins * sizeof(ULONG)); //Zero the storages
}
if (subset1 == 1){
if (subrank1 == 0) {lif=0;}
else {lif=vSegmentosP1[subrank1 - 1];}
lsp=vSegmentosP1[subrank1];}
else{if (subrank2 == 0) {lif=0;}

```

```

else {lif=vSegmentosP2[subrank2 - 1];}
lsp=vSegmentosP2[subrank2];}

delta=d->bins[1][1]-d->bins[1][0];

/*****CONTEO DE PARES DD *****/

fseek(state,0,SEEK_SET);
fprintf(state,"CALCULO DE LA FUNCION DE CORRELACION DE DOS PUNTOS\n");
fprintf(state,"-----\n");

if (subset1 == 0 && subset2==0) {
for (i=lif; i < lsp; i++) {
/* Print some progress info. */

if(whoami==0){df=lsp-lif;
//dfc=i-lsp+1;
//printf("El proceso ve en %lu de un total de %lu\n",i,df);
fprintf(state,"DD: %3.2f\n",(double)i/df*100);
//fseek( state, 0, SEEK_SET );}

for (j=i+1; j < d->data.num_pts; j++) {
sep.x = d->data.pts[i].x - d->data.pts[j].x;
sep.y = d->data.pts[i].y - d->data.pts[j].y;
sep.z = d->data.pts[i].z - d->data.pts[j].z;
l = NORM(sep);
k=floor(l/delta);
if (k < d->num_bins){dd[k]+=2;}
}}}

/*****CONTEO DE PARES DR *****/

if (subset1 == 1) {
for (i=lif; i < lsp; i++) {
/* Print some progress info. */

if(whoami==1){df=lsp-lif;
//dfc=i-lsp+1;
fseek( state, 136, SEEK_SET );
fprintf(state,"DR: %3.2f\n",(double)i/df*100);}

for (j=0; j < c->num_pts; j++) {
sep.x = d->data.pts[i].x - c->pts[j].x;
sep.y = d->data.pts[i].y - c->pts[j].y;
sep.z = d->data.pts[i].z - c->pts[j].z;
l = NORM(sep);

```

```

k=floor(1/delta);
if (k < d->num_bins){dr[k]++;}
}}}

/*#####CONTEO DE PARES RR #####*/

if (subset1 == 0 && subset2==1) {
for (i=lif; i < lsp; i++) {
/* Print some progress info. */

if(whoami==2){df=lsp-lif;
//dfc=i-lsp+1;
fseek( state, 160, SEEK_SET );
fprintf(state,"RR: %3.2f\n", (double)i/df*100);}

for (j=i+1; j < c->num_pts; j++) {
sep.x = c->pts[i].x - c->pts[j].x;
sep.y = c->pts[i].y - c->pts[j].y;
sep.z = c->pts[i].z - c->pts[j].z;
l = NORM(sep);

k=floor(1/delta);
if (k < d->num_bins){rr[k]+=2;}
}}}
fseek( state, 0, SEEK_END );
fprintf(state,"Finalizado el proceso de conteo...\n\n");

/* BEGIN MPI COMMUNICATION */

/* Reduce information for each set:  Processes that belong to the set of
dd; p0, p3, p6, ... reduce by sum to p0
processes that belong to the set of dr; p1, p4, p7, ... reduce by sum to p1
processes that belong to the set of rr; p2, p5, p8, ... reduce by sum to p2 */

fseek( state, 0, SEEK_END );
fprintf(state,"Inicio de comunicaciones...\n\n");

if (subrank1 == 0)
for (i=1; i<p/SUBP; i++) {

if (subset1 == 1) {
MPI_Recv (&drt[0],d->num_bins,MPI_UNSIGNED_LONG,i,i,MPI_SUB_COMM_DR,&status);
dr = SumArray (dr, drt, d->num_bins);}

if (subrank2 == 0)
for (i=1; i<p/4; i++) {
if (subset1 == 0 && subset2==0) {

```

```

MPI_Recv (&ddt[0],d->num_bins,MPI_UNSIGNED_LONG,i,i,MPI_SUB_COMM_DD,&status);
dd = SumArray (dd, ddt, d->num_bins);}

if (subset1 == 0 && subset2==1) {
MPI_Recv (&rrt[0],d->num_bins,MPI_UNSIGNED_LONG,i,i,MPI_SUB_COMM_RR,&status);
rr = SumArray (rr, rrt, d->num_bins);
}}

if (subset1 == 0 && subset2==0)
for (i=1; i<p/4; i++)
if (i == subrank2)
MPI_Send (&dd[0], d->num_bins,MPI_UNSIGNED_LONG,0,i,MPI_SUB_COMM_DD);

if (subset1 == 1)
for (i=1; i<p/SUBP; i++)
if (i == subrank1)
MPI_Send (&dr[0], d->num_bins,MPI_UNSIGNED_LONG,0,i,MPI_SUB_COMM_DR);

if (subset1==0 && subset2== 1)
for (i=1; i<p/4; i++)
if (i == subrank1)
MPI_Send (&rr[0], d->num_bins,MPI_UNSIGNED_LONG,0,i,MPI_SUB_COMM_RR);

MPI_Barrier (MPI_SUB_COMM_DD);
MPI_Barrier (MPI_SUB_COMM_DR);
MPI_Barrier (MPI_SUB_COMM_RR);

/* p2 gathers the information */

if ( whoami == 0 ) MPI_Send (&dd[0], d->num_bins,MPI_UNSIGNED_LONG,
                                2,0,MPI_COMM_WORLD);
if ( whoami == 1 ) MPI_Send (&dr[0], d->num_bins,MPI_UNSIGNED_LONG,
                                2,1,MPI_COMM_WORLD);
if ( whoami == 2 )
{
// Receive from process 0
MPI_Recv (&dd[0],d->num_bins,MPI_UNSIGNED_LONG,0,0,MPI_COMM_WORLD,&status);
// Receive from process 1
MPI_Recv (&dr[0],d->num_bins,MPI_UNSIGNED_LONG,1,1,MPI_COMM_WORLD,&status); }

MPI_Barrier (MPI_COMM_WORLD);

fseek( state, 0, SEEK_END );
fprintf(state,"Finalizada la comunicacion y consolidacion de conteos...\n\n");

/* END MPI COMMUNICATION*/

```

```

if (whoami == 2)
{
fseek( state, 0, SEEK_END );
fprintf(state,"Calculando la correlacion...\n\n");

printf ("Calculando la correlacion...\n");
LDOUBLE ksi, error;
LDOUBLE rat = c->num_pts/(LDOUBLE)d->data.num_pts;
for (k=0; k < d->num_bins; k++) {

if (rr[k] == 0) {
/* XXX: Something like this is what _should_
* be done, however, it would require
* enabling GNU extensions or somesuch. */
/*
ksi = error = NAN;
*/
fprintf(out, "%lg %lg %ld %ld %ld %s %s\n",d->bins[k][0],
          d->bins[k][1], dd[k], dr[k], rr[k],"NaN", "NaN");
}
else {
/* Use the given estimator. */

ksi = cfunc(rat, dd[k], dr[k], rr[k]);

/* For the Poisson errors, we have:
* (1+ksi)/sqrt(x), where x can be DD,
* (N/N_r)*DR or (N/N_r)^2*RR. Of these,
* Martinez & Saar suggest using one of the
* latter two. Of these, again the latter
* is more convenient, as the RR counts should
* be consistently large. */
/*
error = (1 + ksi)/sqrt(dd[k]);
error = (1 + ksi)/sqrt(dr[k]/rat);
*/
error = (1 + ksi)/sqrt(rr[k]/(rat*rat));

fprintf(out, "%lg %lg %ld %ld %ld %lg %lg\n",d->bins[k][0],
          d->bins[k][1], dd[k], dr[k], rr[k],ksi, error);}}
printf("Resultado generado...\n\n");
fseek( state, 0, SEEK_END );
fprintf(state,"CALCULO FINALIZADO...\n\n");}

/* Cleanup. */
MPI_Comm_free(&MPI_SUB_COMM_PS);
MPI_Comm_free(&MPI_SUB_COMM_DD);

```

```

MPI_Comm_free(&MPI_SUB_COMM_DR);
MPI_Comm_free(&MPI_SUB_COMM_RR);

if (whoami == 0) {free(dd);free(ddt);}
if (whoami == 1) {free(dr);free(drt);}
if (whoami == 2) {free(rr);free(dd);free(dr);free(rrt);}
}

/** Create partition *
/*****/
//n: Numero de objetos, p: Numero de procesos entre 3
ULONG * partition (ULONG n, int p, int op) {

ULONG A,B,C,m,L,N;
int pc,i;
ULONG *M;

M=MakeArray (p,1);

if (op == 0 || op == 2) {
L=0;
N=n;
for (i=0; i < p; i++) {
pc=p-i;
A=pc*(1-2*n)*(1-2*n);
B=4*((n*n)-n);
C=4*pc;
//D=(1-2*n)/2;
m=floor(-sqrt((A-B)/C)+n);
n=n-m;
M[i]=m+L;
L=M[i];
//printf("Particion %d de tamaño pc:%d, A:%li, B:%li,
c:%li, D:%li, m:%li, M[%d]=%li, L:%li \n",i,pc,A,B,C,D,m,i,M[i],L);
//printf("Particion de tamaño \n");
}
M[p-1]= N;
}
if (op ==1) {
m=floor(n/p);
for (i=1; i<=p; i++) {
A=m*i;
M[i-1]=A;
}
M[p-1]=n;
}
return (M);}

```

```

/** Make a array for the set of segments *
// *****/
ULONG * MakeArray (int size, int op)
{
int k;
ULONG *temp;
temp=(ULONG *) malloc(size *sizeof(ULONG)); // 0 -> Creates the array
if (op == 1) // 1 -> Creates an array and init with 0.
for (k=0;k<size;k++) temp[k]=0;
return (temp);}

/** Sum two arrays *
// *****/

ULONG * SumArray (ULONG *A, ULONG *B, int LS)
{
ULONG *temp2;
int i;
temp2=MakeArray (LS, 1);
for (i = 0; i<LS; i++ ) temp2[i]= A[i] + B[i];
return (temp2);}

/** Print information *****/
// *****/
void printArray (ULONG *Array, int size, char * message)
{
int i;
printf ("%s\n",message);
for (i=0;i<size;i++) printf ("Array %d [%lu]\n",i,Array[i]);}

```

A.5.2. Código R para la simulación de campos aleatorios

```

C_Bin<-function(N,a,dm=2,graf=FALSE){
#-----
# Esta función genera un campo espacial puntual binomial
# con argumentos:
#   N: Número de puntos a generar
#   a: Lado de la caja que contiene el campo. Por defecto es
#       una caja cuadrada.
#   dm: Dimension del campo, mayor que uno.
#   graf: Variable logica que indica si la función muestra la
#         grafica del campo, por defecto graf=FALSE.
#-----
if(dm>1){
P=runif(N,0,a)

```

```

for(i in 2:dm){
P=cbind(P,runif(N,0,a))
}
P=round(P,5)

if(graf==TRUE){
if(dm==2){
plot(P,
main='CAMPO BINOMIAL',
xlab='X',ylab='Y',
col='blue', asp=1,
xlim=c(0,a), ylim=c(0,a))
#text(0.6,a,expression('N='),cex=1.1)
}
if(dm==3){
require(scatterplot3d)
scatterplot3d(P,
pch=1,
main=expression('CAMPO BINOMIAL'),
xlab='X',
ylab='Y',
zlab='Z',
color='blue',
box=TRUE,
xlim=c(0,a),
ylim=c(0,a),
zlim=c(0,a))
}
if(dm>3){print('No se puede graficas en 4 o más dimensiones...')}
}
}else{
print('El campo debe tener dos dimensiones o más')
}
return(P)
}

C_PH<-function(lambda,a,dm=2,graf=FALSE){
#-----
# Esta función genera un campo espacial puntual poisson
# homogéneo con argumentos:
# lambda: Intensidad del campo por unidad métrica del campo,
#         esto indica que por cada unidad cuadrada o unidad
#         de volumen la intensidad es lambda.
# a: Lado de la caja que contiene el campo. Por defecto es
#    una caja cuadrada.
# dm: Dimensión del campo, mayor que uno.
# graf: Variable lógica que indica si la función muestra la

```

```

# grafica del campo, por defecto graf=FALSE.
#-----
source('A:/TESIS/Programas_R/C_Bin.R', encoding = 'UTF-8')
if(dm>1){
k=0
while(k==0){
N=rpois(1,lambda*a^dm)
if(N>0){break}
}
P=C_Bin(N,a,dm = dm,graf = FALSE)

if(graf==TRUE){
if(dm==2){
plot(P,
main='CAMPO POISSON HOMOGENEO',
xlab='X',ylab='Y',
col='blue', asp=1,
xlim=c(0,a), ylim=c(0,a))
#text(0.6,a,expression('N='),cex=1.1)
}
if(dm==3){
require(scatterplot3d)
scatterplot3d(P,
pch=1,
main=expression('CAMPO POISSON HOMOGENEO'),
xlab='X',
ylab='Y',
zlab='Z',
color='blue',
box=TRUE,
xlim=c(0,a),
ylim=c(0,a),
zlim=c(0,a))
}
if(dm>3){print('No se puede graficas en 4 o más dimensiones...')}
}
}else{
print('El campo debe tener dos dimensiones o más')
}
return(P)
}

C_NS<-function(lambda,a,lambda2,r,dm=2,graf=FALSE){
#-----
# Esta función genera un campo espacial puntual poisson
# no homogéneo con argumentos:
# lambda: Intensidad del campo por unidad métrica del campo,

```

```

#           para los puntos padres
#           a: Lado de la caja que contiene el campo. Por defecto es
#           una caja cuadrada.
#           lambda2: Intensidad del campo por unidad metrica del campo,
#           para los puntos hijos
#           r: Radio del circulo para cada cluster (conj. puntos hijos)
#           dm: Dimension del campo, mayor que uno.
#           graf: Variable logica que indica si la función muestra la
#           grafica del campo, por defecto graf=FALSE.
#-----
source('A:/TESIS/Programas_R/C_PH.R', encoding = 'UTF-8')
source('A:/TESIS/Programas_R/C_BinC.R', encoding = 'UTF-8')
if(dm>1){
P=C_PH(lambda,a,dm=d,graf=F)

nc=dim(P)
N=rpois(nc[1],lambda2*(r^dm)*pi)
PNS=C_BinC(N[1],r,P[1,],dm=d,graf = F)
print('paso')
for(i in 2:nc[1]){
if(N[i]>0){PNS=rbind(PNS,C_BinC(N[i],r,P[i,],dm=d,graf = F))}
}

if(graf==TRUE){

if(dm==2){
plot(PNS,
main='CAMPO DE NEYMAN-SCOTT',
xlab='X',ylab='Y',
col='black', asp=1,
pch=1,cex=1.2,
xlim=c(0,a),
ylim=c(0,a),
xaxs='e')
#xlim=c(round(min(PNS[,1],digits=0)),round(max(PNS[,1],digits=0))),
#ylim=c(round(min(PNS[,2],digits=0)),round(max(PNS[,2],digits=0)))
points(P[,1],P[,2],pch=16,cex=1.2,col='black')
}
if(dm==3){
require(scatterplot3d)

scatterplot3d(PNS,
pch='20',
main=expression('CAMPO DE NEYMAN-SCOTT'),
xlab='X',
ylab='Y',
zlab='Z',

```

```

color='blue',
box=TRUE,
xlim=c(round(min(PNS[,1],digits=0)),round(max(PNS[,1],digits=0))),
ylim=c(round(min(PNS[,2],digits=0)),round(max(PNS[,2],digits=0))),
zlim=c(round(min(PNS[,3],digits=0)),round(max(PNS[,3],digits=0)))
}
if(dm>3){print('No se puede graficas en 4 o más dimensiones...')}
}
}else{
print('El campo debe tener dos dimensiones o más')
}
return(PNS)
}

C_Mat<-function(lambda,a,N,r,dm=2,graf=FALSE){
#-----
# Esta función genera un campo espacial puntual poisson
# no homogeneo con argumentos:
# lambda: Intensidad del campo por unidad metrica del campo,
#         para los puntos padres
# a: Lado de la caja que contiene el campo. Por defecto es
#    una caja cuadrada.
# N: Número de puntos hijos a generar por cada punto padre
# r: Radio del circulo para cada cluster (conj. puntos hijos)
# dm: Dimension del campo, mayor que uno.
# graf: Variable logica que indica si la función muestra la
#       grafica del campo, por defecto graf=FALSE.
#-----
source('A:/TESIS/Programas_R/C_PH.R', encoding = 'UTF-8')
source('A:/TESIS/Programas_R/C_BinC.R', encoding = 'UTF-8')
if(dm>1){
P=C_PH(lambda,a,dm=d,graf=F)
nc=dim(P)

PM=C_BinC(N,r,P[1,],dm=d,graf = F)
for(i in 2:nc[1]){
PM=rbind(PM,C_BinC(N,r,P[i,],dm=d,graf = F))
}
if(graf==TRUE){
if(dm==2){
#par(xaxs='s')
plot(PM,
main='CAMPO DE MATERN',
xlab='X',ylab='Y',
col='black', asp=1,
pch=1,
xlim=c(0,a),

```

```

ylim=c(0,a))
#xlim=c(round(min(PM[,1],digits=0)),round(max(PM[,1],digits=0))),
#ylim=c(round(min(PM[,2],digits=0)),round(max(PM[,2],digits=0)))
points(P[,1],P[,2],pch=20,col='black')

}
if(dm==3){
require(scatterplot3d)

scatterplot3d(PM,
pch='.',
main=expression('CAMPO MATERN'),
xlab='X',
ylab='Y',
zlab='Z',
color='blue',
box=TRUE,
xlim=c(round(min(PM[,1],digits=0)),round(max(PM[,1],digits=0))),
ylim=c(round(min(PM[,2],digits=0)),round(max(PM[,2],digits=0))),
zlim=c(round(min(PM[,3],digits=0)),round(max(PM[,3],digits=0))))
}
if(dm>3){print('No se puede graficas en 4 o más dimensiones...')}
}
}else{
print('El campo debe tener dos dimensiones o más')
}
return(PM)
}

```

A.5.3. Código Python para el cálculo del espectro de potencias

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

plt.rc('text', usetex=True)
plt.rc('font', family='serif')

def f(k, er, r):
return er*np.sin(k*r)*(r**2)/(k*r)

def integral(f, k, er, r):
N = 2*k**3/np.pi
suma = 0
for i in range(len(r)-2):
suma+=(r[i+1]-r[i])*(f(k,er[i+1],r[i+1])+f(k,er[i],r[i]))/2

```

```
return N*suma

df = pd.read_csv('Correlaciones.txt', sep=" ")

lista = ['LS', 'PH', 'DP', 'HM']
rd = 58 # 58, 76, 136, 212
z1 = df[df['Caja']==rd]
#z1 = df[df['Caja']==76]
#z1 = df[df['Caja']==136]
#z1 = df[df['Caja']==212]

r1 = np.array(z1['bs'])

er1 = np.array(z1[lista[0]])
er2 = np.array(z1[lista[1]])
er3 = np.array(z1[lista[2]])
er4 = np.array(z1[lista[3]])

k1 = 2*np.pi/np.array(r1)

f1 = integral(f,k1,er1,r1)
f2 = integral(f,k1,er2,r1)
f3 = integral(f,k1,er3,r1)
f4 = integral(f,k1,er4,r1)

a = 15
b = len(f1)+1

f1 = f1[a:b]
f2 = f2[a:b]
f3 = f3[a:b]
f4 = f4[a:b]
k1 = k1[a:b]
```

Referencias

- [1] Robert J. Alder. *The geometry of random fields*. Jhon Wiley y Sons, 1981.
- [2] F. Bernardeau y col. «Large-scale structure of the universe and cosmological perturbation theory». En: *Physics Report* 367 (sep. de 2002), págs. 1-248.
- [3] Leonardo Castañeda. «Notas de Cosmología». 2018.
- [4] Joel Jose Velasquez Celis. «Perturbaciones cosmológicas en teorías de gravedad escalar-tensor». Magíster en Ciencias - Física. Tesis de mtría. Universidad Nacional de Colombia - Sede Bogotá, 2018. URL: <http://bdigital.unal.edu.co/64796/>.
- [5] A. Cooray y R. Sheth. «Halo Models of Large Scale Structure». En: *Physics Reports* 372 (2002), págs. 1-129.
- [6] Scott Dodelson. *Modern Cosmology*. Academic Press, 2003.
- [7] Santiago I. Ferrero. «Contracción de Halos de Materia Oscura». Universidad Nacional de Cordoba, Facultad de Matemática, Astronomía y Física, Licenciatura en Astronomía. 70 págs. URL: <https://rdu.unc.edu.ar/handle/11086/3>.
- [8] James N. Fry y Enrique Gaztanaga. «Biasing and hierarchical statistics in large scale structure». En: *The Astrophysical Journal* 413 (1993), págs. 447-452. DOI: [10.1086/173015](https://doi.org/10.1086/173015). arXiv: [astro-ph/9302009](https://arxiv.org/abs/astro-ph/9302009) [[astro-ph](https://arxiv.org/abs/astro-ph/9302009)].
- [9] A. Gabrielli y col. *Statistical physics for cosmic structures*. Springer, 2005.
- [10] M. P. Hobson, G. P. Efstathiou y A. N. Lasenby. *General Relativity: An Introduction for Physicists*. Cambridge University Press, 2006, 556–572.
- [11] M. et al. Kerscher. «A comparison of estimators for the two-point correlation function». En: *The Astrophysical Journal* 535.1 (2000).
- [12] Anatoly Klypin y col. «MultiDark simulations: the story of dark matter halo concentrations and density profiles». En: *Monthly Notices of the Royal Astronomical Society* 457 (2016), págs. 4340-4359.
- [13] Hideo Kodama y Misao Sasaki. «Cosmological perturbation theory». En: *Progress of Theoretical Physics Supplement* 78 ().
- [14] A. J. Mead y col. «An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models». En: *Monthly Notices of the Royal Astronomical Society* 454.2 (2015), págs. 1958-1975. URL: <http://arxiv.org/abs/1505.07833><http://dx.doi.org/10.1093/mnras/stv2036>.
- [15] J.F. Navarro, C.S. Frenk y S.D.M. White. «The structure of cold dark matter halos». En: *The Astrophysical Journal* 462 (1996), págs. 563-575.
- [16] J. Neyman y E. L. Scott. «A Theory of the spatial distribution of galaxies». En: *Astrophysical Journal* 116 (1952), págs. 144-163.

- [17] J. Neyman y E. L. Scott. «Statistical approach to problems of cosmology». En: *Journal of the Royal Statistical Society B* 20.1 (1958), págs. 1-43.
- [18] Pihajoki Pauli. *Correlator.c*. 2014. URL: http://www.astro.utu.fi/research/CosmoS/tools/tools_p1.shtml (visitado 14-10-2014).
- [19] J. A. Peacock. *Cosmological Physics*. Cambridge University Press, 1998. DOI: [10.1017/CB09780511804533](https://doi.org/10.1017/CB09780511804533).
- [20] P.J.E. Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980.
- [21] W. H. Press y P. Schechter. «Formation of galaxies and cluster of galaxies by self-similar gravitational condensation». En: *The Astrophysical Journal* 187 (1974), págs. 425-438.
- [22] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2013. URL: <http://www.R-project.org/>.
- [23] Sergio A. Rodriguez. *Función de correlación halo-halo, comparación con la teoría de perturbaciones*. Tesis de Pregrado. Universidad Nacional de Colombia, 2011.
- [24] Robert E. Smith, Román Scoccimarro y Ravi K. Sheth. «Scale dependence of halo and galaxy bias: Effects in real space». En: *Phys. Rev. D* 75 (6 2007), pág. 063512. DOI: [10.1103/PhysRevD.75.063512](https://doi.org/10.1103/PhysRevD.75.063512). URL: <https://link.aps.org/doi/10.1103/PhysRevD.75.063512>.
- [25] Oscar Javier Vija Suarez. «Teoría de perturbaciones cosmológicas en la era de radiación: caracterización de la función de transferencia.» Tesis de mtría. Universidad Nacional de Colombia - Sede Bogotá, 2017. URL: <http://bdigital.unal.edu.co/61237/>.
- [26] Martinez Vincent. y Saar Enn. *Statistical of the galaxy distribution*. Chapman y Hall/CRC, 2001.
- [27] Martínez VincentJ. y col., eds. *Data Analysis in Cosmology*. Lecture Notes in Physics. Springer-Verlag, 2009.