

# Fractal Analysis of Large-Scale Structures

## Análisis Fractal de Estructuras a Gran Escala

Carlos Marrero de la Rosa

---

UNIVERSIDAD DE LA LAGUNA  
FACULTAD DE CIENCIAS  
GRADO EN FÍSICA  
CURSO 2020-2021

---



**Tutores:**

Christopher Bryan Brook, Albano José González Fernández

*“Sometimes the less important things, my dear child, lead us to the greatest discoveries”.*

**The Doctor.**

## **Acknowledgments**

First of all, I would like to thank all the people who have helped me throughout this work, especially Andrés Balaguera-Antolínez who, despite not tutoring this project, has helped me to understand a large part of the concepts involved in it. And on the other hand, thanks to all my classmates and friends from the degree, who have supported me while I was doing this work and have been there at all times to encourage me to continue.

## Resumen

Con el avance de los tiempos se han ido definiendo estructuras o formas que ayudaran al ser humano a comprender mejor su entorno, a aproximarlo de alguna manera a su entendimiento. Es durante los siglos XIX-XX que aparece una nueva forma, lo que se pasaría a llamar un fractal, un objeto matemático cuya aparente irregularidad se repite a diferentes escalas. Un objeto que no sigue la geometría de Euclides. Un objeto que, a pesar de estas curiosas características, se puede vislumbrar en las costas, en las hojas de helecho o en la espuma cuántica. Hausdorff planteó una de las primeras definiciones de dimensión que se podría aplicar a un fractal, abriendo la puerta al cálculo de la dimensión fractal, que será la piedra angular de este trabajo. Se puede entender de muchas formas, pero la que mejor se adapta al interés de este trabajo es que la dimensión fractal proporciona una idea de lo irregular que es una distribución. De cómo se distribuyen los puntos que componen una estructura. Esto indica que puede dar información sobre el agrupamiento de una distribución.

En este trabajo se medirá la dimensión fractal de las estructuras a gran escala del universo, a fin de comprobar si siguen una distribución homogénea. Para ello se emplearán datos provistos por el conjunto de datos de grupos de galaxias BOSS (*Baryon Oscillation Spectroscopic Survey*) que forma parte del SDSS (*Sloan Digital Sky Survey*). En concreto, se trabajará con los datos conjuntos de los dos algoritmos de selección de BOSS, para el casquete galáctico norte: LOWZ, que selecciona objetos hasta un *redshift* tal que  $z \approx 0.4$  y CMASS, que selecciona objetos en un rango de  $0.4 < z < 0.7$ . Este conjunto de ambos se denomina CMASSLOWZTOT\_North, y proporciona datos de unos 953255 objetos.

El objetivo principal será estudiar cómo varía la dimensión fractal de estas estructuras a gran escala con la distancia comóvil, y analizar si los resultados coinciden con aquellos indicados en la literatura. Para lograr este objetivo se medirá la dimensión fractal a través de varios métodos: algoritmos de *box-counting*, la función de correlación de dos puntos y la transformada de Hankel del espectro de potencias.

En primer lugar, para realizar los análisis con los programas de *box-counting*, será necesario tener un mapa de la distribución de los objetos en el cielo. Para ello se empleará la muestra proporcionada por SDSS y, con el lenguaje de programación Python, se dibujará este mapa de distribución. Los primeros métodos de *box-counting* que se emplearán dividirán este mapa en pequeñas cajas bidimensionales, donde solo se tendrán en cuenta para el tratamiento aquellas que tengan, al menos, un objeto en su interior. En uno de los métodos, las cajas no se superpondrán, sino que serán adyacentes unas con otras (método estándar), y en el otro, las muestras se superpondrán entre sí (método *gliding* ; deslizante). Por otra parte, para el tercer método de *box-counting*, se tendrá en cuenta una tercera componente, ya que dividirá el set de datos en cubos. La tercera componente se dará poniendo el mapa de distribución en escala de grises, donde la escala de grises corresponderá a la distancia comóvil. De esta manera se tendrá una medición de la dimensión fractal a través de tres métodos de *box-counting*.

Continuando con los algoritmos de *box-counting*, se realizará una medición del método estándar y del método de escala de grises formando el mapa del cielo con *Healpix*, que reproducirá el cielo en una superficie esférica dividida en píxeles de áreas iguales, permitiendo así una representación más realista del cielo al seguir su geometría.

El siguiente paso corresponderá a emplear la función de correlación de dos puntos para realizar el cálculo de la dimensión fractal. Se utilizará para calcular la función de estructura,  $g(r) = 1 + \xi(r)$ , su gradiente log-log (la función de gradiente),  $\gamma(r) = d \log g(r) / d \log r$ , y la función

de dimensión fractal,  $D(r) = 3 + \gamma(r)$ . En este caso, la función de correlación de dos puntos se obtendrá midiéndola directamente, utilizando conteo de pares. Se empleará para este fin el estimador de Landy & Szalay. Una vez hecho esto, se procederá al cálculo de la función de correlación de dos puntos vía transformada de Hankel del espectro de potencias, y se seguirá el mismo procedimiento anterior, es decir, calcular la función de estructura, su gradiente log-log, etc.

Una vez realizadas todas las mediciones para cada uno de los métodos, se encontrarán los resultados mostrados en la Tabla 0.

Métodos	SBC	GBC	GSBC	HSBC	HGSBC	CF	PS
Dimensión							
Fractal	$1.01 \pm 0.08$	$1.12 \pm 0.08$	$2.42 \pm 0.11$	$1.78 \pm 0.04$	$1.40 \pm 0.11$	$2.25 \pm 0.03$	$2.22 \pm 0.05$
Media							

**Tabla 0:** Resultados obtenidos para la dimensión fractal media en un intervalo de 300 a 2400 [ $Mpc h^{-1}$ ], para cada uno de los métodos. El error se ha estimado como la desviación estándar de las medidas. Además, las siglas se refieren a: SBC- *box-counting* estándar, GBC- *box-counting* deslizante, GSBC-*box-counting* en escala de grises, HSBC- *box-counting* estándar con Healpix, HGSBC- *box-counting* en escala de grises con Healpix, CF - función de correlación, PS- espectro de potencias

Encontrándose que, para todos los métodos, se obtiene un carácter homogéneo de la dimensión fractal, aunque no se puede asegurar un único valor, ya que difieren para cada método. Además, en la literatura se encuentra que en estas escalas  $D \approx 3$ , luego el método que más se acerca sería el que emplea escala de grises, aunque aún estaría lejos de esa cifra.

Se concluirá que se prueba la homogeneidad de las estructuras a gran escala en los intervalos analizados, aunque no con el mismo valor de la dimensión fractal dado por la literatura. A su vez, se propondrá un estudio más detallado para poder localizar la franja en la que se pasa de un universo no homogéneo a uno homogéneo y, también, se propondrá ahondar más en las relaciones entre la geometría fractal y la cosmología siguiendo los pasos de diversos estudios. Así como también se propondrá aumentar la escala en la que se han analizado los datos con el fin de tratar de obtener un resultado más acorde con el mostrado en la literatura.

# Index

<b>I</b>	<b>Introduction</b>	<b>5</b>
<b>II</b>	<b>Data</b>	<b>7</b>
<b>III</b>	<b>Methodology</b>	<b>8</b>
III.1	<b>Box-Counting algorithms</b> . . . . .	9
III.1.1	Standard Box-Counting Method . . . . .	10
III.1.2	Gliding Box-Counting Method . . . . .	11
III.1.3	Gray-scale Box-Counting Method . . . . .	11
III.1.4	Adding Healpix to Standard Box-Counting and Gray-scale Box-Counting Methods . . . . .	12
III.2	<b>Two-point correlation function</b> . . . . .	14
III.3	<b>Power Spectrum</b> . . . . .	15
III.4	<b>Linear fit and error</b> . . . . .	17
<b>IV</b>	<b>Results and Discussion</b>	<b>18</b>
IV.1	Standard, Gliding, and Gray-Scale Box-Counting Methods . . . . .	19
IV.2	Healpix Standard Box-Counting and Gray-scale Box-Counting Methods	22
IV.3	<b>Two-point correlation function</b> . . . . .	23
IV.4	<b>Power Spectrum</b> . . . . .	25
IV.5	<b>Summary</b> . . . . .	27
<b>V</b>	<b>Conclusions and future work</b>	<b>28</b>
<b>VI</b>	<b>References</b>	<b>30</b>

## I Introduction

---

### Resumen

La forma en la que se distribuyen y agrupan los distintos objetos en el universo llama al ser humano a preguntarse por qué. A veces parece que se siguen formas aleatorias, sin sentido aparente. Pero detrás de las intrincadas formas de las costas de los fiordos, de las hojas de los helechos, o incluso de la espuma cuántica, hay un objeto matemático que puede imitar sus formas, los fractales.

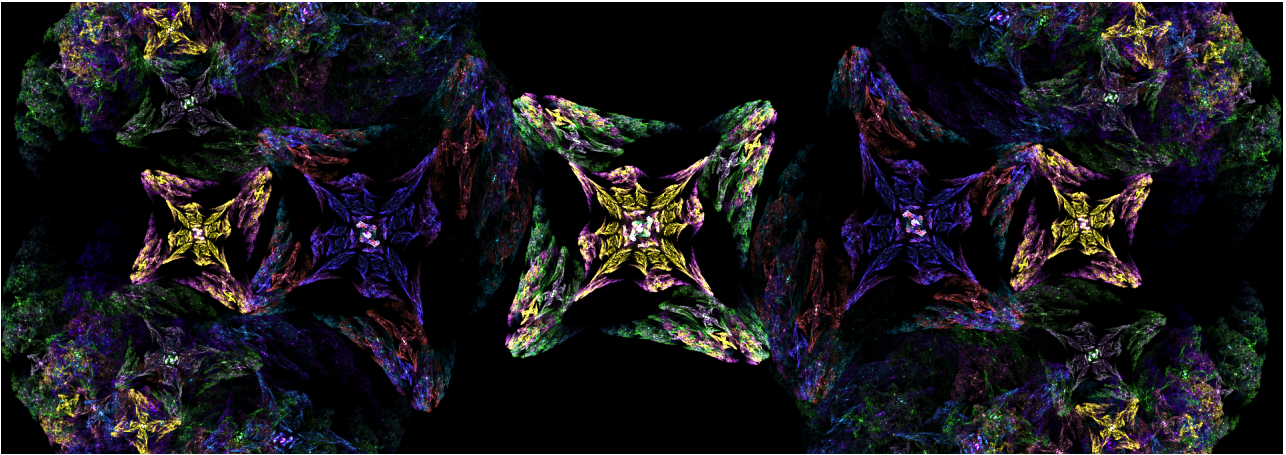
Benoît Mandelbrot, un matemático del siglo XX, empezó a desarrollar las matemáticas de estos objetos. Gracias a ello, hoy se pueden aplicar los fractales a un campo con mucho potencial, la astrofísica. Se pueden utilizar para medir el agrupamiento de los cúmulos de galaxias a lo largo del cielo utilizando una de las propiedades de los fractales, su dimensión. Precisamente eso es lo que se estudiará en este trabajo, se analizará cómo varía la dimensión fractal a lo largo de varios intervalos de distancia comóvil y se comprobará si los resultados concuerdan con los que se observan en la literatura.

---

There have always been structures with a capricious shape, with a shape that humans could not explain how it was formed. It is what is known today as chaos. But the concept of chaos did not begin as another way of saying disorder. According to Greek mythology, Chaos was the primal emptiness that preceded all creation. It was with the passing of the years that more interpretations emerged, such as that of Pherecydes of Syros, which interpreted it as something without concrete form.

It was not until well into the nineteenth century that the possibility of studying chaos began to be considered. One of the first to be interested in what would later become the Chaos Theory was Poincaré, who was interested in studying the problem of the three bodies. With the advance of computation, it became increasingly easier to study these types of problems, which are normally based on recursion relationships. It was during the mid-20th century that Lorenz discovered what we know today, precisely, as the Lorenz attractors. And in parallel, a Polish mathematician named Benoît Mandelbrot would begin the study of fractals, endowing them with solid mathematics. Later, in 1982, he published the work “The Fractal Geometry of Nature” [1], a work where it is shown that fractals can be found in many natural structures and processes, and that they can be studied using fractals.

But, what is a fractal? It can be defined as an object with an irregular shape that repeats at different size scales, is said to have self-similarity. An example of a fractal can be seen in Fig.1, where it has been generated with the Apophysis program. One of the properties of fractals, their dimension, is the most studied for its application when determining the clustering of a sample. One of the first people that defined a precise definition of a dimension that can be applied to a fractal was Hausdorff, around 1918. Its definition could be understood as a measure of the complexity of a set of data, or of its roughness [2]. A mathematical definition is provided in the volume *Fractal Geometry: Mathematical Foundations and Applications* [3]. The problem encountered with this Hausdorff dimension is that sometimes it cannot be easily calculated using computation. For this, there are other definitions that may be more useful. Those are the dimensions calculated through the two-point correlation function or via box-counting algorithms [4].



**Figure 1:** Artistic representation of a fractal.

It is from then on that fractals begin to be applied to the study of nature in various fields, from the study of rocky surfaces [5] to their use in medical applications [6]. Some studies were also started on the fractal dimension of the large-scale structures of the universe. Various authors such as Pietronero, Yuriy Baryshev and Pekka Teerikorpi, to give some examples, have taken advantage of this last characteristic of the fractal dimension to study the relationship between the properties of large-scale structures and fractals [7].

Pietronero et al. [8] explore the relationship of powers between the mass of stellar objects, the distance and the fractal dimension, finding that “Fractal distributions are characterized by long-range power-law correlations”. In another paper by R.Murdzek [4], this power law is ratified and it can be seen, in a concise way, how to calculate the fractal dimension from box-counting methods and through the two-point correlation function. The relationship between the power spectrum and the two-point correlation function can also be seen in a work by István Szapudi et al. [9].

Studies by J.Einasto et al. further analysed how the fractal dimension varies with respect to the distance of the data sample. An important conclusion is drawn in this study: the fractal dimension of the cosmic web is a function of distances. Where the fractal dimension evolves from  $\approx 0$  to  $\approx 2$  at medium separations, i.e, from filaments to sheets, and it reach  $D \approx 3$  for large distances, that is, it behaves like a random distribution [10]. It has also been shown that the fractal dimension is high when clustering is present, but low when voids are found [11], which implies that it can be used as a cluster detector. It can also be used to test the homogeneity of the universe at different scales [12] where it is found, in this particular paper, that homogeneity is reached for scales greater than  $60\text{-}70 [Mpc h^{-1}]$ .

Following in the wake of the work by J.Einasto et al. [10], in this work the fractal dimension at long scales will be measured, at different ranges of comoving distance, and it will be verified if a homogeneous distribution of the fractal dimension is achieved with a value close to  $D \approx 3$ . This will be done by different methods such as box-counting algorithms, from the correlation function and from the power spectrum, where the methodology followed and the data used will be exposed throughout this work.



---

## II Data

---

### Resumen

Los objetivos de este trabajo, que son medir dimensiones fractales de estructuras a gran escala mediante distintos mecanismos (como los algoritmos de box-counting o a través de la función de correlación de dos puntos y del espectro de potencias), hace que sea necesario disponer de una base de datos que aporte información acerca de la distribución en el cielo de los objetos que conforman estas estructuras.

Para ello se emplearán datos provistos por el conjunto de datos de grupos de galaxias BOSS (*Baryon Oscillation Spectroscopic Survey*) que forma parte del SDSS (*Sloan Digital Sky Survey*). En concreto, se trabajará con los datos conjuntos de los dos algoritmos de selección de BOSS, para el casquete galáctico norte: LOWZ, que selecciona objetos hasta un *redshift* tal que  $z \approx 0.4$  y CMASS, que selecciona objetos en un rango de  $0.4 < z < 0.7$ . Este conjunto de ambos se denomina CMASSLOWZTOT\_North, y proporciona datos de unos 953255 objetos.

---

To carry out the measurements required by this work, certain information about the objects will be required, for example, information about the position in the sky, as well as its redshift and its comoving number density for the object's redshift, among other magnitudes. Data provided by the Sloan Digital Sky Survey (SDSS) will be used.

The SDSS is a project whose objective is to observe and take images of the sky, and is named after the Alfred P. Sloan Foundation, which contributed greatly to the project. The goal was to map as much of the sky as possible, providing enough information for cosmological scale studies, as well as other astrophysics studies. It has an immense amount of spectra, color images, images in a range of wavelength bands, and redshift data. All this for about 3 million objects.

For this, a telescope located at the Apache Point Observatory in New Mexico, United States is used. This telescope, called SDSS, is about 2.5 m and it is a Ritchey-Chrétien with f/5. For some measurements, other telescopes have also been used, such as the Irénée du Pont Telescope at Las Campanas Observatory, Chile, which is also a Ritchey-Chrétien 2.5-m f/7.5. This was used for the fourth phase of the SDSS, which consisted of taking measurements from the southern hemisphere. Another telescope used has been the NMSU 1-Meter Telescope, also located at Apache Point Observatory, which is also a Ritchey-Chrétien.

All data collected by this project is publicly available on its website. The observations and measurements began in 2000 and continue even today, where Data Release 16 is being developed. In fact, SDSS -V measurements were started in the summer of 2020, with the intention of mapping the Milky Way and Black Holes.

The data that has been used throughout this work were collected in Data Release 12 (DR12) of the SDSS, which contains all the information collected by the observations of the SDSS through July 2014. This data release maps a third of the celestial vault in different filters that are detailed in the following table.



Filters:	<i>u</i>	<i>g</i>	<i>r</i>	<i>i</i>	<i>z</i>
Wavelengths (Å)	3551	4686	6165	7481	8931
Magnitude limits	22.0	22.2	22.2	21.3	20.5

**Table 1:** Effective wavelength and magnitude limit for the different filters used in the DR12.

The data used in this work is provided by the BOSS galaxy cluster dataset which is part of the SDSS. Specifically, the joint data of the two BOSS selection algorithms, for the northern galactic cap, will be used: LOWZ, which selects objects up to a redshift such that  $z \approx 0.4$  and CMASS, which selects objects in a range of  $0.4 < z < 0.7$ . This set of both is called CMASSLOWZTOT\_North, and it will provide data for about 953255 objects [13], it also provides a catalog of random objects that have the same properties as the objects in the original catalog, but without structure and with a much larger number, in order to perform statistical treatments.

To carry out the measurements, the fiducial cosmology of DR12 BOSS has been chosen [13], which corresponds to cosmological parameters such that:  $\Omega_m = 0.31$ ,  $\Omega_\Lambda = 0.69$ ,  $H_0 = 67.6 [km s^{-1} Mpc^{-1}]$ . That is, assuming flat  $\Lambda$ CDM cosmology. Where  $\Omega_m$  and  $\Omega_\Lambda$  refer to the density of matter in the universe and the density of dark energy respectively, and  $H_0$  is the Hubble constant. Throughout this work  $h$  will appear, which is  $h = H_0/(100[km s^{-1} Mpc^{-1}]) = 0.676$ .

### III Methodology

---

#### Resumen

Como se ha comentado previamente, el objetivo de este trabajo es determinar la dimensión fractal de estructuras a gran escala a distintos intervalos de distancia comóvil, comprobando que se sigue la tendencia dada por la literatura. La determinación de la dimensión fractal se analizará utilizando diversos métodos como: algoritmos de *box-counting*, empleando la función de correlación de dos puntos y a través del espectro de potencias.

Para realizar el análisis con los algoritmos de *box-counting* se precisarán mapas del cielo, tanto en 2D como en 3D, cuando se ejecute el análisis utilizando escalas de grises. Para emplear la función de correlación de dos puntos se necesitará conocer la posición en el cielo de cada galaxia, ya que se puede medir por conteo de pares. Luego, a la hora de calcular el espectro de potencias, se necesitará conocer también la posición en el cielo de las galaxias, ya que el espectro de potencias es la función de correlación de dos puntos en el espacio de Fourier.

Los datos acerca de las posiciones de las galaxias que conforman la estructura a gran escala se obtendrán del Data Release 12 (DR12) de Sloan Digital Sky Survey (SDSS). Una vez se tengan los resultados de la dimensión fractal con cada uno de los métodos, se procederá a estudiar cómo varían los resultados para cada uno de ellos.

---

In this work, the fractal dimension of large-scale structures has been calculated using various methods. In the first place, box-counting algorithms have been used, both in 2D and 3D. Then

the fractal dimension was also calculated using the two point correlation function, and finally by using the power spectrum. For all the methods used, programming codes have been generated with Python, where the *nbodykit* module has been used to work with the two-point correlation function and with the power spectrum. The theoretical background for the calculation using these three methods will now be explained.

But first, it will be explained how the measurement of comoving distances has been carried out. Comoving distance is a cosmological measure that indicates the distance between two points by means of a constant cosmological time curve. And for its measurement the expression given by David W. Hogg. [14] will be used

$$d_c(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{(1+z')^3 \Omega_m + \Omega_\Lambda}}, \quad (1)$$

where  $z$  is the redshift and  $c = 299792458[m/s]$  is the speed of light in vacuum.

### III.1 Box-Counting algorithms

To calculate the dimension of a fractal, imagine a shape lying on a grid, and count how many boxes are required to cover the shape. The box-counting dimension is calculated by seeing how this number changes as the grid becomes finer and finer. The difficulty of calculating a fractal dimension from the definition given by Hausdorff is quite noticeable. To solve this problem and to obtain an estimate of the fractal dimension, the idea is that, measuring at a scale  $\delta$  on an F curve, and being  $M_\delta(F)$  the measurement of a pair of divisor elements of size  $\delta$  that traverse F, an estimator of the fractal dimension can be constructed from the following power law [3]:

$$M_\delta(F) \sim c\delta^{-D}, \quad (2)$$

where, by taking logarithms and making the size of the divisors tend to zero, the estimation of the dimension is obtained as

$$D = \lim_{\delta \rightarrow 0} -\frac{\log M_\delta(F)}{\log \delta}. \quad (3)$$

The idea of the box-counting algorithms is to study how a number of points  $N_0$  is distributed in a certain survey. To do this, the data survey is divided into  $N$  squares of  $\delta$  side that cover the entire map. From this, one can define the box-counting dimension as

$$D_{BC} = \lim_{\delta \rightarrow 0} \frac{\log N(\delta)}{\log \left( \frac{1}{\delta} \right)}. \quad (4)$$

Then the fractal dimension can be inferred by plotting  $\log N(\delta)$  against  $\log(1/\delta)$  where, when a least squares adjustment is computed, the slope of the fit line will correspond to the fractal dimension [4]. The key point will be to determine how the number of boxes  $N(\delta)$  is calculated.

To implement this method, it is first necessary to create a sky map where the position of all objects is included. The comoving distance is included, as the grey level of each pixel, to use it during the treatment with one of the methods (Fig.2).

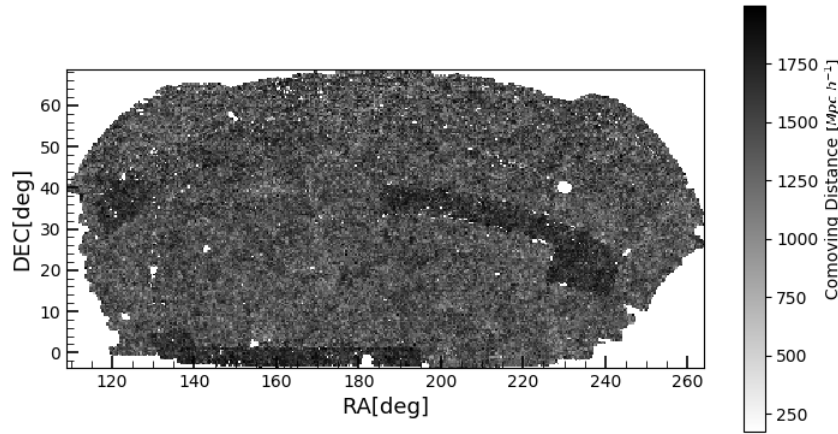


Figure 2: A sky map with the comoving distance as grayscale can be seen.

As a result, a map like that shown in the image will be divided into different numbers of boxes  $N(\delta)$ . The key will be in how  $N(\delta)$  is calculated. For this, three different methods will be used. An example of a linear fit used to measure the fractal dimension is shown in Fig.3.

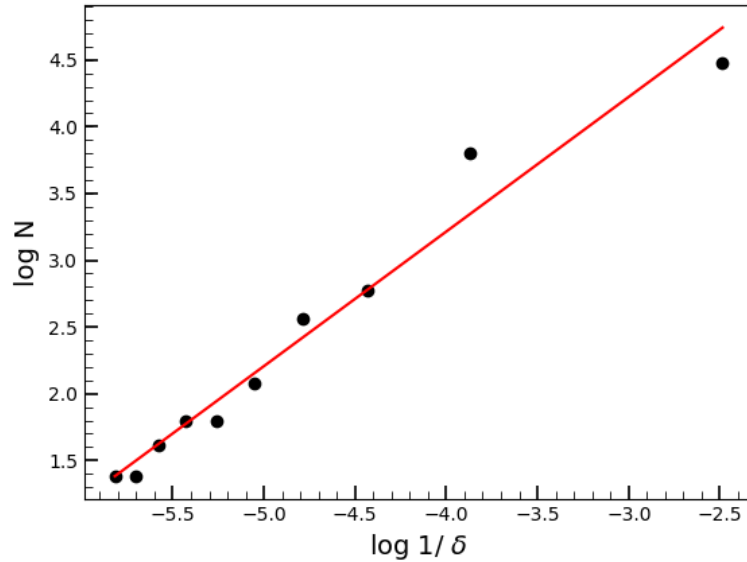
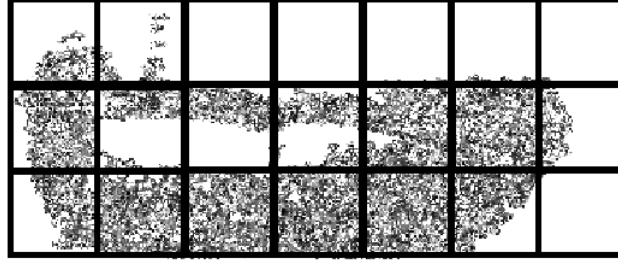


Figure 3: A linear fit of  $\log N(\delta)$  against  $\log(1/\delta)$  is shown .

### III.1.1 Standard Box-Counting Method

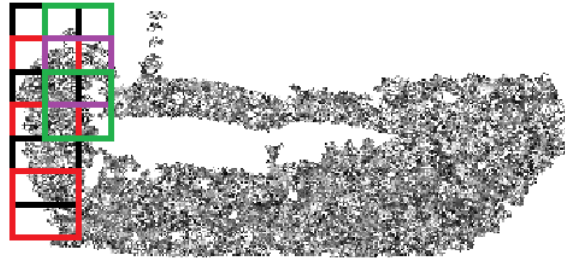
This method is based in partitioning the image into adjacent squares of size  $\delta$  and, for each object of the image that falls within one of the squares, one unit is added to the number of boxes  $N(\delta)$  (see Fig.4). Then, the procedure will be to choose several sizes of  $\delta$ , from highest to lowest or viceversa, and record the number of total boxes, that contain any element, for each of the sizes. In our case, sizes range from  $\delta_i$  to  $\delta_n$  where  $\delta_{i+1} = \delta_i + n$  will be chosen,  $n$  is a parameter to be set and it will depend on the size of the boxes. Once this is done, linear adjustments can be carried out, plotting  $\log N(\delta)$  against  $\log(1/\delta)$ , and calculating the fractal dimension from the slope.



**Figure 4:** Example of dividing a sky map into boxes for the standard box-counting method. The size of the boxes has been exaggerated for better appreciation.

### III.1.2 Gliding Box-Counting Method

The process is similar to the standard box-counting, except for one difference. This time, the boxes will not be adjacent to each other, but will overlap. In this work it has been considered that the boxes overlap on the middle of the previous box (see Fig.5). Once the number of boxes has been calculated using this procedure, the fractal dimension can be computed in the same way as in the case of standard box-counting, i.e. by plotting  $\log N(\delta)$  against  $\log(1/\delta)$  and calculating the slope.



**Figure 5:** Example of dividing a sky map into boxes for the gliding box-counting method. The size of the boxes has been exaggerated for better appreciation.

### III.1.3 Gray-scale Box-Counting Method

This method differs slightly from the previous ones and is based on the work of Chinmaya Panigrahy et al. [15]. The idea continues to be to divide the image in a grid, in the same way as in the standard box-counting method, the image will be divided into adjacent meshes of size  $\delta \times \delta$ . But this time, the size of  $\delta$  has a restriction. It must be in the interval  $(2, P/2)$ , for an image of  $P \times P$  pixels.

Now, within each cell of size  $\delta \times \delta$  there will be a certain number of cubes of size  $\delta \times \delta \times h$ , being  $h$  the height of each cube, so that this number of cubes counts the level of gray in that cell  $\delta \times \delta$  (see Fig.6). In the work of Chinmaya Panigrahy et al. [15] a value of  $h$  that best fits the measures is proposed, and it is given by the expression:

$$h = \begin{cases} \frac{\delta}{\lceil \log_2(P) \times (\log_2(P) - 2) \rceil} & \text{if } \log_2(P) \geq 3 \\ \frac{\delta}{\lceil \log_2(P) \rceil} & \text{Otherwise} \end{cases} . \quad (5)$$

So, now that the value of  $h$  has been calculated, the number of cubes of size  $\delta \times \delta \times h$  needed to account for the gray level can be inferred. Then the number of cubes for a mesh  $(i, j)$  is defined as

$$n(i, j) = \lceil \frac{g_{max}}{h} \rceil - \lceil \frac{g_{min}}{h} \rceil + 1, \tag{6}$$

where  $\lceil \cdot \rceil$  symbolizes the ceiling function,  $g_{max}$  and  $g_{min}$  are the values of the gray scale, maximum and minimum respectively, present in the corresponding cell  $\delta \times \delta$ . Finally, the total number of boxes  $N(\delta)$  can be calculated following

$$N(\delta) = \sum_{i,j} n(i, j). \tag{7}$$

And then, the fractal dimension can be computed by applying Eq.(4).

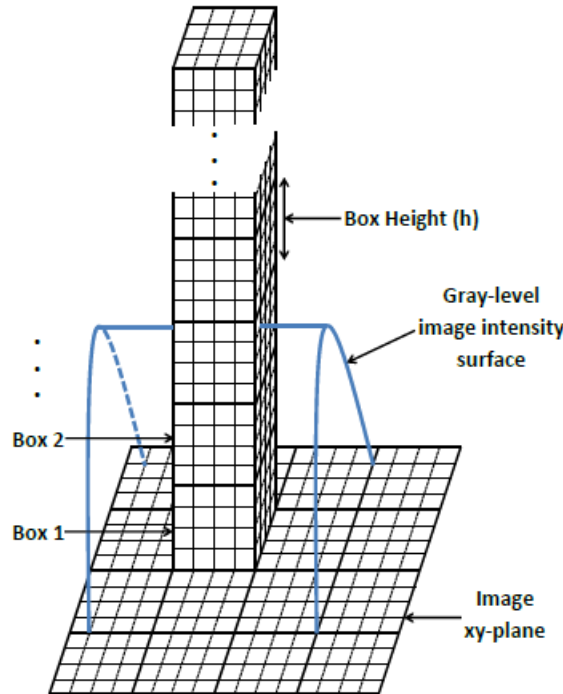
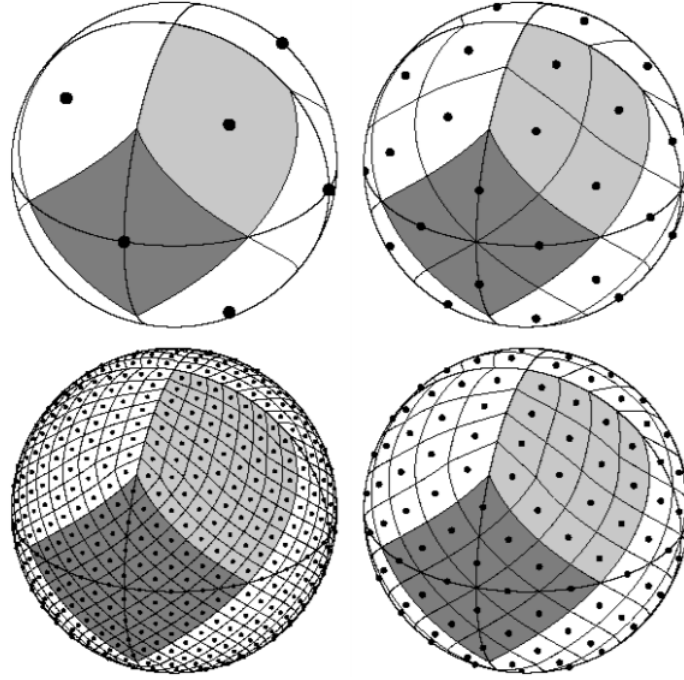


Figure 6: Image taken from [15] which shows a schematic of the number of cubes  $n(i, j)$ .

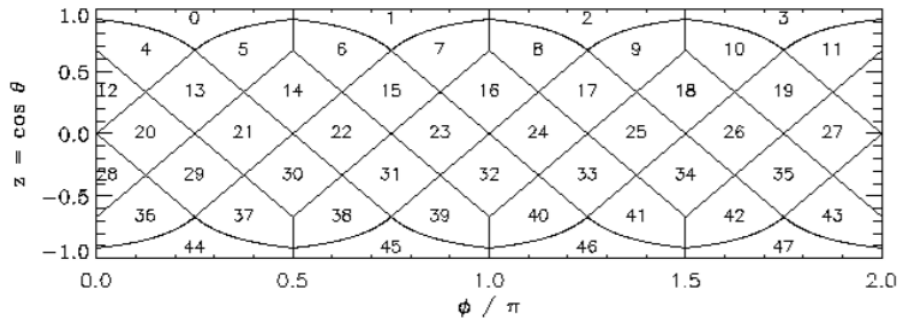
### III.1.4 Adding Healpix to Standard Box-Counting and Gray-scale Box-Counting Methods

In order to carry out a more realistic analysis of the sky, that is, try to better reflect the geometry of the observations, it is proposed below to use the Python Healpix module [16]. In essence, this module allows us to make a map of the sky and divide it into pixels as shown in Fig.7. This allows each object to be assigned a pixel on the map according to its right ascension and declination. Then, these maps will be made for different ranges of comoving distance, taking different spherical shells where it will be assumed that the galaxies are distributed on that surface [17].



**Figure 7:** Example of dividing a sky map into pixels using Healpix, taken from [15].

Once the pixel is assigned to each galaxy, a histogram of the pixels can be made to see how many objects fall within each one of them. This allows the box-counting method to be implemented, being  $N_{side}$  the resolution of the grid, or in other words, the number of divisions along one of the sides of the reference pixel. This quantity is taken as the  $\delta$  parameter used in Eq.(4). Where the total number of pixels given by Healpix is such that:  $N_{pix} = 12N_{side}^2$ . The way these pixels are distributed, the way they are listed, follows what is called in Healpix as RING scheme. This scheme distributes the pixels from north to south pole along each iso-latitude ring.



**Figure 8:** Cylindrical projection of the RING scheme, being  $\theta$  the co-latitude, and  $\phi$  the longitude, taken from [15].

Following the same scheme, Healpix will be incorporated into the gray-scale box-counting treatment, where several approximations will have to be taken into account. First of all, it must be taken into account that the gray-scale box-counting method is designed to work on a flat 2D image, on an image of  $P \times P$  pixels. Therefore, the first approximation is the fact that  $P \approx \sqrt{N_{pix}}$ . And, as it was approximated in the standard method, for the squares  $\delta \times \delta$ ,  $\delta = N_{side}$  will be taken.

On the other hand, for the gray scale, the comoving distances of the objects will be used, where these distances will be scaled to the gray scale that follows values from 0 to 255. In this way, it is no longer assumed that in spherical shells the galaxies are found on the shell surface, but



rather the roughness due to the different values of the comoving distance of each object is taken into account.

### III.2 Two-point correlation function

The two-point correlation function is a commonly used tool in astrophysics to study clustering phenomena, and in general to study the distribution of galaxies on a large scale. To be more precise, what is going to be used throughout this work is called a spatial two-point correlation function. Which depends on the distance between pairs of points and which will be denoted as  $\xi(r)$ . This function indicates the excess probability of, having a number  $N$  of points, to find a pair separated by a certain distance  $r$ , with respect to the probability of finding said pair at a distance  $r$  in an unstructured sample. Hence the importance of the random catalogs. If the average density of observed galaxies is  $\bar{n}$ , the fluctuations in the density field are defined as [18]

$$\delta(\mathbf{x}) = (n(\mathbf{x}) - \bar{n})/\bar{n}, \quad (8)$$

being  $\mathbf{x}$  the distance from an observer to an object. Besides, the two-point correlation function is defined as follow, where the average is carried out on a statistical ensemble [19]

$$\xi(\mathbf{x}, \mathbf{x}') \equiv \langle \delta(\mathbf{x})\delta(\mathbf{x}') \rangle, \quad (9)$$

where  $\mathbf{r} = \mathbf{x}' - \mathbf{x}$ . Here a new assumption is made, that has not been taken into account when treating with the box-counting algorithms, and it is the fact that the correlation function of two points is isotropic, that is,  $\xi(\mathbf{r}) = \xi(r)$ , where  $r$  is the distance between two points, the magnitude of the vector  $\mathbf{r}$ . Then, to calculate the fractal dimension from the correlation function, the structure function has to be calculated first [10], which is defined as

$$g(r) = 1 + \xi(r). \quad (10)$$

Then, its log-log gradient, the gradient function, should be calculated. And, from this, the fractal dimension can be computed as

$$\gamma(r) = \frac{d \log g(r)}{d \log r} \rightarrow D(r) = 3 + \gamma(r). \quad (11)$$

Now that how to calculate the fractal dimension from the correlation function is known, the correlation function can be measured.

The method that will be used to measure the two-point correlation function will be through calculating the average number of neighbors at a given distance. To perform this pair count, several tools will be needed. Galaxy surveys are required first, along with their random catalogs, because as will be discussed below, the pair count requires these quantities. The random catalogs follow the same geometry as the survey and the objects it contains follow the same characteristics, but the objects are placed randomly, so there is no structure.

As shown in the literature [20] the estimators for the two-point correlation function most used in the literature are

$$\text{Peebles \& Hauser: } \xi(r) = \frac{DD}{RR} - 1 \quad (12)$$

$$\text{Hewett: } \xi(r) = \frac{DD - DR}{RR} \quad (13)$$



$$\text{David \& Peebles: } \xi(r) = \frac{DD}{DR} - 1 \quad (14)$$

$$\text{Hamilton : } \xi(r) = \frac{DD \times RR}{DR^2} - 1 \quad (15)$$

$$\text{Landy \& Szalay: } \xi(r) = \frac{DD - 2DR + RR}{RR}, \quad (16)$$

where the terms that appear in each estimator refer to the following quantities:

- $DD(r) = \frac{dd(r)}{N(N-1)/2}$
- $RR(r) = \frac{rr(r)}{N_r(N_r-1)/2}$
- $DR(r) = \frac{dr(r)}{NN_r}$ .

Here  $N$  is the number of survey objects and  $N_r$  is the number of randoms, where the quantities  $dd(r)$ ,  $rr(r)$ ,  $dr(r)$  correspond to the pair spacing distributions in the catalog, in the random catalog, and between the object catalog and the random catalog, respectively.

In this work, the Landy & Szalay estimator will be used to perform the pair count. This decision is due to the fact that it is, as mentioned in [20], the estimator that has the least amount of bias. An optimized estimator is also shown in this paper, but for the purpose of this study the calculations made with the Landy & Szalay estimator is considered sufficient.

### III.3 Power Spectrum

The last method that will be used to measure the fractal dimension requires the measurement of the power spectrum, it is the correlation function of two-points in Fourier space. To explain that, from Eq.(8),  $\delta(\mathbf{k})$  is introduced, which is  $\delta(\mathbf{x})$  in Fourier space, where for the Fourier transform the convention of adding the negative sign in the exponent is taken [21]

$$\delta(\mathbf{k}) = \int \delta(\mathbf{x}) e^{-i\mathbf{k}\mathbf{x}} d\mathbf{x} \quad ; \quad \delta(\mathbf{x}) = \int \delta(\mathbf{k}) e^{-i\mathbf{k}\mathbf{x}} \frac{d\mathbf{k}}{(2\pi)^3}. \quad (17)$$

Then, if another point is taken into account with its respective  $\delta(\mathbf{k}')$  there is the following relation

$$\delta(\mathbf{k})\delta^*(\mathbf{k}') = \int d\mathbf{x} \delta(\mathbf{x}) e^{-i\mathbf{k}\mathbf{x}} \int d\mathbf{x}' \delta(\mathbf{x}') e^{i\mathbf{k}'\mathbf{x}'}, \quad (18)$$

where, if an average is performed in the statistical ensemble

$$\begin{aligned} \langle \delta(\mathbf{k})\delta^*(\mathbf{k}') \rangle &= \int d\mathbf{x} e^{-i\mathbf{k}\mathbf{x}} \int d\mathbf{x}' e^{i\mathbf{k}'\mathbf{x}'} \langle \delta(\mathbf{x})\delta(\mathbf{x}') \rangle = \int d\mathbf{x} e^{-i\mathbf{k}\mathbf{x}} \int d\mathbf{r} e^{i\mathbf{k}'\cdot(\mathbf{r}+\mathbf{x})} \langle \delta(\mathbf{x})\delta(\mathbf{r} + \mathbf{x}) \rangle = \\ &= \int d\mathbf{x} e^{-i(\mathbf{k}-\mathbf{k}')\mathbf{x}} \int d\mathbf{r} e^{i\mathbf{k}\mathbf{r}} \xi(\mathbf{r}) = (2\pi)^3 \delta_D^3(\mathbf{k} - \mathbf{k}') P(\mathbf{k}). \end{aligned} \quad (19)$$

So, the Fourier transform of the correlation function is given the name of power spectrum

$$P(\mathbf{k}) = \int d\mathbf{r} e^{i\mathbf{k}\mathbf{r}} \xi(\mathbf{r}). \quad (20)$$

As in the case of the correlation function, the isotropy will be assumed, so that  $P(\mathbf{k}) = P(k)$ . Therefore, the following expression remains, as it appears in the literature[18]

$$\langle \delta(\mathbf{k})\delta^*(\mathbf{k}') \rangle = (2\pi)^3 \delta^3(\mathbf{k} - \mathbf{k}') P(k). \quad (21)$$

In order to measure the power spectrum, the estimator given in the literature by Feldman et al. (hereafter FKP)[22, 23] will be used. First a weighted density fluctuation field is defined as

$$F(\mathbf{r}) = w_{FKP}(\mathbf{r})[n_c(\mathbf{r}) - \alpha n_r(\mathbf{r})], \quad (22)$$

where  $n_c(\mathbf{r})$  and  $n_r(\mathbf{r})$  are the number densities for the real galaxy catalog and the random catalog, respectively.  $\alpha$  is a renormalization factor because the random objects have a number much higher than the galaxy catalog [23], this parameter is introduced in order to force the fluctuations to have an average equal to zero,  $\langle F(\mathbf{r}) \rangle = 0$ , so

$$\alpha = \frac{\int w_{FKP}(\mathbf{r}) n_c(\mathbf{r}) d^3r}{\int w_{FKP}(\mathbf{r}) n_r(\mathbf{r}) d^3r}. \quad (23)$$

The *FKP* weights come from Feldman, Kaiser, and Peacock [24], and correspond to a set of statistical weights, written as Eq.(24) that minimize the variance in the power spectrum. That is, a weighted density field is generated, Eq.(22) in such a way that its variance is minimal

$$w_{FKP}(\mathbf{r}) = \frac{1}{1 + n_c(\mathbf{r})P_0}. \quad (24)$$

In the derivation followed in the literature [22],  $P(\mathbf{k})$  appears instead of  $P_0$ . But since an estimate of  $P(\mathbf{k})$  is required, this term in Eq.(24) is replaced by  $P_0$ , which is an estimate of the expected amplitude of the measured power spectrum.

Then, from this, the estimator for the power spectrum is constructed as [25]

$$P_{est}(\mathbf{k}) = \frac{1}{A} \int \frac{d\Omega_k}{4\pi} |F(\mathbf{k})|^2, \quad (25)$$

$A$  being a normalization factor written as:

$$A = \int d\mathbf{r} [n_c(\mathbf{r})w_{FKP}(\mathbf{r})]^2. \quad (26)$$

This estimator does not provide the power spectrum directly, but rather provides the power spectrum plus the constant shot noise component due to statistical fluctuations in the measurement, see Eq.(27). This shot noise is subtracted by following the expression in Eq.(28) [24]

$$P_{est}(\mathbf{k}) \approx P(\mathbf{k}) + P_{Shot} \quad (27)$$

$$P_{Shot} = (1 + \alpha) \int d\mathbf{r} n_c(\mathbf{r}) w_{FKP}^2(\mathbf{r}), \quad (28)$$

where it is remembered that the isotropy condition  $P(\mathbf{k}) = P(k)$  will be assumed.  $\Omega_k$  in Eq.(25) is the solid angle in Fourier space. Once the shot noise is subtracted, and averaged over spherical shells, the bin-averaged power spectrum is obtained,  $P_{est}(k)$ . The bin-averaged power spectrum is the convolution product between the cluster power spectrum with  $|W(\mathbf{k})|^2$ ,

which is the square of the Fourier transform of the window function [23], given by

$$W(\mathbf{k}) = \int n_c(\mathbf{r})w_{FKP}(\mathbf{r})e^{-i\mathbf{k}\cdot\mathbf{r}}d\mathbf{r}, \quad (29)$$

where

$$P_{est}(\mathbf{k}) = \int \frac{d\mathbf{k}'}{(2\pi)^3}P(\mathbf{k}')|W(\mathbf{k}-\mathbf{k}')|^2 + (1+\alpha)\frac{\int d\mathbf{r}n_c(\mathbf{r})w_{FKP}^2(\mathbf{r})}{\int d\mathbf{r}n_c^2(\mathbf{r})w_{FKP}^2(\mathbf{r})}. \quad (30)$$

It has been chosen to use *nbodykit*, a program written in Python language, to carry out the necessary calculations. The tool that *nbodykit* provides for the above measurements is *ConvolvedFFTPower* [24].

When working with *nbodykit* the first thing to do is to group the available data into two well-differentiated catalogs, one corresponding to the selected data and the other to the random data. In these catalogs the positions of the objects in Cartesian coordinates will be included together with their corresponding average density of objects. Once this is done, the next step is to create a mesh from the input data, the input data being the previous catalogs. To build this mesh, *nbodykit* provides the tool *to\_mesh*, where it will indicate the number of cells into which the mesh will be divided. Here it can also be selected what are known as the mass assignment. These functions are in charge of determining which cell each object will contribute to, and there are a number of them. First it is the *Nearest Grid Point*, its mechanism is simple and it is that the object contributes to the closest cell. Then it is the *Cloud In Cell* or the *Triangular Shaped Cloud*. A number of cells corresponding to  $N_{cell} = 256^3$  and the *TSC* function has been used as was done in [23].

Therefore, the correlation function can be inferred from this, to carry out the necessary operations for the calculation of the fractal dimension. For this, the Eq.(20) is inverted [21], getting that

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}}P(\mathbf{k}). \quad (31)$$

To carry out this calculation, a Hankel transform will be used, which is the isotropic version of a Fourier transform [21]

$$\xi(r) = \int_0^\infty \frac{k^3 P(k)}{(2\pi)^2} j_0(kr) \frac{dk}{k}, \quad (32)$$

where  $j_0$  is the Bessel function of the first kind, which is written as  $j_0(x) = \sin(x)/x$ .

It must be taken into account that this case deals with a continuous spectrum, and what has been measured during this work is a discrete spectrum. Therefore, an interpolation of the discrete spectrum was performed to be able to apply Eq.(32). It should be clear from this point on, although it will be emphasized later during the discussions, that this procedure is not correct, it is not rigorous. Although, as will be seen later, the results obtained will not differ much from those obtained for the method that measures the correlation function by counting pairs.

### III.4 Linear fit and error

All measurements carried out have an associated error, either due to the characteristics of the instruments used for their measurement or introduced by the different methods applied. In order to give an estimate of the error that may be occurring when measuring the fractal

dimension, the error associated with the slope of the linear fit will be used. As indicated in the paper [10].

This decision is based on the fact that to propagate the errors in the box-counting algorithm, different methods are required, such as perturbative treatments, and that the propagation of errors associated with the correlation function and the power spectrum can be complicated. Therefore, this error will be considered in principle, discussing later whether in some cases it has been a sufficient consideration or not.

The way in which the values of the fractal dimension will be calculated is through a Vandermonde matrix, since the Python command `numpy.polyfit` will be used [26]. In the case of this work, where it will fit a line, polynomials of degree 1 will be used to adjust the measurements, such that

$$P_1(x_n) = y_n = a_1x_n + a_0, \quad (33)$$

where  $n$  is the total number of measurements and 1 the degree of the polynomial. From here the Vandermonde matrix is constructed as

$$V = \begin{bmatrix} 1 & x_0 \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}. \quad (34)$$

Then, being  $\mathbf{a} = (a_0, a_1)^T$  and  $\mathbf{y} = (y_0, y_1, \dots, y_n)^T$ , the coefficients can be calculated such that  $V\mathbf{a} = \mathbf{y}$  [27]. And, being  $\mathbf{m}(x_n) = (1, x_n)^T$  the interpolating polynomial can be calculated as  $P_1(x_n) = \mathbf{a}^T \mathbf{m}(x_n)$ .

Once the coefficients of the polynomial have been calculated, the square root of the corresponding element of the diagonal of the covariance matrix will be taken as the error in the slope of the fit.

## IV Results and Discussion

---

### Resumen

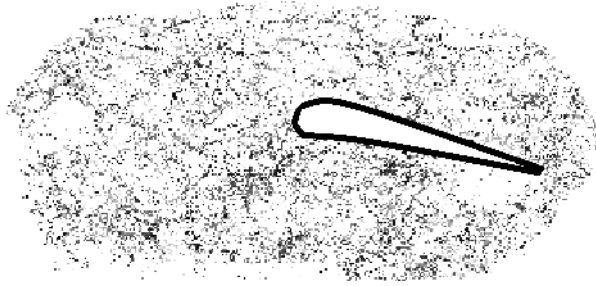
En esta sección se expondrán los distintos resultados obtenidos para cada uno de los métodos anteriores, primeramente se analizarán los resultados de cada conjunto de métodos por separado, y luego al finalizar se hará una valoración general de todos ellos.

Se obtendrá que, para todos los métodos empleados, se observa un comportamiento compatible con una constante, es decir, homogéneo. No obstante, para cada método esta homogeneidad es alcanzada en valores distintos y, por otra parte, alejados del valor para la dimensión fractal dado por la literatura a estas escalas.

Se discutirá sobre las posibles razones por las cuales los valores para la dimensión fractal difieren de un método a otro y, a su vez, se analizarán los posibles errores que hayan podido ser minusvalorados, de acuerdo a que no se ha realizado propagación de los mismos.

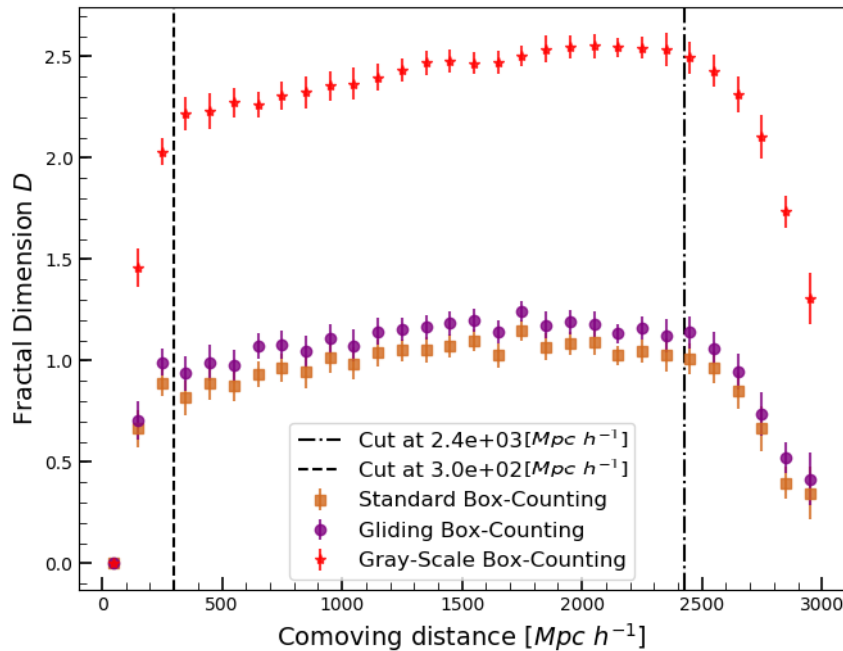
## IV.1 Standard, Gliding, and Gray-Scale Box-Counting Methods

Once the data to be used is available and the programs for the different methods have been built, it is possible to proceed to the measurement of the fractal dimension for each case. First, the sky maps in 2D are obtained for the simplest cases, standard, gliding and gray-scale box-counting, these maps are the same as those discussed in Fig.2. In these maps, one can sometimes see the voids in some areas of the sky at a glance, as the one that can be seen in Fig.9.



**Figure 9:** Distribution of the sky for a cut-off of 500 to 600 [ $Mpc h^{-1}$ ] in the sample of *galaxyDR12v5CMASSLOWZTOTNorth*.

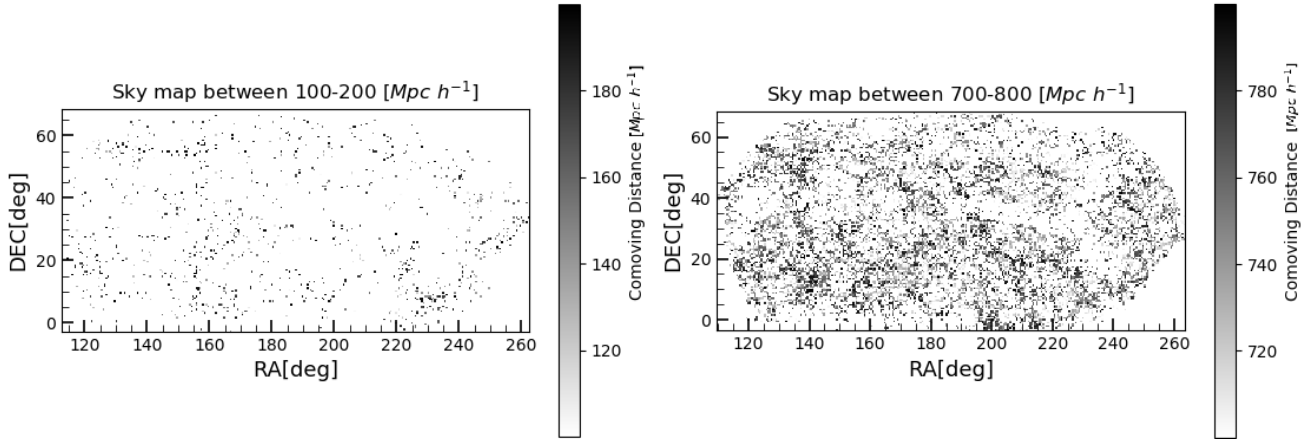
For the measurement of the fractal dimension in these cases, a total of 30 sky maps have been made, objects that were in an interval from 0 to 3000 [ $Mpc h^{-1}$ ] have been selected, where each map represents a 100 [ $Mpc h^{-1}$ ] interval. Once the maps are made, the fractal dimension is measured, obtaining the results shown in Fig.10.



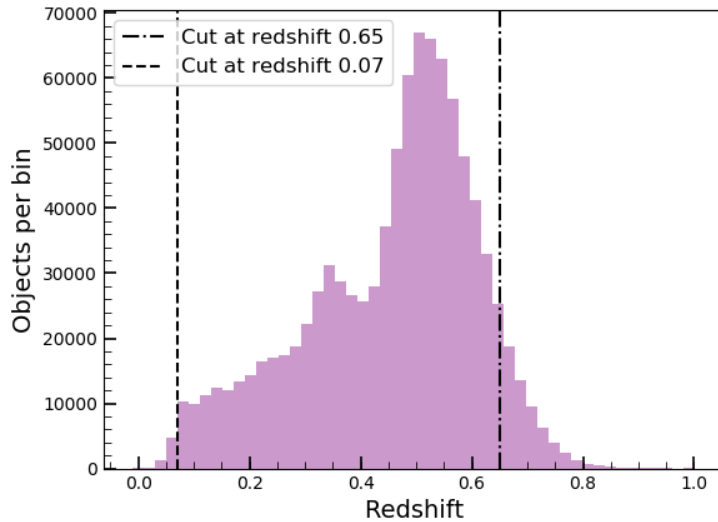
**Figure 10:** Results for the fractal dimension using the standard, gliding and gray-scale box-counting methods. The axis of comoving distance represents the midpoint of each 100 [ $Mpc h^{-1}$ ] bin.

In light of the results obtained for the simplest methods, a clear difference is seen between those that only handle the 2D image and the one that uses the gray scale for the calculation. Despite the difference in the order of magnitude, it is observed that the distribution follows the same trend for the different methods. A certain deviation of the behavior can also be observed

for the fractal dimension below and above certain thresholds, this is possibly due to a bias related to the number of objects present on each map. Considering the redshift histogram of the objects, it can be seen that it is precisely near these cuts where a decrease of the objects per bin is located. An example of one of the maps for that area and the histogram can be seen in Fig.11 and Fig.12.



**Figure 11:** On the left, a map of the sky in the range of 100 to 200 [ $Mpc h^{-1}$ ] and on the right in the range of 700 to 800 [ $Mpc h^{-1}$ ]. It is clear that in the latter there is a greater number of objects.



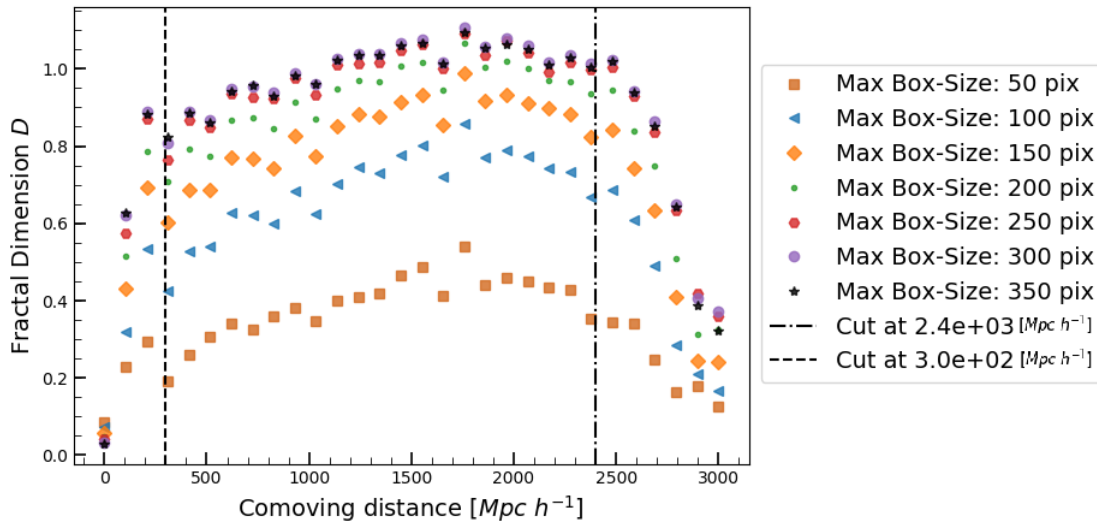
**Figure 12:** Histogram of the redshift values of the objects in the catalog, showing the cuts from which the number of objects changes significantly.

Then, it must be taken into account the fact that objects do not appear in the limits of the catalog does not mean that there are no objects in those areas in the universe. Then the fractal dimension values given for those zones are strongly skewed by this fact.

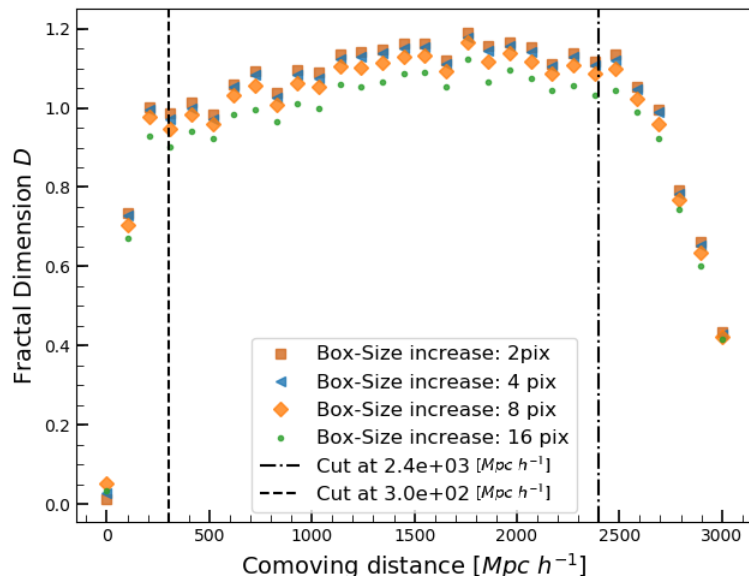
On the other hand, the difference of the fractal dimension values given by each method will now be discussed. The difference between the standard and the gliding box-counting is not particularly significant, but nevertheless, when looking at the results obtained by the gray-scale box-counting, there is a very noticeable difference.

The reason for this difference may lie in the number of boxes taken into account when counting, and in the sizes chosen for them. The method that uses the fewest boxes is the standard

box-counting method, followed by gliding box-counting and ending with the gray-scale box-counting, that is, it follows the distribution seen in Fig.10. Besides, for the standard box-counting method, the dependence of the results with respect to the number of boxes and their size was studied (Fig.13).



**Figure 13:** Graph for the standard box-count method using different numbers of boxes with different maximum limits of the box sizes.



**Figure 14:** Graph for the standard box counting method, where the limit of the maximum box size is 504 pixels. The fractal dimension measurement is displayed for several different box size increments, that is, the higher the increment, the fewer the total number of boxes.

Several conclusions can be drawn from the previous figures. It is seen that the result strongly depends on the maximum size that the boxes used to analyze the image can reach, and that it has a dependency, in turn, with both the number of boxes and the increase in their size (Fig.14). These results seem to reinforce the hypothesis why there are such notable differences in the results given for the standard, gliding, and gray-scale box-counting methods seen in Fig.10. That is, that the variation was given by the number of boxes used by each method.



## IV.2 Healpix Standard Box-Counting and Gray-scale Box-Counting Methods

The results obtained when implementing Healpix to the standard and gray-scale box-counting methods will now be displayed. First, for the standard method, it was decided to perform the analysis on spherical shells with a radius of  $100 [Mpc h^{-1}]$  each, where galaxies are assumed to be on the surface of that shell [17]. On the other hand, in the case of the gray-scale method, it was also decided to do the treatment in spherical shells of  $100 [Mpc h^{-1}]$ . But in this case, when taking into account the comoving distance of each object, it is no longer considered that the galaxies are exactly on the surface of the spherical shell, as commented in section III.1.4. The obtained results can be seen in the following figure.

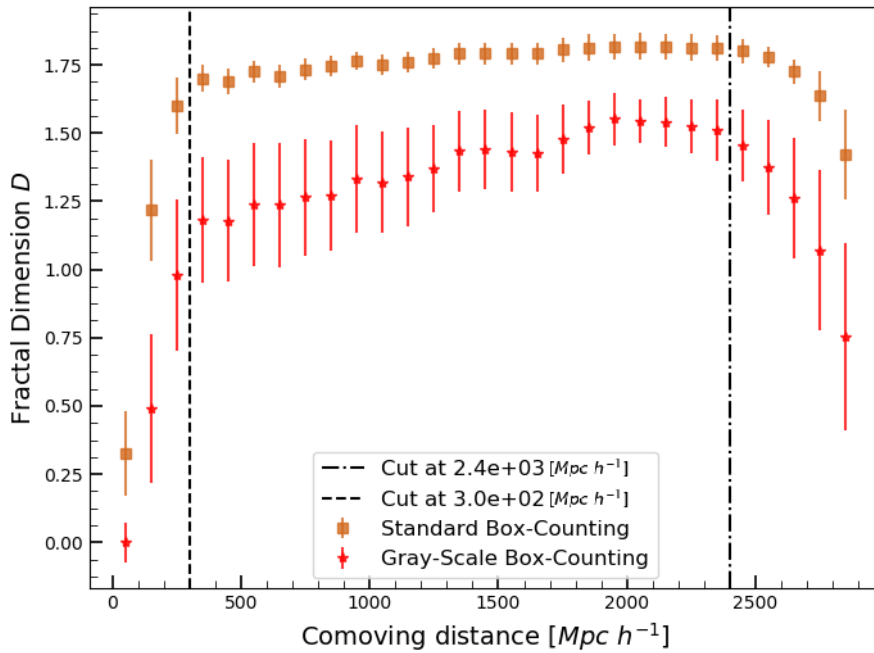


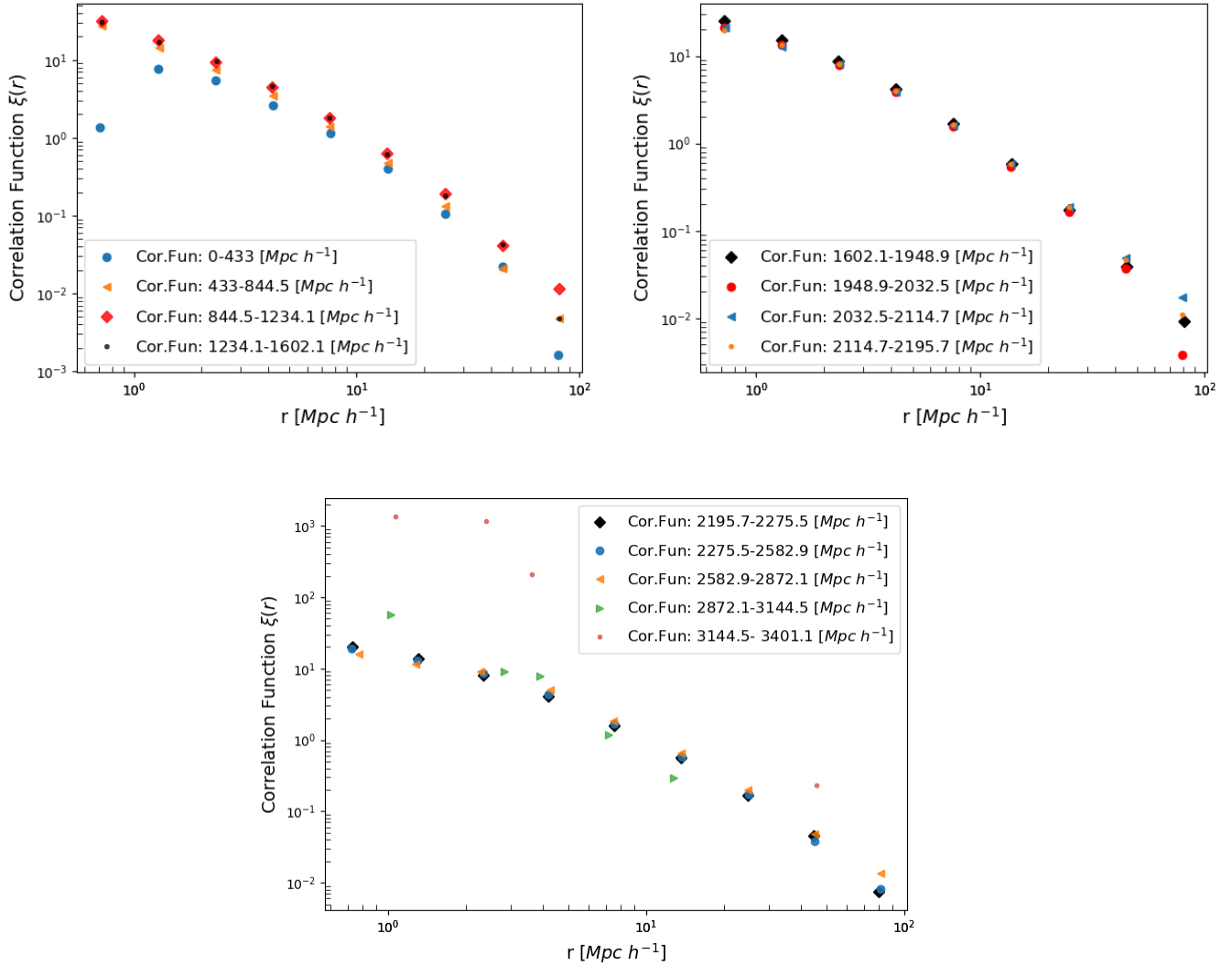
Figure 15: Graph for standard and gray-scale box-counting methods using Healpix.

Again a difference is seen in the obtained results, but the same trend is followed in both cases. Unlike in the previous cases, this time the standard method provides higher values than the gray-scale method for the fractal dimension. This can happen due to the fact that in this case the fractal dimension has been calculated for spherical shells, where when doing it by the simple method we are grouping the galaxies of  $100$  in  $100 [Mpc h^{-1}]$  in a shell. In other words, the fractal dimension is being overestimated by grouping objects on the shell surface. On the other hand, when applying the grayscale method, must be taken into account that the objects are not exactly distributed along the spherical shell.

But, apart from this, another aspect must be taken into account. And it is the fact that the method used for gray scales is specifically designed to work with 2D images, and in this case spherical shells are being used. Then, as seen in Fig.6, the cubes used to reach the gray level should not be cubes, but should take into account the curvature of the surface being worked on. This reason can be the main problem when trying to implement Healpix to the gray-scale method, which would mean that it now provides a worse result than the standard method and, in addition, with larger error bars, which shows that the adjustments are being made to measures that do not have a very linear behavior.

### IV.3 Two-point correlation function

The results obtained for the correlation function will now be presented as well as the fractal dimension measured from them.



**Figure 16:** Correlation function for different comoving distance intervals. The x-axis symbolizes the separation bin edges between pairs.

From the results obtained for the different correlation functions, one result can be highlighted. And it is the fact that the correlation functions measured for the intervals below  $300 [Mpc h^{-1}]$  and above  $2400 [Mpc h^{-1}]$  present a different correlation from the rest. It is recalled that in previous sections a cut was made from these figures because from those limits there was a deficiency of objects, here the consequences of this fact are observed once again, and the choice of said cuts is ratified as correct. Then, with this confirmation, it can be assumed that the results obtained outside these limits should not have any consideration. The results obtained for the data that are within the limits seem to have similar values to those observed in the literature [10].

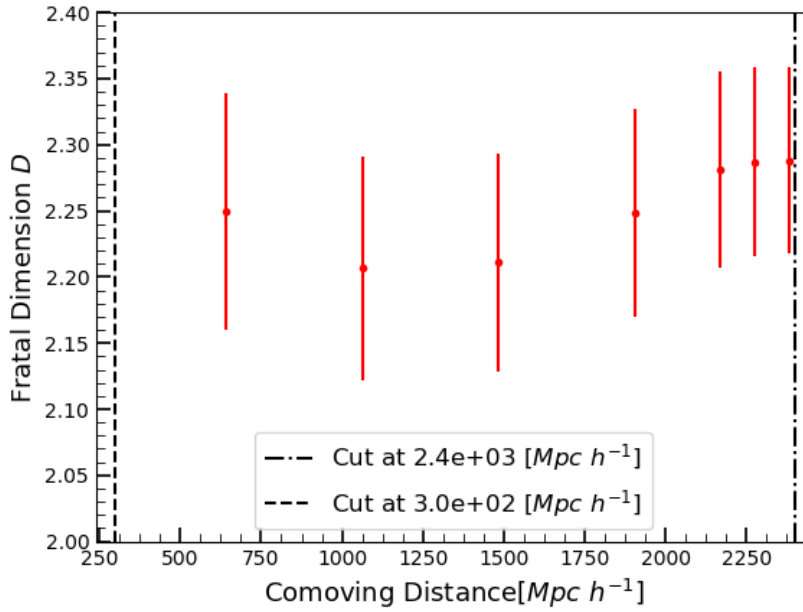


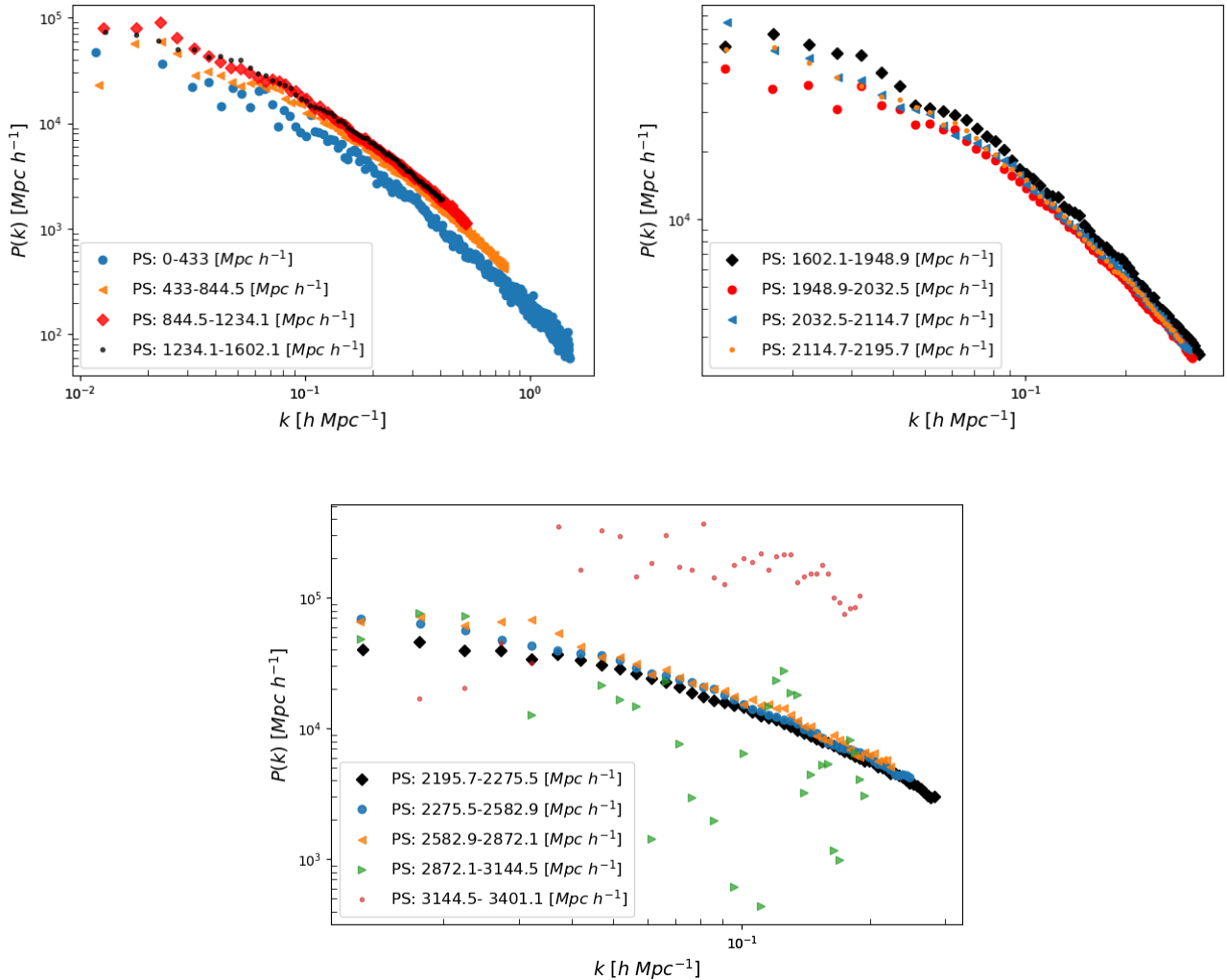
Figure 17: Fractal dimension measured from the correlation function.

Now, if the results obtained for the fractal dimension are analyzed using the correlation function, it can be observed that there is a difference in the values obtained with respect to the box-counting methods (except gray-scale method), although these methods also had notable differences among themselves. It is observed that the error bars obtained together with the results are compatible with a constant, that is, it tends to a homogeneous value of the fractal dimension [10]. Then, despite the fact that a completely different method from the one used for the box-counting algorithms has been used, a homogeneous trend has been obtained in the studied range, as has been observed for the cases of box-counting. However, although the values obtained are compatible with a constant, they do not reach the value of the fractal dimension given by the literature.

It must be taken into account that the correlation function has been measured up to a maximum separation between pairs of  $r \approx 80 [Mpc h^{-1}]$ , and the acoustic peak for the data used is reached close to  $\sim 150[Mpc h^{-1}]$  [13]. That is, it has been measured on a relatively small scale, and that can have consequences on the results obtained for the fractal dimension, since the pairs are being counted with fewer objects.

## IV.4 Power Spectrum

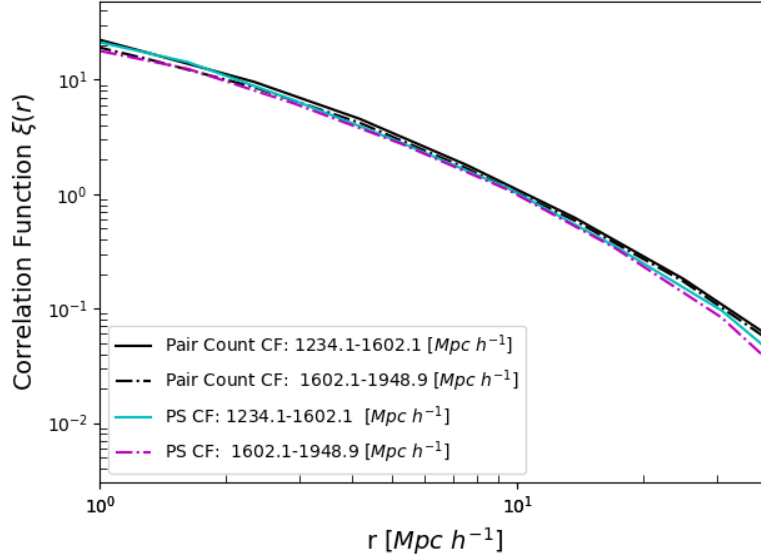
The results obtained for the power spectrum will now be presented in the same interval in which the results for the correlation function were shown.



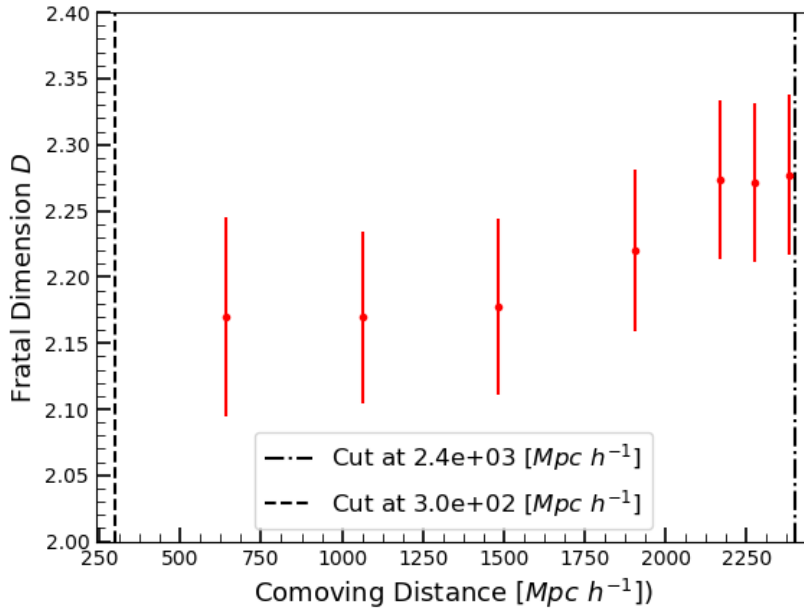
**Figure 18:** Results obtained for the power spectrum at different ranges of comoving distance. Where the x-axis symbolizes the separation bin edges, along the  $k$  space, between pairs.

Once again, the strange behavior of the results can be appreciated when they leave the imposed cuts of 300 and 2400  $[Mpc h^{-1}]$ , where beyond 2800  $[Mpc h^{-1}]$  the results are meaningless. The results obtained for the power spectrum have turned out to be of the order of magnitude of those obtained according to [28]. Therefore, within the range, it seems that the results obtained are in accordance.

The results obtained for the fractal dimension will now be presented. First, the relation Eq.(20) has to be inverted, and then the structure function will be calculated and a linear adjustment will be made as indicated in section III.2. In Fig.19 a comparison between two correlation functions measured in the same interval is shown, both for counting pairs and from the power spectrum. Where it can be seen that, in both procedures, they follow a similar trend. The results for the fractal dimension are shown in Fig.20.



**Figure 19:** Results obtained for the correlation function at different intervals from pair counts and the power spectrum.



**Figure 20:** Fractal dimension measured from the correlation function obtained from the power spectrum.

In the case of the fractal dimension, it is observed that it follows the same behavior and have similar values as when the correlation function measured directly was used. However, the same discussion is followed as in the previous section, and it is that a homogeneous trend has been obtained in the range studied, as has been observed for the cases of box-counting. However, it must be taken into account that the method used has not been entirely rigorous, although similar values have been obtained. And it is that a discrete power spectrum has been measured, while the method followed is used for continuous spectra. Although an attempt has been made to solve this problem, the method should not be used, as there may be problems when transforming.

## IV.5 Summary

For each of the proposed methods various values of the fractal dimension have been obtained, although all the methods seemed to reach a homogeneous state. However, there is a significant difference between the values for which homogeneity is reached in this work and the values that can be found in the literature [10, 29, 30], where it is detailed that homogeneity is reached for values greater than 60-70 [ $Mpc h^{-1}$ ] and that, in addition, the value that the fractal dimension reaches for said homogeneous state is of the order of  $D \approx 3$ . On the other hand in works like [17] it can be observed some values of the fractal dimension similar to those obtained during this work, although, the value obtained for the rest of the publications will be taken as a reference. Then it can be said that the homogeneity at large scales has been proven, but due to the disparity of the results for the different methods, it is not possible to give a single value of what fractal dimension is reached when this occurs.

The following table will show the different mean fractal dimensions measured for each method.

Methods	SBC	GBC	GSBC	HSBC	HGSBC	CF	PS
Mean Fractal Dimension	$1.01 \pm 0.08$	$1.12 \pm 0.08$	$2.42 \pm 0.11$	$1.78 \pm 0.04$	$1.40 \pm 0.11$	$2.25 \pm 0.03$	$2.22 \pm 0.05$

**Table 2:** Results of the mean fractal dimension obtained for the different methods for the intervals from 300 to 2400 [ $Mpc h^{-1}$ ] are shown here. The error has been given as the standard deviation of the measurements.

In addition, the acronyms refer to: SBC-Simple Box-Counting, GBC- Gliding Box-Counting, GSBC- Gray-Scale Box-Counting, HSBC- Healpix Simple Box-Counting, HGSBC- Healpix Gray-Scale Box-Counting, CF- Correlation Function, PS- Power Spectrum

From here it can be seen that the method that is closest to the homogeneous value of  $D \approx 3$  is the gray-scale box-counting method. Although it must be remembered that it seems that the box counting methods are biased by the number of boxes measured, and that the methods used using the correlation function may have a dependence on the scale used. Emphasize, once again, that the last method used was not completely rigorous, which casts doubt on the value obtained from it.

It should also be noted that the errors shown throughout the work referred to the error presented in the slope of the fit curve, and not to a propagation of errors. Therefore, in some cases, key errors may be omitted for the discussion of the results, leading to a discussion that is further from reality.

## V Conclusions and future work

---

### Resumen

A partir de los resultados mostrados en la sección anterior y su discusión, se puede inferir que no se ha logrado el objetivo de obtener un valor único y consistente de la dimensión fractal en los intervalos dados. No obstante, para una buena parte de los métodos, se ha logrado observar un cierto comportamiento homogéneo y, en uno de ellos, con un valor cercano al dado por la literatura. Se ha visto también que los resultados obtenidos para los métodos de *box-counting* mostraban una cierta dependencia con el número de cajas empleado para el análisis y que, para el caso de los métodos que involucran a la función de correlación, se ha podido minusvalorar el error cometido al no realizar propagación de errores. Se propone como trabajo futuro el continuar con esta línea de investigación, pero en rangos más pequeños, a fin de localizar el punto de transición desde la no-homogeneidad a la homogeneidad y, también, investigar la relación que tiene la geometría fractal con la *web* cósmica.

---

It is clear, in light of the results obtained in the previous section that can be seen in Table 2, that the objective of measuring a value of the fractal dimension for the established scales, that coincided with what was given in the literature [10, 29, 30], has not been reached. Although, it is true that for the methods, especially the one that uses gray scale, a homogeneous behavior has been observed, although a bit biased by the number of objects. Due to the dependence on the number of objects, it would be interesting to carry out all this analysis, either with a more complete sample, or from simulations where this type of bias could be avoided. A future work perspective would be to correctly implement Healpix to the box-counting method that uses gray scale, taking into account the curvature of the spherical shells in the treatment.

In the previous section, the dependence of the box-counting methods with the number of boxes and their size was analyzed, although this seems to introduce an important bias, it seems that at a point it converges to a certain value, as shown in the analysis of the standard box-counting, the problem is that it converges to a value quite different from the one given in the literature. However, it must be taken into account that the standard and gliding methods only study an image in 2D, where much information about the objects studied is lost, since the cosmology in these methods only intervenes when making the different maps to different distances. On the other hand, in the method that uses gray scale, a third component is studied, adding information to the method and, as it can be seen, bringing the result closer to the results of the literature.

Regarding the methods that involve the correlation function, both provide very similar results. However, as has been severely emphasized during the work, the method that applies power spectrum to correlation function was not entirely rigorous. Despite this, a homogeneous trend was obtained for both methods, although this result may have been altered due to the maximum scale used, where the separation between pairs could not exceed  $r \approx 80[Mpc h^{-1}]$ . This limitation is mainly due to the fact that the code used, being made in Python, consumed too much computing time, sometimes causing the computer to automatically stop the process. A future work perspective would be to make this code in C++, trying to optimize it, while also increasing the maximum scale. Trying to check if indeed the result has been altered by this fact. Another work perspective, related to scale, could be to study the fractal dimension in smaller areas of the sky, such as clusters or super-clusters, in order to analyze how matter is



distributed in them, as was done in the literature, for example [11].

Another aspect that has not been taken into account in this work is that it has been assumed that all the objects that are being studied are equal, and this does not have to be the case. As shown in [31] the fractal dimension varies according to the type of galaxy, an interesting study would be to study the distribution of galaxies, separating them by type, and seeing how they vary from one to another. In fact, it was suggested that if time allowed, this work would include a small analysis of the fractal dimension on a galactic scale, that is, of individual galaxies, finally this could not be carried out, but in the future, with the work done during this project, it could be carried out.

With this work various aspects of fractal geometry and its link to large-scale structures have been understood, in the future it is proposed to delve further into the relationship between the cosmological properties of the universe and the fractal dimension, as indicated in [32] and try to carry out multifractal analysis and see how the results vary for different cosmologies. Another option would be to work with different stellar objects such as quasars, as was done at work [30].

## VI References

- [1] Benoit B. Mandelbrot and John A. Wheeler. “The Fractal Geometry of Nature”. In: *American Journal of Physics* 51.3 (Mar. 1983), pp. 286–287. DOI: [10.1119/1.13295](https://doi.org/10.1119/1.13295). URL: <https://ui.adsabs.harvard.edu/abs/1983AmJPh..51..286M>.
- [2] Tilmann Gneiting, Hana Ševčíková, and Donald B. Percival. “Estimators of Fractal Dimension: Assessing the Roughness of Time Series and Spatial Data”. In: *arXiv e-prints* (Jan. 2011). Provided by the SAO/NASA Astrophysics Data System. arXiv: [1101.1444](https://arxiv.org/abs/1101.1444). URL: <https://ui.adsabs.harvard.edu/abs/2011arXiv1101.1444G>.
- [3] Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. eng. Hoboken: Wiley, 2013. ISBN: 1-119-94239-X.
- [4] R. Murdzek. “The Box-Counting Method in the Large Scale Structure of the universe”. In: (Aug. 2006).
- [5] T Ai et al. “Box-counting methods to directly estimate the fractal dimension of a rock surface”. eng. In: *Applied surface science* 314 (2014), pp. 610–621.
- [6] Ivana Konatar, Tomo Popovic, and Natasa Popovic. “Box-Counting Method in Python for Fractal Analysis of Biomedical Images”. eng. In: (2020), pp. 1–4.
- [7] Yuriy Baryshev and Pekka Teerikorpi. “Fractal Approach to Large-Scale Galaxy Distribution”. In: (2005).
- [8] F Sylos Labini, M Montuori, and L Pietronero. “Statistical analysis of the Perseus-Pisces redshift survey: spatial and luminosity properties”. eng. In: *Physica A* 230.3-4 (1996), pp. 336–358. ISSN: 0378-4371.
- [9] István Szapudi et al. “Fast Edge Corrected Measurement of the Two-Point Correlation Function and the Power Spectrum”. In: 631.1 (2005). ISSN: 0004637X.
- [10] J. Einasto et al. “Correlation function: biasing and fractal properties of the cosmic web”. In: 640, A47 (Aug. 2020), A47. DOI: [10.1051/0004-6361/202037683](https://doi.org/10.1051/0004-6361/202037683). arXiv: [2002.02813](https://arxiv.org/abs/2002.02813) [[astro-ph.CO](https://arxiv.org/abs/2002.02813)]. URL: <https://ui.adsabs.harvard.edu/abs/2020A&A...640A..47E>.
- [11] Vimal Raj et al. “Fractal study on Saraswati supercluster”. In: *Indian Journal of Physics* 93.11 (Nov. 2019), pp. 1385–1390. DOI: [10.1007/s12648-019-01400-2](https://doi.org/10.1007/s12648-019-01400-2). URL: <https://ui.adsabs.harvard.edu/abs/2019InJPh..93.1385R>.
- [12] Jaswant Yadav et al. “Testing homogeneity on large scales in the Sloan Digital Sky Survey Data Release One”. In: 364.2 (Dec. 2005), pp. 601–606. DOI: [10.1111/j.1365-2966.2005.09578.x](https://doi.org/10.1111/j.1365-2966.2005.09578.x). arXiv: [astro-ph/0504315](https://arxiv.org/abs/astro-ph/0504315) [[astro-ph](https://arxiv.org/abs/astro-ph/0504315)]. URL: <https://ui.adsabs.harvard.edu/abs/2005MNRAS.364..601Y>.
- [13] Shadab Alam et al. “The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample”. In: 470.3 (Sept. 2017), pp. 2617–2652. DOI: [10.1093/mnras/stx721](https://doi.org/10.1093/mnras/stx721). arXiv: [1607.03155](https://arxiv.org/abs/1607.03155) [[astro-ph.CO](https://arxiv.org/abs/1607.03155)]. URL: <https://ui.adsabs.harvard.edu/abs/2017MNRAS.470.2617A>.
- [14] David W. Hogg. “Distance measures in cosmology”. In: *arXiv e-prints*, [astro-ph/9905116](https://arxiv.org/abs/astro-ph/9905116) (May 1999), [astro-ph/9905116](https://arxiv.org/abs/astro-ph/9905116). arXiv: [astro-ph/9905116](https://arxiv.org/abs/astro-ph/9905116) [[astro-ph](https://arxiv.org/abs/astro-ph/9905116)]. URL: <https://ui.adsabs.harvard.edu/abs/1999astro.ph..5116H>.
- [15] Chinmaya Panigrahy et al. “An Approximated Box Height for Differential-Box-Counting Method to Estimate Fractal Dimensions of Gray-Scale Images”. eng. In: *Entropy (Basel, Switzerland)* 19.10 (2017), p. 534. ISSN: 1099-4300. URL: <https://doi.org/10.3390/entropy19100534>.

- [16] “The HEALPix Primer”. In: (June 2021). URL: <https://healpix.sourceforge.io/pdf/intro.pdf>.
- [17] Jenna Bergevin. “Fractal Analysis of the Universe”. In: Vol.1 (2012). URL: <https://via.library.depaul.edu/depaul-disc/vol1/iss1/12>.
- [18] Scott Dodelson and George Efstathiou. “Modern Cosmology”. In: *Physics Today* 57.7 (July 2004), pp. 60–61. DOI: [10.1063/1.1784308](https://doi.org/10.1063/1.1784308). URL: <https://ui.adsabs.harvard.edu/abs/2004PhT....57R..60D>.
- [19] “Clustering Statistics”. In: (Mar. 2014). URL: <https://ned.ipac.caltech.edu/level5/March14/Percival/Percival1.html>.
- [20] M. Vargas-Magaña et al. “An optimized correlation function estimator for galaxy surveys”. In: 554, A131 (June 2013), A131. DOI: [10.1051/0004-6361/201220790](https://doi.org/10.1051/0004-6361/201220790). arXiv: [1211.6211](https://arxiv.org/abs/1211.6211) [astro-ph.CO]. URL: <https://ui.adsabs.harvard.edu/abs/2013A&A...554A.131V>.
- [21] Hannu Kurki-Suonio. “Galaxy Survey Cosmology, part 1”. In: (2021).
- [22] Hume A. Feldman, Nick Kaiser, and John A. Peacock. “Power-Spectrum Analysis of Three-dimensional Redshift Surveys”. In: 426 (May 1994), p. 23. DOI: [10.1086/174036](https://doi.org/10.1086/174036). arXiv: [astro-ph/9304022](https://arxiv.org/abs/astro-ph/9304022) [astro-ph]. URL: <https://ui.adsabs.harvard.edu/abs/1994ApJ...426...23F>.
- [23] A. Balaguera-Antolinez et al. “The REFLEX II galaxy cluster survey: power spectrum analysis”. In: 413.1 (May 2011), pp. 386–400. DOI: [10.1111/j.1365-2966.2010.18143.x](https://doi.org/10.1111/j.1365-2966.2010.18143.x). arXiv: [1012.1322](https://arxiv.org/abs/1012.1322) [astro-ph.CO]. URL: <https://ui.adsabs.harvard.edu/abs/2011MNRAS.413..386B>.
- [24] Nick Hand et al. “nbodykit: An Open-source, Massively Parallel Toolkit for Large-scale Structure”. In: 156.4, 160 (Oct. 2018), p. 160. DOI: [10.3847/1538-3881/aadae0](https://doi.org/10.3847/1538-3881/aadae0). arXiv: [1712.05834](https://arxiv.org/abs/1712.05834) [astro-ph.IM]. URL: <https://ui.adsabs.harvard.edu/abs/2018AJ...156..160H>.
- [25] Nick Hand et al. “An optimal FFT-based anisotropic power spectrum estimator”. In: 2017.7, 002 (July 2017), p. 002. DOI: [10.1088/1475-7516/2017/07/002](https://doi.org/10.1088/1475-7516/2017/07/002). arXiv: [1704.02357](https://arxiv.org/abs/1704.02357) [astro-ph.CO]. URL: <https://ui.adsabs.harvard.edu/abs/2017JCAP...07..002H>.
- [26] “Numpy Library”. In: (June 2021). URL: <https://numpy.org/doc/stable/reference/generated/numpy.polyfit.html>.
- [27] W. Gander. “Change of basis in polynomial interpolation”. eng. In: *Numerical Linear Algebra with Applications* 12.8 (2005), pp. 769–778. ISSN: 1070-5325.
- [28] “Final BAO and RSD Measurements”. In: (2021). URL: <https://www.sdss.org/science/final-bao-and-rsd-measurements/>.
- [29] Marie-Noelle Celerier and Reuben Thieberger. “Fractal dimensions of the galaxy distribution varying by steps?” In: (2005).
- [30] Prakash Sarkar et al. “The many scales to cosmic homogeneity: Use of multiple tracers from the SDSS”. In: (2016).
- [31] Jorge de la Calleja, Elsa M. de la Calleja, and Hugo Jair Escalante. “Fractal dimension analysis for automatic morphological galaxy classification”. In: (2017).
- [32] J Gaité. “The Fractal Geometry of the Cosmic Web and Its Formation”. English. In: *Advances In Astronomy* 2019 (2019). ISSN: 1687-7969.