Simulating the quasar distribution in the Cosmic Web

by

Jose María Coloma Nadal

Master's thesis supervisors:

Francisco-Shu Kitaura Joyanes

Ginevra Favole

July 2023

Resumen

En este trabajo investigamos un modelo de sesgo que describe con precisión la distribución de quásares en el Universo. En particular, este modelo relaciona la distribución de materia oscura con la de los quásares y se basa en la clasificación de la red cósmica a diferentes escalas. La clasificación relativa a las escalas más grandes se hace mediante los autovalores del tensor de mareas, es decir, depende del potencial gravitatorio. Por otro lado, la clasification relativa a las escalas más pequeñas está relacionada con el campo de densidad de materia oscura. Bajo el propósito the encontrar los parámetros de nuestro modelo que mejor definan la distribución de quásares en la red cósmica, utilizamos un catálogo de referencia que se ha generado mediante el método SHAM. Este método asume que los quásares más luminosos (o masivos) se encuentran en los halos de materia oscura más masivos (con mayores velocidades circulares) de la simulación cosmológica de N cuerpos UNITSIM. La función de luminosidad de quásares está basada en observaciones del Sloan Digital Sky Survey. Este catálogo de referencia es una caja cúbica cuyo lado mide 1000 h^{-1} Mpc.

Para definir y estudiar el modelo de sesgo, necesitamos también un catálogo de imitación. Este se genera con una simulación de materia oscura que se pobla con quásares según la ley de sesgo. La simulación consiste en una caja cúbica de las mismas dimensiones que el catálogo de referencia y que, a partir de la evolución de las mismas condiciones inciales, evoluciona las partículas de materia oscura en base a teoría de perturbaciones lagrangianas. Más concretamente, hemos utilizado el método Augmented Lagrangian Perturbation Theory para definir el campo de desplazamiento de las poisiciones iniciales hasta el desplazamiento al rojo donde centramos el estudio ($z = 1.5$). De esta forma, conseguimos obtener la red cósmica de materia oscura a gran escala. El potencial gravitatorio nos permite clasificar la red cósmica en nodos, filamentos, paredes y vacios. Además, proponemos por primera vez una subclasificación de estas regiones en términos del campo de densidad que nos proporciona información sobre escalas más pequeñas. Una vez disponemos de este campo de materia oscura, estudiamos en detalle las componentes deterministas (no locales y no lineales) y estocásticas del modelo de sesgo. La descripción de los efectos no locales, es decir que la distribución de quásares en una región no dependa solo de su entorno local, puede hacerse mediante una ley de potencias. El exponente de esta ley crece linealmente con el sesgo del espectro de potencias de los quásares respecto al de la materia oscura. De forma genérica, las galaxias necesitan una masa mínima para su formación. Esto puede regularse imponiendo una función escalón que suprima las densidades muy bajas. En este caso, el sesgo del espectro de potencias crece conforme se suprimen más regiones. Por otro lado, se conoce que la distribución quásares guarda relación con las de las galaxias con líneas de emisión (ELGs), y estas no suelen encontrarse en los centros de las regiones de más altas densidad. Hemos modelado esto con una función exponencial negativa que suprime los picos de densidad más altos que un determinado límite. Al contrario que en el caso anterior, conforme más picos de densidad se supriman el sesgo será menor, es decir, la distribución de quásares será más similar a la de la materia. Comprobamos que la supresión de bajas y altas densidades puede modelarse de forma mucho más suave con funciones exponenciales dependientes de dos parámetros libres en lugar de uno. La normalización de la ley de sesgo controla la densidad de objetos. Comprobamos, como cabía esperar, que densidades más bajas dan lugar a sesgos mayores. Por último, esta ley contiene una componente estocástica asociada al hecho de que estamos trazando poblaciones discretas de galaxias (quásares) a partir una distribución continua de materia oscura. Se puede tener encuenta esto generando realizaciones poissonianas o siguiendo una distribución binomial negativa.

Una vez conocemos la repercusión de los diferentes posibles parámetros del modelo que define la distribución de quásares se procede a reproducir la estadística del catálogo de referencia. Encontramos que, una vez se ha aplicado la clasificación de la red cósmica de forma jerárquica en las diferentes escalas, el modelo de sesgo se puede reducir a una ley de potencias con exponente α, una densidad numérica $\langle n \rangle$ en cada región y una desviación poissoniana β que tiende a hacerse mínima. En este trabajo demostramos que la estadísica hasta tercer orden del catálogo de referencia se puede reproducir con gran precisión. Concretamente, reproducimos con una precisión menor al 5 por ciento el espectro de potencias consiguiendo, además, ajustes precisos tanto en la función de densidad de probabilidad (PDF) como en el biespectro. Se encuentran fuertes dependencias en paredes y vacíos con la distribución de materia oscura mientras que, para nodos y filamentos, las regiones de mayor densidad, la depencia se vuelve débil. Esto podría ilustrar el hecho de que el ratio de formación estelar de las galaxias, al interacionar con la red cósmica, disminuye progresivamente. En base a futuro trabajo, preveemos todavía mejoras en la precisión de estos ajustes por medio de clasificaciones de la red cósmica en base a términos adicionales como podrían ser aquellos que definen el sesgo no local a segundo orden. Por otra

parte, para estimar incertidumbres en los parámetros estamos preparando un ajuste por medio de cadenas de Márkov por medio de la función de densidad de probabilidad y del espectro de potencias. Los resultados que se han obtenido en este trabajo son potencionalmente relevantes para llevar a cabo análisis cosmológicos de los expeimentos DESI y EUCLID.

Abstract

This work investigates the effective assembly bias that accurately describes the distribution of quasars in the Universe. In particular, it relies on a cosmic web classification at different scales which is directly related to non-local bias: the long-range classification is based on the gravitational potential meanwhile the short-range is based on the density field. The reference quasar catalogue has been generated using a Sub-Halo-Abundance Matching (SHAM) scheme through the luminosity function of quasars at redshift 1.5, as obtained from the Sloan Digital Sky Survey, and applied to the dark matter only UNITSIM N-body simulation. This catalogue is a cubic box of 1000 h^{-1} Mpc side. The mock catalogue, in order to populate it with quasars and study the bias prescription, have been done by evolving the corresponding initial conditions, in a mesh of 256^3 cells, with Augmented Lagrangian Perturbation Theory to obtain the large-scale dark matter cosmic web. The gravitational potential lead us to first classified this cosmic web in knots, filaments, sheets and voids. We also propose fort he first time a subclassification based on the dark matter density field, and analogously to the gravitational potential, which provides information on the cosmic web at smaller scales. Then, we study in detail the non-linear and non-local deterministic and stochastic bias components to populate the simulated dark matter density field with quasars. We find that after applying our hierarchical cosmic web classification the bias can be reduced to a power-law bias with exponent α, the specific number density in each region $\langle n \rangle$ and a vanishing deviation from Poissonity β. We demonstrate that the statistics until third order of the reference catalogue can be reproduced with high fidelity. Specifically, we reproduce within a few percentage accuracy (< 5 per cent) the reference power spectrum with accurate univariate PDF and bispectra. We can foresee further improvements continuing the cosmic web classification with additional terms, such as one based on second order non-local bias, or based on the velocity shear. The results obatained in this thesis are potentially relevant to perform a cosmological analysis from galaxy redshift surveys such as DESI and EUCLID.

Contents

1 Introduction

To shed light on the formation of the large-scale structure (LSS) of the Universe (i.e. the observed distribution of galaxies and clusters of galaxies) is one of the foremost goals of modern cosmology. The Λ Cold Dark Matter (ΛCDM) cosmological model – where Λ stands for the cosmological constant and Cold Dark Matter represents the preferred candidate for dark matter particles – is nowadays the most accepted theory due to its explanations of the Cosmic Microwave Background (CMB), the LSS in the distribution of galaxies, the baryon acoustic oscillations (BAO) or the accelerated expansion of the Universe. In the ΛCDM model, structures are formed by hierarchical growth lead by the gravitational instability of the initial density fluctuations. The dark matter potential wells then accrete baryonic matter which forms galaxies. Despite its great achievements, the main components of the Universe, i.e. dark energy ($\sim 70\%$) and dark matter ($\sim 25\%$), are still not well understood.

Since galaxies and clusters can be observed out to cosmological distances, they constitute a very important source of information about the Universe. Nevertheless, understanding how the spatial distribution of the different tracers (i.e. halos, galaxies, clusters) is related to that of the underlying mass is still plenty of unknowns. What do we really know about galaxy formation? The Planck Collaboration results demonstrate that the initial distribution of density fluctuations in the Universe is nearly Gaussian (Planck Collaboration et al. (2020)). When dark matter gravitationally evolves from such an initial field, a very characteristic pattern with filaments of galaxies separated by under dense regions, named voids, appears both in observational surveys and simulations of the large scales. This complex large-scale pattern was dubbed *cosmic web* (van de Weygaert & Schaap (2008)). In the standard scenario of structure formation, gravity makes dark matter clumping into virialized halos. Baryonic matter, that interacts with dark matter only through gravity, falls into this potential wells, cools and condensates forming galaxies. This is the reason why galaxies are not faithful tracers of the underlying dark matter, and this effect is commonly known as *bias*. In addition, not all types of galaxies cluster in the same way, hence they trace the matter distribution differently. Understanding the relation between galaxies and dark matter becomes a key ingredient to interpret the observed LSS of the Universe, to enhance our knowledge of its composition, and the processes behind the initial seeds of structures.

The present work focuses on studying the interplay between quasars and dark matter. Quasars are commonly defined as active galactic nuclei (AGN) with bolometric luminoities $L_{bol} > 10^{38}$W. These are some of the most luminous astrophysical objects and are believed to be powered by accretion onto supermassive black holes at the centers of galaxies (Salpeter (1964)). From an observational perspective, such a high luminosity places quasars among the most important tracers but, unfortunately, our knowledge of their physical properties and distribution is poor. In fact, they are are extremely rare and present very low densities. Several observational studies have been done in this field: White et al. (2012) point out that quasars are very biased tracers of dark matter. Uchiyama et al. (2017) suggest that they do not occupy the most over dense regions of the Universe. When comparing observations with the SIMBA hydrodynamic simulation, Sorini et al. (2020) demonstrate that real quasars show similarities with the ones identified in SIMBA as the most rapidly-accreting black holes. These are just newborn features that quasars seem to present but a lot of work has to be done in order to understand how they distribute in the cosmic web and their relation with the underlying dark matter field.

To study the LSS, both theory and astronomical observations are needed. Past cosmological surveys, such as SDSS-III/BOSS, SDSS-IV/eBOSS or DES, have collected the light of the large-scale structure. Nowadays, most of the attention relies on the Dark Energy Spectrocopic Instrument (DESI; Levi et al. 2013), that seeks to map the LSS by observing about 40 million galaxy spectra over $14000\,\mathrm{deg}^2$ of sky to constrain the growth rate of structures and the Universe expansion history. DESI data demand mock catalogues to analyse theoretical models of galaxy clustering and covariance estimates. N-body simulations have been used to this end, but simulating such large cosmic volumes is costly in terms of computation. As an alternative, approximate gravity solvers are also commonly used with the advantage of being computationally fast. They provide large-scale dark matter fields on a mesh which can then be populated with different tracers following a bias prescription, a technique known as *bias mapping*.

The goal of this master thesis is twofold. Firstly, we review the Lagrangian Perturbation Theory (LPT) for its application, in particular the approximated gravity solver, to create a dark matter density field in a cubic volume of $1000\,h^{-3}\mathrm{Mpc}^3$ divided into 256^3 cells. Specifically, we rely on the Augmented LPT (ALPT)

method proposed by Kitaura & Heß (2013) to evolve initial positions to the desired redshift. The dark matter solver code, written by the supervisor of this work, F.-S. Kitaura, first classifies the cosmic web based of the gravitational potential in knots, filaments, sheets and voids (Kitaura et al. (2022)). Then, as a state-of-the-art method, it performs a subclassification of the above regions in terms of the density field. To carry out this study, we choose the redshift z=1.5 as a target, that is where the DESI quasar density peaks (see Figure 1 in Yè che et al. 2020). Secondly, we study in detail a non-linear and stochastic bias prescription to describe the relation between quasars and the underlying dark matter field. We analyse through the power spectrum the effect of different terms that can be added, here the advantage of such analytical bias prescriptions, based on the known properties of quasars. To constrain the model, we rely on a quasar galaxy mock catalogue provided by the supervisor of this work G. Favole who, through a *sub-halo abundance matching technique* (SHAM; Favole et al. 2016), matches the observational luminosity function to the halos and sub-halos of the N-body cosmological simulation UNITSIM.

The work is structured as follows: In section 2 we review the ALPT, which is behind the dark matter field we simulate, and we explain how the tidal field can be used to classify the cosmic web. In section 3 we describe the bias of dark matter tracers in terms of perturbation theory. In section 4 we build and analyse term by term an analytical model relating quasar and dark matter densities. A brief description of the reference catalogue based on SHAM is provided in section 5. The results of constraining the bias parameters with this catalogue are shown in section 6 and the conclusions are discussed in section 7. Finally, we include an appendix with the definition of some statistical tools frequently used in the LSS context.

2 Approximated gravity solver to simulate a dark matter density field

In this section we review the derivation of a dark matter fluid equation describing the cosmological structure formation in order to understand the LSS of the Universe. For the evolution of the mean density field we use the Friedan equations describing the Universe as an ideal fluid, solve General Relativity equations and assume the cosmological principle (homogenous and isotropic metric). In the gravitational instability scenario commonly accepted since more than 40 years, density fluctuations are so small that can be well described, especially on large scales, by perturbation theory (PT; Bernardeau et al. 2002). This is the method we will use to solve the cosmological dynamic equation. In particular, we follow a variation of the Lagrangian perspective proposed by Kitaura & Heß (2013) under the name of Augmented Lagrangian perturbation theory (ALPT). This novel approach provides a very fast path, compared to N-body simulations, to solve cosmic dynamics and simulate dark matter density fields at any epoch.

The most accepted explanation for the LSS seen in galaxy surveys, i.e. that small primordial fluctuations had been amplified by the interaction among collisionless cold dark matter (CDM) in an expanding Universe (Peebles (1982)), is assumed hereafter. Hence, dark matter can be considered a non-linear fluid described by a distribution function written in terms of position, velocity and time $f(r, v, t)$. Due to probability conservation (Liouville theorem), the continuity equation must be satisfied and therefore leads to the Vlasov equation (also known as Bolztmann equation for a collisionless fluid):

$$\frac{\partial}{\partial t} f + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} f + \mathbf{g}\frac{\partial}{\partial \mathbf{v}} f = 0\,, \tag{1}$$

where $\mathbf{v} = \partial\mathbf{r}/\partial t$ and $\mathbf{g} = \partial\mathbf{v}/\partial t$. A careful derivation can be found in Bernardeau et al. (2002). We are interested in solving the evolution of the spatial distribution, rather than the full phase-space dynamics, and this can be done through the moments of the Vlasov equation. In particular:

1. The zeroth moment, which is obtained multiplying equation 1 by the particle mass m and integrating over $\mathrm{d}\mathbf{v}$, represents the continuity equation:

$$\frac{\partial}{\partial t}\rho + \nabla\cdot(\rho\mathbf{u}) = 0\,, \tag{2}$$

   where ρ is the density and $\mathbf{u} = \langle\mathbf{v}\rangle$ the velocity field.

2. The first moment is obtained multiplying equation 1 by the momentum $m\mathbf{v}$ and integrating over $\mathrm{d}\mathbf{v}$. The result is the Euler equation:

$$\frac{\partial}{\partial t}\rho\mathbf{u} + \frac{\partial}{\partial \mathbf{r}}\left(\rho\mathbf{u}\mathbf{u} + \rho\langle\mathbf{w}\mathbf{w}\rangle\right) = \rho\mathbf{g}\,. \tag{3}$$

   The term containing the random velocity $\mathbf{w} \equiv \mathbf{u} - \mathbf{v}$ is a tensor which is generally decomposed into a trace tensor describing pressure and a traceless viscous tensor $\rho\langle\mathbf{w}\mathbf{w}\rangle = \mathbf{P} - \mathbf{T}$ (Kitaura et al. 2023).

A collisionless fluid is assumed to describe dark matter and therefore the pressure term is neglected. By combining equations 2 and 3, one can obtain the so-called Navier-Stokes equation:

$$\rho\left(\frac{\partial}{\partial t}\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}\right) = \nabla\cdot\mathbf{T} + \rho\mathbf{g}\,. \tag{4}$$

We want to make explicit the dependence on the cosmological scale factor $a(t)$ in order to consider the departures from the homogeneous Hubble expansion. To this end, it is convenient to write the physical coordinates $\mathbf{r}$ in terms of the comoving coordinates $\mathbf{x}$, $\mathbf{r} = a(t)\mathbf{x}$, and the cosmic proper time t in terms of the conformal time τ, $dt = a(t)d\tau$. The proper velocity can be re-defined as the sum of the Hubble expansion velocity and the peculiar velocity: $\mathbf{u} \equiv \dot{\mathbf{r}} = \dot{a}(t)\mathbf{x} + a(t)\dot{\mathbf{x}} = \mathcal{H}\mathbf{x} + \mathbf{v}$. [1]

We now re-write equation 4 in comoving coordinates. The gravity force term can be expressed as the gradient of the gravitational potential $\mathbf{g} = -a^{-1}\nabla\phi$, [2] which is given by Poisson equation $\nabla^2_{\mathbf{r}}\phi = 4\pi G\rho$. Let us define

[1] $\mathcal{H} \equiv d(\ln a)/d\tau = Ha = \dot{a}$ is the conformal expansion rate and H the Hubble constant.

[2] The scale factor appears by the gradient transformation: $\nabla_{\mathbf{r}} \Rightarrow a^{-1}\nabla_{\mathbf{x}}$.

the dimensionless overdensity parameter δ:

$$\delta \equiv \frac{\rho - \langle\rho\rangle}{\langle\rho\rangle}\,, \tag{5}$$

where $\langle\rho\rangle$ is the mean density. The Poisson equation in terms of the density contrast and in comoving coordinates reads:

$$\nabla^2\Phi = 4\pi G\langle\rho\rangle\delta a^2\,, \tag{6}$$

where a new definition of the potential, $\Phi = \phi - \frac{\partial(Ha)}{\partial\tau}\mathbf{x}^2$, including the Hubble kinetic term has been introduced for convenience. The Navier-Stokes equation (4), in terms of peculiar motions v, is finally rewrited as

$$\frac{\partial}{\partial\tau}\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v} = \frac{1}{a\rho}\nabla\cdot\mathbf{T} - \nabla\Phi - \mathcal{H}\mathbf{v}\,, \tag{7}$$

where $\nabla \equiv \nabla_{\mathbf{x}}$ instead of $\nabla \equiv \nabla_{\mathbf{r}}$. We now look for solutions of this equation. The stress tensor $\mathbf{T}$ characterizes the deviation of particle motions from a single stream defined by the first term in Eq. 3, ρuu. Thus, in the first stages of gravitational instability, when structures did not have time yet to collapse and virialize, we can assume $\mathbf{T} \approx 0$. In the context of large-scale structure, this approximation provides good results even at present times and largely simplifies the calculations.

2.1 Lagrangian perturbation theory

The perturbation theory is a commonly used method to find solutions to the equation of motion. The Eulerian perspective is a well studied method consisting in expanding both the density and the gradient of the velocity fields from their linear solutions to look for solutions of the continuity and Euler equations at different orders (Bernardeau et al. 2002). Another possible path, the one we follow here, is to develop the Lagrangian Perturbation Theory (LPT). In the Lagrangian description of cosmological fluid dynamics the fluid motion is traced by following fluid elements through space and time instead of looking at specific locations in space as Eulerian perspective does. Therefore, the first step is to re-write the Navier Stokes equation (Eq. 4) without the stress tensor and in Lagrangian coordinates, that is using the material derivative $d/d\tau = \partial/\partial\tau + \mathbf{v}\cdot\nabla$. We find:

$$\frac{d}{d\tau}\mathbf{v} = -\nabla\Phi - \mathcal{H}\mathbf{v}. \tag{8}$$

Making the position dependence explicit, $\mathbf{v} = d\mathbf{x}/d\tau$, and re-defining the gravitational potential as $\tilde{\Phi} = -[\frac{3}{2}\Omega_m\mathcal{H}^2]^{-1}\Phi$, we obtain:

$$\frac{d^2}{d\tau^2}\mathbf{x} + \mathcal{H}\frac{d}{d\tau}\mathbf{x} = -\frac{3}{2}\Omega_m\mathcal{H}^2\nabla\tilde{\Phi}\,. \tag{9}$$

The object of interest in LPT is the displacement field $\mathbf{\Psi}$ defining a mapping relation between the initial Lagrangian position $\mathbf{q}$ of a fluid element and its final Eulerian (comoving) position $\mathbf{x}$,

$$\mathbf{x}(\mathbf{q},\tau) = \mathbf{q} + \mathbf{\Psi}(\mathbf{q},\tau)\,, \tag{10}$$

which contains the information of how gravity acts during cosmic evolution (Kitaura & Heß (2013)).

Assuming that the Universe is homogeneous at its initial state, mass conservation between Eulerian and Lagrangian coordinates reads

$$\bar{\rho}(\tau)d\mathbf{q} = \rho(\mathbf{x},\tau)d\mathbf{x} = \bar{\rho}(\tau)(1+\delta(\mathbf{x},\tau))d\mathbf{x}\,, \tag{11}$$

and hence a relation between the Eulerian density contrast and the displacement field can be obtained:

$$1 + \delta(\mathbf{x},\tau) = \left|\frac{d\mathbf{q}}{d\mathbf{x}}\right| = J(\mathbf{q},\tau)^{-1}\,, \tag{12}$$

where $J(\mathbf{q},\tau)$ is the Jacobian of the transformation between Eulerian and Lagrangian space:

$$J(\mathbf{q},\tau) = \det\left(\delta^K_{ij} + \Psi_{i,j}(\mathbf{q},\tau)\right) \equiv \det\left(\mathcal{D}_{ij}(\mathbf{q},\tau)\right)\,. \tag{13}$$

Note that in the above equation δ^K_{ij} is the Kronecker delta, $\Psi_{i,j} = \partial\Psi_i/\partial q_j$, and $\mathcal{D}_{ij}(\mathbf{q},\tau)$ is the tensor of deformation. In this way we see that the Lagrangian description is non-linear in the density field. Taking the divergence of Eq. 9 and multiplying it by the Jacobian, we can obtain

$$J(\mathbf{q},\tau)\left[\delta^K_{ij} + \Psi_{i,j}(\mathbf{q},\tau)\right]^{-1}\nabla\cdot\left[\frac{d^2\mathbf{\Psi}}{d\tau^2} + \mathcal{H}\frac{d\mathbf{\Psi}}{d\tau}\right] = \frac{3}{2}\Omega_{\rm m}\mathcal{H}^2(J(\mathbf{q},\tau)-1)\,, \tag{14}$$

where we used the chain rule to write $\nabla_{x_i} = [\delta^K_{ij}+\Psi_{i,j}(\mathbf{q},\tau)]^{-1}\nabla_{q_j}$. Note that Eq. 14 is a second order non-linear equation for $\mathbf{\Psi}(\mathbf{q})$ written in Lagrangian coordinates, thus $\nabla \equiv \nabla_{\mathbf{q}}$. The LPT finds solutions to this equation by perturbatively expanding the displacement field around its linear solution $\mathbf{\Psi}(\mathbf{q},\tau) = \mathbf{\Psi}^{(1)}(\mathbf{q},\tau)+\mathbf{\Psi}^{(2)}(\mathbf{q},\tau)+\cdots$.

2.2 Linear solution: Zel'dovich Approximation

Considering the linear displacement field $\mathbf{\Psi}(\mathbf{q},\tau) = \mathbf{\Psi}^{(1)}(\mathbf{q},\tau)$, we obtain the Jacobian

$$J(\mathbf{q},\tau) \approx 1 + \mathbf{\Psi}_{i,i}(\mathbf{q},\tau), \tag{15}$$

and its inverse is given by the binomial approximation:

$$J(\mathbf{q},\tau)^{-1} \approx 1 - \mathbf{\Psi}_{i,i}(\mathbf{q},\tau)\,. \tag{16}$$

By substituting the last equation into equation12, we obtain:

$$\delta^{(1)}(\mathbf{x},\tau) \equiv D^{(1)}(\tau)\delta^{(1)}(\mathbf{x},0) = -\nabla\cdot\mathbf{\Psi}^{(1)}(\mathbf{q},\tau)\,, \tag{17}$$

where the linear growth factor $D^{(1)}(\tau)$ describes the temporal evolution of the initial density field $\delta^{(1)}(\mathbf{x},0)$ which is imposed by the initial conditions (Peebles 1980) . As we have assumed that vorticity vanishes, Eq. 17 determines completely the displacement field to linear order. It is interesting to note that the evolution of fluid elements at this order is local because it does not depend on the behaviour of the rest of fluid elements.

The velocity field related to equation 10 is defined as the derivative of the the displacement field, $\mathbf{v}=\partial\mathbf{\Psi}/\partial\tau$, therefore using equation 17 we obtain:

$$\nabla\cdot\mathbf{v}^{(1)}(\mathbf{q},\tau) = -\dot{\delta}^{(1)}(\mathbf{q},\tau)\,. \tag{18}$$

The *Zel'dovich Approximation* (Zel'dovich 1970) consist of extrapolating the linear LPT into the non-linear regime, i.e., where density perturbations $\delta\rho$ are not negligible. To determine its eigenvectors, one can diagonalize the deformation tensor and, following equation 12, find the next expression for the overdensity:

$$\delta(\mathbf{x},\tau) = \frac{1}{[1-\lambda_1 D_1^{(1)}(\tau)][1-\lambda_2 D_2^{(1)}(\tau)][1-\lambda_3 D_3^{(1)}(\tau)]} - 1\,, \tag{19}$$

where λ_i are the eigenvalues of the tidal tensor $\mathbf{\Psi}_{i,j}(\mathbf{q},\tau)$. As time goes on and structures grow, D increases. When it reaches $1/\lambda_1$, the shortest axis collapse and a cosmic sheet is formed. Afterwards, the growth factor would reach $1/\lambda_2$ and a filament will be formed; subsequently, spherical collapse will occur at $1/\lambda_3$ showing knots. There is just one more possibility: if the three eigenvalues satisfy λ_i <0, $\delta \approx$ -1, corresponding to an under dense region (void). A very important result has been shown: the Zel'dovich Approximation is able to predict the cosmic web.

2.3 Second-order Lagrangian Perturbation Theory

Expanding the displacement field until second order, i.e. $\mathbf{\Psi}(\mathbf{q},\tau) = \mathbf{\Psi}^{(1)}(\mathbf{q},\tau) + \mathbf{\Psi}^{(2)}(\mathbf{q},\tau)$, has been demonstrated to improve considerably the properties of density and velocity fields in comparison with the above described Zel'dovich Approximation (Buchert et al. 1993). On the other hand, the improvement of third

order LPT (3LPT) with respect to 2LPT is not remarkably. To understand this, we study the second order solution of equation 14 describing the correction to the linear displacement field as a consequence of gravitational tidal effects (Bernardeau et al. 2002) and thus taking into account that gravitational instability is non local. For curl-free fields it is demonstrated that (Catelan 1995):

$$\nabla \cdot \mathbf{\Psi}_{\rm 2LPT} = -D^{(1)}\delta^{(1)} + D^{(2)}\delta^{(2)} \,. \tag{20}$$

Particular expressions (power laws of the density parameter Ω) for the linear and second-order growth factors are derived also in Catelan (1995), whereas the second-order overdensity depends on the first and second-order potentials which are obtained by solving the corresponding Poisson equations $\nabla^2\phi^{(i)} = \delta^{(i)}(\mathbf{q})$ for i =1, 2 (for further details see Kitaura & Heß 2013):

$$\delta^{(2)}(\mathbf{q}) = \sum_{i>j} \left(\phi^{(1)}_{,ii}(\mathbf{q})\phi^{(1)}_{,jj}(\mathbf{q}) - [\phi^{(1)}_{,ij}(\mathbf{q})]^2 \right) \,. \tag{21}$$

2.4 Augmented Lagrangian perturbation theory

We can go even further and split the displacement field $\mathbf{\Psi}(\mathbf{q},\tau)$ into a long- ($\mathbf{\Psi}_{\rm L}(\mathbf{q},z)$) and a short-range ($\mathbf{\Psi}_{\rm S}(\mathbf{q},z)$) component. This is the theory behind the code used in this work to generate the dark matter box. Proposed by Kitaura & Heß (2013), the method was dubbed Augmented Lagrangian Perturbation Theory (ALPT). The displacement field is written as:

$$\mathbf{\Psi}(\mathbf{q},z) = \mathbf{\Psi}_{\rm L}(\mathbf{q},z) + \mathbf{\Psi}_{\rm S}(\mathbf{q},z) \,, \tag{22}$$

where:

- The long-range component relies on 2LPT using a Gaussian filter $\mathcal{K}(\mathbf{q},r_s)$=exp($-|\mathbf{q}^2|/(2r_s^2)$) ($r_s \equiv$ smoothing radius) to improve it at small scales, where it fails:

$$\mathbf{\Psi}_{\rm L}(\mathbf{q},z) = \mathcal{K}(\mathbf{q},r_s) \circ \mathbf{\Psi}(\mathbf{q},z)_{\rm 2LPT} \,. \tag{23}$$

- The short-range component relies on the spherical collapse approximation (SC Fosalba & Gaztañ aga 1998) corrected not to fail on large-scales:

$$\mathbf{\Psi}_{\rm S}(\mathbf{q},z) = \mathbf{\Psi}_{\rm SC}(\mathbf{q},z) - \mathcal{K}(\mathbf{q},r_s) \circ \mathbf{\Psi}_{\rm SC}(\mathbf{q},z) \,. \tag{24}$$

The APLT method, which is computationally very fast, improves the above approximations preserving power on the large scales with just one additional parameter: the smoothing radius defining the transition between both regimes. We simulate a dark matter density field relying on the ALPT method to define the displacement field.

2.5 The dark matter density field box

To have a sufficiently large volume to provide good statistics, we have simulated in only a few seconds a $1000\,h^{-1}{\rm Mpc}^3$ cubic volume divided into 256^3 cells, meaning that the resolution is $\sim 3.9\,h^{-1}$Mpc per cell. We rely on ΛCDM cosmology with parameters { $\Omega_M = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_b = 0.048$, $\sigma_8 = 0.823$, $w = -1$, $n_s = 0.95$, and a Hubble parameter $h = 0.68$ ($H_0 = 100\,h\,{\rm km\,s^{-1}Mpc^{-1}}$). The initial conditions are 2LPT based at redshift $z = 99$, and we run from there down to z=1.5 (i.e., where the DESI quasar distribution peaks, see Yè che et al. (2020)) in one step. In a follow-up work we will span from $z = 0$ up to $z \sim 4$. Figure 1 shows a box slice; here we can distinguish the high density regions (known as knots) connected through filaments surrounded by large under dense regions (voids).

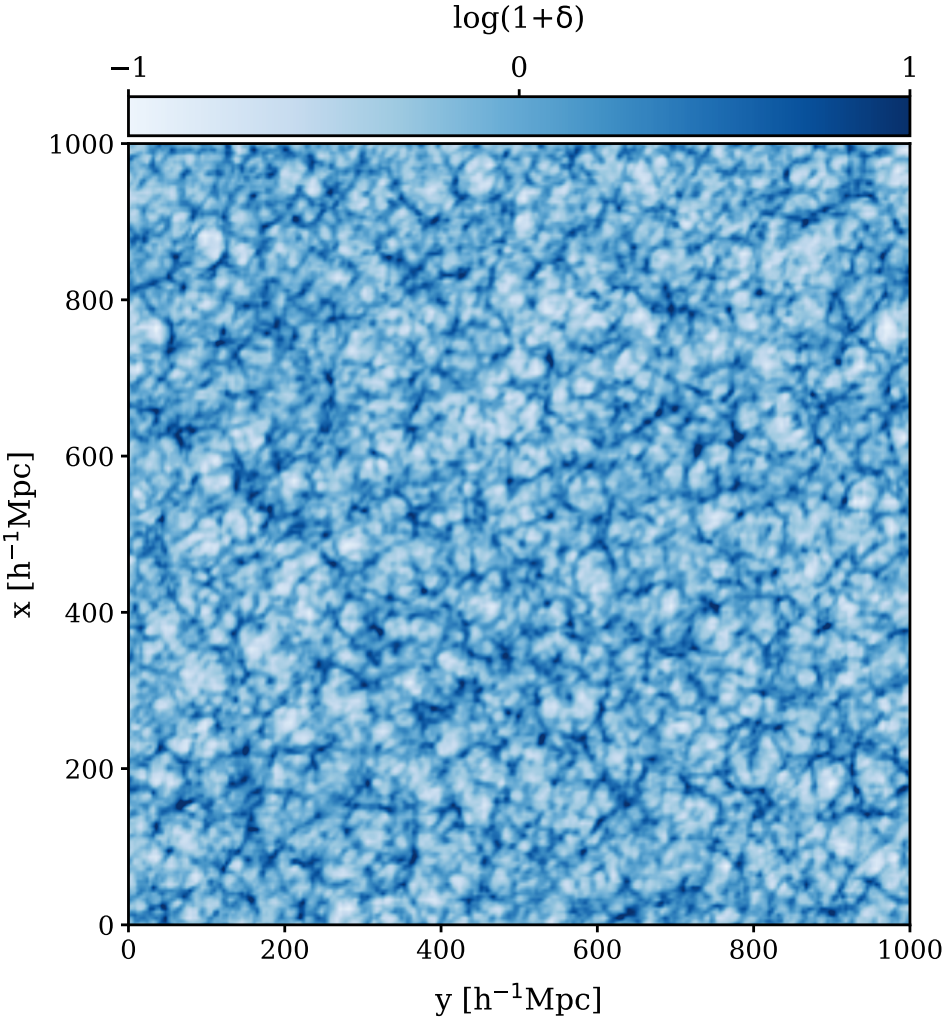

Figure 1: Dark matter density slice 20 h^{-1} Mpc thick and 1000 h^{-1} Mpc side at redshift $z = 1.5$ simulated with the ALPT method.

2.6 The cosmic web classification

Numerical simulations have shown that the gravitational growth of the initial density perturbations leads to a characteristic pattern of structures, similar to the observations of the large-scale distribution of galaxies, dubbed *cosmic web* (van de Weygaert & Schaap 2008). Most of the Universe is be composed of low-density regions (voids) surrounded by sheets of matter with higher density. Filaments, i.e. the main contribution to the mass distribution, generate the web-pattern and its intersections (knots) have the higher densities. Hahn et al. (2007) propose a theoretical criterion to distinguish between these regions by considering the equation of motion of a particle moving in the gravitational potential of a halo: $\ddot{x} = -\nabla\phi$. The center of the halo is an extremum, i.e. $\nabla\phi = 0$, hence the equation of motion can be linearized to determine the particle acceleration in terms of the tidal field $T_{ij} \equiv \partial_i\partial_j\phi$. Therefore, the eigenvalues of the tidal tensor $\lambda_1 > \lambda_2 > \lambda_3$ determine the linear dynamics in a local extremum of a potential well. Following Zel'dovich (1970), one can define:

- Knots (λ_1, λ_3, $\lambda_3 > \lambda_{th}$): regions where gravitational collapse causes matter inflow.
- Voids (λ_1, λ_3, $\lambda_3 < \lambda_{th}$): expanding regions just opposite to knots.
- Filaments (λ_1, $\lambda_2 > \lambda_{th}$; $\lambda_3 < \lambda_{th}$): regions closer to knots.

- Sheets ($\lambda_1 > \lambda_{th}$; λ_2, $\lambda_3 < \lambda_{th}$): regions closer to voids.

The eigenvalue threshold $\lambda_{th} \neq 0$ was introduced by Forero–Romero et al. (2009) to reduce the dependence of the classification on the smoothing scale, mesh resolution and mass assignment scheme. Kitaura et al. (2022) investigate the relation between the cosmic web and the halo distribution through the invariants of the tidal field, defined by its eigenvalues, which provide very complete galaxy bias descriptions. Figure 2 presents our cosmic web classification based on the dark matter simulation box. The density values on the color bar represent the *knots*(1), *filaments*(2), *sheets*(3), and *voids*(4). The left panel shows the ϕ-Tweb classification, i.e. based on the tidal field eigenvalues. In the right panel, we propose a new subclassification of the cosmic web now in terms of the dark matter over density δ. It is based on the similar approach of Sinigaglia et al. (2022) to map the ionized gas on the dark matter field by considering also the invariants of the tidal field tensor based on delta. Compared to the potential classification, the δ-Tweb provides information about smaller scales.

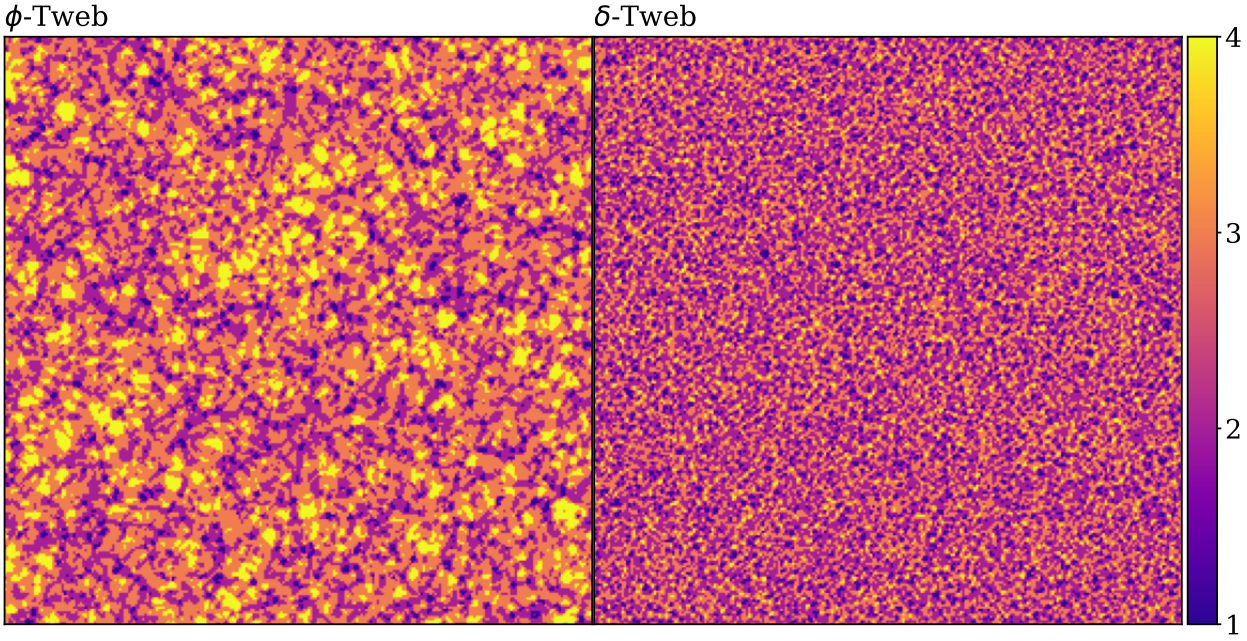

Figure 2: Different scale cosmic web classifications for a dark matter slice at redshift $z = 1.5$ simulated with the ALPT method in a 1000 h^{-1} Mpc side box with a mesh of 256^3 cells. The classification is done following the points of section 3.6 for the dark matter potential (ϕ) and its first and second order overdensity parameters δ and δ^2.

3 The bias of dark matter tracers

The consensus scenario in structure formation (see e.g. White & Rees 1978) establishes that dark matter aggregates into virialized clumps, known as dark matter haloes thanks to gravitational collapse. The baryonic matter falls onto these potential wells and, after cooling, gas condensate to form galaxies. Almost four decades ago, Bardeen et al. (1986) suggested that galaxies could correspond to peaks in the initial field and therefore they should be more clustered than the underlying mass (effect commonly known as *galaxy bias*).

There are two main bias prescriptions for galaxy clustering based on density. On one hand, *Lagrangian bias models* relate galaxy formation to the initial density field (Bardeen et al. 1986). On the other hand, *Eulerian bias models* (Mann et al. 1998) study the bias properties of an evolved mass density field. None of the descriptions above is realistic at all because galaxies can merge while structures grow, but they are reasonable approaches to study galaxy clustering at a given redshift (Narayanan et al. 2000). Furthermore, the bias relation is known to be non linear, stochastic and scale dependent (see Kitaura et al. (2013) and references therein).

We adopt here an Eulerian perspective and define the galaxy number overdensity $\delta_{\rm g}$ as a function of the local value of the evolved mass density field, ρ. Fry & Gaztanaga (1993) suggested to describe such relation using perturbation theory assuming that large-scale density fluctuations in the Universe are small. Therefore, the galaxy density is defined as a Taylor expansion of the local dark matter density

$$\delta_{\rm g}(\mathbf{x}) = f(\delta(\mathbf{x})) = \sum_{i=0}^{\infty} \frac{b_i}{i!}\delta^i(\mathbf{x})\,, \tag{25}$$

where the coefficients b_i are free bias parameters. McDonald & Roy (2009) propose an extension of this model by describing the bias terms with the tidal field tensor $\mathcal{T}$, which coincides with the velocity shear in the linear regime and in the absence of vorticity of the velocity field. To complete a non-local description they also consider a short-range component related to hydrodynamics and non-linear galaxy formation. They also include a variable, ϵ, representing the white noise to allow for stochasticity and shot-noise, since the halo distribution is a discrete realisation of the underlying dark matter field that generates a stochastic uncertainty component. The resulting expanded bias dependence of Kitaura et al. (2022) based on McDonald & Roy (2009) can be described as:

$$\delta_{\rm h}(\vec{r}) = \underbrace{\overbrace{c_\delta\delta(\vec{r})}^{\text{local}}}_{\text{first order}} + \underbrace{\overbrace{\frac{1}{2}c_{\delta^2}\left(\delta^2(\vec{r}) - \langle\delta^2\rangle\right)}^{\text{local}} + \overbrace{\frac{1}{2}c_{s^2}\left(s^2(\vec{r}) - \frac{2}{3}\langle\delta^2\rangle\right)}^{\text{non-local}}}_{\text{second order}} + \underbrace{\overbrace{\frac{1}{3!}c_{\delta^3}\delta^3(\vec{r})}^{\text{local}} + \overbrace{\frac{1}{2}c_{\delta s^2}\delta(\vec{r})s^2(\vec{r}) + \frac{1}{3!}c_{s^3}s^3(\vec{r})}^{\text{non-local}}}_{\text{third order}} + \underbrace{\overbrace{\mathcal{O}(F_{\delta,\mathcal{T}(\delta)}|^4)}^{\text{local \& non-local}}}_{\text{fourth order+}}$$

$$\underbrace{\overbrace{+\underbrace{F_{\delta^{\rm SR}}(\partial_i^l\partial_j^l\delta(\vec{r}))}_{\substack{\text{short range } \theta=\delta \text{ terms } l\in\mathbb{N}\\ \text{first order+}}} + \underbrace{F_{\rm shear}(\vec{v}(\vec{r})|_{\rm curl-free}) + F_{\rm shear}(\vec{v}(\vec{r})|_{\rm div-free}) + F_{\rm curl}(\vec{v}(\vec{r}))}_{\substack{\theta\neq\delta \text{ terms} \qquad \text{vorticity } \theta=\delta \ \& \ \theta\neq\delta \text{ terms}\\ \text{third order+}}}}^{\text{non-local}} + \underbrace{\overbrace{F_\epsilon(\epsilon(\vec{r}))}^{\text{local \& non-local}}}_{\substack{\text{noise terms}\\ \text{first order+}}}}_{\text{curl-free \& } \theta=\delta \text{ terms}} \,.$$

This is a halo bias model with several dependencies: $\delta_h(\vec{r}) = F(\delta, \mathcal{T}, \Gamma, \Sigma, \omega, \epsilon)$. In particular, δ is the dark matter overdensity, $\mathcal{T}$ is the tidal field tensor, Γ gathers the short range non-local terms such that $\Gamma^l_{ij} = \partial^l_i\partial^l_j\delta$, Σ is the velocity, ω the curl in the velocity field and ϵ accounts for stochasticity. The cs are unknown bias factors and θ is the velocity divergence. The terms s^2 and s^3 are scalars derived from the tidal field tensor and can be related to its invariants, that are defined by its eigenvalues $\lambda_1, \lambda_2, \lambda_3$ (Kitaura et al. (2022)). We have shown in Section 2.6 how to classify the cosmic web through these eigenvalues. Therefore, there is a relation between the tidal tensor and the distribution of dark matter tracers. This particular behaviour will be better understood by fitting our model considering the different parts of the cosmic web.

3.1 The large scale bias parameter

The large-scale bias b_{LS} is a parameter describing the clustering of galaxies on linear scales. In Fourier space it is defined as the square root of the ratio of the galaxy and mass power spectra[3] (see e.g. Mann et al. (1998)):

$$b_{\rm LS}^{\rm FS}(k) \equiv \sqrt{\frac{P_{\rm gal}(k)}{P_{\rm DM}(k)}}\Bigg|_{k>k_{\rm th}}. \tag{26}$$

Narayanan et al. (2000)) show that $b(k)$ tends to a constant value on scales larger than $\sim 4h^{-1}$Mpc and different values are found for each galaxy type. In particular, early-type galaxies are always positively biased, as opposed to to spirals. Kaiser (1984) justified this clustering behaviour proposing that over dense regions on small scales also reside in large-scale regions with high density. Observational evidence of large-scale bias was shown in Peacock & Dodds (1994). On this line, Kitaura et al. (2020) propose a large-scale bias parameter definition in configuration space computed through a convolution of the overdensity with a Gaussian kernel that softens the small-scale contribution:

$$b_{\rm LS}^{\rm CS}(z) \equiv \sqrt{\frac{\sigma^2_{K_g}(z)}{\sigma^2_{K_{\rm DM}}(z)}}. \tag{27}$$

In the analysis of the next section we use this last description (equation 27), and referer to it just as large-scale bias, assuming a smoothing scale of 100 h^{-1}Mpc.

[3]See appendix for a brief review of some LSS statistics.

4 Analytical bias model to populate the cosmic web with quasars

Here we aim at constructing and analyzing a bias model term by term to describe the general features of galaxies, together with some particular considerations for quasars.

4.1 Deterministic bias

4.1.1 The non-linear bias parameter

A non-linear bias prescription is constructed by defining a power-law dependence between the number of quasars in a cell i, $N_i^{\rm QSO}$ and the dark matter overdensity in the same cell, $\delta_i^{\rm DM}$:

$$N_i^{\rm QSO} \propto \left(1+\delta_i^{\rm DM}\right)^{\alpha} \ . \tag{28}$$

A power-law bias is one of the simplest models one can think of, and provides accurate descriptions of different tracers (de la Torre & Peacock 2013). Here it is defined by the exponent α which describes non linearity. Note that such a power law can be Taylor expanded following the binomial theorem, hence it preserves the 'infinity' behaviour of equation 3, but avoids negative density values (i.e. non-physical values) which can appear when truncating the expansion at a given order. Figure 3 show the change in the power spectrum obtained by increasing the α parameter. It is clear that the higher the exponent, the higher the large-scale bias. We obtain a linear relation between large-scale bias and α (see inset plot).

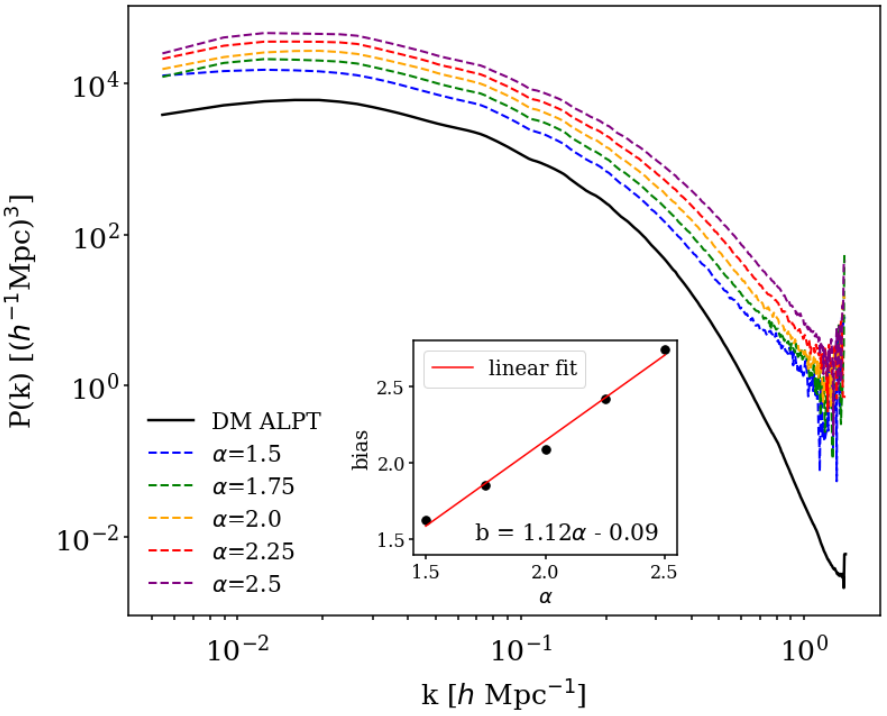

Figure 3: Power spectra comparison for the power-law bias model (Eq. 28) with different values for the exponent α and enough number density ($\bar{N}$=0.01 h^3 Mpc^{-3}) to have good statistics. The inset plot shows a linear correlation between the large-scale bias (27) and α.

A different perspective to understand the effect of the power law is presented in Figure 4. From left to right, we show slices of the quasar overdensity field for increasing values of the exponent α. In the right panel, that

has the highest α values, we can appreciate how galaxies are preferentially distributed on the density peaks in contrast with the other panels, where objects are more spread out over the entire density field.

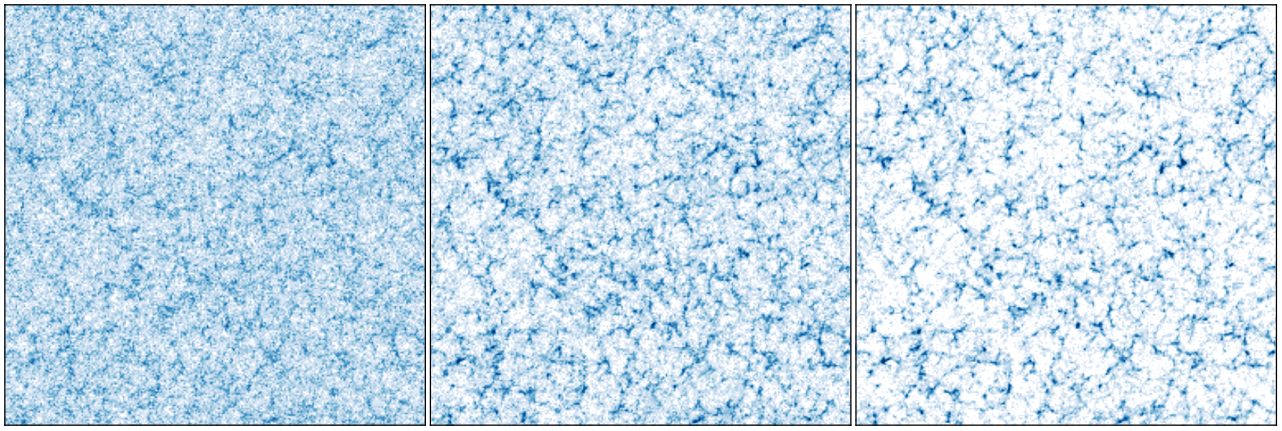

Figure 4: Log-density slices of our 1000 h^{-1} Mpc box, with a thickness of 20 h^{-1} Mpc, generated using the power-law bias model (Eq. 28) with $\bar{N}$=0.01 h^3 Mpc^{-3}. From left to right we show α = 1.5, 2.0 2.5. Structures appear tighter as we move from left to right in the figure.

Different results in the literature lead to the same conclusion: quasars are highly biased tracers of the dark matter. Measures of the quasar two-point correlation function over the redshift range $2.2 < z < 2.8$ return a large-scale bias factor of $b \sim 3.5$ in SDSS-III/BOSS data (White et al. 2012). From the catalogue of the 2dF QSO Redshift Survey (2QZ), Croom et al. (2005) obtained $b(z = 1.55) = 2.41$. From the SDSS-IV/eBOSS survey, Laurent et al. (2017) found a compatible value, $b(z = 1.55) = 2.45 \pm 0.05$, by fitting the auto-correlation function with a flat ΛCDM model. It is also interesting to see the good agreement between previous values of the linear bias parameter and the ones obtained from simulations of the DR14 quasar sample with the EZmock method as shown in Figure 8 of Ata et al. (2017). We have demonstrated that such a high power can be reached by increasing α. We will further check that the power-law exponent is not the only free parameter that can model this important feature.

4.1.2 Threshold bias to suppress under dense regions

Quasars, and galaxies in general, can only form in regions with a minimum mass. Therefore, adding a threshold bias term as suggested by Kaiser (1984), seems reasonable to avoid the under dense regions. This additional terms can be described by a step function (Kitaura et al. 2013), which suppresses overdensities smaller than a threshold $\delta_{\rm th}$. The bias model now reads:

$$N_i^{\rm QSO} \propto \left(1 + \delta_i^{\rm DM}\right)^{\alpha} \times \theta\left(\delta_i^{\rm DM} - \delta_{\rm th}\right) . \tag{29}$$

where the step function $\theta(\delta_i^{\rm DM} - \delta_{\rm th})$ takes the value 1 for $\delta_i^{DM} \geq \delta_{th}$ and 0 otherwise, for instance in the lowest density regions of the dark matter density field.

Figure 5 shows how the power spectrum increases for higher thresholds values, $\delta_{th} = \{-1, -0.5, -0.25, 0.0, 0.5\}$, that is when under denser regions are suppressed. For each $\delta_{\rm th}$ value above, the percentage of suppressed cells in relation to the total is $\{0, 3.9, 20.3, 59.7, 89.5, 97.2\}$. This means that less than 3 per cent of the cells have over densities higher than unity (i.e. densities higher than twice the mean density, Eq. 5). Moreover, about 40 per cent of the cells have higher density than the mean density. Avoiding under dense regions without changing the number density implies placing the objects where the dark matter density field presents high densities, thus the distributions are more highly biased. Figure 6 helps understanding the effect of a low-threshold bias. It shows the quasar density fields based on our dark matter simulated box for three different over density thresholds; the δ_{th} values increase from left to right. The left panel exhibits both under and over dense regions populated with quasars. In the central panel all the regions where the dark matter is smaller than the mean density over the full volume are suppressed. In the right panel only the density peaks survive.

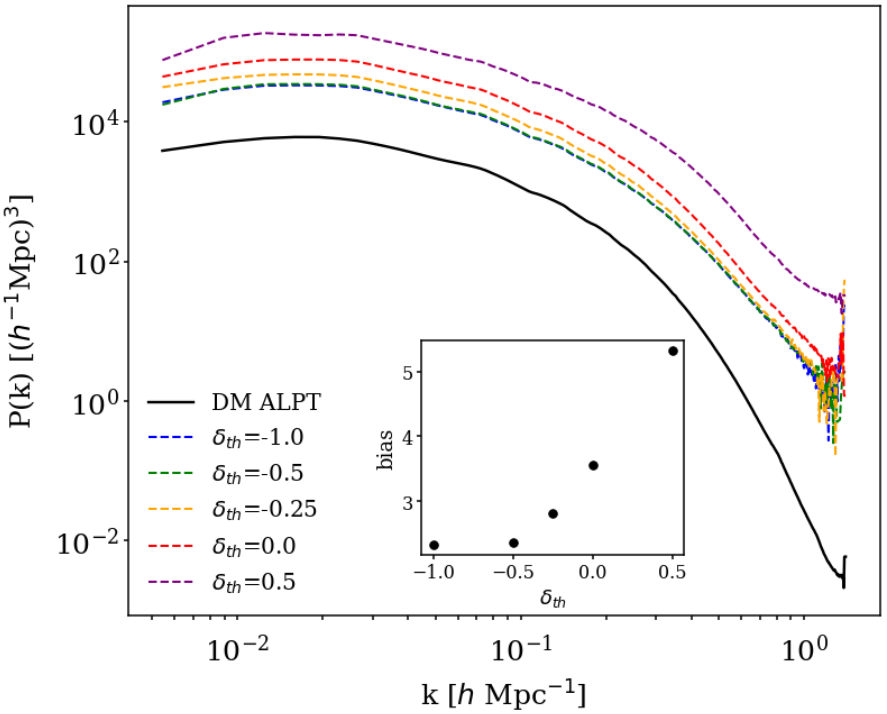

Figure 5: Power spectra of the bias model (Eq. 29) with different low-threshold over densities δ_{th} to avoid the under dense regions in the underlying dark matter density field. The model is defined by $\alpha = 2.2$ and $\bar{N}$=0.01 h^3Mpc^{-3}. The inset plot shows that higher thresholds imply higher large-scale bias.

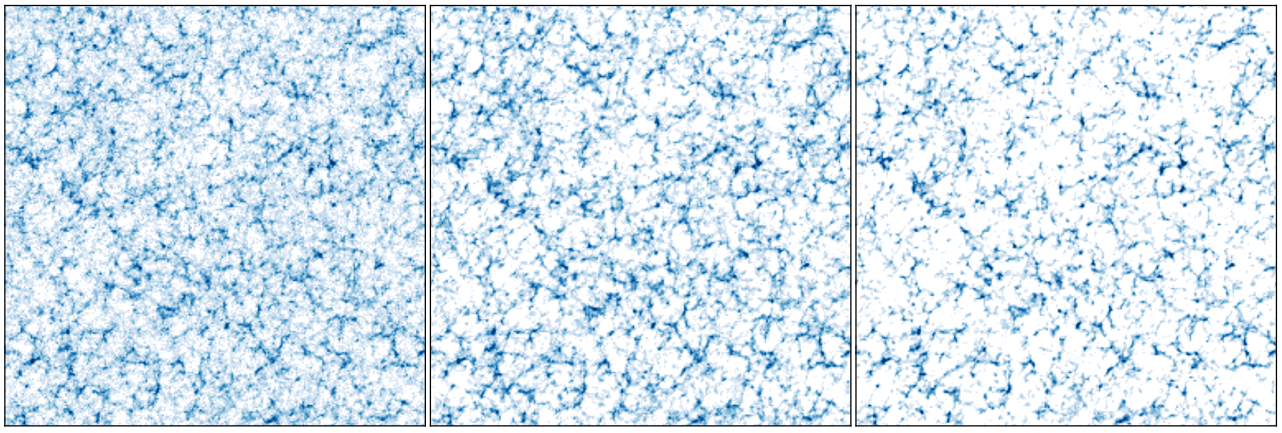

Figure 6: Log-density slices of thickness 20 h^{-1} Mpc and 1000 h^{-1} Mpc side through the bias model with $\alpha = 2.2$, $\bar{N}$=0.01 h^3 Mpc^{-3} and δ_{th}=$\{-0.5, 0.0, 0.5\}$ from left to right respectively. As the threshold is more restrictive, moving to the right, objects in the under dense regions disappear and only the ones in the peaks survive.

4.1.3 Exponential cutoff for density peaks

At high redshift (z∼4) quasars are believed not to occupy the most over dense regions of the Universe (Uchiyama et al. 2017). This suggests a possible connection between quasars and emission-line galaxies (ELGs), which are known to populate the outskirts of less massive halos compared to those hosting luminous red galaxies (LRGs), which are instead distributed around a broader radial range of the hosting halo (Orsi & Angulo 2018).

We therefore included in our model a negative exponential to cutoff the density peaks above a high-density threshold δ_ϵ,

$$N_i^{\rm QSO} \propto \left(1+\delta_i^{\rm DM}\right)^{\alpha} \times \theta\left(\delta_i^{\rm DM}-\delta_{\rm th}\right) \times \exp\left(-\left(1+\delta_i^{\rm DM}\right)\theta\left(\delta_\epsilon-\delta_i^{\rm DM}\right)\right), \tag{30}$$

where the step function now is 1 for $\delta_i^{\rm DM} \leq \delta_\epsilon$. Figure 7 shows the power spectrum for different thresholds $\delta_\epsilon = \{1.0, 1.5, 2.0, 2.5, 3.0\}$, together with the bias evolution for such values. The number of suppressed cells, in percentage with respect to the total (256^3), is { 2.8, 0.84, 0.28, 0.10, 0.039} for each threshold, respectively. We can clearly appreciate a gain of power when less peaks (i.e. higher δ_ϵ) are suppressed. Avoiding objects in higher density regions in the cosmic web translates into a lower bias with respect to dark matter. The shape of the power spectrum is preserve until $k = 0.6\,h\,{\rm Mpc}^{-1}$, where the configuration with lower bias adquires power. Figure 8 shows slices of our simulated box for an increasing number of suppressed regions as we move from left to right. This result helps us to understand the effect of a change in δ_ϵ, which regulates the tendency of objects to cluster in the high-density regions. Including a negative exponential in the bias prescription allows us to manage the percentage of cells related to the density peaks we want to avoid.

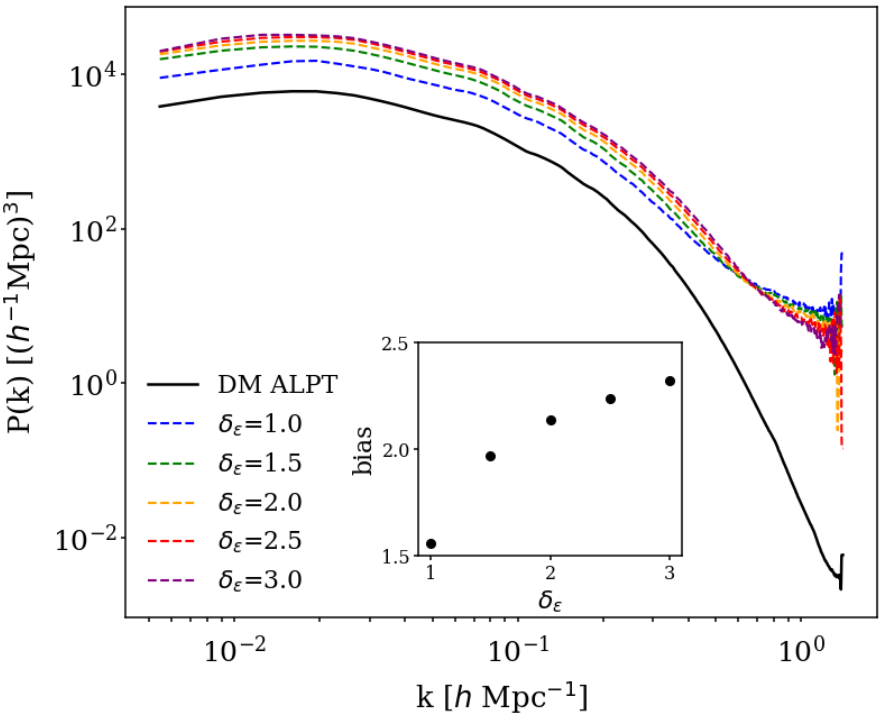

Figure 7: Power spectrum of the bias model (Eq. 30) with different high-threshold over densities δ_ϵ to suppress the density peaks of the dark matter density field. The model is defined with $\alpha = 2.2$, $\bar{N}$=0.01 $h^3{\rm Mpc}^{-3}$ and $\delta_{th} = -0.5$. The inset shows a positive correlation between the large-scale bias and δ_ϵ. The density peaks enhance the bias.

Neyrinck et al. (2014) found good fits for a similar damping exponential but with two free parameters instead of one. Such a model provides a soft decay in the tail of the distribution in contrast to the sharp decay when only one parameter is consider in the exponential. For this reason, here we propose to change the step function suppressing low densities with another two-parameter exponential. In this case the bias model is:

$$N_i^{\rm QSO} \propto \left(1+\delta_i^{\rm DM}\right)^{\alpha} \times \exp\left(-\left[\frac{1+\delta_i^{\rm DM}}{\rho_\eta}\right]^{\eta}\right) \times \exp\left(-\left[\frac{1+\delta_i^{\rm DM}}{\rho_\epsilon}\right]^{\epsilon}\right), \tag{31}$$

where $\eta < 0$ and $\epsilon > 0$ provide a fine-tuning of the dependency. Figure 9 shows a comparison between both models for the quasar $\delta_{\rm QSO}$ and dark matter $\delta_{\rm DM}$ over densities.

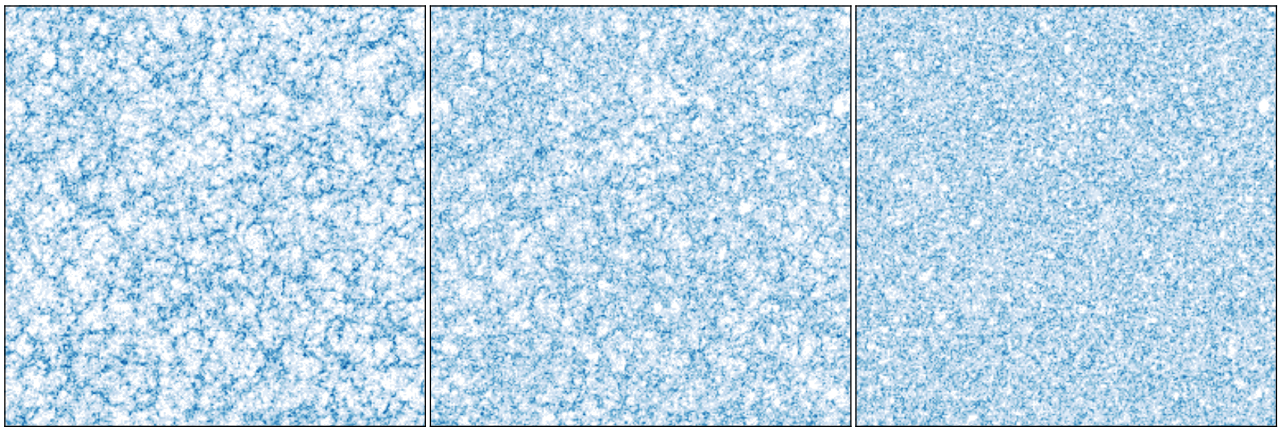

Figure 8: Log-density slices of thickness 20 h^{-1} Mpc and $1000h^{-1}$ Mpc side for the bias model with $\alpha = 2.2$, $\bar{N}$=0.01 $h^3\,\mathrm{Mpc}^{-3}$, δ_{th}=−0.5. From left to right we show δ_ϵ=2, 1 and 0. The smaller the threshold, the more density peaks are suppressed.

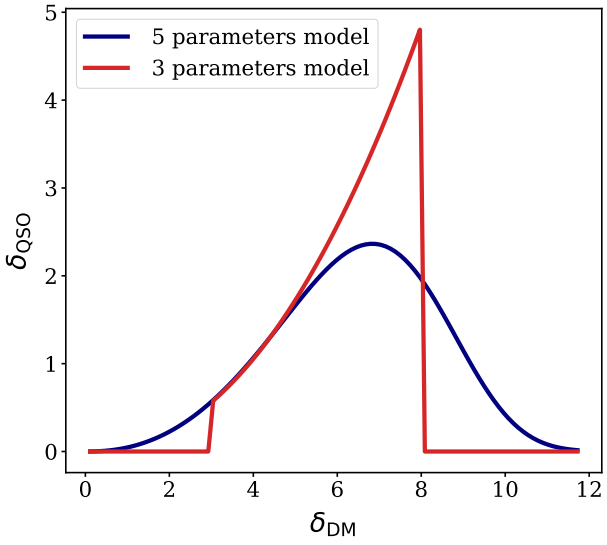

Figure 9: Comparison between the 3 (equation 30) and the 5 (equation 31) free-parameters models. The 5 parameter model (blue line) provides a soft transition both to high- and low-density regions.

4.1.4 The number density of quasars

The number density of quasars ($\bar{N}$) is not yet a well constrained parameter. Conroy & White (2012) estimate a value of $\approx 5\times10^{-7}$ Mpc^{-3}($\approx 1.6\times10^{-6}\,h^3\,\mathrm{Mpc}^{-3}$ for $h = 0.6777$) by means of the auto-correlation function. Measures from the DR14 quasar sample of the the SDSS-IV extended Baryon Oscillation Spectroscopic Survey (eBOSS) show that the quasar number density peaks in proximity of redshift $z = 1.5$, similar to what expected for DESI, with values between 1-2 $\times10^{-5}\,h^3\,\mathrm{Mpc}^{-3}$ (Ata et al. 2017; Laurent et al. 2017). The estimate from DESI quasar target selection is of 6 $\mathrm{deg}^{-2}/(0.05)\Delta z$ (see Figure 20 in Chaussidon et al. (2023)), that is equivalent to $7.8\times10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$ in UNITSIM cosmology with fixed redshift $z = 1.5$ and considering a total area of $14000\,\mathrm{deg}^{-2}$.

In the model we are building step by step, the number density $\bar{N}$ appears when substituting the relation of

proportionality between the quasar number density and the dark matter over density by an equality including the normalization f_N defined as

$$f_N = \frac{\bar{N}}{\left\langle (1+\delta_i^{\rm DM})^\alpha\, \theta(\delta_i^{\rm DM} - \delta_{\rm th}) \times \exp\left(-(1+\delta_i^{\rm DM})\theta(\delta^\epsilon - \delta_i^{\rm DM})\right)\right\rangle_V}\,, \tag{32}$$

where $\langle ... \rangle_V$ denotes the ensemble average in a sufficiently large volume V. A similar normalization can be defined for the 5 free-parameter model, and it controls the quasar number density $\bar{N} = \langle N_i^{QSO} \rangle_V$. In what follows, we study the properties of the model as a function of $\bar{N}$. In figure 10 we notice that the power increases as the number density diminishes. This reflects the fact that a smaller number of quasars, or galaxies in general, leads to a higher bias in its distribution with respect to dark matter. In the inset plot we show the value of the large-scale bias (equation 27) for each number density, which gets very high values for low densities.

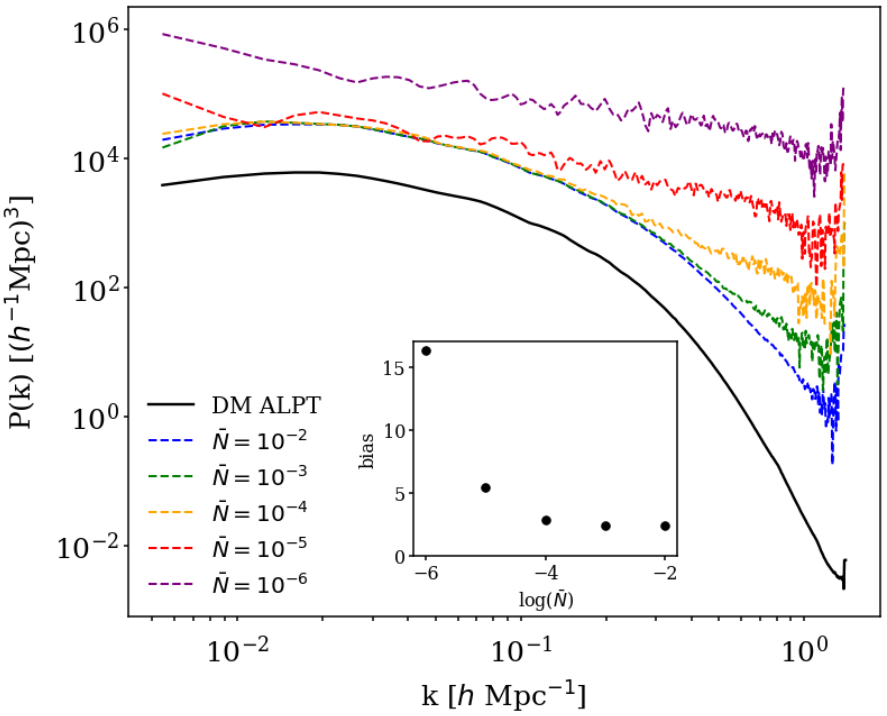

Figure 10: Power spectrum at z=1.5 for different number densities in units of h^3 Mpc^{-3}. The model used here is the power law (Eq. 28) with exponent $\alpha = 2.2$ and neglecting $\{\delta_\epsilon, \delta_{th}\}$. The dark matter power spectrum (black solid line) at the same redshift is shown as reference. The insets show how the large-scale bias (27) changes with density.

4.2 Stochactic biasing

The process of drawing discrete samples of the quasar distribution with respect to the continuous underlying dark matter field is stochastic. Therefore, there is a discrete noise contribution to the bias (Kitaura et al. 2013). A common assumption to account for is to generate Poissonian realizations of the quasar density field in each cell, i.e, through the Poisson distribution:

$$P(N_i|\lambda_i) = \frac{\lambda_i^{N_i}}{N_i!} \exp\left(-\lambda_i\right). \tag{33}$$

Such Poissonian behaviour does not hold always as we will demonstrate later on. Therefore, a Negative Binomial (NB) distribution is often used:

$$P(N|\lambda_i, \beta) = \frac{\lambda_i^{N_i}}{N_i!} \frac{\Gamma(\beta + N_i)}{\Gamma(\beta)(\beta+\lambda)^{N_i}} \frac{1}{(1+\lambda/\beta)^\beta}\,. \tag{34}$$

This depends on the β parameter, which defines the deviation from Poissonity. The NB tends to the Poisson distribution for $\beta \to \infty$. The effect of this stochastic processes is illustrated in figures 11 and 12. The power spectra comparison (11) shows how the Poissonian distribution $P(k)$ (black line) preserves the shape of $P_{\rm DM}(k)$ (i.e. dark matter) in contrast to the non-stochastic distribution (red line) which looses power starting from $k > 0.1\,h\,{\rm Mpc}^{-1}$. In figure 11, three power-spectra computed after sampling from the NB distribution are shown for different values of Poissonian deviations: $\beta = 500$ (orange line), $\beta = 10$ (green) and $\beta = 1$ (blue). Based on these results, in what follows we will consider $\beta > 1000$ as the regime where the NB is equivalent to the Poissonian distribution; note that this happens rigorously when $\beta \to \infty$. We have found another way of gaining power in the bias description, now at small scales. The over densities slices shown in Fig. 12 from left to right correspond to the non-stochastic, the Poissonian, and the NB ($\beta = 1$) distributions, respectively. The objects in the left panel are less clustered that in the others. In the NB slice with higher deviation from Poissonity we can appreciate even less objects spread out on small scales.

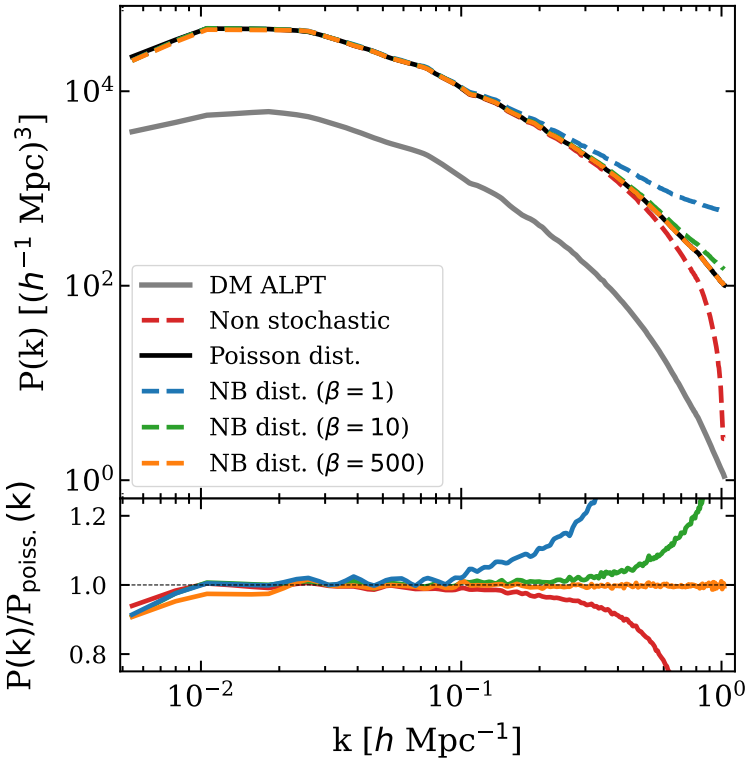

Figure 11: Power spectrum comparison between a stochastic (Poisson and NB) and a non stochastic distribution of the quasar sample. The Poissonian realization $P(k)$ (black line) preserves the shape of the dark matter $P(k)$ (grey line). The NB distribution (blue, green and orange lines) gains power towards small-scale but tends to the Poisson distribution when the deviation parameter β is big enough. The ratio between the different configurations to the Poissonian power spectra is shown on the bottom; the grey shadow indicates a 5 per cent deviation.

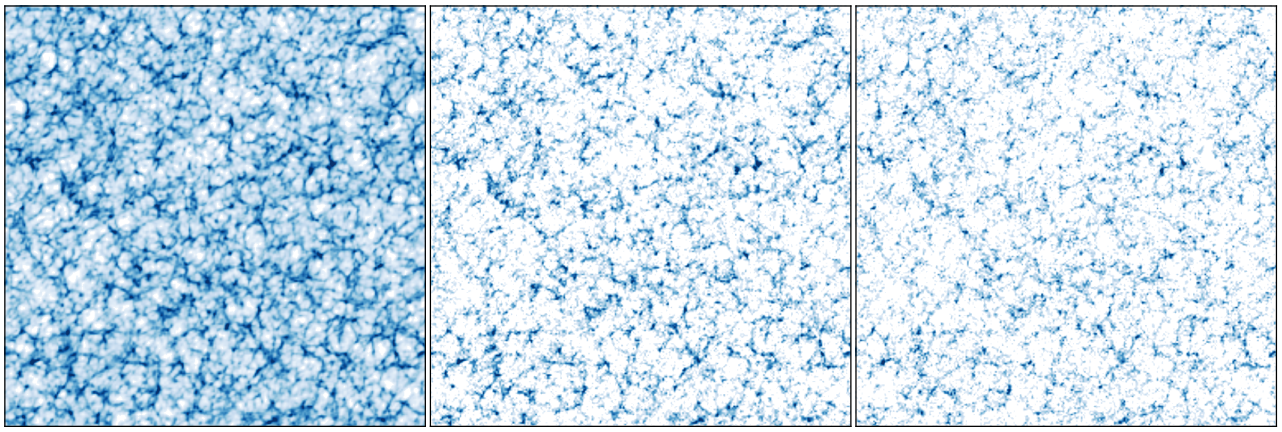

Figure 12: Log-density slices of thickness 20 h^{-1} Mpc and 1000 h^{-1} Mpc side through the bias model for a non-stochastic distribution (left), a Poissonian distribution (center) and a NB distribution with parameter $\beta = 1$ (right). Clustering grows from left to right.

4.3 Do quasars reside in voids?

In a recent study, Pierce et al. (2023) conclude that, in the local Universe, galaxy mergers could be the main mechanism which ignites quasar activity. This suggests that quasars may not populate the voids of the cosmic web. In section 2.6, we classify each cell of our simulation in knots, filaments, sheets and voids. That is why we check which are the effects of suppressing the distribution of quasars in the voids. Figure 13 shows the power spectrum comparison between the distribution of quasars over the full Cosmic Web and its distribution when the voids are suppressed, that is when they only populate knots, filaments and sheets. Despite the shape of the power spectrum is preserved, a clear difference emerges: when the voids are suppressed, the power increases at most scales, in particular on the largest ones, where the ratio exceeds the 5 per cent deviation region (shaded area). In our simulation the voids occupy just a 0.25 per cent of the total cells, which returns small differences, but in larger simulations they can constitute an important part of the cosmic web.

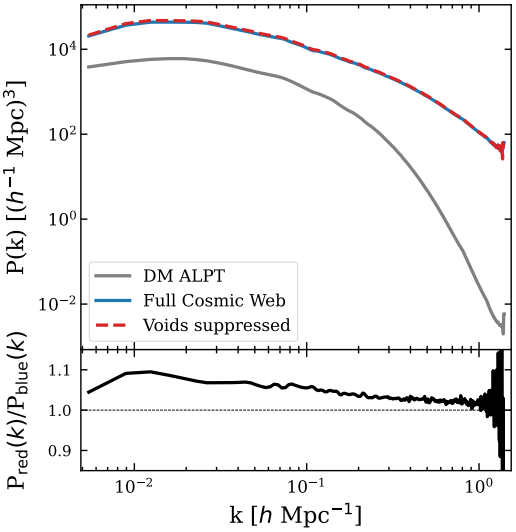

Figure 13: Power spectrum comparison when the voids are suppressed. The amplitude of the clustering increases when the low-density regions in the cosmic web are avoided. The bottom panel shows the ratio of such difference with the grey shaded area representing deviation of 5%.

5 Reference catalogue based on SHAM

The mock catalogue we use as reference has been generated by G. Favole by coupling the SHAM method (see e.g. Favole et al. 2021)) with the UNITSIM N-body cosmological simulation. In what follows, we will refer to it either as 'reference catalogue' or 'SHAM catalogue'. The SHAM prescription is based on the simple assumption that more luminous (or massive) quasars reside in more massive haloes with higher circular velocities. We choose the luminosity and the maximum circular velocity as proxies for quasars and haloes, respectively, we rank order them and perform the matching allowing a fixed Gaussian scatter between them to be physical. The quasar luminosities are drawn from the fit to the observed luminosity function computed by Pei (1995). The quasar number density in the SHAM catalogue is fixed to the DESI target selection value at redshift $z = 1.5$ from Chaussidon et al. (2023). In their Figure 20, they show a target density of 6 $deg^{-2}/(0.05\Delta z)$ at $z = 1.5$. This value, converted into comoving galaxy number density, considering a total area of 14000 deg^2 in the UNITSIM reference cosmology is $7.8{\times}10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$.

The SHAM catalogue covers the same volume of the dark matter simulated box (1000 h^{-1} Mpc side), it contains the halo positions (x, y, z), peculiar velocities (v_x, v_y, v_z), maximum circular velocities, virial masses M_{vir}, M_{200}, as well as the luminosities of 780000 mock quasars. As we are interested in densities and over densities, we need to count the number of objects in each cell. Therefore, we calculate an index for each spatial direction as, e.g. for x direction, $i_x = x/r$ where r is the resolution of our simulation (3.9 h^{-1} Mpc per cell). The index of the global 1D array is just $i_z + 256 i_y + 256^2 i_x$. Iterating over all the positions of the SHAM catalogue we count the number of objects in each cell. In Figure 14 we show how quasars are distributed in the over density field (left panel) together with the PDF and the power spectrum (right panel). There is clearly a bias between the obtained power spectrum and the reference $P_{\rm DM}(k)$. In fact, the large scale bias for the reference catalogue is $b = 2.15$. The PDF of the mock catalogue shows that most of the cells have a small number of objects and just a few of them contain 5, 6 or 7 objects maximum.

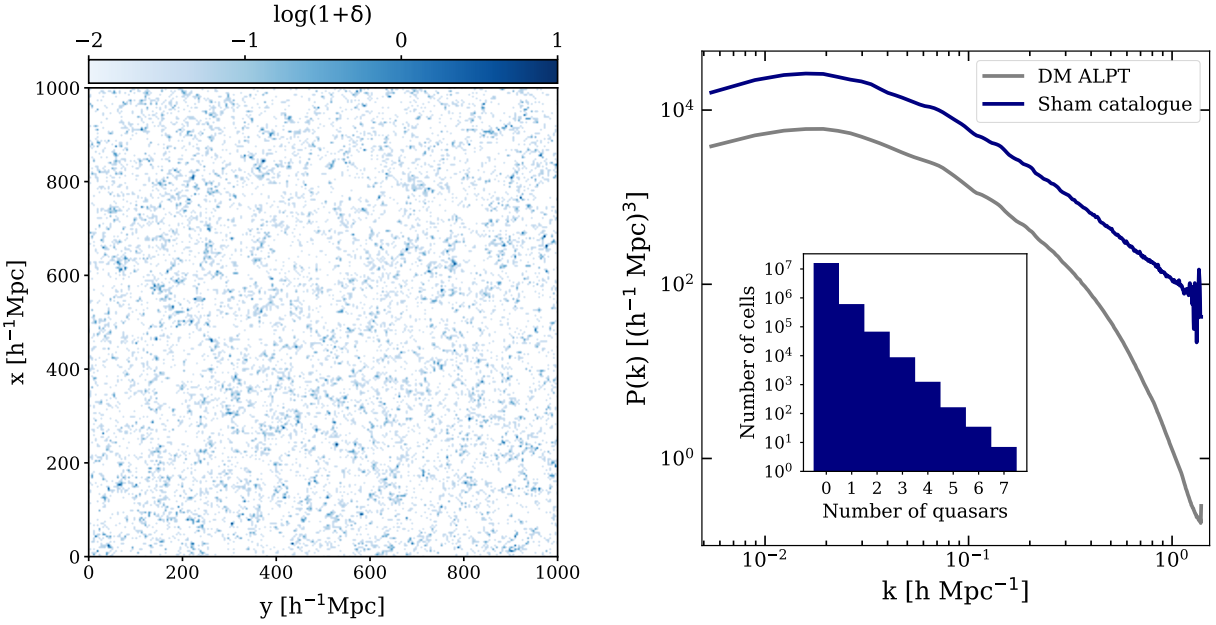

Figure 14: SHAM mock catalogue for quasars. *Left*: Quasar log-density slice of thickness 20 h^{-1} Mpc and 1000 h^{-1} Mpc side of the reference mock catalogue. *Right*: Comparison between the power spectrum of quasars (blue line) and dark matter (grey line). The inset plot shows the PDF.

6 Results: constraining the bias parameters of the model

In Section 4, we have proposed an analytical model with four free parameters $\{\alpha, \beta, \delta_{th}, \delta_\epsilon\}$ (or six in the case of adding a fine-tuning to the low- and high-thresholds), based both on galaxy properties in general, and quasar features specifically, to populate the dark matter density field with quasars. We now aim at constraining the bias parameters to understand whether all of them are necessary through the reproduction of both one- and two-point statistics of the reference SHAM catalogue.

We find that after applying our hierarchical cosmic web classification, and fixing the specific number density in each region, the bias prescription can be reduced into a two parameter model $\{\alpha, \beta\}$, respectively controlling the non-local behaviour and the deviations from Poissonity. The statistics up to third order of the reference catalogue can be reproduced with high fidelity. Specifically, we reproduce within a few percent accuracy (<5 per cent) the reference power spectrum with accurate univariate PDF and bispectra. Figures 15, 16 and 17 show excellent agreement between our model and the SHAM reference catalogue in the one- (PDF) and two-($P(k)$) point statistics for the different regions of the cosmic web. We first (figure 15) reproduce the reference catalogue based on the ϕ-Tweb classification, that separates the 780000 objects in knots (28.3%), filaments (58.5%), sheets (12.9%) and voids (0.3%). We achieve very high precision in our results: the ratios of the power spectra almost lie within the 5 per cent confidence region (grey shaded area) and the PDFs are almost identical to the reference. Note that the large-scale (i.e. small wave numbers k) deviations in the $P(k)$ ratios are a consequence of the cosmic variance and can be smoothed by generating a high number of realizations. On very small scales (i.e. $k > 1\,h\,\mathrm{Mpc}^{-1}$) we can appreciate some noise due to the limited number of objects. The knots, as classified by the gravitational potential, can be divided based on the over density in knots (67.1%), filaments (30.6%), sheets (2.3%) and voids (0.01%). The same subclassification is done for the filaments obtaining a subdivision in knots (24%), filaments (63.5%), sheets (11.8%) and voids (0.1%). Figures 16 and 17 show the statistical results of such subclassifications, i.e. the small scale cosmic web where the ϕ-knots and ϕ-filaments are both divided into δ-knots, δ-filaments and δ-sheets. We have here achieve also to reproduce the reference catalogue with percentage accuracy.

The parameters α and β obtained to reproduce this results are summarized in Table 1. We have considered the previous results of Section 4 to check which one fits better the reference. The ϕ-Tweb classification presents soft dependencies on the dark matter density for knots and filaments ($\alpha < 1$), while strong correlations are found for sheets and voids. The α parameter has suffer have flatten in the high density regions (knots and filaments) with respect to sheets and voids. This behaviour could illustrate the quenching effect in star formation rates. Calvo et al. (2019) show how the interaction of a galaxy with a cosmic web feature, e.g. the accretion of a galaxy by a filament, could perturb the gas trajectory and mark a turning point in its star formation history. Regarding the deviations from Poissonity, we see large deviations (i.e. small β values) in all cosmic web types, based on the potential, except for sheets, which tend to be Poissonian. On the other hand, when the small-scale subclassification is performed, all cosmic web types behave in a Poissonian fashion.

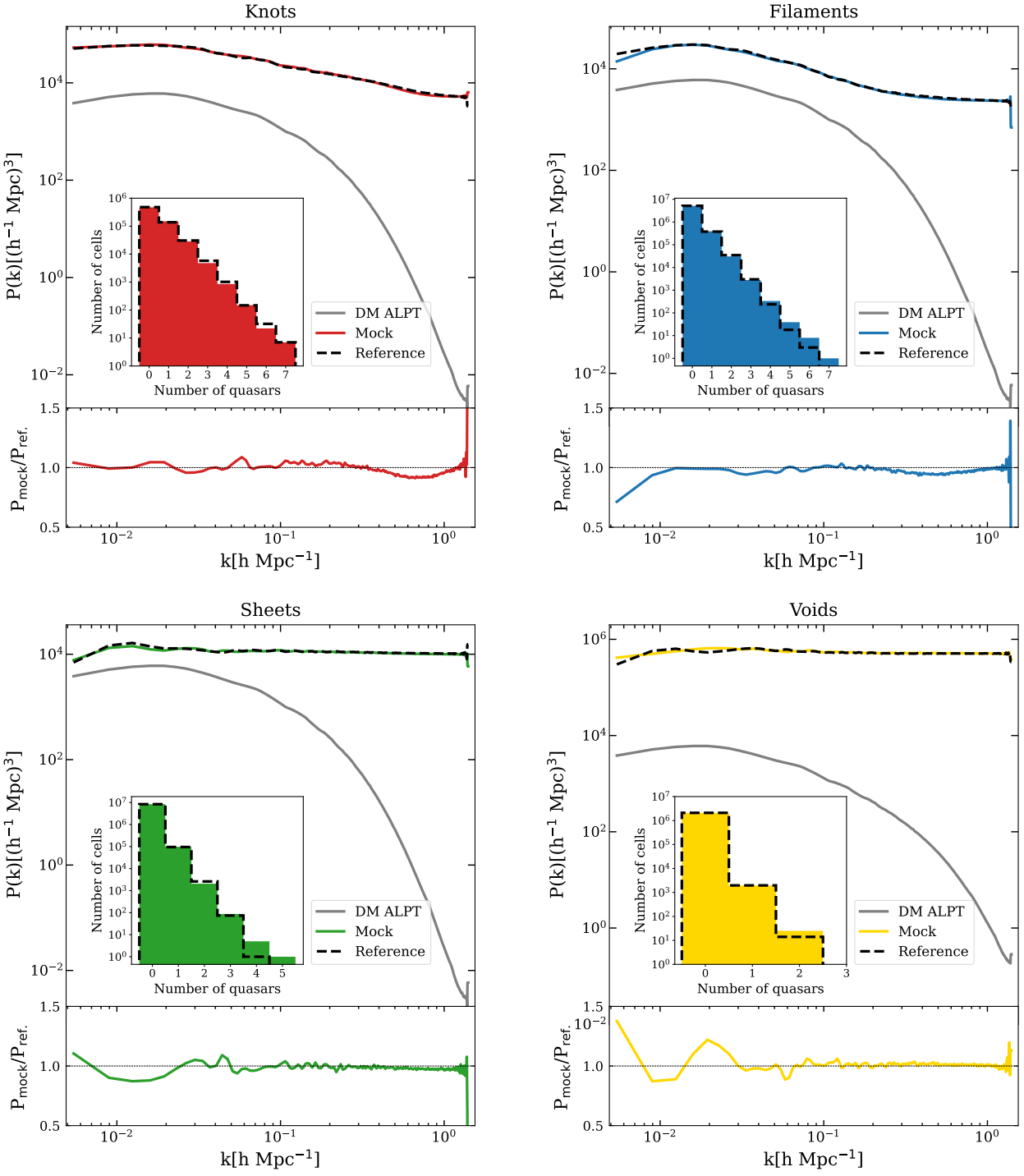

Figure 15: ϕ-Tweb classification. Comparison of the mock and the reference power spectra and PDF functions (inset plots) for the quasars in knots (28.3%), filaments (58.5%), sheets (12.9 %) and voids (0.25 %) classified by the potential. The grey lines are the dark matter power spectra all at the same redshift ($z = 1.5$). In each plot, the bottom panel shows the ratio between the mock and the reference power spectra, which in all cases lies almost within the 5 per cent deviation region (grey shadow area). Fluctuations on large scales (k <0.02 h Mpc^{-1}) are a consequence of cosmic variance. The free parameters α and β used in these results are summarized in Table 1.

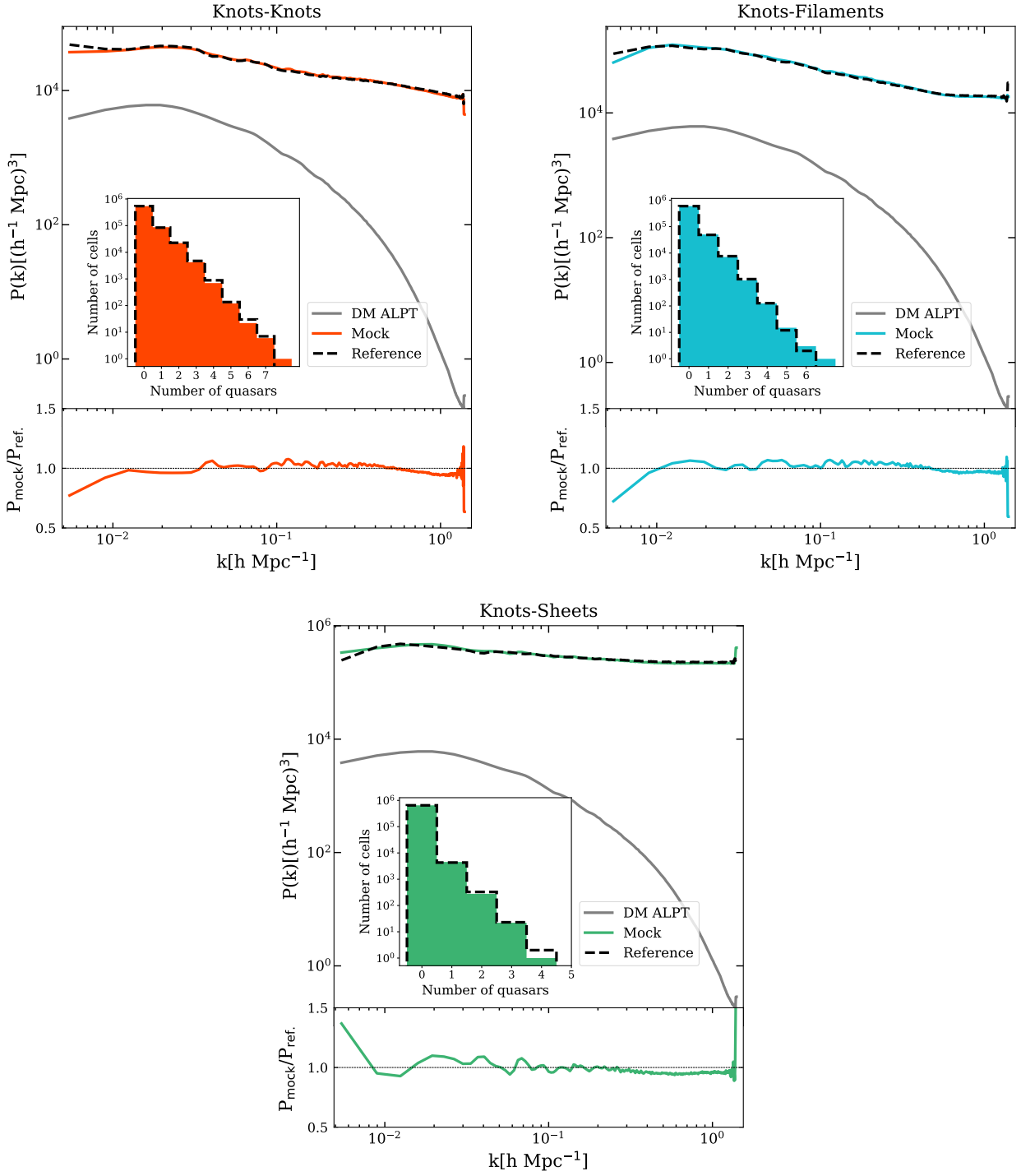

Figure 16: δ-Tweb subclassification. We compare the mock against the reference power spectra and PDF functions (inset plots) for the quasar subclassification inside the ϕ-knots into δ-knots (67.1%), -filaments (30.6 %) and -sheets (2.3 %). Voids are omitted as they constitute a negligible part ($\sim$ 0.01 per cent) of the subclassification. The grey line is the dark matter power spectrum at the same redshift ($z = 1.5$). In each plot, the bottom panel shows the ratio between the mock and the reference power spectra, which in all cases lies almost within the 5 per cent region (grey shadow). Fluctuations on large scales are due to cosmic variance. The free parameters α and β used in this plot are reported in Table 1.

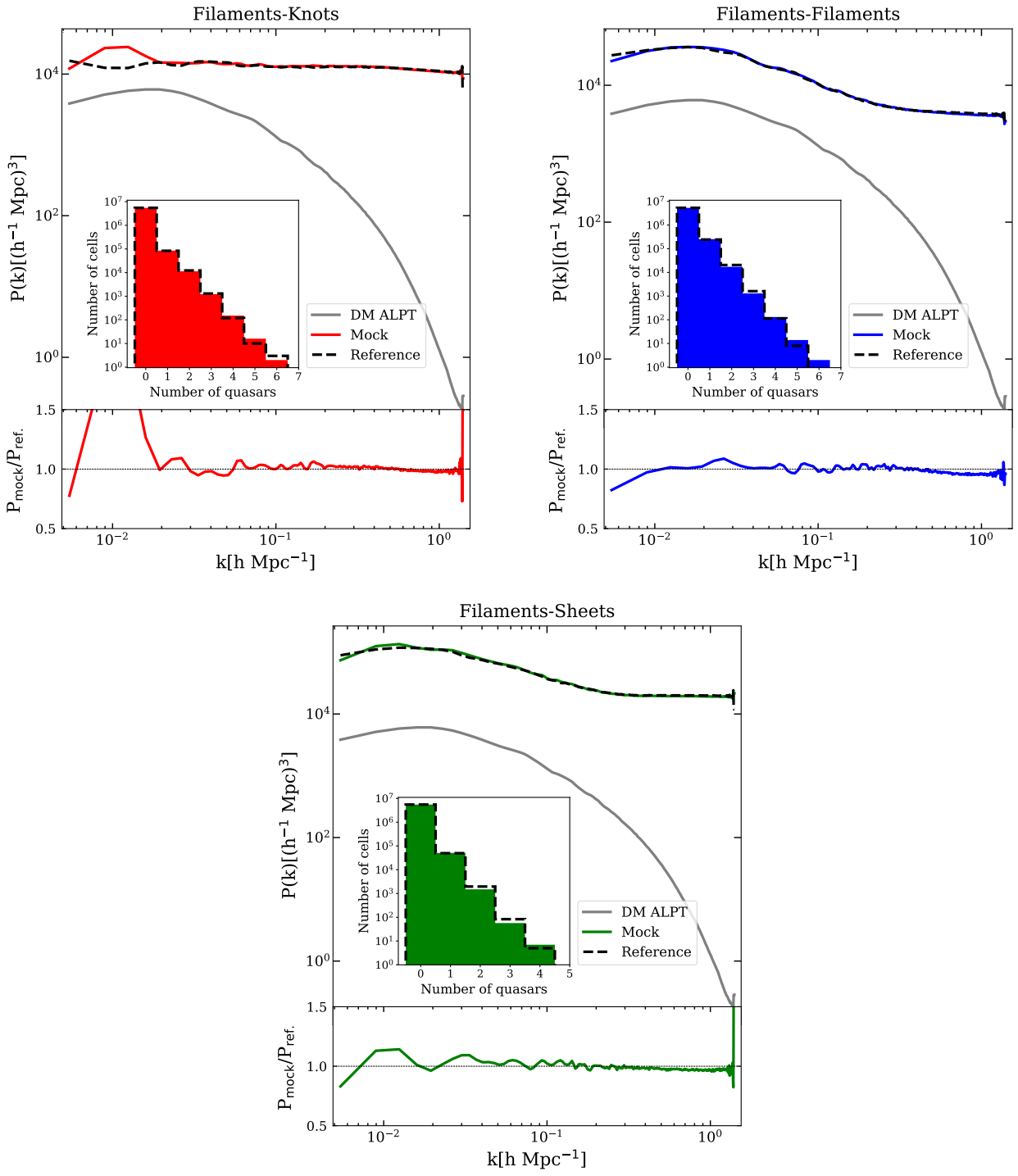

Figure 17: Analogous plot to figure 16 now for the the quasar subclassification inside the ϕ-filaments into δ-knots (24.6%), -filaments (63.5 %) and -sheets (11.8 %). Voids are also omitted here ($\sim 0.1\%$)

		Knots	Filaments	Sheets	Voids
ϕ-Tweb	α	0.58	0.9	3	5
	β	7.8	1.6	700	5
δ-Tweb (knots)	α	0.85	1.3	2.2	-
	β	10^4	10^4	10^4	-
δ-Tweb (filaments)	α	1.1	1.6	2.5	-
	β	10^4	10^4	10^4	-

Table 1: Non-local bias , α, and Poissonian deviation , β, parameters to reproduce the reference catalogue in the ϕ-Tweb classification and in the δ-Tweb subclassification of the cosmic web with percentage accuracy.

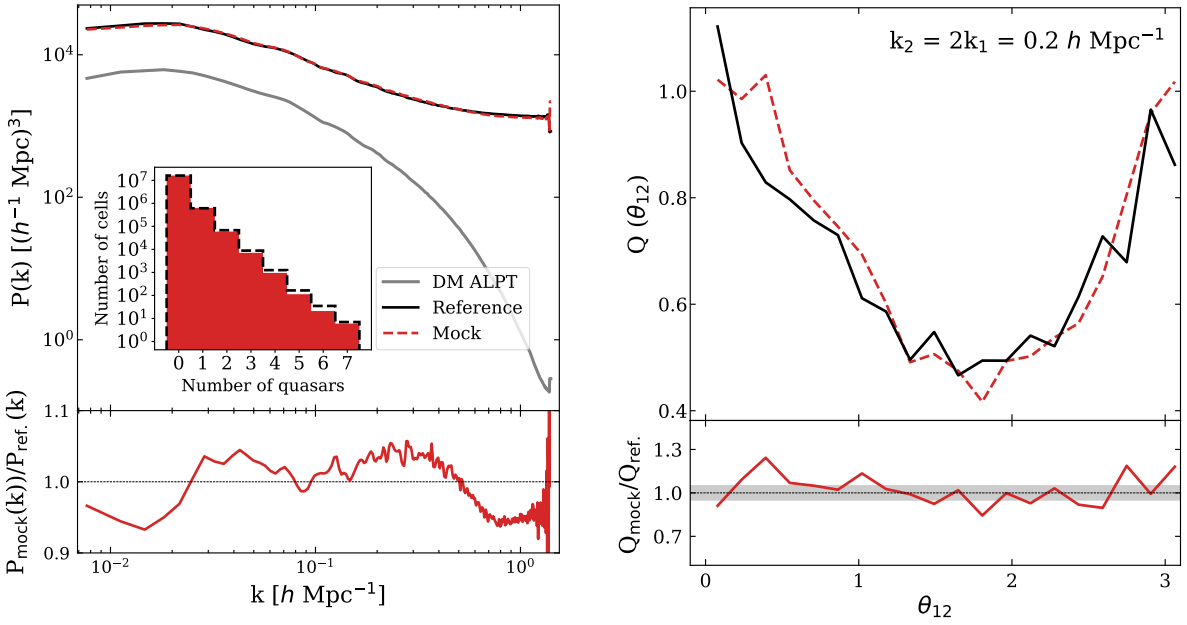

Figure 18: Global statistics up to third order: the PDF(inset left plot), the power spectrum (left plot) and the bispectrum (right plot) computed with table 1 parameters reproduce the reference catalogue with percentage precision.

7 Summary and conclusions

In this work, we have reviewed how solutions to the Navier-Stokes equation describing the motion of the dark matter as a fluid can be obtained through perturbation theory in a Lagrangian perspective. The displacement field, object of interest of this theory, describes the evolution of initial Lagrangian coordinates to evolved, at any epoch, Eulerian positions which define the a dark matter density field. Such displacement field is separated into a long-range component relying on second order Lagrangian Perturbation Theory(LPT), and in a short-range component based on spherical collapse. This method, proposed and dubbed ALPT by Kitaura & Heß (2013), has been used to simulate a dark matter density field using a code written by F-S. Kitaura. A dark matter box of L = $1000\,h^{-1}$ Mpc on a side has been created and populated with quasars at redshift $z = 1.5$. The dark matter structures have been classified in knots, filaments, sheets and voids based on the eigenvalues of the tidal tensor (and thus on the derivatives of the gravitational potential). Analogously, we have proposed a classification based on the density field which provides small-scale information (see figure 2).

The perturbation theory is also used to write the galaxy density as an infinite expansion of the dark matter density. A complete bias prescription, following the work of McDonald & Roy (2009), can be obtained up to third order with local and non-local terms. Based on this theory, we propose a non-local and stochastic bias model with 4 free-parameters $\{\alpha, \delta_{th}, \delta_\epsilon, \beta\}$ (also two exponential fine-tuning parameters can be added) and the number density $\bar{N}$, which is fixed in order to have good statistics. The effect of each parameter in the power spectrum and in the field distribution leads us to the following conclusions:

- The exponent α, which controls the non-local bias, grows linearly with the large-scale bias. A stronger dependence on the dark matter density (higher α), makes objects cluster in high density regions increasing the power spectrum but maintaining its shape.
- The fact that a minimum mass is needed to form a galaxy can be modelled by avoiding the under dense regions below a certain threshold $\delta_{\rm th}$. We find that the clustering increases with $\delta_{\rm th}$. Avoiding such regions, we exclusively force the objects in the density peaks to survive.
- The possible connection between the distribution of quasars with ELGs suggests to include an exponential suppression of the highest density peaks (i.e. above a certain limit δ_ϵ), where they are suppose not to not. We find that the bias increases with this limit, i.e., a higher clustering when less density peaks are suppressed.
- The low- and high- thresholds $\delta_{\rm th}$, δ_ϵ have a strong physical background, but lead to a very sharp description of the quasar density (9). We therefore propose to model both effects with two-parameter exponential functions, which provide soft decays both in low- and high-density regimes. The minimum mass to form a galaxy and the distribution of quasars in the high-density peaks are probably not abrupt phenomena.
- The normalization of the bias model controls the number density of the distribution. Small number densities provide a rise of the bias and a deformation of the power spectrum shape due to the shot-noise. Precision statistics require a minimum number of objects. We also analyse the effect of the shot-noise correction of the power-spectrum due to the Fourier analysis. An inversely proportional component to the number of objects must be substracted to the computed power spectrum and generates important distortions on small scales (19).
- The comparison between the Poisson and the NB distributions leads us to see that high deviations from Poissonity, small β, significantly increase the bias towards small scales. Meanwhile, values of $\beta > 500$ tend to preserve the Poisson behaviour which maintains the shape of the dark matter power spectrum (Figure 11).
- We consider avoiding the distribution of quasars in voids based on the fact that galaxy mergers are supposed to be the main mechanism to ignite quasars. The power spectrum increases above 5 per cent on scales k <0.1 $h\,\mathrm{Mpc}^{-1}$, when the objects are distributed between knots, filaments and sheets instead of considering the full cosmic web.

The importance of the analyzed bias model is that it can be generalized and adapted to other tracers, in particular to emission-line galaxies (ELGs). Understanding the impact of the different parameters on our results is the main driver of our fitting routine applied to the PDF and the power spectrum of the reference catalogue. The latter was built on a state-of-the-art SHAM prescription applied to the UNITSIM N-body cosmological simulation. After some attempts of reproducing the reference catalogue with the different parameters of the bias model, we achieved the following conclusions:

- A simple model controlling non-locality and stochasticity (α and β) is enough to reproduce with high precision the one- and two-point statistics if the cosmic web is separated in knots, filaments, sheets and voids accordingly to the gravitational potential ϕ.
- We find α and β values that fit the the power spectrum at almost all scales within 5% fidelity and the PDF for the ϕ-Tweb (figure 15) is in excellent agreement with the reference.
- We propose a new cosmic web classification based on the dark matter density field: the δ-Tweb. In figure 2 we show how this classification provides small-scale information which is not present on the ϕ-Tweb.
- The δ-Tweb allows us to subclassify the knots and filaments of the ϕ-Tweb, where most of the quasars of the reference simulation reside ($\sim$ 87 per cent), obtaining α and β parameters such that the respective power spectra are still reproduced with percentage accuracy ($\sim$ 5 per cent) (see Figures 16 and 17).
- Combining the different parameters we find, that are summarized in Table 1, we reproduce the global power spectrum of the reference catalogue within 5 per cent deviation on all scales (see Fig. 18). Note that the small discrepancy on large scales is due to cosmic variance. The global PDF and bispectrum are also matched with high precision.

We have managed to deconstruct assembly bias in modelling quasars and ELG-like galaxies. The parameters (table 1) providing this excellent agreement with the reference catalogue can be interpreted as follows:

- In the ϕ-Tweb, the dependence of the expected number counts of objects on the dark matter density is very strong in voids, while it flattens out in knots and filaments (α <1). This illustrates the quenching effect (Calvo et al. 2019) in the star formation rates of galaxies when entering the high density regions of the cosmic web.
- The distribution of objects present a vanishing deviation from Poissonity when small scales are classified in terms of the δ-Tweb.

We can foresee further improvements continuing the cosmic web classification with additional terms based, e.g., on second order non-local bias. The results we have presented in this master thesis are directly relevant to perform cosmological analysis of galaxy surveys such as DESI and EUCLID.

References

Ata, M., Baumgarten, F., Bautista, J., et al. 2017, Monthly Notices of the Royal Astronomical Society, 473, 4773

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, , 304, 15

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, Physics Reports, 367, 1

Buchert, T., Melott, A. L., & Weiss, A. G. 1993 [[arXiv]astro-ph/9309056]

Calvo, M. A. A., Neyrinck, M. C., & Silk, J. 2019, The Open Journal of Astrophysics, 2

Catelan, P. 1995, Monthly Notices of the Royal Astronomical Society, 276, 115

Chaussidon, E., Yè che, C., Palanque-Delabrouille, N., et al. 2023, The Astrophysical Journal, 944, 107

Conroy, C. & White, M. 2012, The Astrophysical Journal, 762, 70

Croom, S., Boyle, B., Shanks, T., et al. 2005, Monthly Notices of the Royal Astronomical Society, 356, 415

de la Torre, S. & Peacock, J. A. 2013, Monthly Notices of the Royal Astronomical Society, 435, 743

Favole, G., Comparat, J., Prada, F., et al. 2016, Monthly Notices of the Royal Astronomical Society, 461, 3421

Favole, G., Montero-Dorta, A. D., Artale, M. C., et al. 2021, Monthly Notices of the Royal Astronomical Society, 509, 1614

Forero–Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, Monthly Notices of the Royal Astronomical Society, 396, 1815

Fosalba, P. & Gaztañ aga, E. 1998, Monthly Notices of the Royal Astronomical Society, 301, 503

Fry, J. N. & Gaztanaga, E. 1993, The Astrophysical Journal, 413, 447

Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007, Monthly Notices of the Royal Astronomical Society, 375, 489

Jing, Y. P. 2005, The Astrophysical Journal, 620, 559

Kaiser, N. 1984, , 284, L9

Kitaura, F.-S., Ata, M., Rodríguez-Torres, S. A., et al. 2020, Monthly Notices of the Royal Astronomical Society, 502, 3456

Kitaura, F.-S., Balaguera-Antolí nez, A., Sinigaglia, F., & Pellejero-Ibáñez, M. 2022, Monthly Notices of the Royal Astronomical Society, 512, 2245

Kitaura, F.-S. & Heß, S. 2013, Monthly Notices of the Royal Astronomical Society: Letters, 435, L78

Kitaura, F. S., Sinigaglia, F., Balaguera-Antolínez, A., & Favole, G. 2023, arXiv e-prints, arXiv:2301.03648

Kitaura, F.-S., Yepes, G., & Prada, F. 2013, Monthly Notices of the Royal Astronomical Society: Letters, 439, L21

Laurent, P., Eftekharzadeh, S., Goff, J.-M. L., et al. 2017, Journal of Cosmology and Astroparticle Physics, 2017, 017

Levi, M., Bebek, C., Beers, T., et al. 2013, The DESI Experiment, a whitepaper for Snowmass 2013

Mann, R. G., Peacock, J. A., & Heavens, A. F. 1998, Monthly Notices of the Royal Astronomical Society, 293, 209

McDonald, P. & Roy, A. 2009, Journal of Cosmology and Astroparticle Physics, 2009, 020

Narayanan, V. K., Berlind, A. A., & Weinberg, D. H. 2000, The Astrophysical Journal, 528, 1

Neyrinck, M., Aragón-calvo, M., Jeong, D., & Wang, X. 2014, Monthly Notices of the Royal Astronomical Society, 441, 646

Orsi, Á. A. & Angulo, R. E. 2018, Monthly Notices of the Royal Astronomical Society, 475, 2530

Peacock, J. A. 1998, Cosmological Physics (Cambridge University Press)

Peacock, J. A. & Dodds, S. J. 1994, , 267, 1020

Peebles, P. J. E. 1980, The large-scale structure of the universe

Peebles, P. J. E. 1982, , 263, L1

Pei, Y. C. 1995, , 438, 623

Pierce, J. C. S., Tadhunter, C., Ramos Almeida, C., et al. 2023, Monthly Notices of the Royal Astronomical Society, 522, 1736

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6

Salpeter, E. E. 1964, , 140, 796

Sinigaglia, F., Kitaura, F.-S., Balaguera-Antolínez, A., et al. 2022, The Astrophysical Journal, 927, 230

Sorini, D., Davé, R., & Anglés-Alcázar, D. 2020, , 499, 2760

Uchiyama, H., Toshikawa, J., Kashikawa, N., et al. 2017, Publications of the Astronomical Society of Japan, 70

van de Weygaert, R. & Schaap, W. 2008, in Data Analysis in Cosmology (Springer Berlin Heidelberg), 291–413

White, M., Myers, A. D., Ross, N. P., et al. 2012, Monthly Notices of the Royal Astronomical Society, 424, 933

White, S. D. M. & Rees, M. J. 1978, , 183, 341

Yè che, C., Palanque-Delabrouille, N., Claveau, C.-A., et al. 2020, Research Notes of the AAS, 4, 179

Zel'dovich, Y. B. 1970, , 5, 84

A Large scale statistics

We briefly describe here some statistical concepts to analyse the large-scale structure which are use in this work and are fundamental to its understanding. For further analysis see classical references as Peebles (1980) or Peacock (1998). Let us start by introducing the concept of density perturbation field or, as named in the text, the over density field:

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \langle\rho\rangle}{\langle\rho\rangle}. \tag{35}$$

The cosmological principle suggest that the statistical properties of δ should be homogeneous at large-scales. The Fourier analysis is the preferred tool to build up the overdensty field by the superposition of many modes, i.e., by a superposition of plane waves. It is specially convenient for a flat comoving geometry but can be also used in curved space. Considering the universe is periodic in some large rectangular volume of side L, then

$$\delta(\mathbf{x}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}. \tag{36}$$

By periodicity, the wavenumbers take the values

$$k_{\mathrm{x}} = \mathrm{n}\frac{2\pi}{\mathrm{L}}, \qquad \mathrm{n} = 1, 2, ... \tag{37}$$

and exactly the same in the y and z directions. The power spectrum $P(k)$ is related to $\delta(\mathbf{k})$ by

$$P(\mathbf{k}) \equiv \langle|\delta_{\mathbf{k}}(\mathbf{k})|^2\rangle \tag{38}$$

where $\langle ...\rangle$ is an ensemble average. Essentially, defines the variance of the amplitudes at a given scale k. In our physical context, where δ_k is the over density field in Fourier space, the power spectrum on large scale is a direct measure of the primordial density fluctuations and, at small scales, gives information of non-linear evolution in the growth of structures. It serves to distinguish between different theoretical models. Note that in an isotropic universe the power spectrum must satisfy $\langle|\delta_{\mathbf{k}}(\mathbf{k})|^2\rangle = |\delta_{\mathrm{k}}(\mathrm{k})|^2$.

In the limit of large volume, the sum in equation 36 changes to an integral and the relation between the fields in configuration and in Fourier spaces are defined by the well known Fourier transform:

$$\begin{aligned} \delta(\mathbf{x}) &= \left(\frac{\mathrm{L}}{2\pi}\right)^{\mathrm{n}} \int \delta_{\mathbf{k}}(\mathbf{k}) \exp\left(-\mathrm{i}\mathbf{k}\cdot\mathbf{x}\right) \mathrm{d}^{\mathrm{n}}\mathrm{k}, \\ \delta_{\mathbf{k}}(\mathbf{k}) &= \left(\frac{1}{\mathrm{L}}\right)^{\mathrm{n}} \int \delta(\mathbf{x}) \exp\left(\mathrm{i}\mathbf{k}\cdot\mathbf{x}\right) \mathrm{d}^{\mathrm{n}}\mathrm{x}. \end{aligned} \tag{39}$$

The two-point correlation function $\xi(r)$ of the density field, usually defined as the excess probability of finding an object at a distance r from a chosen random one in the ensemble, is

$$\xi(r) \equiv \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle, \tag{40}$$

By homogeneity and isotropy, it depends only on the module of the distance r. Using equation 36, we can write last equation in terms of the Fourier expansions,

$$\xi(r) = \left\langle \sum_k \sum_{k'} \delta_{\mathbf{k}} \delta^*_{\mathbf{k}'} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{x}} e^{-i(\mathbf{k}\cdot\mathbf{r}\cdot\mathbf{x}} \right\rangle. \tag{41}$$

The quantities $\delta_{\mathbf{k}}$ are complex random varibles but $\delta(\mathbf{x})$ is real and therefore $\delta(\mathbf{k}) = \delta^*(-\mathbf{k})$. Periodic conditions made cross terms $\mathbf{k}' \neq \mathbf{k}$ average to zero. Finally, changing the sum by an integral we get

$$\xi(r) = \frac{V}{(2\pi)^3} \int |\delta_{\mathbf{k}}|^2 e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k. \tag{42}$$

One can also take ensemble averages at each side of the equation and the relation holds. Therefore, the correlation function is the Fourier transform of the power spectrum.

Just as the power spectrum is the Fourier transform of the autocorrelation function, the bispectrum B is defined as the three-point correlation function in Fourier space:

$$B(\mathbf{k}_1\mathbf{k}_2) \equiv \langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle \tag{43}$$

where the k-vectors must satisfy $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$. Instead of the bispectrum, the reduced bispectrum Q has been historically used and is defined as

$$Q(\theta_{12}|\mathbf{k}_1, \mathbf{k}_2) \equiv \frac{B(\mathbf{k}_1, \mathbf{k}_2)}{P(k_1)P(k_2) + P(k_2)P(k_3) + P(k_1)P(k_3)} \tag{44}$$

where θ_{12} is the angle between $\mathbf{k}_1$ and $\mathbf{k}_2$. In LSS context, the bispectrum is usually computed to search for nonlinear interactions between galaxies.

A.1 The shot noise in the power spectrum

We are measuring power spectra for simulation data, where the continuous density field $\rho(\mathbf{x})$ is sampled by a discrete number of objects N. As a consequence, there is a noise component arising from the Poisson shot noise. Its effect is to introduce an additional term $1/N$ to the true power spectrum such that

$$\langle |\delta^d(\mathbf{k})|^2\rangle = \langle |\delta(\mathbf{k})|^2\rangle + \frac{1}{\mathrm{N}}, \tag{45}$$

where the superscript d refers to the measured power spectrum. This additional term must be removed. In the case of using the Fast Fourier Transform (FFT) to compute $P(k)$, Jing (2005) demonstrate that a normalization constant C is introduced to the shot noise term $1/N$. We define such normalization constant as $C = (256^3)^2$, i.e., in terms of the grid size of the simulation, and remove this contribution to the FFT computed $P(k)$. Figure 19 shows the effect of removing the shot noise to the computed power spectrum.

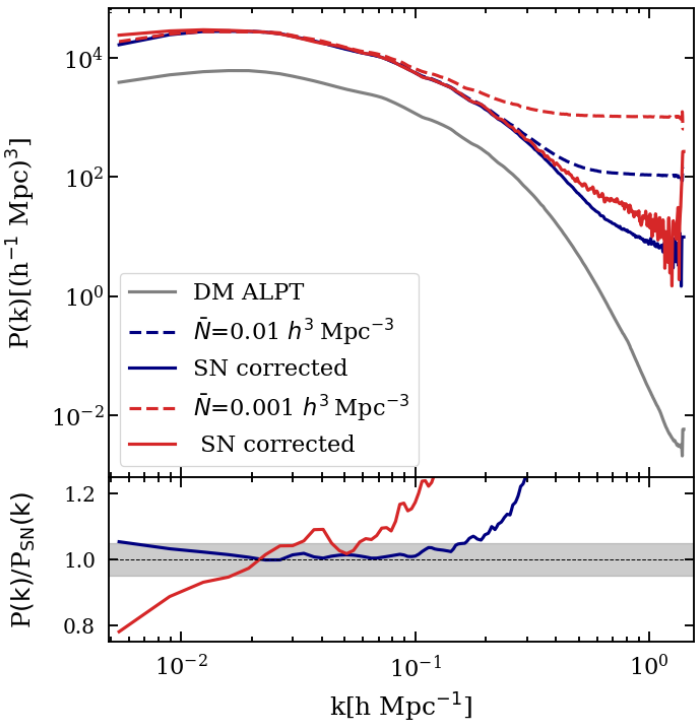

Figure 19: The effect of the shot-noise correction in the power spectrum for densities $\bar{N} = 0.01$ and $\bar{N} = 0.001$. Smaller densities have a greater impact as the deviation goes as $1/N$. The blue line (highest density) shows a discrepancy just for the smallest scales meanwhile the red one (smallest density) shows non correlation in all scales.