A new Monte Carlo approach for galaxy clustering analysis

Philippe Baratta

To cite this version:

Philippe Baratta. A new Monte Carlo approach for galaxy clustering analysis. Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]. Aix-Marseille Université; Physique et Sciences de la Matière (352), 2021. English. NNT : . tel-03597229

HAL Id: tel-03597229
https://hal.science/tel-03597229

Submitted on 4 Mar 2022

Aix*Marseille université
Initiative d'excellence

NNT/NL : 2021AIXM0029/005ED352

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 15 janvier 2021 par

Philippe BARATTA

A new Monte Carlo approach for galaxy clustering analysis

Discipline
Physique et Sciences de la Matière

Spécialité
Astrophysique et Cosmologie

École doctorale
ED352

Laboratoires
Centre de Physique des Particules de Marseille
Centre de Physique Théorique

Composition du jury

Alain Blanchard IRAP, Toulouse	Rapporteur
Sandrine Codis IAP, Paris	Rapporteure
Cristinel Diaconu CPPM, Marseille	Examinateur
Jean-Paul Kneib EPFL, Lausanne	Examinateur
Christian Marinoni CPT, Marseille	Examinateur
Valeria Pettorino CEA, Gif-sur-Yvette	Examinatrice
Julien Bel CPT, Marseille	Directeur de thèse
Anne Ealet IP2I, Villeurbanne	Directrice de thèse

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE
CPPM

CPT
Centre de Physique Théorique

euclid

Je soussigné, Philippe Baratta, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Julien Bel et Anne Ealet, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 15 janvier 2021

Résumé en français

La cosmologie moderne se donne pour objectifs de décrire la gravitation et d'identifier le contenu énergétique de l'univers. Majoritairement fondé sur la théorie de la relativité générale, le modèle standard de la cosmologie présente une vision cohérente de l'évolution de l'univers et de ses structures, des âges primordiaux jusqu'à aujourd'hui.

Néanmoins quelques ombres obscurcissent le tableau. Bien que l'on ait mis en évidence au cours du siècle dernier la présence d'un secteur sombre ayant pour effet une expansion accélérée de l'univers (énergie sombre) et un taux de structuration anormalement haut (matière noire), leur cause demeure jusqu'à présent un mystère. De nombreux modèles cosmologiques proposant diverses origines, au-delà du standard, ont été proposés dans la littérature. Chacun d'entre eux produisant des prédictions observables les différenciant le plus souvent dans des phénomènes de clustering[1] aux petites échelles. Ainsi l'objet des futurs sondages de galaxies tels qu'Euclid, LSST ou DESI est de produire une grande quantité de données précises sur une large période cosmique et intervalle d'échelles. Celles-ci pourront être utilisées dans une analyse de clustering des galaxies afin de contraindre les divers modèles cosmologiques.

Confronter les prédictions théoriques aux observations devra se faire au moyen d'analyses statistiques robustes, non biaisés, systématiques pour chaque modèle cosmologique testé, mais aussi impliquant une parfaite connaissance des erreurs sur les données (instrumentales et théoriques), appelées matrices de covariance. Prédire de telles matrices de covariance théoriques, dépendante de la cosmologie testée, constitue aujourd'hui l'une des plus grandes difficultés des analyses de sondage de galaxies. Afin de contourner le problème, les matrices de covariances sont plutôt estimées (plutôt que prédites) sur de larges échantillons de simulation de catalogue de galaxies.

Générer de tels catalogues reste une discipline nouvelle. Nécessitant une lourde infrastructure CPU et de stockage, les méthodes s'appuient majoritairement sur des simulations à N-corps. Nombreuses approches ont été développées pour accélérer le processus d'estimation, tout en gardant une haute fiabilité des covariances produites, appelées méthodes semi-analytiques. Cette thèse a donc pour objet de présenter une méthode originale d'estimation Monte Carlo rapide de matrices de covariance pour la statistique des paires de galaxies, pour un modèle cosmologique arbitraire, tout en minimisant la dépendance de la méthode aux résultats

---

[1] Cet anglicisme difficilement traduisible en français, désigne l'étude de la croissance des structures cosmiques à travers l'évolution de la distribution des galaxies.

des simulations à N-corps.

Chapitre 1

Le premier chapitre a pour objet de présenter les fondements de notre compréhension de l'univers. Basés sur la théorie de la relativité générale et additionnés d'hypothèses de symétrie (le principe cosmologique), je discute les principaux résultats décrivant l'évolution de l'univers dans son ensemble (équations de Friedmann) et leurs implications. En particulier, il apparaît que l'univers est doté d'une structure flexible, dont l'expansion ou la contraction sont dictées par le contenu énergétique de l'univers. Ces derniers, constituant en partie les paramètres cosmologiques pouvant être contraints par l'observation, renseignent directement sur l'histoire de l'univers.

Ceci étant posé, j'énonce les différents apports observationnels qui ont participé à construire notre modèle standard de la cosmologie via la mise en évidence de

- l'expansion de l'univers par l'observation de Céphéides
- d'un univers primordial chaud par la détection du fond diffus cosmologique (CMB) et de l'oscillation acoustique des baryons (BAO)
- l'existence d'une matière noire froide (CDM) dynamisant la formation des grandes structures et responsable, entre autres, d'une courbe de rotation spécifique des galaxies
- l'accélération de l'expansion de l'univers par l'analyse de supernovae de type I-a (SNI-a), menant à l'hypothèse d'une constante cosmologique Λ.

Combinées, ces diverses sondes (CMB, BAO, SNI-a) offrent une vision cohérente en convergeant vers des proportions d'environ 70%, 25% et 5% du budget énergétique total respectivement associées à Λ, à la CDM et à la matière baryonique.

Parlant donc d'un modèle de concordance ΛCDM, il est cependant à noter que quelques insuffisances persistent. La première est théorique et relève de notre inhabilité à identifier l'origine physique des deux constituants principaux : la matière noire et l'énergie sombre[2]. La seconde est davantage observationnelle. Avec le développement de nouvelles sondes, qui indépendamment mesurent les mêmes paramètres cosmologiques que les sondes historiques, je résume les différentes tensions

- sur l'incompatibilité des mesures du paramètre d'Hubble H_0 (quantifiant l'expansion actuelle de l'univers)
- sur l'incompatibilité des mesures du paramètre $\sigma_{8,0}$ (quantifiant l'amplitude de clustering actuel)
- sur la potentielle incompatibilité du paramètre de courbure de l'univers.

Ceux-ci étant source de développement théorique de modèles alternatifs, je discute les résultats principaux de la théorie effective des champs pour l'énergie sombre

---

[2] Incluant Λ, plusieurs modèles couverts du terme énergie sombre peuvent produire l'accélération cosmique observée. Une alternative est de modifier les lois gouvernant la gravitation : les modèles de gravitation modifiée.

(EFT), combinant dans un unique formalisme nombreux modèles d'énergie sombre et de gravitation modifiée dont l'accélération cosmique est conduite par un seul degré de liberté scalaire. Notamment, j'insiste sur l'imposante diversité des prédictions théoriques qui, à l'heure actuelle, ne peuvent être contraintes.

C'est pourquoi, après discussion dans un tout autre registre de l'intérêt des mesures cosmologiques sur les contraintes de la masse des neutrinos, je présente la mission spatiale Euclid. Répondant aux besoins de la cosmologie de précision, ce futur sondage de galaxies disposera de toutes les spécificités pour tester les différents modèles cosmologiques.

Chapitre 2

Le contexte étant posé, je discute dans ce second chapitre la sonde de clustering des galaxies plus en détails. Relevant de l'étude des fluctuations de densité de la matière $\delta \equiv \rho(x,t)/\bar{\rho}(t) - 1$ (ρ étant la densité locale et $\bar{\rho}$ la densité moyenne de la matière), je décris en premier lieu le paradigme inflationnaire à l'origine de ces inhomogénéités. Ensuite je décris leur évolution dans un univers jeune et chaud gouverné par des phénomènes radiatifs dominants, et responsables d'oscillations dans le plasma primordial (BAO). Une fois la domination énergétique laissant place à la matière non relativiste, j'étudie l'évolution de δ aux échelles linéaires, insistant sur la difficulté de résoudre les équations d'évolution non linéaires aux petites échelles.

Je montre par la suite l'intérêt d'une étude statistique des fluctuations de densité via la construction du spectre de puissance en espace de Fourier $\delta^{\mathcal{D}}(k+k')P(k) = \langle \delta_k \delta_{k'} \rangle$, dont l'estimation permet de contraindre la majorité des paramètres cosmologiques.

Confronter les prédictions théoriques d'une telle sonde avec les données observées n'est pas trivial. Ainsi je présente la méthode d'inférence bayésienne, capable de déterminer l'ensemble des paramètres cosmologiques décrivant au mieux les données, au sein d'un modèle cosmologique testé. Pour ce faire, cela requiert l'estimation d'une figure de vraisemblance (likelihood) dépendante d'une matrice de covariance décrivant les erreurs théoriques et instrumentales associées aux données et à l'observable analysée.

Me focalisant ainsi sur la prédiction de la matrice de covariance théorique associée au spectre de puissance, je dresse la problématique de cette thèse. En effet, sa prédiction via une approche totalement analytique apparaît hautement complexe ; les non linéarités couplant les diverses échelles (contribution non gaussienne de la matrice de covariance) se simulent plutôt que se prédisent. Il est donc courant, voir exclusif, de recourir aux simulations à N-corps pour troquer la prédiction à l'estimation des matrices de covariance. Afin d'obtenir une estimation non bruitée et suffisamment précise, cette méthode se révèle être extrêmement gourmande en ressources numériques, si bien que l'estimation d'une matrice de covariance exploitable dans une analyse de données réalistes peut prendre, pour un seul modèle cosmologique, plusieurs mois.

Pour y remédier, cette dernière décennie a vu émerger de nombreuses méthodes approximées, ou méthodes semi-analytiques, accélérant ce processus d'estimation. Je synthétise et catégorise l'ensemble de ces alternatives en identifiant une approche peu étudiée dans la littérature. En conséquence, je propose dans ce manuscrit d'étudier l'impact du spectre de puissance et de la fonction de distribution de champ de densité (PDF) sur la forme de la matrice de covariance. Pour ce faire, la méthode à réaliser consiste à générer des catalogues de particules (galaxies, CDM, etc.) dont le spectre de puissance et la PDF du champ de densité sous-jacent soient arbitraires et contrôlés, pour conclure sur la fiabilité des covariances produites via une méthode Monte Carlo.

Chapitre 3

Dans ce chapitre, je construis une procédure théorique originale permettant de générer des champs non gaussiens dont la PDF peut être choisie analytique ou bien numérique, tout comme son spectre de puissance associé.

Pour ce qui est de la PDF, une transformation locale non linéaire sur un champ gaussien ν suffit à obtenir un champ non gaussien de contraste de densité δ. Pour que cette application cible une forme particulière de PDF, je montre que l'identification de $\mathcal{L}$ dans la relation

$$\delta = \mathcal{L}(\nu) = C_\delta^{-1}[C_\nu(\nu)] \tag{0.1}$$

(où les C_X définissent les fonctions cumulatives des deux PDF), permet d'atteindre un tel objectif.

Par la suite, mettant en évidence l'impact de $\mathcal{L}$ sur le spectre de puissance, l'idée est donc en premier lieu de générer un champ gaussien dont le spectre de puissance est correctement conçu, de telle sorte qu'une fois la transformation non linéaire $\mathcal{L}$ localement appliquée pour cibler la PDF donnée, le spectre de puissance suive lui aussi le modèle attendu. Je montre qu'un tel spectre de puissance d'entrée s'obtient via la procédure

$$P_\nu(\vec{k}) = \mathcal{F}\left\{\lambda^{-1}\mathcal{F}^{-1}\left[P_\delta(\vec{k})\right]\right\} \ , \tag{0.2}$$

où $\mathcal{F}$ symbolise la transformée de Fourier et $\mathcal{F}^{-1}$ sa fonction inverse, P_X le spectre de puissance du champ X et ou λ est défini à travers l'application $\mathcal{L}$ par

$$\xi_\delta \equiv \langle\delta_1\delta_2\rangle = \iint \mathrm{d}\nu_1\mathrm{d}\nu_2\mathcal{L}(\nu_1)\mathcal{L}(\nu_2)P_\nu(\nu_1,\nu_2,\xi_\nu) \equiv \lambda(\xi_\nu) \ , \tag{0.3}$$

avec $\xi_X \equiv \mathcal{F}^{-1}P_X$. La difficulté d'un tel formalisme réside dans la résolution de l'intégrale 2-D 0.3. Je présente en conséquence l'intérêt de l'expansion de Mehler, échangeant cette intégrale 2-D par un nombre fini (car de convergence rapide) d'intégrales 1-D (via les coefficients c_n) telles que $\xi_\delta = \sum_{n=0}^{\infty} n!c_n^2\xi_\nu^n$.

L'application numérique d'une telle procédure implique une utilisation intensive d'algorithmes de transformée de Fourier (FFT) appliqués aux divers champs échantillonnés sur grille. J'insiste donc sur l'intérêt primordial de prendre en considération dans la méthode 0.2 les biais numériques associés aux FFT, à savoir l'*aliasing*[3].

Ainsi complétée, j'applique la méthode au cas d'une PDF Log-Normale. L'intérêt d'un tel exercice réside dans le fait que λ^{-1} peut être déterminé analytiquement, facilitant l'identification d'éventuel biais de la méthode. Après discussions sur l'aspect non gaussien de la distribution des spectres de puissance produits, Je mets en évidence la fiabilité de la méthode, réalisant des PDF parfaitement contrôlées ainsi que des spectres de puissance produits avec une précision meilleure que le pourcent.

Enfin, je tente une première comparaison des matrices de covariance du spectre de puissance estimées par la méthode Monte Carlo avec une prédiction provenant de la théorie des perturbations. Cette dernière ne révélant pas de convergence, la covariance Monte Carlo produite ne peut, pour l'instant, être totalement validée.

Chapitre 4

L'objectif majeur de ce chapitre est de valider la méthode au niveau des covariances produites avec celles estimées sur un ensemble particulier de simulations à N-corps, appelées `DEMNUni_cov`.

Étant donné que la procédure 0.2 permet de générer des champs continus de contraste de densité non gaussiens échantillonnés sur la grille, je présente plusieurs procédures de discrétisation du champ δ. En effet, par la définition d'une densité moyenne $\bar{\rho}$, une grille de nombre de particules par unité de volume (par voxel de la grille) peut être reconstruite. Ces nombres n'étant pas des entiers, je leur applique une distribution de Poisson pour associer à chacun des voxels de la grille le nombre de particules à placer. Deux schémas d'interpolation sont présentés. Le premier, d'ordre 0, consiste à placer les particules dans le voxel associé suivant une distribution uniforme. La seconde, d'ordre 1 et assurant une meilleure continuité, place les particules suivant une distribution tri-linéaire définie par l'amplitude des densités aux 8 sommets de chaque voxel. Je montre ensuite que les catalogues de particules, dont la nature dépend du spectre de puissance d'entrée (galaxies, CDM, halos, etc.), sont caractérisés par des spectres de puissance altérés par ces processus de discrétisation poissonniens. Je montre qu'un tel effet peut être prédit, quel que soit le schéma d'interpolation employé, à nouveau au pourcent près.

Pour une comparaison rigoureuse entre les covariances produites par `DEMNUni_cov` et la méthode Monte Carlo, l'ensemble des réglages et données statistiques susceptibles d'influencer la forme de la matrice de covariance doivent être similaires. C'est pourquoi je clone à la fois le spectre de puissance et la PDF estimés sur les

---

[3] Cet anglicisme correspond au recouvrement de bande, largement étudié dans la théorie du traitement du signal.

simulations à N-corps, à l'aide de divers processus de filtrage. Je montre qu'en agissant ainsi, le spectre de puissance d'entrée du champ gaussien ν est produit avec une quantité non négligeable de valeurs négatives, non physiques, empêchant la procédure d'aboutir. Je propose et compare par conséquent deux méthodes correctives, permettant la production de catalogues, mais au prix d'un nouveau filtrage (dont l'impact sur l'amplitude du spectre de puissance résultant peut être prédit de façon exacte). Je montre alors que la PDF produite par 1000 catalogues Monte Carlo dans une cosmologie standard avec neutrinos massifs est parfaitement reproduite, tandis que les spectres de puissances sont compatibles à $1-\sigma$ jusqu'à $k \sim 0.2h/\mathrm{Mpc}$ pour un tel exercice.

Concernant les matrices de covariance produites, j'identifie et corrige un biais dans leur estimations, afin de les comparer avec celles de `DEMNUni_cov`. Bien que les spectres Monte Carlo soient fiables jusqu'à $k \sim 0.2h/\mathrm{Mpc}$, il apparaît que les éléments de la matrice de covariance sont eux compatibles à $1-\sigma$ jusqu'à de plus petites échelles ($k \sim 0.8h/\mathrm{Mpc}$).

L'étude précédente, réalisée en espace comobile, n'est toutefois pas adaptée à des sondages de galaxies réalistes. N'ayant accès par l'observation qu'à l'espace des redshifts, pollué par les vitesses propres des particules (via effet Doppler), la seconde partie de ce chapitre est ainsi exclusivement dédiée au traitement des vitesses dans la méthode Monte Carlo.

Je commence donc par montrer que dans un tel espace des redshifts, le spectre de puissance est distordu sous l'action des vitesses (un effet appelé RSD). Après introduction des théories sous-jacentes (modèle de Kaiser, modèle de diffusion, modèle de dispersion), j'applique ces notions pour associer à chacune des particules des catalogues produits une vitesse. Pour ce faire, j'applique une procédure similaire à 0.2 pour générer un champ de vitesses. Afin d'assurer les corrélations du champ de densité avec celui des vitesses, la subtilité réside dans le partage de phases dans l'espace de Fourier entre le champ de densité et celui du champ de divergence des vitesses $\theta \equiv \vec{\nabla} \cdot \vec{v}$. Une telle construction permet de simuler les effets RSD aux échelles linéaires, décrite par l'effet Kaiser. Pour simuler les effets non linéaires, je propose une procédure d'assignation des vitesses suivant le modèle de distribution

$$\mathcal{P}^i \equiv \mathcal{N}\left[\vec{w}(\vec{x}_i), \Sigma^2(\rho_i)\right] \ , \tag{0.4}$$

où $\mathcal{N}$ est la distribution normale, $\vec{w}$ est la valeur du champ des vitesses interpolé aux coordonnées $\vec{x}$ de la particule i considérée et $\Sigma^2(\rho)$ est la relation de dispersion des vitesses en fonction de la densité locale.

Après inspection de cette dernière relation sur les simulations à N-corps, je la modélise sous forme de loi de puissance, permettant d'imposer de façon exacte la variance du champ des vitesses discrétisé. Ainsi la procédure 0.4 permet de simuler l'effet des RSD aux échelles non linéaires, communément appelé *Finger-of-God*.

Fort de ce développement, je produis 1000 catalogues dans l'espace des redshifts dans une cosmologie standard avec neutrinos massifs et compare le spectre de puissance et tous ses sous-produits (monopole, quadrupole et hexadecapole) avec ceux produits par `DEMNUni_cov`. Il apparaît que l'effet RSD sur le spectre de puissance est bien reproduit, de façon certaine jusqu'à $k \sim 0.6h/\mathrm{Mpc}$, bien qu'à nouveau, le spectre de puissance soit fiable jusqu'aux échelles $k \sim 0.2h/\mathrm{Mpc}$.

Concernant finalement les matrices de covariance produites dans l'espace des redshifts, celles-ci présentent une haute fiabilité jusqu'à $k \sim 0.6h/\mathrm{Mpc}$, bien que des corrélations non physiques couplant les petites et grandes échelles apparaissent. Je propose une possible solution pour les corriger, actuellement sous investigation.

Chapitre 5

Pour terminer, le dernier chapitre s'extrait de l'étude des fonctions de corrélation en espace de Fourier pour se focaliser sur l'espace des harmoniques sphérique.

Ce dernier, jusqu'à présent fort peu exploité dans l'étude du clustering des galaxies, présente pourtant de nombreux avantages sur l'espace de Fourier. En effet, en rappelant que l'estimation seule du spectre de puissance nécessite de connaître la distance comobile entre les particules, un modèle cosmologique doit par conséquent être postulé. Au contraire, dans le cas des harmoniques sphérique, le spectre de puissance angulaire, appelé C_ℓ, peut être estimé au moyen d'angles d'observation et de redshifts, des observables directes indépendantes de la cosmologie. Aussi, les C_ℓ étant définis en coordonnées sphériques, une analyse combinée avec d'autres observables aussi définies sur la sphère est facilitée (CMB, WL, carte H-α, etc.).

Ce chapitre présente donc diverses méthodes d'adaptation des catalogues Monte Carlo à l'étude des C_ℓ, prenant soit la forme de méthodes de reconstruction de cônes de lumière, soit celle de réorganisation spatiale de l'information. La première méthode, appelée *shell-method*, reconstruit le cône de lumière en accolant des volumes comobiles (des coquilles) extraits de N_{shell} simulations Monte Carlo définies à des redshifts différents ($z \in [z_{\min}, z_{\max}]$). La seconde, appelée *cell-method*, d'abord génère une réalisation de champ à un redshift donné, puis l'évolue linéairement en utilisant le facteur de croissance linéaire $D(z)$ afin d'introduire une évolution temporelle. Seulement adaptée aux spectres de puissance linéaires, cette seconde méthode est développée afin de calibrer la première qui, après analyse de simulations de cônes de lumière définis par une PDF Log-Normale, présente un haut niveau de reproductibilité de spectre (<1%) pour un nombre de coquilles dépassant les $N_{\mathrm{shell}} = 200$ sur un intervalle en redshift de 0.2-0.3. Enfin, une analyse des matrices de covariance produites suggère une gaussianisation des corrélations lorsque comparées à celles en espace de Fourier.

Pour une validation directe de la méthode Monte Carlo en harmoniques sphérique et pour des spectre non linéaires, j'emploie à nouveau les catalogues `DEMNUni_cov`. Ayant a disposition des simulations très séparées dans le temps ($z \sim [0.5, 1, 1.5, 2]$), l'emploie de la shell-method s'avère proscrite. C'est pourquoi je présente une méthode de re-modélisation des catalogues donnés à un redshift (des *snapshots*), et adapté à l'estimation des C_ℓ, appelé *snap-shell method*. Je monte alors que les résultats des C_ℓ estimées et de leur matrices de covariance associées sont compatibles pour les deux méthodes, jusqu'à des échelles limites comparables à celles précédemment citées en espace de Fourier (sans tenir compte des vitesses). Un tel exercice est mené à la fois pour les auto et cross-C_ℓ (étude des corrélations autos ou croisées des coquilles).

J'ouvre ce chapitre sur la présentation d'un projet en cours, tentant d'appliquer une analyse bayésienne d'estimation de paramètres cosmologiques impliquant les C_ℓ estimés après emploi de la méthode snap-shell sur les catalogues `DEMNUni_cov`. Bien que les covariances utilisées dans cette analyse soient compatibles avec les mesures, il semble qu'une faiblesse accompagne la prédiction théorique des C_ℓ associées pour retrouver la vraie valeur des paramètres cosmologiques. Ainsi des études plus approfondies sont nécessaires.

Ouverture

Enfin, j'insiste sur les diverses potentielles applications d'une telle méthode Monte Carlo pour

- étudier comment le bruit dans l'estimation de la matrice de covariance se propage dans les erreurs d'estimations des paramètres cosmologiques
- étudier l'impact d'une distribution non gaussienne d'une observable dans une inférence cosmologique, supposant une forme de likelihood gaussienne
- produire des prévisions de matrices de covariance réalistes en agrémentant les catalogues de masques angulaires et radiaux et d'un biais liant le champ de matière à celui des galaxies, mais aussi en estimant la covariance supra relevé (SSC)
- estimer des cross-covariances entre différentes sondes telles que les fonctions de corrélation dans différents espaces (configuration, Fourier et harmoniques sphérique), leurs sous-produits (monopole, quadrupole, hexadecapole), ou, plus prometteur, estimer les corrélations entre les C_ℓ du clustering des galaxies et celui du lentillage gravitationnel.

à toi, à notre future première rencontre

Contents

Abstract

The conceptual revolution brought by Einstein's theory of General Relativity over the last century has resulted in a spectacular development of Cosmology. Being defined as the science that studies the general laws and the content of the Universe, Cosmology is from then on endowed with the field equations linking the geometry of the Universe to its content. This reinterpretation of gravitation involves a modular and dynamical structure of the Universe, allowing its expansion or contraction, but also constituting a background for new physical phenomena.

The popularity of the theory and its extensions lies in the large number of predictions successively validated. These started in 1920 with the demonstration of the curvature of light rays near a massive body by Eddington (Dyson et al., 1920), then in the mid-century by the observation of the fossil radiation by Penzias & Wilson (1965) or more recently the detection of gravitational waves, resulting from the fusion of two black holes by the Ligo and Virgo collaborations (Abbott et al., 2016).

Modern cosmology is born in the last century with some other new observables, often called cosmological probes, that have participated to build and consolidate what has become the standard model of Cosmology. With the Hubble's discovery of the expansion of the Universe (Hubble, 1929a) using Cepheids, cosmologists then focused on understanding its contents. First discoveries show the presence to a large extent of an unknown matter, only sensitive to gravitation and both inferred from cosmological and astrophysical observations (e.g. Zwicky, 1937; Rubin & Ford, 1970), called *cold dark matter*. This matter anyway is expected to exclusively exert an attractive power, that would naturally have prevailed over global dynamics. However recent observations of supernovae (Perlmutter et al., 1999; Riess et al., 1998) went far beyond this speculation: the Universe is expanding at an accelerating pace.

This unforeseen event turns to (re)introduce a cosmological constant Λ in the field equations. As this does not seem to fit with a natural solution, this has often been called *the cosmological constant problem* and opened a remarkable call for new hypotheses on the nature of such expansion. Among them the vacuum energy of quantum mechanics or a new exotic negative pressure fluid called *dark energy*. Furthermore, another possibility would be to revisit the structure of the field equations themselves (modify gravity models).

Today, this question remains an open problem even if the final picture gives a coherent model of background and cosmological structures evolution, called the ΛCDM model, which is considered so far as the standard model of Cosmology. According to it, the energy budget of the Universe contains today 95% of dark components whose the inner nature remains unknown and is one of the most intriguing question of the decade.

This is why probing and understanding the dark sector constitutes the main goals

of future large galaxy surveys like Euclid (Laureijs et al., 2011), LSST (Abate et al., 2012) or DESI (Levi et al., 2019) that have been designed to provide a large amount of accurate data able to disentangle and constrain alternative models, through galaxy clustering properties. Indeed, the spacial distribution of galaxies in the Universe represents a mine of physical information as baryonic abundance, dark matter halo distribution, neutrino mass, initial conditions, expansion rate of the Universe and so on. All of these cosmological features can be traced through the correlation function of galaxy pairs, a key probe to link theory to observation.

But even with infinitely precise and voluminous data, discriminating cosmological models is a challenging issue as it requests very specific tools well suited to the theoretical predictions. In particular the complexity of galaxy clustering data analysis mainly lies on the fact that it should combine several types of correlated probes, identifying many systematic and biases, whereas it must lay on robust statistical tools including a good estimation of the data correlations/errors. Often based on Bayesian inference, the cosmological model extraction is based on the comparison of the theoretical predictions to the data distribution by the evaluation of a likelihood, requesting an accurate *covariance matrix* (matrix of errors). Preparing these tools is one of the most complex tasks of future surveys.

As a matter of fact, the covariance matrices exploited in survey analysis today rely on a large sample of cosmological simulations. Above the fact that their construction is highly CPU and time consuming, the major problem is that the matrix is model-dependant. Different approaches have been investigated to bypass these issues, but no fully satisfactory solution has been found yet.

However semi-analytic approaches are currently developed to speed up the process of covariance matrix estimation, and significant developments are still needed. This thesis is then proposing new solutions to this challenging issue. We have developed a new method based on a fast Monte Carlo approach, where the main goal is to be able to estimate in a fast and systematic way the covariance matrix of galaxy pairs statistics, for any cosmological model and adapted to future galaxy clustering surveys.

In chapter 1, after reviewing the main pillars of the standard model of Cosmology, I will discuss the current situation that leads to consider alternative models. In particular, I will focus on data inconsistencies and tensions in the framework of the standard model. I will then describe Euclid, a space mission dedicated to measure millions of galaxies to understand the nature of dark matter and dark energy.

In chapter 2, I will describe the clustering probe, its properties on both the theoretical and observational sides. I will then detail the method to extract the cosmological information through *the matter power spectrum* estimated on data, when comparing it to its theoretical predictions. It will help me to show the needs for a reliable covariance matrix estimation, and motivate the development of a semi-analytical approach.

In chapter 3, the theoretical basics of the Monte Carlo method will be considered, notably by showing how to generate a non-Gaussian density field with a target power spectrum and distribution function (PDF), essential for the production of reliable covariance matrices. The method will be validated when considering a numerical application on Log-Normal PDF, where any bias will be identified and treated.

In chapter 4, I will extend the method by showing some discretisation procedures of these generated continuous fields. This will be done first in comoving space, then in redshift-space by presenting a numerical modelling of peculiar velocities. A method validation test on the basis of power spectrum and associated covariance matrix comparison with the ones obtained using the `DEMNUni` suit of N-body simulations (Carbone et al., 2016; Castorina et al., 2015) will be conducted in both spaces.

Finally, chapter 5 will be devoted on the adaptation of the Monte Carlo process to the harmonic space, a more convenient approach for observational data analysis. This way, a first application of the method on cosmological parameters estimation will be presented.

Chapter 1

An overview of modern cosmology

In 1900, Lord Kelvin proclaimed to the London Royal Institution the imminent end of the progress of science: "In the clear blue sky of physics there remained on the horizon just two small clouds of incomprehension that obscured the beauty and clearness". Whereas a few years later, once solved, these two clouds led to two major scientific revolutions, the Quantum Mechanics and the Special Relativity (followed by General Relativity), we are standing once again in front of a formidable description of the laws of nature, nevertheless incomplete.

In particular the standard model of Cosmology, built over the last century, is able to make robust predictions for gravitationally related phenomena, covering small to large scales on their whole cosmic ages. It is made from a collection of models, each seeking to account for distinct observational features. Among them the primordial nucleosynthesis, which succeed to predict the right abundance of elements filling the Universe, the inflationary paradigm, partly explaining the initial seed at the origin of astrophysical structures or the hypothesis of a surrounding dark sector, independently reporting the way large scale structures are evolving and the evidence for a cosmic acceleration. In that respect, it is more common to present it as the *concordance model*, relying on some solid pillars allowing to match the observation.

On the other hand, some aspects tarnish its solidity. Whether it is from the inability to understand the nature of dark energy or dark matter, which are supposed to be the major components of the universe, or the observational and persistent inconsistencies, many clouds point toward an unsuspected physics.

In this chapter, I will first give a quick overview of General Relativity and the cosmological principle (in section 1.1), from the theoretical foundations to the cosmological observations which have constantly exchange with model developments to build the current concordant cosmological standard model.

This construction is far from being linear. It took nearly 100 years and many types of observations and theoretical interpretations, gradually getting our hands on the many pieces of a cosmic puzzle. Even if there is now a large consensus on this model, which made it possible to assemble all the pieces in a coherent vision of the evolution and the content of the Universe, it is worth mentioning that the set is not complete and many questions are still opened. In particular we still don't understand the nature of the dark sector (dark matter and dark energy) and the model seems at some level to need an extension.

That is why in section 1.2, I will go through these questions and will give an overview of current observational tensions laying between what we infer from primordial and late probes of the Universe. These questions are opening the door to new interpretations and I will present some beyond standard model propositions attempting to refine our global comprehension of gravitation.

Converging towards a need for more data to reach the precision needed to answers these questions, I will finally introduce in section 1.3 the Euclid mission, a space mission dedicated to the understand of the dark energy nature, with a galaxy survey responding to the requirements in the constraining of such alternative cosmological models.

1.1 General Relativity overview

For more than 200 years and until the beginning of the 20th century, the theory of the Universe was exclusively described by Newtonian gravitation (Newton, 1687) and within the framework of Euclidean geometry. Plagued by many instabilities and paradoxes (Bentley's paradox, Olbers' paradox) due to the infinite space and time assumption, General Relativity theory radically redefined the admitted principles to address these issues. In this first section, I will give an overview of the description of gravitation in a cosmological framework based on General Relativity.

1.1.1 The Einstein field equations

Space and time are no longer absolute when considering the invariance principle of the speed of light and form a space-time entity, leading to a theory of relativity of velocities, called Special Relativity (Einstein, 1905). This is the introduction of the *equivalence principle*, claiming that in a fairly small region of space-time, the laws of a static particle laying in a gravitational field are the same as for a uniformly accelerated reference frame, that made Einstein consider generalising his theory to the relativity of accelerations. Including the *Mach principle* (inertia occurs when a body is accelerated relatively to a global distribution of matter in the universe), all the ingredients were gathered to establish the theory of General Relativity (Einstein, 1916)[1].

In a purely geometrical approach, the master equations of the theory are called *field equations*, and read in non-Euclidean space

$$G_{\mu\nu} = \kappa \, T_{\mu\nu} \; . \tag{1.1}$$

By setting κ to a constant and $T_{\mu\nu}$ to a source term, the above expression decomposes as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \qquad \text{Einstein tensor} \tag{1.2}$$

$$R_{\mu\nu} = \partial_\rho \Gamma^{\rho}{}_{\mu\nu} - \partial_\nu \Gamma^{\rho}{}_{\rho\mu} + \Gamma^{\rho}{}_{\rho\lambda} \Gamma^{\lambda}{}_{\mu\nu} - \Gamma^{\rho}{}_{\nu\lambda} \Gamma^{\lambda}{}_{\rho\mu} \qquad \text{Riemann tensor} \tag{1.3}$$

$$\Gamma^{\lambda}{}_{\mu\nu} = \frac{1}{2} g^{\lambda\kappa} \left(\partial_\mu g_{\kappa\nu} + \partial_\nu g_{\kappa\mu} - \partial_\kappa g_{\mu\nu} \right) \qquad \text{Christoffel symbol} \tag{1.4}$$

$$R = R^{\mu\nu} g_{\mu\nu} \; , \qquad \text{Ricci scalar} \tag{1.5}$$

---

[1] The designation "general" comes from the assumption that the theory is valid regardless the observer, *i.e.* invariant under a general coordinate transformation (covariant theory).

by mean of a metric $g_{\mu\nu}$ of the curved manifold admitting as fundamental properties

$$g^{\alpha\nu} g_{\alpha\mu} = \delta^{\nu}_{\mu} \ , \quad g^{\mu\nu} V_{\nu} = V^{\mu} \ , \quad \nabla_{\alpha} g_{\mu\nu} = 0 \ . \tag{1.6}$$

When identifying the manifold to our gravitational field (the non-Euclidean 4D space-time), $T_{\mu\nu}$ takes the form of the stress-energy tensor of a cosmic fluid, which for simplicity is set to a perfect fluid (shear-less) in thermodynamic equilibrium

$$T^{\mu\nu} = (\rho + P)\, U^{\mu} U^{\nu} - g^{\mu\nu} P \ , \tag{1.7}$$

with ρ, P and U^{μ} defining the fluid's energy density, its pressure and its 4-vector velocity, respectively. Moreover with such application, κ can be obtained in a Newtonian limit to the value[2] $\kappa = 8\pi G$ where G is the usual Newton's constant.

Regardless of any particular symmetry of the universe, Einstein's equations 1.1 reduces to 10 non-linear and non-homogeneous equations ($R_{\mu\nu}$, $g_{\mu\nu}$ and $T_{\mu\nu}$ being symmetrical). Such a non-linear system is a wealth of informations to appreciate the strong link between energy spread throughout the universe (right side of 1.1), and curvature (left hand side) redefining gravitation : "Space tells matter how to move, Matter tells space how to curve " (Wheeler in Misner et al., 1973).

1.1.2 The cosmological principle

Symmetries in the universe

Facing the impossibility of analytically solving the field equations for an arbitrary geometry $g_{\mu\nu}$, cosmologists quickly turned to the identification of symmetries to reduce the number of degrees of freedom to be considered. First locally by spherical symmetry around a massive body (Schwarzschild, 1916), then globally by the hypothesis of a uniform universe.

Stated in 1933 by Milne, the *cosmological principle* directly concerns the matter distribution. It is the hypothesis of a homogeneous and isotropic universe at large scales ($\gtrsim$ 100 Mpc), *i.e.* invariant under translation and rotation for a *comoving observer* (defined below). It can be seen as the combination of the local isotropy to the *Copernicus principle* (there is no privileged location in the universe).

The notion of comoving observer is recurrent in Cosmology; related to a fundamental and privileged reference frame, its geodesic[3] motion $x^{\mu} = (t, \vec{r}) = (t, r, \theta, \phi)$ in the global gravitational field of the universe is such that the surrounding fluid seems motionless. Under these conditions, the surrounding matter distribution is characterised by a 4-velocity vector $U^{\mu} = (1, 0, 0, 0)$.

Technically, the consideration of such symmetries imposes a specific modelling of the metric of the universe. Indeed the consideration that space-time is a "$3+1$" structure in the description of a pseudo-Riemanian manifold (the three spatial dimensions have an opposite signature to the time signature), the cosmological principle implies that the spatial structure is of constant curvature and independent of time and space. It follows from the study of these symmetries that the most general

---

[2] The speed of light is conventionally set to unity in natural units.

[3] The geodesic is defined as the path used by a particle in free-fall with respect to the global gravitational field.

metric $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$ in accordance with the cosmological principle is diagonal and reads

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}R^2}{1-kR^2} + R^2(\mathrm{d}\theta^2 + \sin^2(\theta)\mathrm{d}\phi^2)\right) \ , \tag{1.8}$$

called the *Friedman-Lemaitre-Robertson-Walker* (FLRW) metric ($\sim$ 1930) and involving a free time-dependent parameter $a(t)$. On the other hand, the spatial curvature k, due to symmetries, can adopt the only three discrete values[4]

$$k = \begin{cases} +1 & \text{spherical, closed universe} \\ 0 & \text{opened Minkowski universe} \\ -1. & \text{opened universe} \end{cases} \tag{1.9}$$

The open/closed nature of the universe is an abstract notion. If the symmetry group associated with the manifold is compact (the manifold associated to the group is of finite volume), then the universe is said to be endowed by a closed geometry. For a non-compact group, the universe will have an opened geometry.

The cosmological implications

A direct consequence of the cosmological principle is the separation between space and time coordinates. While $a(t)$ represents the time-parametrised scale factor of the spacial coordinates, thus the universe itself, it also involves the uniqueness of a cosmic time associated to a cosmic observer. These features allow the definition of an abstract comoving reference frame $\vec{x}$ related to the physical space coordinates $\vec{r}$ as

$$\vec{r}(t) = a(t)\vec{x} \ , \tag{1.10}$$

which has the advantage to remain time invariant. Taking the derivative of eq. 1.10 gives

$$\vec{V}(t) = \frac{\dot{a}}{a}\vec{r} \equiv H(t)\vec{r} \ , \tag{1.11}$$

a linear relation between galaxy distances and their corresponding recessing/approaching velocities can be obtained, called *Hubble–Lemaître* equation. This velocity is related to the *Hubble flow* and comoving coordinates are blind to this effect. Thus measuring the sign and amplitude of the expansion rate $H(t)$ today (in the very local universe), called the *Hubble parameter* H_0, provides information on the current expansive (or contracting) behaviour of the universe.

Managing such measurement may use another implication of the FLRW metric, namely the most direct cosmological observable called *redshift*. Witnessing expansion or contraction of the space between galaxies, electromagnetic wavelength are stretched or contracted in physical space following eq. 1.10.

Be λ_e and λ_r respectively the wavelengths emitted from an astrophysical object (for which $a(t) = a(t_e)$) and received by an observer (for which $a(t) = a_0$). It can

---

[4] Note that a slightly different parametrisation of this metric allows to normalise $a(t)$ today at $a_0 = 1$, while verifying the following discrete values for k.

be shown that the stretching/contracting rate of wavelength is equal to those of the scale factor, such that $\lambda_e/a(t_e) = \lambda_0/a_0$. Defining the redshift as

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e} \ , \tag{1.12}$$

and recalling that we set $a_0 = 1$, the scale factor can be related to it as

$$a(t) = \frac{1}{1+z} \ . \tag{1.13}$$

In the local universe, it can be assumed that any $H_0 \neq 0$ propagates up to $|\vec{V}|$ in a classical way, thanks to the *Doppler–Fizeau* effect only affecting the line-of-sight velocity (explicitly putting back the light velocity c)

$$z \sim |\vec{V}|/c \ . \tag{1.14}$$

Using Cepheids[5] as a tracer to estimate distances $|\vec{r}|$ in eq. 1.11 and the Doppler–Fizeau effect 1.14 for velocity, Hubble notified the systematic red-shifting of extragalactic nebulae[6], showing in figure 1.1 a positive H. Assuming the Copernican principle, it constitutes the first evidence for an expanding universe.

Finally, from equation 1.13 can be derived the relation $\mathrm{d}z/\mathrm{d}t = -\left[(1+z)H\right]^{-1}$ relating the cosmic time t to z. Since z is a direct observable to the detriment of t, which is theory/model/cosmology dependent through H (discussed in the following), it turns the redshift as the default time parameter for Cosmology ordering events. This remark also applies to distances inferred from redshift measurements (different from the Cepheid way). Indeed, combining the $\mathrm{d}z/\mathrm{d}t$ relation to the photon path $\mathrm{d}s = 0$, the comoving distance reads

$$D_c(z) = c\int_0^z \mathrm{d}z H^{-1}(z) \ . \tag{1.15}$$

Thus one has to be careful when dealing with cosmological probes whose the estimation is exploiting relation 1.15 (in particular the matter power spectrum, see chapter 2).

In conclusion, this few sample of relations, simply inheriting the cosmological principle, is extremely powerful to quantify the expansion rate of the universe. Making the best use of them to gain more informations about the universe is to investigate how H behaves. To do so, equations 1.1 is the best way to understand its dynamics.

1.1.3 The Friedmann equations

Introducing the metric 1.8 and the expression of the stress-energy tensor of a perfect fluid 1.7 in the field equations 1.1 gives rise to two independent equations (instead of

---

[5] Cepheids are a class of stars populating galaxies whose the radius size is periodically changing. As a consequence, it has been demonstrated by Leavitt & Pickering (1912) that the pulsation period can be related to the absolute luminosity. Measuring an apparent magnitude, a distance can be inferred.

[6] A controversy surrounded the beginning of the century as to whether or not these objects should be interpreted as galaxies.

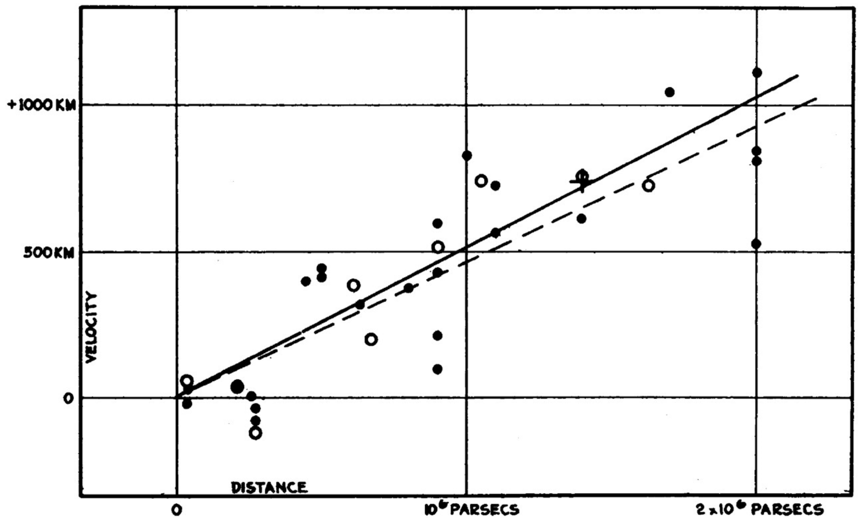


Figure 1.1: "Velocity-Distance Relation among Extra-Galactic Nebulae. Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually". Credit: Hubble (1929b).

10 as a result of imposed symmetries) called *Friedmann–Lemaître equations* (1922). They govern the overall (background) behaviour of the universe through the evolution of its scale factor $a(t)$

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{k}{a^2} = \frac{8\pi G}{3}\sum_i \rho_i \qquad \text{cinematic behaviour} \tag{1.16}$$

$$\frac{\ddot{a}}{a} + \frac{\dot{a}^2 + k}{a^2} = -8\pi G \sum_i P_i \ , \qquad \text{dynamic behaviour} \tag{1.17}$$

where i labels the non-interacting cosmic species at the level of the background. These two equations, a synthesis of the FLRW model, can be combined in a more explicit way to give

$$\dot{\rho}_i = -3H(\rho_i + p_i) \qquad \text{continuity equation} \tag{1.18}$$

$$\frac{\ddot{a}}{a} = -\frac{4}{3}\pi G\Big(\sum_i \rho_i + 3\sum_i P_i\Big) \ . \qquad \text{Raychaudhuri equation} \tag{1.19}$$

The former equation means that species are diluted in different ways in an expanding universe while in second one, the negative sign suggests that the presence of standard fluids (radiation or matter) slows down the expansion. Making the hypothesis that all fluids in the universe are barotropic, that is to say that each energy density is completely determined by the pressure it causes by mean of an equation of state parameter w, *i.e.*

$$P_i(t) = w_i(t)\rho_i(t) \ , \tag{1.20}$$

it allows to solve equation 1.18 by introducing the unknown initial density constants $\rho_{0,i}$ as

$$\rho_i(a) = \rho_{0,i} a^{-3(1+w_i)} \ . \tag{1.21}$$

In the thermodynamic limit, w_i can be found using the distribution function of each specie in phase space $f(p)^{-1} = \exp\left[(E(p)-\mu)/T\right] \pm 1$, verifying the on-shell limit $E^2 = p^2 + m^2$ where E is the total energy of the particle, μ its chemical potential, m its mass, p its momentum and T the equilibrium temperature of the fluid. While the $+$ and $-$ signs denote respectively the Fermi-Dirac (relativistic fermions) and the Bose-Einstein statistics (relativistic bosons), the Maxwell-Boltzmann statistics (non-relativistic particles) on the other hand is reached in the limit $E/T \approx m/T \gg 1$. For all of the types of species, energy density and pressure are given by

$$\rho_i(T) = g_i \int_0^\infty \frac{\mathrm{d}^3\vec{p}}{(2\pi)^3} f_i(p) E(p) \ , \tag{1.22}$$

$$p_i(T) = g_i \int_0^\infty \frac{\mathrm{d}^3\vec{p}}{(2\pi)^3} f_i(p) \frac{p^2}{3E(p)} \ , \tag{1.23}$$

$$\tag{1.24}$$

where g_i is the spin degeneracy of the specie.

When applying the cosmological principle symmetries in the integrations 1.22 and 1.23, it gives in the case of non-relativistic baryonic matter and radiation (the only two species whose existence is actually certain)

$$\rho_b(T) = g_b \frac{m^{5/2}}{2\pi^{3/2}} T^{3/2} e^{(\mu-m)/T} \quad , \quad P_b(T) = g_b \frac{m^{3/2}}{2\sqrt{2}\pi^{3/2}} T^{5/2} e^{(\mu-m)/T} \ , \tag{1.25}$$

$$\rho_r(T) = g_r \frac{\pi^2}{30} T^4 \quad , \quad P_r(T) = g_r \frac{\pi^2}{90} T^4 \ , \tag{1.26}$$

that can be combined to obtain the equation of state parameters

$$w_b = \frac{P_m}{\rho_m} \xrightarrow{m \gg T} 0 \ , \qquad w_r = \frac{P_r}{\rho_r} = \frac{1}{3} \ , \tag{1.27}$$

and justifying the usual practice to approximate them to a constant, *i.e.* $w_i(t) \to w_i$.

Thus, coming back to relation 1.21, it shows that radiation dilutes faster than baryonic matter. Anticipating the following of the section, a third component characterised by $w_\Lambda = -1$ (non-dilutable) should also be mentioned. In figure 1.2 is displayed the relative density as a function of time between these three densities. It shows that (of course depending on initial densities which are not randomly set in this figure) the universe went through several stages of domination (separated by equivalence times), each described by different physical phenomena discussed in next section.

Furthermore, as depicted in equations 1.16 and 1.17, the spacial curvature k is expected to participate to the expansion of the universe. In particular, eq. 1.16 teaches us that to this intrinsic geometrical feature can be associated an energy

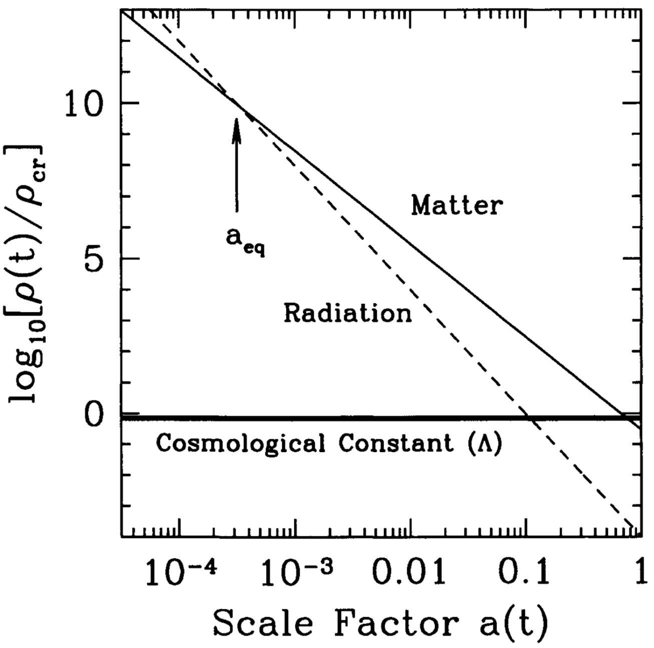


Figure 1.2: Energy density (divided by a constant ρ_{cr}, hereafter defined) as a function of the scale factor for different species in a flat universe ($k = 0$). Picked from Dodelson (2003).

density, diluting in a^{-2} when interpreting it as a cosmic fluid of equation of state parameter $w_k = -1/3$. On the other hand, it should be mentioned that the presence of curvature alone cannot produce acceleration or deceleration of cosmic expansion following equations 1.19 and 1.20. That's why in the following, it will be treated slightly differently, allowing to write equation 1.16 in a more explicit form. Indeed, dividing eq. 1.16 by H^2 gives

$$\frac{8\pi G}{3H^2}\sum_i \rho_i - \frac{k}{(aH)^2} = 1 \ , \tag{1.28}$$

and shows that when

$$\sum_i \rho_i = \rho_{cr} \equiv \frac{3H^2}{8\pi G} \ , \tag{1.29}$$

a specific density is reached, called the *critical density*, which corresponds to a null curvature. This quantity can be used in order to define the dimensionless energy densities

$$\Omega_i \equiv \frac{\rho_i}{\rho_c r} = \frac{8\pi G}{3H^2}\rho_i \ , \tag{1.30}$$

$$\Omega_k \equiv -\frac{k}{a^2H^2} \ , \tag{1.31}$$

which rearrange the first Friedmann equation in its simplest form

$$\sum_i \Omega_i(t) + \Omega_k(t) = 1 \ . \tag{1.32}$$

Finally, applying eq. 1.21, it yields

$$H^2 = H_0^2 \left[\sum_i \Omega_{i,0} a^{-3(1+w_i)} + \Omega_{k,0} a^{-2} \right] \equiv H_0^2 E(a)^2 \ , \tag{1.33}$$

showing that the expansion rate of the universe behaves exclusively following its content. Identifying and constraining these present *cosmological parameters*, that are not predicted by the theory and in consequence must be fitted with observation, is therefore essential to reconstruct the history of the universe. The next axes of discussion will therefore be oriented in the theoretical and observational strategies to quantify them.

1.1.4 The ΛCDM concordance model

The construction of the standard model over the last century, called the ΛCDM concordance model, has resulted from a perpetual dialogue between the previous general relativity framework for theory and some different sets of observational data. In particular, data was used to constraint the cosmological parameter values of eq. 1.33 and lead to the current concordance model. In the following, I review the main observational probes that gave an inflection in the construction of this as a standard model.

The Big Bang model

The Hubble's discovery has been the starting evidence for an expanding universe. Hence a natural extrapolation is to consider that the universe was smaller in the past[7] and presumably hotter. Being a container for thermodynamics species that are responding to global dilution (see eq. 1.21), a thermal history can be drawn without speculating any kind of new physics. This model, called *Big Bang*, is very powerful in the sense that it produces billion years observable predictions still submittable to the observation.

Thus from the simple knowledge of an expanding universe, I'll present in the following a brief history of the primordial universe during the earlier stages of the universe (thus in the radiation domination epoch, see figure 1.2) as it is predicted by the Big Bang model. This thermal history will be related as a function of the bounded species temperature ($\equiv$ temperature of the universe), decreasing as densities are diluting.

Let me underline first of all that at temperature $\gtrsim 13$ TeV, any proposed mechanism (SUSY breaking, Great Unification but also inflation) is in the field of speculation as the standard model of particle physics has been so far tested up to this typical energy at the *Large Hadron Collider*. Contrarily, a certainty is that no stable matter can be formed and the universe consists in a plasma of elementary particles (quarks, leptons) when associated with energy of few TeV.

At the temperature of $\gtrsim 1$ TeV occurs a baryonic symmetry breaking allowing quarks to associate in baryons (protons, neutrons) in a mechanism still poorly understood, called *Baryogenesis*. In particular, the reason why only baryons and not

---

[7] In fact it can be shown that an expanding universe cannot be infinite in the past (Borde et al., 2003).

anti-baryons remains at the end of the process is still a mystery[8]. At the end of this phase, baryons and leptons (electron, positrons, neutrinos, etc) are forming a plasma kept in thermal equilibrium.

Then comes the *electro-weak transition* at ~ 150 GeV, where the weak interaction becomes a force in its own right by decoupling itself from electromagnetism. This symmetry breaking, manifesting when the temperature cross the Higgs mass, is the mechanism that associate a mass to all particles belonging to the hot plasma.

As the scale of weak interaction is significantly reduced, when reaching $\sim 10^2$ MeV, strong interaction imposes quarks to confine and produce hadrons and mesons in a phase called *Quark-Hadron transition.* After this process, the plasma is made of e^+e^- doublets, neutrinos, protons, neutrons and photons in thermal equilibrium.

Weak interaction cross-section is still decreasing and while temperature reduces to ~ 1 MeV (~ 1s), neutrinos decouple from the rest of the fermionic matter and are released in a quasi free streaming. This constitutes one of the first early universe observable, called the *cosmic neutrino background* which has not been yet detected (mainly due to their weak interaction with baryonic matter and their expected small number density today). Shortly after (around ~ 200 keV), doublets e^+e^- annihilate each other with a matter-antimatter asymmetry in favour of electrons, leaving a photon-baryon-electron plasma in equilibrium.

Reaching the temperature of 0.1 MeV (~ 200s), the strong interaction cross-section becomes sufficiently important to bound baryons and forms the lightest nuclei (D, ^{3}He, ^{4}He, ^{7}Li) in a phase called *Big Bang Nucleosynthesis* (BBN). For a couple of minutes, the universe is transformed into a huge Hydrogen bomb where about 25% of the baryonic matter present in the universe is transformed into Helium nuclei[9]. Such relative nuclei abundance predictions is one of the most striking consequences of the Big Bang theory. In agreement with observation, it can be used in order to give some constrains on the baryon density Ω_b parameter.

For a long period of time and while the photons are still too energetic, they prevent the formation of the first atoms because of Compton and Thomson scatterings (matter-radiation interaction), even when crossing the matter radiation equality at the time of *equivalence* (~ 1 eV, $\sim 10^{11}$s). We finally have to wait for the thermal photon bath to cool down enough down to ~ 0.1 eV (~ 380.000 yrs), to let electrons bind to nuclei and form neutral atoms, in a period called *recombination.* Photons decouples and, as for neutrinos earlier, form the free-streaming *Cosmic Microwave Background* (CMB) in a period called *Last Scattering Surface* (LSS). Their initial temperature of about 3000 K is presently detectable at the temperature $T_\gamma^0 = 2.73$ K due to the stretch of wavelength between $a(t = \mathrm{LSS})$ and a_0, indicating that the redshift of the last scattering surface is $z_{\mathrm{LSS}} \sim 1090$.

This first image of the primordial universe (the universe was opaque before recombination), that have been measured by several experiments as COBE (Smoot et al., 1992), WMAP (Spergel et al., 2003) and Planck (Aghanim et al., 2018), is described by a highly isotropic temperature map (see top panel of figure 1.3) extremely close to the black-body spectrum. The sole knowledge of the averaged temperature

---

[8] Explaining this asymmetry can be addressed exploiting the CP violation, see Sakharov (1991) and Christenson et al. (1964).

[9] As a comparison, stellar nucleosynthesis for about ten billion years has only succeeded in converting only 2% of the total Hydrogen of the universe into Helium.

T_γ^0 provides an accurate estimation of the radiation density today : $\Omega_{r,0}h^2 \sim 10^{-4}$.

Moreover, small temperature anisotropies of the level $\Delta T/T \sim 10^{-5}$ have been detected and accurately probed by Planck. An angular correlation function[10] $\mathcal{C}_\ell$ of these anisotropies can be estimated (see centre panel of figure 1.3) and constitutes today one of the most precise information able to constrain a large sample of cosmological parameters today (see bottom panel of figure 1.3). In particular, the background parameters already introduced in the previous section (H_0, Ω_b, Ω_r) but also other that I didn't yet introduced, affect the shape of the C_ℓ. These highly informative wiggles in the C_ℓ are described below.

Baryon acoustic oscillations

Between BBN and recombination, out of equilibrium phenomena are taking place and have significant consequence, today detectable in several probes. Indeed, during this long period ($\sim 10^6$ yrs), gravitation has enough time to act on matter, attempting to grow some local fluctuations. Energetic photons instead, acting toward a uniformisation of the plasma, systematically counterbalance local gravitational collapse (through Compton and Thomson scatterings). These exchanges, seeming like a collection of springs, generate oscillations propagating in the baryon-photon plasma as sound waves as long as radiation is coupled to baryons, called *Baryon Acoustic Oscillations* (BAO). Statistically, each primordial over-density is thus surrounded by an over-dense growing shell of baryons and photons (Eisenstein & Hu, 1998a; Eisenstein et al., 1998).

During the release process of the CMB, the transition from the plasma to a baryonic gas has the consequence to freeze out the BAO as it stands. The propagating shell stops at a radius given by the sound horizon, leaving a statistical trace in the CMB, witness of the environment conditions of the early universe. As a consequence, the C_ℓ representation (figure 1.3) featured the fundamental and harmonic frequencies of the BAO, allowing the estimations of the cosmological parameters.

Way after recombination, matter is free to fall into over-densities and progressively forms large scales structure that we see today. In particular, the BAO imprint must also be exhibited in the correlation function of galaxies. This will be described in detail in chapter 2.

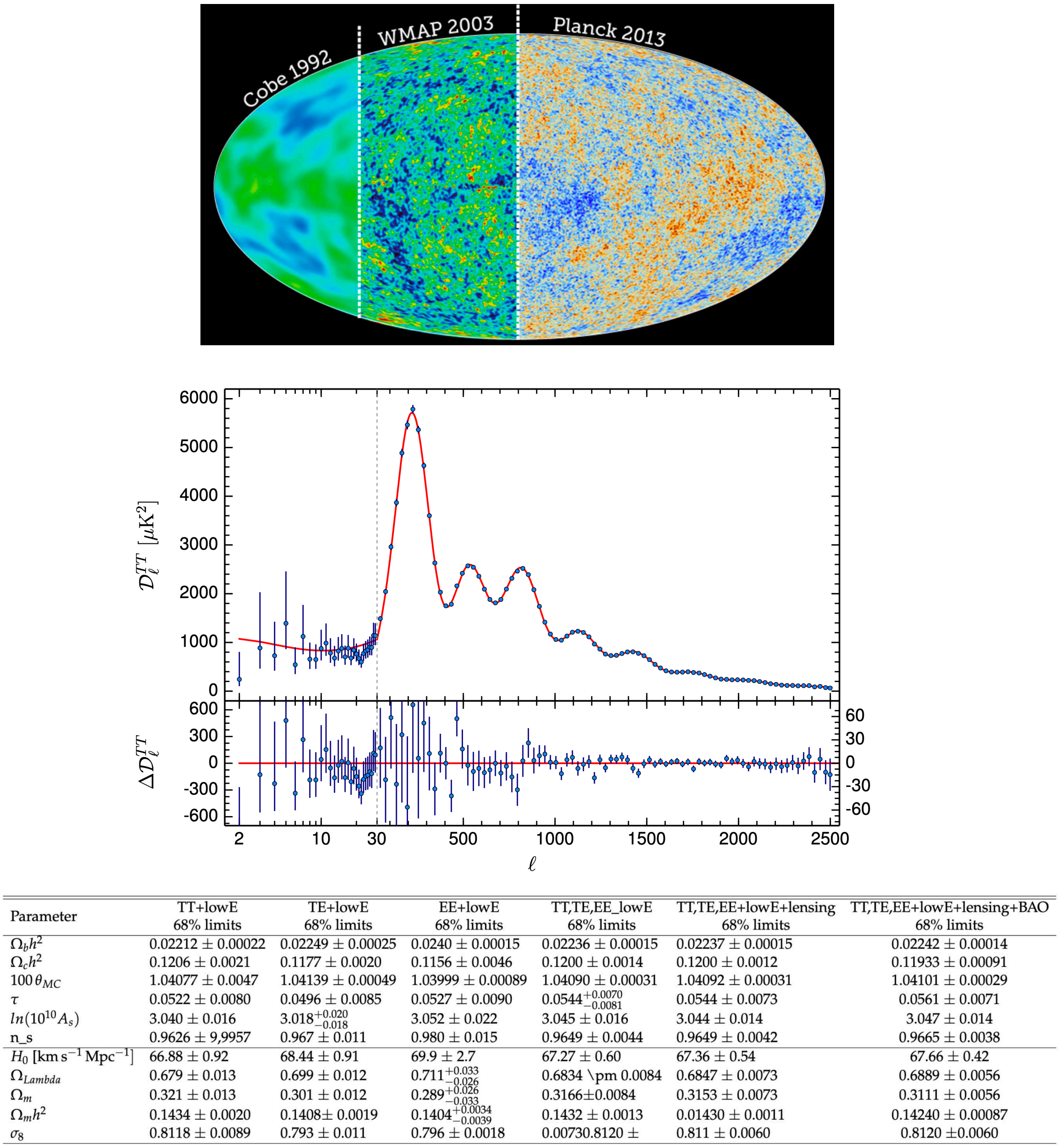
Inflation

The temperature anisotropies observed in the CMB are essential since without them, it shouldn't have been possible to start the formation of large scale structures. As a matter of fact, a temperature perturbation means a local curvature fluctuation counterpart, felt by matter. Undergoing gravitational instabilities, these primordial inhomogeneities transformed into matter density fluctuations will give rise to the first cosmological structures.

In order to explain why the universe was not born and did not evolved in a perfectly uniform universe, theoreticians has imagined a primordial mechanism in which quantum fluctuations are at the origin of such seed. This commonly accepted paradigm, called *primordial inflation* (nevertheless without conclusive evidence),

---

[10] A formal introduction to correlation functions will be the object of chapter 2.



Parameter	TT+lowE 68% limits	TE+lowE 68% limits	EE+lowE 68% limits	TT,TE,EE_lowE 68% limits	TT,TE,EE+lowE+lensing 68% limits	TT,TE,EE+lowE+lensing+BAO 68% limits
$\Omega_b h^2$	0.02212 ± 0.00022	0.02249 ± 0.00025	0.0240 ± 0.00015	0.02236 ± 0.00015	0.02237 ± 0.00015	0.02242 ± 0.00014
$\Omega_c h^2$	0.1206 ± 0.0021	0.1177 ± 0.0020	0.1156 ± 0.0046	0.1200 ± 0.0014	0.1200 ± 0.0012	0.11933 ± 0.00091
$100\,\theta_{MC}$	1.04077 ± 0.0047	1.04139 ± 0.00049	1.03999 ± 0.00089	1.04090 ± 0.00031	1.04092 ± 0.00031	1.04101 ± 0.00029
τ	0.0522 ± 0.0080	0.0496 ± 0.0085	0.0527 ± 0.0090	$0.0544^{+0.0070}_{-0.0081}$	0.0544 ± 0.0073	0.0561 ± 0.0071
$ln(10^{10} A_s)$	3.040 ± 0.016	$3.018^{+0.020}_{-0.018}$	3.052 ± 0.022	3.045 ± 0.016	3.044 ± 0.014	3.047 ± 0.014
n_s	$0.9626 \pm 9{,}9957$	0.967 ± 0.011	0.980 ± 0.015	0.9649 ± 0.0044	0.9649 ± 0.0042	0.9665 ± 0.0038
H_0 [km s^{-1} Mpc^{-1}]	66.88 ± 0.92	68.44 ± 0.91	69.9 ± 2.7	67.27 ± 0.60	67.36 ± 0.54	67.66 ± 0.42
Ω_{Lambda}	0.679 ± 0.013	0.699 ± 0.012	$0.711^{+0.033}_{-0.026}$	0.6834 \pm 0.0084	0.6847 ± 0.0073	0.6889 ± 0.0056
Ω_m	0.321 ± 0.013	0.301 ± 0.012	$0.289^{+0.026}_{-0.033}$	0.3166 ± 0.0084	0.3153 ± 0.0073	0.3111 ± 0.0056
$\Omega_m h^2$	0.1434 ± 0.0020	0.1408 ± 0.0019	$0.1404^{+0.0034}_{-0.0039}$	0.1432 ± 0.0013	0.01430 ± 0.0011	0.14240 ± 0.00087
σ_8	0.8118 ± 0.0089	0.793 ± 0.011	0.796 ± 0.0018	0.00730.8120 ±	0.811 ± 0.0060	0.8120 ± 0.0060

Figure 1.3: *top panel:* temperature fluctuations maps (of the order $\Delta T/T \sim 10^{-5}$) in the CMB for the three experiments COBE, WMAP and Planck. *centre panel:* Angular correlation function in the Planck temperature map $\mathcal{D}_\ell^{TT} = \ell(\ell+1)C_\ell^{TT}/(2\pi)$ as a function of the multipols ℓ (inversely related to angles). The measurements are plotted in blue with error bars, while in red is represented the fitted correlation function. *bottom panel:* table of estimated cosmological parameters from the Planck measurements.

consists in an extraordinary accelerated expansion phase that lasted a fraction of second and in which quantum fluctuations of a hypothetical scalar field become classical. Also more detailed in chapter 2, inflation constitute another hypothesis nevertheless not yet part of the standard model.

Cold Dark Matter

Given that the Big Bang model seems highly robust when confronting it to observations, it may be conceivable to rely on it as a starting point to understand the evolution of structures up to their current state (galaxies, cluster of galaxies, etc). However, assuming that the matter content is exclusively filled with standard baryonic matter, itself described by tiny fluctuations related to those observed in the CMB, theory is unable to predict the current level of clustering without adding extra mass component to the total budget of the universe. Unless the fluctuation evolution models are not adapted, we have to postulate the existence of an only-gravitationally interacting, cold and non-baryonic matter, whose initial perturbation are adiabatic and Gaussian (Blumenthal et al., 1984), named *cold dark matter* (CDM).

These three properties (non baryonic + cold + only gravitationally interacting) are motivated by the fact that the observed electromagnetic spectrum is blind to it, prohibiting any electromagnetic and in consequence strong interactions as well[11] while additional astrophysical clues, as the mismatch of kinetic and potential energy of visible matter in the Coma cluster (Zwicky, 1937) or the study of rotation curves of galaxies (Rubin & Ford, 1970; Roberts, 1969; Bosma, 1981), point in the direction of a strong gravitational power (see a review in Drees & Gerbier, 2012).

Moreover, this component should not be baryonic for (among others) cosmological purposes. Not subject to Thomson or Compton scatterings before decoupling, it is free to start its own clustering way before baryons are allowed. In doing so, baryonic gravitational collapse is enhanced after their release from radiation, explaining their actual state of clustering.

The last property is more controversial and result from the disentangling of two paradigms: the *bottom-up* and the *top-down* scenarios (see Peebles, 1980a; Peebles, 1982). In the former, masses the size of star collapse first, then assemble into galaxies, clusters of galaxies, and super-clusters. It requires dark matter particles to behave as non-relativistic particles (cold) at sufficiently early time to quickly assemble into small scales haloes. In the contrary, the top-down scenario build large scales structure prior to locally collapse into astrophysical objects, by delaying or totally avoiding the transition from a relativistic to a non-relativistic dark matter. Such hot dark matter particles could have a mass of a few tens of eV (Primack & Gross, 2000). It seems however that simulations and observations privilege the bottom-up scenario, embedding the CDM in the standard model while participating about 85% of the total matter, according to CMB observation (see parameter estimations in figure 1.3).

---

[11] Note that it leaves an eventual door open for weak interacting dark matter, see Jungman et al. (1996).

Supernovae observation and the accelerated expansion

Considering an expanding universe filled by pressure-less matter (baryonic or CDM) and radiation, a natural consequence dictated by Friedmann equations 1.16 and 1.17 is to slow down the expansion rate of the universe, due to the action of gravity. The cosmological purpose in the end of the XX$^{\text{th}}$ century was therefore to measure the associated deceleration parameter of the universe.

At the end of the last century, type-Ia Supernovae (SN-Ia) have been used as standard candles by two independent teams (*Supernova Cosmology Project* and *High-Z Supernova Search Team*) exploiting different data set, in order to draw the luminosity distance as a function of the redshift (see eq. 1.36. Their results, published in Riess et al. (1998) and Perlmutter et al. (1999), revealed against all odds a deceleration parameter with a negative sign; witness of a current accelerated expansion of the universe. For instance in the Riess results, displayed in figure 1.4, it shows that at redshift parity, the objects appear further away from what theory predicts in the case of a decelerating scenario, preferring a universe filled by a relatively low matter (baryons + CDM) contribution ($\sim 25\%$) and an additional component which dominates ($\sim 75\%$) the energy budget of the universe today.

Technically, this is equivalent to observationally estimate a deceleration signal, related to $\ddot{a}$ thanks to the Taylor expansion (up to second order)

$$a(t) \approx a_0 + H_0(t - t_0) - \frac{1}{2} q(t_0) H_0^2 (t - t_0)^2 \ , \tag{1.34}$$

where the time-dependent deceleration parameter is defined as

$$q(t) \equiv -\frac{\ddot{a}(t) a(t)}{\dot{a}^2(t)} \ . \tag{1.35}$$

Measuring this quantity can be done noticing that, in the approximation of small redshifts, the luminosity distance of distant objects can be put in the form (see Perlmutter & Schmidt, 2003b)

$$d_L \approx \frac{cz}{H_0} \left(1 + \frac{1 - q(t_0)}{2} z \right) \ . \tag{1.36}$$

To reach such a conclusion, a relation between deceleration parameter and cosmic constituent features can be obtained combining relations 1.35, 1.19, 1.21 and 1.30,

$$q_0 = \frac{1}{2} \sum_i \Omega_i (1 + 3w_i) \ . \tag{1.37}$$

Putting all together in the analysis, authors of the study, as well as more recent SN-Ia observations, converges toward postulating a specie component whose equation of state parameter is suitable to drive a cosmic acceleration ($w < -1/3$, see relation 1.19), consistent with a cosmological constant Λ characterised by $w_\Lambda = -1$ and a non-diluting energy density (in natural units)

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} \ , \tag{1.38}$$

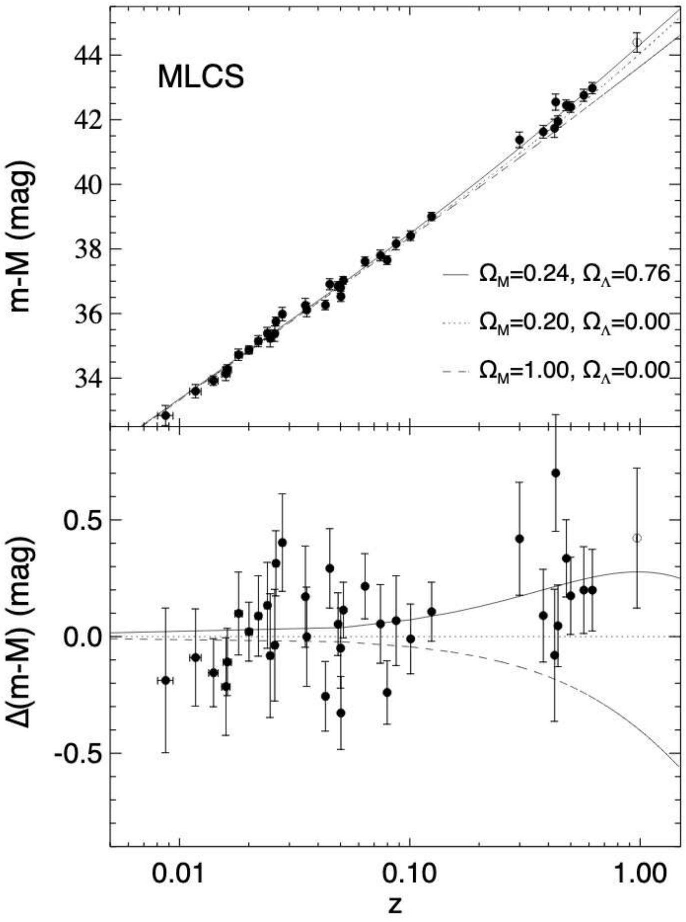


Figure 1.4: *upper panel:* difference between the apparent m and absolute M magnitudes, called distance modulus and related to d_L in 1.36, for a sample of 34 type-Ia supernovae. The solid, dotted and dashed lines represent the theoretical predictions for different values of cosmic constituents Ω_i (see definition 1.30). The dashed one corresponds to the expected scenario where matter is the dominant constituent, the dotted line when adding curvature (see eq. 1.32) and the solid line to the best fit when accounting for observation (taking $\Omega_k = 0$). *bottom panel:* relative difference between the prediction ($\Omega_m = 0.2, \Omega_\Lambda = 0$) and data and the other two predictions. Taken from Riess et al. (1998).

that start to dominate currently[12] (see figure 1.2).

The cosmological constant Λ, an integration constant allowed in the field equations 1.1 was first introduced by Einstein (1917) to counterbalance gravitation and achieve a static universe, at that time much more reasonable and conventional when trying to go one step further than Newton universe. Given the evidence for an expanding universe, this constant was then dropped but finally reintroduced as a possible candidate driving the acceleration of the expansion of the universe.

On the other hand, let me mention that adding Λ to the total budget of the universe reach some concerns. If it is interpreted as a geometrical feature of the

---

[12] The fact that $\rho_m \sim \rho_\Lambda$ recently is whether indicating that we are living in a special time of the cosmic history or rather, that data are misinterpreted. This potential questioning is usually called the *coincidence problem*.

universe, it may seem on some level unnatural. Instead, when trying to interpret it as a energy constituent of quantum origin, it appears difficult to match the prediction to the observation[13]. Λ is then a questioning solution and is not anyway the unique way to fit the observation. Several approaches can be theoretically investigated as :

- the addition of negative-pressure components called *dark energy*, whose equation of state parameter is free to vary,
- the test of the validity of the field equations themselves at large scales, leading to develop extension of the general relativity. In such approaches, the geometrical response to energy is revised, and this is often called *modified gravity models*. Some examples will be presented in section 1.2.2,
- a relaxation of the Copernican principle, placing us rather in a very special place of the universe (as for instance in a huge cosmic void), that makes us interpret data in a wrong way,
- considering that inhomogeneities neglected in the FLRW model may lead to a *backreaction* over large scales, provoking the background accelerated expansion (Buchert, 2000; Rasanen, 2006).

The cosmological parameters extraction

The previous probe measurements, namely SN-Ia, CMB and BAO, can be combined in a joint analysis. The results, presented in figure 1.5, show that the analysis puts some tight constrains on the Ω_Λ, Ω_m and w_Λ parameters, revealing that the different probes are all converging toward an interpretation with a cosmological constant Λ and an extra dark matter constituent (Ω_b being observationally incompatible with ~ 0.25).

Due to the simplicity of Λ (see Occam's razor in chapter 2) and its tendency to fit observations, the hypothesis of a cosmological constant is currently embedded in the standard model, as well as the hypothesis of a cold dark matter. Their are finally summarised in the denomination ΛCDM concordance model.

The specificity of this model is that it allows to combine in one global view observations of different probes, whether they concerns early universe or late time universe. This model is described by a few set of independent parameters, each requiring to be fitted by the observation.

Note that many other probes can be added to this constrains (as cluster analysis, strong and weak lensing analyses, gravitational waves, etc). Specifically, the analysis of the clustering properties of galaxies is a probe that can be seen as an extension of the previous BAO measurement and is often shortly called *galaxy clustering*. This approach consists in statistically studying the galaxy cartography in the universe, also sensitive to a large set of cosmological parameters beyond the BAO primordial information. This probe will be intensively describe in the next chapters of this dissertation.

---

[13] An approach was to relate the to the vacuum energy, unsuccessfully so far due to the difficulty to predict it with an acceptable expectation value (see Adler et al., 1995; Weinberg, 1989).

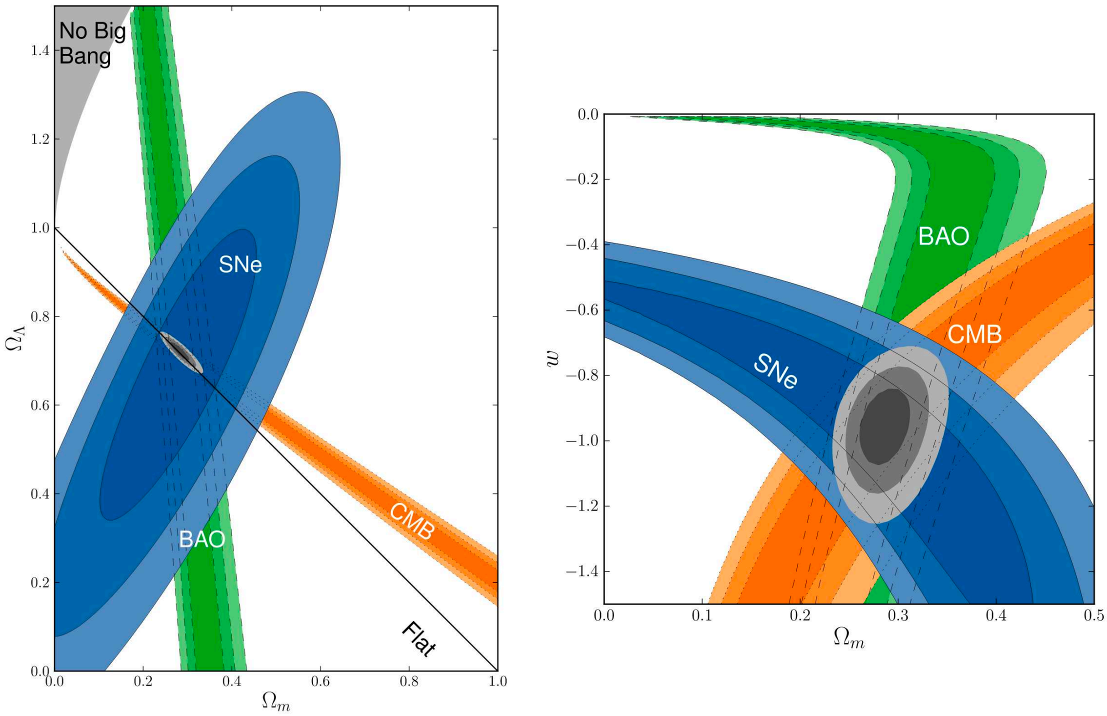


Figure 1.5: Constraint of the $(\Omega_\Lambda, \Omega_m)$ (left panel) and $(\Omega_\Lambda,\Omega_m)$ (right panel) planes using joint analysis between SN-Ia, CMB and BAO. Taken from Suzuki et al. (2012).

However, even if the ΛCDM concordance model seems robust, combining many probes in a coherent way is mandatory to avoid misinterpretations. In fact, adding probes is not only a way to constrain the cosmological models, but it is also a way to revel sources of incoherence in data, a conundrum eventually paving the way to a new physics.

1.2 Limit and extension of the standard model

As previously stated, if so far the standard model succeed in predicting most of the observational measurements (SN, CMB and BAO in particular), some issues both on the theoretical and the observational side must be addressed.

First of all, the standard model is missing the inner nature of dark energy[14] and of dark matter, as well as predicting the existence of the primordial fluctuations. Inflation is still subject to debate several decades after its introduction.

Secondly, even if the concordance model appears to fit well most of the data, some data measurement are in tension within the context of this standard model, opening puzzling questions.

In particular, tensions are appearing in the measurements of the Hubble parameter H_0 and in the measurement of a parameter quantifying the clustering amplitude, called σ_8. Even if this second parameter has not been yet formally introduced (see chapter 2 for details), no prerequisites are necessary to follow the rest of this section.

---

[14] This denomination include Λ.

In this section, I will first focus on these data tensions, leading to the belief that the ΛCDM model is either incomplete (as seen as an approximation of a more general theory) or needs to be rebuilt (by addressing the foundation of such theory). The discussion will then be followed by the presentation of some theoretical aspects, specifically going beyond the standard model, seeking to solve some of these open questions with a better description of the universe.

1.2.1 Some observational data tensions

One of the most significant data inconsistency is related to the local measurements[15] of the Hubble parameter H_0, in tension with the value inferred from the CMB data, both when assuming the standard model.

- Concerning the early universe probes, the Planck CMB measurement (Aghanim et al., 2018) or independently the combined ACT+WMAP (Aiola et al., 2020) provides respectively $H_0 = (67.27 \pm 0.60)$ km/s/Mpc and $H_0 = (67.6 \pm 1.1)$ km/s/Mpc. When looking at another CMB polarization probe, the STPpol collaboration (Henning et al., 2018), a non significant tension is observed with the value of $H_0 = (71.3 \pm 2.1)$ km/s/Mpc.
  Moreover, when combining CMB+BAO, a similar low-H_0 measurement is still preferred with $H_0 = (67.0 \pm 3.2)$ km/s/Mpc in the 6dFGS data analysis (Beutler et al., 2011) or $H_0 = (67.35 \pm 0.97)$ km/s/Mpc with SDSS-IV (Alam et al., 2020).

- For the late universe probes, they are mainly coming from HST SN-Ia local measurements (standard distance ladder) finding $H_0 = (73.5 \pm 1.4)$ km/s/Mpc (Reid et al., 2019) in the SH0ES experiment or $H_0 = (74.03 \pm 1.42)$ km/s/Mpc (Riess et al., 2019) when using different objects. $H_0 = (73.3 \pm 4.0)$ km/s/Mpc is obtained for the MIRAS experiment (Huang et al., 2019) or $H_0 = (73.3^{+1.7}_{-1.8})$ km/s/Mpc by measuring the time delays on six gravitationally lensed quasars in the H0LiCOW experiment (Wong et al., 2019). For a more exhaustive list, see Di Valentino et al. (2020a).

Thus two H_0 attractors seem to be identified, the high-z and low-z probes that respectively tends toward low and high-H_0 values, a tension that can reach 4.4σ when comparing results from Planck and SH0ES.

We should ask for the origin of such tensions. Does it come from a biased analysis or is it a direct witness of new physics? In the first case, an instrumental systematic error can be discarded as data are coming from different telescopes. Moreover the systematic errors can come from the data analysis itself, as non accounting for the eventual fact that the survey window (small fraction of the Hubble volume) could be in a under-dense or over-dense region when dealing with local probes. Here an under-dense local region could increase the value of H_0 and has to be taken into account

---

[15] It is not uncommon to find in literature the denomination of local/late/low-z and early/high-z probes. Here I follow the same definitions: early probes for CMB, BBN and BAO, and local probes for SN-Ia, WL.

by computing the additional super sample covariance (first discovered for cluster counts by Hu & Kravtsov (2003)). Unfortunately, when accounting for it, the effect appears too small ($\sigma(H_0^{loc}) = 0.31$km/s/Mpc when one requires $\sim$ 6km/s/Mpc) to reduce the tension (Wu & Huterer, 2017; Kenworthy et al., 2019).

Putting data and analysis reliability aside, reducing the tension may come from the fiducial cosmology itself, rising a considerable list of theoretical models drastically reducing these tensions. Some dark energy model could solve the anomalies either involving $w < -1$ (Di Valentino et al., 2020b), or through emergent dark energy model (Li & Shafieloo, 2019), interacting dark energy models (Yang et al., 2018) or early dark energy models (Poulin et al., 2019). Several other non dark energy solutions have been investigated as well, as modified gravity models (Solà Peracaula et al., 2019), interacting neutrinos (Kreisch et al., 2020), decaying dark matter (Vattis et al., 2019), modification of the fundamental constants (Hart & Chluba, 2018), modified early time recombination (Jedamzik & Pogosian, 2020) and so on. Note that a model-independent method will be able in the near future to discriminate between the high and low values of H_0, namely multi-messenger observations of gravitational waves with the electromagnetic spectrum counterparts, that is expected to provide a $< 2\%$ uncertainty on H_0 in half a decade (Chen et al., 2018).

An other tension has been observed,once again between low-z and high-z measurements, estimated at $z = 0$, packed in the reduced quantity $S_8 = \sigma_8\sqrt{\Omega_m/0.3}$, between CMB measurements and weak lensing measurements + redshift surveys. A $\sim 3\sigma$ is reported, confronting and assuming ΛCDM.

- From one side there is the CMB measurement of Planck and ACT+WMAP that measure respectively $S_8 = 0.834 \pm 0.016$ and $S_8 = 0.84 \pm 0.03$.

- From the other side low-z probes as gravitational weak lensing and/or cosmic shear gives $S_8 = 0.783^{+0.021}_{-0.025}$ in DES (Abbott et al., 2018; Troxel et al., 2018), $S_8 = 0.745 \pm 0.039$ in KiDS-450 (Hildebrandt et al., 2017), $S_8 = 0.766^{+0.020}_{-0.014}$ in KiDS-1000 (Heymans et al., 2020). When combining them with other lensing probes in KiDS-450 + 2dFLenS (Joudaki et al., 2018) we get $S_8 = 0.742 \pm 0.035$ and $S_8 = 0.737^{+0.040}_{-0.036}$ in KiDS+VIKING-450 (Hildebrandt et al., 2020). Finally, $S_8 = 0.703 \pm 0.045$ is inferred from the redshift-space power spectrum of BOSS (Ivanov et al., 2020). For a more exhaustive list, see Di Valentino et al. (2020c).

Here also, two attractors can be identified, high-z that tend toward high S_8 values while low-z prefer smaller ones. A puzzling fact is that, when applying the same theoretical solutions used to fix the H_0 tension, σ_8 discrepancies are getting worse. For example when using the CMB priors for H_0, galaxy cluster count in the STP-SZ collaboration (de Haan et al., 2016) gives a $S_8 = 0.797 \pm 0.031$, incoherently compatible with low-z probes. Solutions can be found for example exploiting other types of decaying dark matter (Abellan et al., 2020), modified gravity models (Sola et al., 2020) or interacting dark matter (Di Valentino et al., 2020d).

To these anomalies, we should potentially add another one, on the curvature Ω_k parameter. This parameter should be extremely close to zero to comply the foundations of the inflation paradigm (see chapter 2). Indeed, the 2018 Planck results analysed by Di Valentino et al. (2019) gives an $\Omega_k < 0$ (closed universe) at about 3.4σ[16], while ACT+WMAP gives $\Omega_k = -0.001^{+0.014}_{-0.010}$, compatible with a flat universe. In general, combining other probes is providing a flat curvature, as with BAO+BBN+H0LiCOW (Nunes & Bernui, 2020) that gives $\Omega_k = -0.07^{+0.14}_{-0.26}$. Once again, assuming ΛCDM, H_0 and σ_8 tensions are both exacerbate when relaxing the curvature parameter and no theoretical model can solve all of these three anomalies.

In conclusion, solving the previous data tension is extremely puzzling. The solution must be brought both from the side of the observation/analysis and from the theory side. One should improve the observational accuracy for H_0 and σ_8, by a better control of systematic errors, with an estimation of the bias in the analysis, but also by measuring independently, the same observable with several experiments. On the theory side, we should remove model instabilities or non physical effect to reduce the parameter space to be tested, but also it should be useful to develop (or improve) more cosmological tests or model-independent observables.

1.2.2 Modifications of the standard model

As previously emphasised, the standard model provides a good fit of the data but open many other questions. Looking for theoretical alternatives to solve data tensions or theoretical questions, must lead to a phenomenological equivalent, but with different prediction of observable values. In other terms, the challenge of these theoretical developments is to preserve the predictive success of the standard model (BBN, BAO, CMB), while the physics of the perturbed sector[17] must be rewritten to match the observation (see σ_8 tensions for instance). Such developments can be grouped in two approaches.

- The first one, while keeping the field equations 1.1 unchanged, acts on the nature and/or the characteristics of the energy components of the right-hand side. This can take the popular form of replacing the cosmological constant by hypothetical components like scalar fields with variable equation of state parameters $w \to w(t)$, called quintessence models (Sahni & Starobinsky, 2000; Caldwell et al., 1998).

- The second option lays on geometrical aspects. By coupling the Ricci scalar to a scalar field (or other), it modifies the response of the geometry of the universe to its content. They are called *modified gravity theories*. Although offering infinite possibilities of clustering predictions, it turns out to be difficult to recover gravitation at the well constrained scales (the solar system). To correct for these unpleasant effects, some artefacts have to be implemented such as screening mechanisms (Vainshtein, 1972), artificially imposing to recover classical General Relativity at these local typical scales.

A general formalism unifying both dark energy and modified gravitation theories, whose accelerating expansion of the universe is driven by a scalar field and called the *Effective Field Theory of Dark Energy* (EFT), will be discussed in this section.

---

[16] This reinterpretation is also subject to debates, see Efstathiou & Gratton (2020).

[17] I refer here to the sector of cosmological structure evolution (clustering). The next chapter will go through this notion in more details.

Then, in a totally different domain, I will explore another beyond standard model aspect where large scales structures meets particle physics. In particular, I will discuss how associating a non standard mass to neutrinos can affect cosmological probes.

Modifying the Einstein Field equations

First developed for primordial inflation (Creminelli et al., 2006; Cheung et al., 2008), then extended to late time cosmology to the dark energy problem, the *effective field theory for dark energy* (EFT) (Gubitosi et al., 2013) represents a class of cosmological models whose dark energy takes the form of a single scalar degree of freedom. Called the *Goldstone boson*, the last is associated with the spontaneous symmetry breaking of time diffeomorphism in a flat space[18].

Indeed, the hypothesis is based on the assumption that in its early stage, the universe underwent a decrease of symmetry number from a de Sitter universe (De Sitter, 1917) (10 dimensions symmetry group) to a FLRW universe (see eq. 1.8) (6 dimensions symmetry group). The four Goldstone bosons resulting from this symmetry breaking can be associated with a single scalar field perturbation π, representing the scalar degree of freedom added to the fields equations 1.1.

EFT, a reformulation of the Horndeski's models (Horndeski, 1974), is then a unified formalism where the modification of gravity is related to the presence of a scalar field as general as possible with a second order equation of motion. Moreover this reformulation sets aside the whole Horndeski's models space of unhealthy theories (*i.e.* avoiding any nonphysical effects), imposing the theory to be stable (no-ghost conditions[19] or gradient instabilities[20]) and prohibiting any super-luminal scalar or tensor perturbation propagation. Also, the EFT formalism, depending on its setting, is equivalent to work with usual scalar dark energy models. Among them the quintessence model (Sahni & Starobinsky, 2000; Caldwell et al., 1998), k-essence (Armendariz-Picon et al., 2000), Brans-Dicke (Brans & Dicke, 1961; Boisseau et al., 2000), $f(R)$ (De Felice & Tsujikawa, 2010; Sotiriou & Faraoni, 2010), Kinetic braiding (Deffayet et al., 2010), DGP (Dvali et al., 2000), Galileon Cosmology (Chow & Khoury, 2009), $f(G)$-Gauss-Bonnet (Nojiri & Odintsov, 2005) or Galileons (Nicolis et al., 2009; Deffayet et al., 2009).

The general EFT action[21], written conventionally in unitary gauge, implies a basis in which the perturbations of the scalar field π are absorbed, identifying the scalar field with the time coordinate $\pi \equiv t$. It reads[22]

$$\begin{aligned} S = S_m[g_{\mu\nu},\Psi_i] + \int \mathrm{d}^4x \,\sqrt{-g}\frac{M^2(t)}{2}\Bigg[& R - 2\lambda(t) - 2\mathcal{C}(t)g^{00} \\ & \mu_2^2(t)\delta(g^{00})^2 - \mu_3(t)\delta K\delta g^{00} + \epsilon_4(t)\left(\delta K^{\mu}{}_{\nu}\delta K^{\nu}{}_{\mu} - \delta K^2 + \frac{{}^{(3)}R\delta g^{00}}{2}\right)\Bigg], \end{aligned} \tag{1.39}$$

---

[18] The flatness being motivated by observation.

[19] Bounding from below the kinetic term of π in order to avoid macroscopic instabilities.

[20] Avoiding exponentially growing of perturbations.

[21] Whose the variational principle leads to a modified form of equations 1.1.

[22] In this gauge, the action is not covariant. It can become covariant by leaving the unitary gauge via the *Stueckelberg procedure* consisting in applying the time diffeomorphism: $t \to t + \pi$.

where S_m is the action for the matter field, $g \equiv \det g_{\mu\nu}$, ${}^{(3)}R$ the spatial Ricci scalar and R its time extension, $K_{\mu\nu}$ the extrinsic curvature over a hypersurface of fixed time defined by

$$K_{\mu\nu} = h_{\mu}^{\ \sigma} \nabla_\sigma n_\nu \ , \qquad h_{\mu\nu} \equiv g_{\mu\nu} + n_\mu n_\nu \ , \qquad K = K^{\mu}_{\ \ \mu} \ . \tag{1.40}$$

Finally, n_μ is the unit vector orthogonal to the hypersurface defined by a constant time. In this action we can see that the first contribution in R is nothing else but the Einstein–Hilbert action, leading to equations 1.1 when explicitly setting the varying Planck mass $M(t)$ to a constant $M^2 = (8\pi G)^{-1}$ (in natural units).

With such formalism, the expansion of the universe and structures evolution are thus captured in the six structural parameters

$$M^2(t), \lambda(t), \mathcal{C}(t), \mu_2^2(t), \mu_3(t), \epsilon_4(t) \ , \tag{1.41}$$

where it can be shown that the background evolution is only constrained by the functions $M(t)$, $\lambda(t)$ and $\mathcal{C}(t)$ while the others participate exclusively to the perturbed sector (Perenon et al., 2015). As shown in figure 1.6, a large sample of dark energy theories, previously quoted, are embedded in this formalism. Activating or not the structural parameters allows to discretely travel in the theory space spanned by the EFT while providing it a continuous value allows to appreciate the degrees of freedom of the considered model.

	$\mu = \frac{d \log M^2(t)}{dt}$	λ	$\mathcal{C}$	μ_2^2	μ_3	ϵ_4
ΛCDM	0	const.	0	0	0	0
Quintessence	0	✓	✓	0	0	0
k-essence [77]	0	✓	✓	✓	0	0
Brans-Dicke [78, 79]	✓	✓	✓	0	0	0
$f(R)$ [19]	✓	✓	0	0	0	0
Kinetic braiding [80]	0	✓	✓	✓	✓	0
DGP [16]	✓	✓	✓	✓	✓	0
Galileon Cosmology [82]	✓	✓	✓	✓	✓	0
$f(G)$ -Gauss-Bonnet [23]	✓	✓	✓	✓	✓	✓
Galileons [83, 84]	✓	✓	✓	✓	✓	✓
Horndeski [75, 76]	✓	✓	✓	✓	✓	✓

Figure 1.6: Explicit dark energy models covered by 1.39, when are activated or not the structural parameters 1.41. Credit: Piazza et al. (2014).

More specifically, applying the least action principle $\delta S = 0$ gives the modified Einstein's equations

$$\mathcal{C} = \frac{1}{2}\left(H\bar{\mu} - \dot{\bar{\mu}} - \bar{\mu}^2\right) + \frac{1}{2M^2}\Big(\rho_D + P_D + \sum_i \left[\rho_i + P_i\right]\Big) \ , \tag{1.42}$$

$$\lambda = \frac{1}{2}\left(5H\bar{\mu} + \dot{\bar{\mu}} + \bar{\mu}^2\right) + \frac{1}{2M^2}\Big(\rho_D - P_D + \sum_i \left[\rho_i - P_i\right]\Big) \ , \tag{1.43}$$

where $\bar{\mu} \equiv \mathrm{dlog}M^2/\mathrm{d}t$. In these expressions, ρ_D and P_D characterised the dark energy fluid and are defined thanks to the modified cinematic and dynamic Friedmann equations (analogous to equations 1.16 and 1.19)

$$H^2 = \frac{1}{3M^2(t)}\Big(\rho_D + \sum_i \rho_i\Big) \ , \tag{1.44}$$

$$\dot{H} = -\frac{1}{2M^2(t)}\Big(\rho_D + P_D + \sum_i \left[\rho_i + P_i\right]\Big) \ , \tag{1.45}$$

and showing that the modification of gravity can be both generated by the time dependence of the Planck mass (similar to take $G \to G_{\text{eff}}(t)$) and/or by the addition of a dark energy component minimally coupled to the metric. When combined, equations 1.44 and 1.45 gives a similar continuity equation as 1.18 for regular species, while modified for the dark energy component by introducing a non zero source term related to the modification of the Planck mass

$$\dot{\rho}_i + 3H(\rho_i + P_i) = 0 \ , \tag{1.46}$$

$$\dot{\rho}_D + 3H(\rho_D + P_D) = 3\mu M^2 H^2 \ . \tag{1.47}$$

Quantifying the impact of these modifications on cosmological quantities has been done in a phenomenological way in Perenon et al. (2015) for three quantities:

- the modification of the Newton constant $G \to G_{\text{eff}}(t, \Theta)$, where Θ represents the six structural functions of the EFT (see eq. 1.41)
- the modification of the clustering weighting function $\sigma_8(t) \to \sigma_8(t, \Theta)$
- the modification of the gravitational slip parameter (constant in a standard description) $\eta \to \eta(t, \Theta)$ where η is defined as the ratio of the two Newtonian potentials Φ and Ψ perturbing the flat FLRW metric such that (in Newtonian gauge)

$$\mathrm{d}s^2 = -(1 + 2\Phi)\mathrm{d}t^2 + a^2(1 - 2\Psi)\delta^K_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{1.48}$$

These three predictions are compared to the standard ones (ΛCDM) in figure 1.7, phenomenologically showing a wide and continuous variety of predictions in the perturbed sector. In particular, the various predictions of the quantity $f\sigma_8$, directly measurable in galaxy surveys (see chapter 4), are difficult to disentangle given the actual data. An improvement of statistics and probe combination are nevertheless expected to provide better constraints on these models. In section 1.3, I will present an example of observational strategy to achieve it.

Finally, let me mention that a recent observation has significantly reduced the total holding space of EFT models, involving a multi-messenger observation where both the gravitational wave and the electromagnetic spectrum counterpart have been characterised from the same source; a binary neutron star merger (Abbott et al., 2017). Assuming simultaneous emission of these two propagations, the two independent collaborations have detected the signal with a time-delay of 1.74 ± 0.05 seconds, tightly constraining the tensorial perturbation propagation parameter $c_T^2 = (1+\epsilon_4)^{-1}$ to $-3 \times 10^{-15} < c_T/c - 1 < 7 \times 10^{-16}$ (Perenon & Velten, 2019), definitely fixing $\epsilon_4 = 0$. Moreover, recent papers as Creminelli et al. (2020) exploit the gravitational wave frequency to severely constrain EFT models down to k-essence, while a

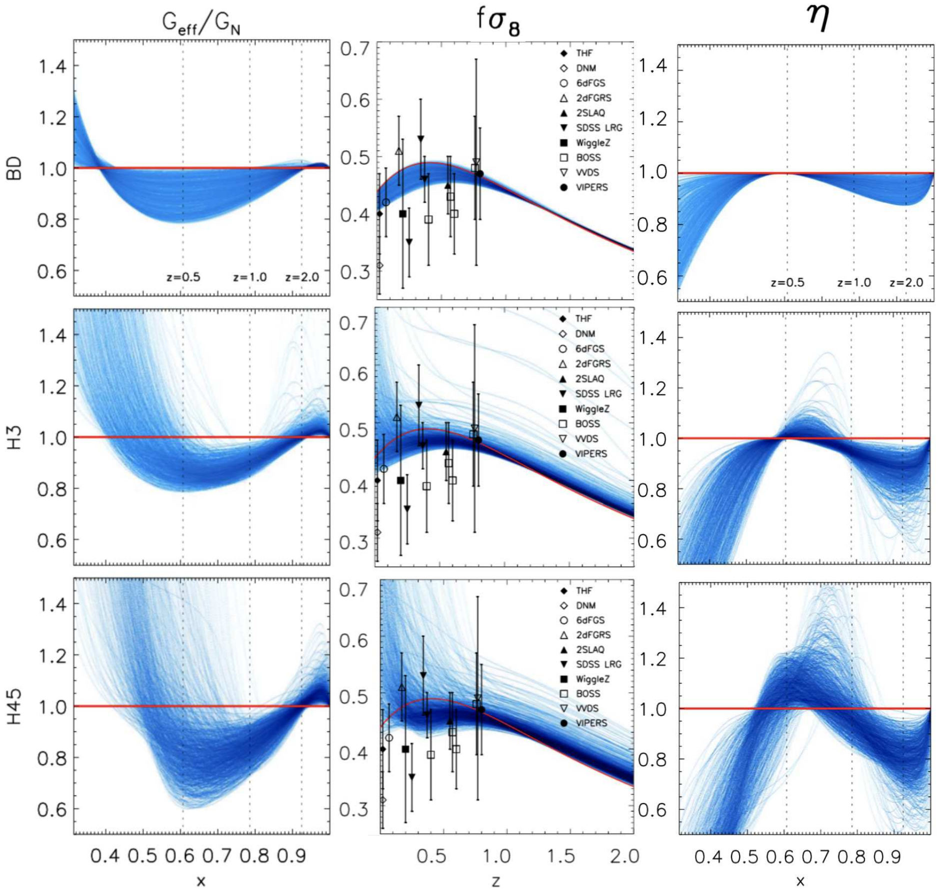


Figure 1.7: Joint results from Perenon et al. (2015) for the predictions of the modified Newton constant $G_{\rm eff}$ in the left panels, the $f\sigma_8$ parameter in the centre panel (f is the growth rate of structure in linear theory, defined in chapter 2) and the gravitational slip parameter η in the right panel. Note that some $f\sigma_8$ results pour various experiments (when fitting the ΛCDM model) are plotted. Each panel are time-parametrised either by the redshift or by $x \in [0.3, 1]$ (x running between the early time $x = 1$ and today $x = 0.3$). Three sub-spaces of EFT models are considered here: Brans-Dicke with non zeros μ and λ structural parameters (BD), cubic Galileon and Horndeski theories with non zeros λ and μ_3 (H3) and Horndeski theories with non zeros λ and ϵ_4 (H45), *i.e* each of them are described by the ΛCDM model with one additional EFT parameter. The red solid lines are representing the ΛCDM predictions.

much more optimistic conclusion that keeps alive EFT models has been advanced in de Rham & Melville (2018).

In conclusion, this non exhaustive description of modified gravity/dark energy theories gives an insight of the many possibilities to go beyond the standard model. These models are testable by observation, but they failed today to be enough constrained due to a lack of data. This will require new observations as in particularly precise galaxy clustering (and weak lensing) measurements to probe the perturbed sector through, for example, the $f\sigma_8$ parameter. In addition, further work needs to be done on the models and their theoretical viability to compare them to data. In fact, analysis, as it will be discussed in chapter 2, is not able to test an infinite number of cosmological models (related to the covariance matrix problem) and reducing the model zoology to solid ones is of paramount importance.

Massive neutrino

In this section, I will show another example of modification of the standard cosmological model, that can be tested with large scales structures of the universe. It introduces massive neutrinos and can participate to develop the standard model of particle physics by constraining the neutrino mass.

Setting aside the dark energy and cosmic acceleration problems, it has been noticed that probing the perturbed sector helps improving knowledge on particle physics. In particular some non cosmological evidences tend to converge toward a non-zero neutrinos mass and in consequence is expected to affect the clustering behaviour of the universe. Indeed, it can be shown that such property has the direct consequence of damping the structure formation inside the particle horizon (maximal travelling distance since their creation) of such massive neutrinos (Bond et al., 1980), an effect that can be detected with galaxy redshift surveys (Hu et al., 1998). In this section, I'll discuss the basics of this field of study and its cosmological implications.

Neutrinos are leptonic particles (spin ½) described by the standard model of particle physics as massless and chargeless; only subject to weak interactions. Postulated by Pauli (Pauli, 1930) then experimentally discovered by Cowan et al. (1956), they come in $N_\nu = 3$ flavor states $\nu_\alpha = (\nu_e, \nu_\mu, \nu_\tau)$, respectively the electronic, muonic and taunic states. A number that has been successfully confirmed by the LEP/SLC measurement when fitting the standard model: $N_\nu = 2.996 \pm 0.007$ (Zyla et al., 2020).

On the other hand, what the standard model of particle physics did not include, is that these three leptonic states can be seen as a quantum superposition of three light mass states $\nu_i = (\nu_1, \nu_2, \nu_3)$ as

$$\nu_\alpha = \mathcal{U}_{\alpha i}\ \nu_i\ , \tag{1.49}$$

where $\mathcal{U}_{\alpha i}$ is a unitary mixing matrix, parametrised in terms of three mixing angles θ_{12}, θ_{13} and θ_{23}. These mass states being considered as solutions of the Schrodinger equation, their evolution is governed by a time evolution operator implying that ν_α can oscillate from one flavor to another, depending on their distance of propagation.

Such feature was initiated by Pontecorvo in the 50's (Pontecorvo, 1957, 1958) and experimentally confirmed in solar neutrino experiments, by observing a reduction of the atmospheric neutrino flux in Super Kamiokande (Fukuda et al., 1998) and demonstrating that one initial solar neutrino is converted into other flavours in the SNO experiment (Ahmad et al., 2002).

Recent experiments of solar, atmospheric, reactor, and accelerator neutrino combined in the NuFit 4.0 analysis (Esteban et al., 2019) are jointly capable of giving the relative mass difference between the three mass states. In the context of standard model, the best-fit values of neutrino mass splittings are currently

$$\Delta m^2_{21} = (7.39^{+0.21}_{-0.20}) \times 10^{-5}\ \mathrm{eV}^2\ \text{(NH or IH)} \tag{1.50}$$

$$\Delta m^2_{31} = (2.525^{+0.033}_{-0.032}) \times 10^{-3}\ \mathrm{eV}^2\ \text{(NH)} \tag{1.51}$$

$$\Delta m^2_{32} = (-2.512^{+0.034}_{-0.032}) \times 10^{-3}\ \mathrm{eV}^2\ \text{(IH)} \tag{1.52}$$

where $\Delta m_{ij}^2 = m_i^2 - m_j^2$.
Constraining these mass differences allows to consider two hierarchical configurations:

- the normal hierarchy (NH) for which we have one heavier neutrino and two lighter ones
- the inverted hierarchy (IH) in the case we have two heavier neutrinos and a lighter one.

Note that for these two configurations, one massless neutrino is allowed while on the other hand, an equally shared (or degenerate) mass between the three neutrino species is an excluded scenario. Thus, previous results show that in normal hierarchy the total neutrino mass $\sum m_\nu$ is bounded from below at ~ 0.06 eV while in inverse hierarchy, this bottom limit is about ~ 0.1 eV.

As particle physics experiment can provide only an upper limit for the lightest neutrino mass (KATRIN (Aker et al., 2019) constraining $m_{\text{Lightest}} \lesssim 1$ eV).
Cosmology can be a powerful tool for probing neutrino physics. Indeed, even if cosmological measurement won't be able to measure individual neutrino mass states (Archidiacono et al., 2020), a measurement of the total neutrino mass can be inferred. As already mentioned in section 1.1.4, neutrino decoupling occurs at the early stage of the universe at ~ 1 MeV, before the photon decoupling. During this time interval, neutrinos evolve in an independent way respectively to photons, kept in thermal equilibrium with the rest of the cosmic plasma. In consequence, it can be shown that after the last scattering surface (the release of the CMB), neutrinos and photons temperatures are linearly related through $T_\nu = T_\gamma \left(4/11\right)^{1/3}$ (*e.g.* Dodelson, 2003; Lesgourgues & Pastor, 2006), expected today at $T_\nu^0 \sim 1.95K \sim 1.68 \times 10^{-4}$ eV when measuring $T_\gamma^0 = 2.73K$. Considering that neutrinos start to behave as non relativistic species when the universe sufficiently cools down such that $T_\nu \ll m_\nu$, the previous neutrino to photon temperature relation can be combined with the Bose-Einstein density distribution (see equations 1.25) which can be converted in adimensional density using relations 1.30 as

$$\Omega_\nu = \frac{\sum m_\nu}{93.14 h^2 \text{eV}} \ , \tag{1.53}$$

where $h \equiv H_0/100$.

Relation 1.53 is obtained in the degenerate scenario. However, given the current Δm_{ij}^2 results, at least two neutrino species are expected to be non-relativistic today. In the case where the third neutrino is either of mass $m_{\text{Lightest}} \ll T_\nu^0$ or massless, such relativistic species does contribute insignificantly to the global energy density budget. Therefore even in non-degenerate hierarchies, relation 1.53 is still valid and shows that inferring Ω_ν from galaxy surveys does not help in any way to constraint individual mass state.

Associating a non zero mass to neutrinos has a direct and measurable impact on the CMB power spectrum, and consequently on the galaxy distribution (see Lesgourgues & Pastor, 2006). Indeed, depending on the mass amplitude, the relativistic to non-relativistic transition may occur at different epoch $z_{\text{nr}}(m_\nu)$, obtained when the temperature $T \propto a^{-1}$ (assuming conservation of entropy during the radiation epoch) sufficiently drops to reach $m_\nu = \langle p \rangle$ (the average momentum). To do so, we use the non-relativistic approximation that neutrino velocity is thermal, *i.e.* $\langle p \rangle \approx 3T_\nu^{\rm nr}$. Recalling that $T_\nu^0 \sim 1.68 \times 10^{-4}$ eV, it reads

$$1 + z_{\rm nr} = 1980 \left[\frac{m_\nu}{\rm eV}\right] \ . \tag{1.54}$$

Furthermore, according to the time dependent neutrino propagation velocity

$$v_\nu = \frac{\langle p \rangle}{m_\nu} \approx \frac{3T_\nu}{m_\nu} \ , \tag{1.55}$$

one can define a typical neutrino free streaming scale $\lambda_{\rm FS}$, whose $k_{\rm FS} = 2\pi/\lambda_{\rm FS}$ is its Fourier analogue, only able to increase over time. Thus the specific mode $k_{\rm nr} = k_{\rm FS}(z = z_{\rm nr})$, representing the causal/particle horizon at the transition time, can be seen as a reference when describing the influence of massive neutrinos on the clustering behaviour. Given that the total matter is subject to the gravitational potential through the Poisson equation as (ϕ standing for the gravitational potential)

$$\vec{\nabla}^2\phi = 4\pi G\bar{\rho}_{\rm tot}a^2\left((1 - f_\nu)\delta_{\rm cdm} + f_\nu\delta_\nu\right) \tag{1.56}$$

where the specific baryonic contribution is not written for simplification, $f_\nu \equiv \rho_\nu/\rho_{\rm tot}$ ($\rho_{\rm tot}$ is the total density) and where I anticipate the use of local fluctuation of densities $\delta_i \equiv \delta\rho_i/\bar{\rho}_i - 1$ ($\bar{\rho}$ defined as the background/average density of the specie). Two cases can be discussed:

- for $k < k_{\rm nr}$ (large scales), it corresponds to wave modes reachable by neutrinos in their relativistic phases, a condition prohibiting any clustering behaviour, neither enhancing nor suppressing it. In other terms, these scales cannot be affected by non relativistic free streaming and neutrino behaves as the dark matter itself : $\delta_\nu = \delta_{\rm cdm}$ (they cannot escape from gravitational well potentials since their velocities do not allow it). Regarding the Poisson equation using such result, the gravitational potential felt by baryonic matter (if we are interested in visible structure) is only induced by dark matter and no impact of m_ν on clustering is expected.

- for $k > k_{\rm nr}$ (small scales), it corresponds to the scales inside the neutrino particle horizon in their non-relativistic phase. But due to their low masses, thermal velocity spread them in an uniform manner : $\delta_\nu = 0$. Thus eq. 1.56 tells us that massive neutrinos in facts reduce the gravitational potential. We may expect that clustering is damped for these typical scales.

Additionally, coming back to relation 1.54, a larger mass mechanically increases $z_{\rm nr}$. Prematurely removing from the total radiation budget neutrinos for the benefit of non-relativistic matter lead to modify in a non trivial way the first acoustic peak amplitude of the BAO observable in the CMB and galaxy probes (as the various correlation functions, see Lesgourgues & Pastor, 2006). Since this effect is not detected when comparing to observations, it gives a tight cosmological constraint by setting the upper bound at $\sum m_\nu < 1.67$ eV (using eq. 1.54).

The combination of all of these constraints coming from Cosmology and particles physics communities is displayed in figure 1.8. It first shows that joint cosmological

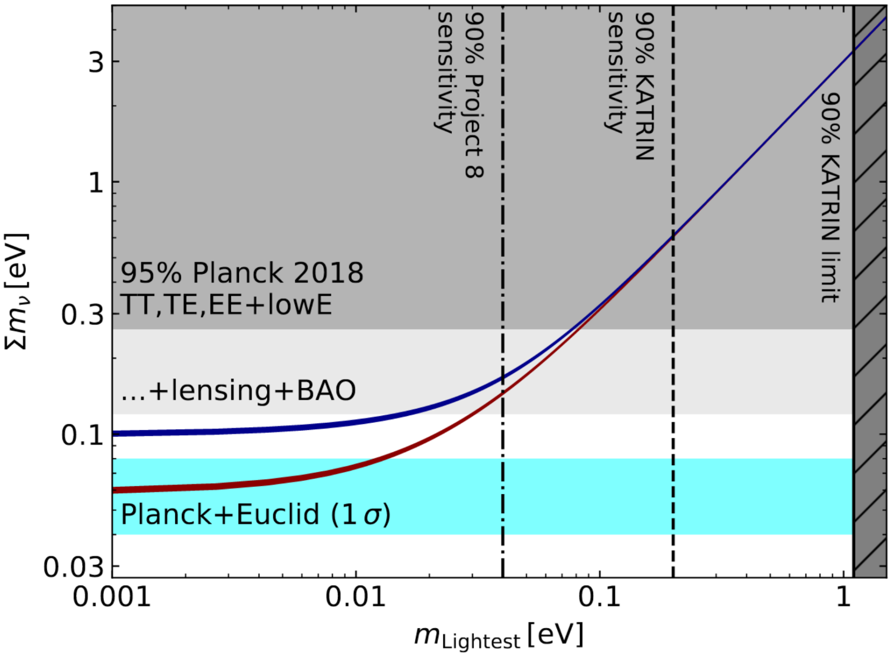


Figure 1.8: Total neutrino mass as a function of the lightest neutrino mass. The red and blue lines depict respectively the normal and inverse hierarchy scenarios. In horizontal grey area are displayed exclusion regions from cosmological data (Aghanim et al. (2018)), in blue is the 2σ error expected from a combination of Planck and Euclid (central value arbitrarily set to $\sum_\nu = 0.06$ eV and $\sigma(\sum m_\nu) = 0.02$ eV, see Sprenger et al. (2019)). The vertical dashed references represent the expected sensitivity of KATRIN and Project 8 (Ashtari Esfahani et al., 2017). Credit: Archidiacono et al. (2020).

probes exclude at 95% confidence level (Aghanim et al., 2018) $\sum m_\nu > 0.26$ eV when assuming a degenerated hierarchy. Moreover, a Euclid (Laureijs et al., 2011) (hereafter introduced) + Planck combination is expected to provide an error bar allowing to eventually exclude normal hierarchy if one considers that the sum of the neutrino masses is centered on the proposed (and arbitrary set) position.

In conclusion, as in the case of modified gravity models, some theoretical predictions in the perturbed (clustering) sector are expected to slightly deviate for the standard model onesn and can be measured with an accurate galaxy survey.

In the next section, I'll present Euclid, one of the next generation galaxy surveys whose the ambition is to focus on understanding the dark energy questions.

1.3 The Euclid survey

As already mentioned in section 1.2.1, disentangling the large variety of cosmological models claiming to describe gravity will have to go through a coordinated effort between theoreticians and observers. Whereas the former is expected to propose only viable models (see section 1.2.2), observational tests and accurate predictions of the perturbed sector, the second requires several observational aspects, whether they are related to the measurements or to the data analysis. Most of the future galaxy surveys, on top of investigating several probes, will have to cover a large

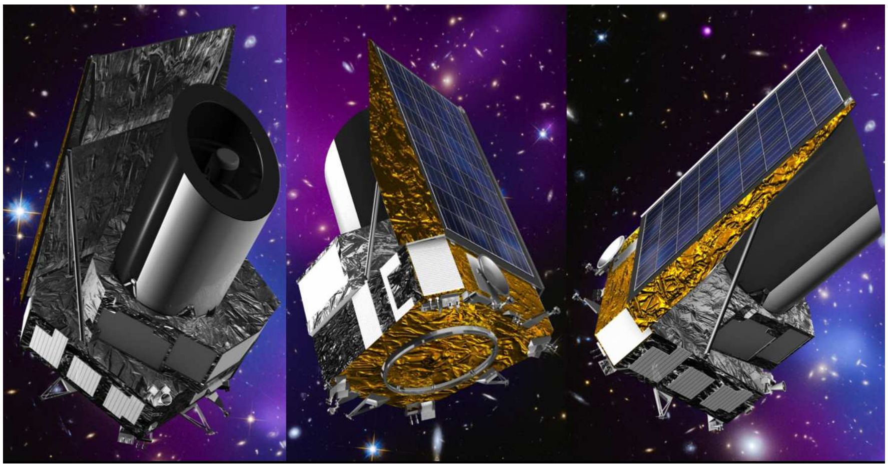


Figure 1.9: Euclid satellite, artist image , Credit : euclid-ec.org

sky area and redshift range in the interest of both a statistical improvement and an overview of the cosmological structures time evolution.

Addressing all of these aspects, Euclid[23] (Laureijs et al., 2011) is the next space Cosmology mission after Planck from the European Space Agency (ESA), and one of the most promising survey to understand the physical origin of the dark sector of the universe: the dark energy and the dark matter.

The satellite consists in a 1.2 m Korsch telescope (see figure 1.9) and two instruments; the VIS and the NISP. The first will be used for imaging in the visible spectrum (VIS) while the other is a photo-spectroscopic instrument working in the near-infrared wavelength domain specifically suited to accurate redshift measurements, called the Near Infrared Spectrometer and Photometer (NISP). Expected to be launched in 2022, Euclid will be transferred to the L2 Lagrange point of the Sun-Earth system, for a lifetime of about 6 years. Thanks to the wide survey of 15000 deg^2 (+ 40 deg^2 in a second deep field survey), Euclid plans to image a billion of galaxies and measure tens of millions of galaxy redshifts for a total redshift range of $z \in [0.7 - 2]$. This will allow to study the evolution of large structures of the Universe up to a cosmic time of about 10 billion years (*i.e.* more than 75% of the age of the Universe).

In particular, the mission is optimised for two cosmological probes, the galaxy clustering and the weak lensing (low gravitational shear). First, the galaxy clustering probe will use the NISP instrument to cartography with high precision the redshift-space position of more than 50 million galaxies. The distances are obtained by a spectroscopic sky survey based on a slitless spectrograph specifically set to allow the detection of $H\alpha$ emission lines for more than 45% of the detected objects up to a limit flux of 3×10^{-16} erg $\mathrm{s}^{-1}\mathrm{cm}^{-2}$ and for objects extended by 1 arcsecond. This represents 3000 galaxies per deg^2 ($0.001(1 + z)$ of redshift accuracy and 0.3 arcseconds of angular pixel resolution) detected in the wave-length range $900 - 2000$ nm through a broad band Y, J, H of filters. This way, several aspects of galaxy clustering will be analysed, as the BAO signature and the redshift-space distortion effect (see chapter 4) on correlation functions.

---

[23] euclid-ec.org

On the other side, the shear measurement will be derived from the measurement of the shape of distant galaxies and their distance. The visible imager (VIS) instrument will measure the ellipticity (with an accuracy better than 2×10^{-4} and controlled spatial variations better than 2×10^{-3}) of 30 galaxies per arcmin^2 in a wavelength range of $550 - 900$ nm thanks to wide visible band R+I+Z filters. It will evaluate with high resolution (0.1 arcseconds) the shape distortion (and the photometric redshift with $0.005(1 + z)$ accuracy) of a billion of distant background galaxies induced by gravitational lensing effects (Weak Lensing). In doing so, a dark matter map distribution and its time evolution will be reconstructed, a direct witness allowing to understand its influence on the growth rate of structure across the cosmic epochs, but also its nature and properties.

Each probe being a mine of common cosmological information and featured by various complementary aspects (measurement of gravitational potential, matter density and velocity fields), their measurement will be embedded in a combined analysis, ensuring tighter constraints on cosmological parameters. For instance, Euclid aims at measuring the equation of state parameter for the dark energy fluid parametrised as $w(a) = w_0 + w_a(1 - a)$ (Chevallier & Polarski, 2001; Linder, 2003) to an accuracy of 2% for the constant component w_0, and 10% for a possible variation as a function of redshift, encoded in w_a.

Note that since these two probes are partly measuring the same objects in the same volume of the sky with different instruments, it will allow the verification and control of systematic errors in the final interpretation. Moreover, by implementing a deep field twice as deep as the wide field and covering 40 deg^2, Euclid will be able to monitor the stability of the instruments by returning to the same objects throughout the mission.

Thanks to these measurements, many cosmological and astrophysical tests will be possible. Non exhaustively can be quoted the reconstruction analysis of galaxy clusters, cosmic voids, integrated Sachs Wolfe effect that are very promising on the cosmological side. Also, the wide astrophysical database will allow the study of galaxy evolution, galaxy structure, planetary searches (exoplanets) and other transient objects such as distant supernovae (particularly interesting on the cosmological point of view as well).

Finally, it is worth pointing out that breaking statistical records is not an end in itself. Before carrying out a combined analysis, a control of systematic errors and theoretical bias for each of them is already one of the key challenging issue.

Testing General Relativity at different scales involves beforehand to predict the theoretical errors accompanying such measurements. In particular, this thesis attempts to quantify the puzzling effect of non-linearities related the the prediction of theoretical errors on the estimation of correlation function. In the next chapter, I will review in more details the clustering probes, and the way they can be estimated and analysed to extract the cosmological parameters values.

Chapter 2

Galaxy clustering : theory and observation

In the previous chapter, we saw how the cosmological principle combined to the Einstein field equations can describe the universe as a whole. I briefly introduced the link between cosmological parameters and quantities characterising the evolution of the universe as if it was uniformly filled with cosmic fluids.

However, this picture has to be completed by noticing that the universe does not appear uniform when considering smaller scales than $\sim 150h^{-1}$Mpc (Marinoni et al., 2012). Indeed, the cosmic structures such as galaxies or cluster of galaxies has to be formed thanks to gravity, through the evolution of initial and hypothetical tiny perturbations present in the matter fluid. Also called inhomogeneities, they constitute the large scale structures of the universe and can be probed with the study of the dark matter clustering.

In this chapter, I will detail how one can take advantage of the dark matter clustering in order to constrain cosmological parameters. To do so, I will first review in section 2.1 how the inflationary paradigm is able to provide a mechanism generating those initial fluctuations in cosmic fields. Then, after presenting the key steps of the perturbation theory allowing to describe dark matter inhomogeneities, I will describe in section 2.2 how the theoretical predictions of the growth of cosmic structures can be compared to observations through statistical analysis.

As stressed in chapter 1, cosmological models are set up to describe gravitation on various scales. I discuss in section 2.3 the way they can be disentangled based on observations in a Bayesian analysis framework. It will lead me to introduce a well known problem of such analysis, related to the way observational data errors, together with their correlations, are estimated. Being one of the main motivation of this thesis, it will lead me to present my own contribution to this topic.

2.1 The inhomogeneous universe

In the standard paradigm, large scale structures of the universe have reach their current state by the gravitational amplification of initial tiny inhomogeneities. Before studying the physical laws ruling this evolution, let me detail the mechanism allowing to generate initial inhomogeneities into cosmic fluids.

2.1.1 Primordial fluctuations

Motivations for an inflationary phase of the universe

After the success of the standard model based on the Planck measurements, a clear consensus then developed around the Big Bang model. Leaving further questions such as the origin of the apparent anisotropies, the analysis of the CMB actually turned out to be a great opportunity to develop and orient theories aiming at explaining them.

On top of this point, the inflationary process stems from the need to solve fundamental cosmological issues as the horizon and the flatness problems. Indeed, due to observational evidences for a flat universe (or at least nearly flat, see section 1.2.1), the inflation formalism makes the assumption that the intrinsic curvature of the universe is zero. In this case the FLRW metric 1.8 is conformaly Minkowskian by mean of the introduction of the conformal time $\tau = \int \mathrm{d}t/a(t)$,

$$\mathrm{d}s^2 = a(t)^2 \left[\mathrm{d}\tau^2 - \mathrm{d}\vec{r}^2\right] \ . \tag{2.1}$$

In such a case, the geodesic path of a photon ($\mathrm{d}s^2 = 0$) is linear in a conformal spacetime diagram. This allows to define the particle horizon as a virtual boundary of the region that has been in the past light cone likely to influence a given point at a given time. It delimits the causality domain

$$\chi_p(\tau) = \tau - \tau_i = \int_{t_i}^{t} \mathrm{d}t'/a = \int_0^a \mathrm{d}\ln a \ H^{-1} \ , \tag{2.2}$$

where we set the initial time $t_i = 0$ and decompose $\mathrm{d}t = \mathcal{H}^{-1}\mathrm{d}a$ to make apparent the quantity $\mathcal{H}^{-1} \equiv (aH)^{-1}$ which has the dimension of a length, called the *the conformal expansion rate* or *causal horizon*.

Formally, explaining how the universe reach its actual state confront us to the *Cauchy problem*. It shows our inability to determine the initial conditions of Einstein field equations in order to expect a global existence and uniqueness of the solutions. Given such lack of information, theoretical models need to extrapolate toward fine tuned initial conditions. More specifically, let me discuss two concrete examples.

- The CMB measurements show a nearly isotropic temperature map with tiny relative fluctuations of the order of 10^{-5}. It indicates that the universe at the time of recombination ($T_u \sim 0.1$ eV or $\sim 10^5$ years) had already reached a high level of homogeneity, witness of a global thermal equilibrium happening on very large scales. In other terms, putting aside any fine tuned initial conditions, it means that all regions were in thermal contact in the past. Furthermore, noticing that the time evolution of the causal horizon can be written

  $$\mathcal{H}^{-1} \propto a^{\frac{1}{2}(1+3w)} \ , \tag{2.3}$$

  it shows that during a matter ($w = 0$) or radiation ($w = 1/3$) domination epoch, the horizon monotonically increases over time. Thus assuming that the universe was filled by standard energetic content and that the recombination happens when the species temperature allows it, one can show that about

4.4×10^3 regions were causally disconnected (representing a causal angular region comparable to the angular size of the moon, see Simon, 2016). The horizon problem is raised: how can causally disconnected regions be in thermal equilibrium if they had never been in causal contact in the past ?

- If we assume that the spatial curvature k of the universe is not known, then $\Omega_k \equiv -k\mathcal{H}^{-2}$ is contributing to the total energy budget of the universe. Once again, $\mathcal{H}^{-1}$ is an increasing function of time when considering standard radiation or matter, implying that the curvature should be increasing with cosmic time. However, given the observational evidences for a universe nearly flat today (see section 1.2.1), the curvature at primordial epochs must be infinitely close to zero, but not completely null. Thus, the upper limit of the value $|\Omega_k|$ at the time of the Big Bang Nucleosynthesis (3 min post Big Bang) must be of the order of $\sim 10^{-16}$, while dropping to $\sim 10^{-61}$ at the time of the Planck era (see Baumann, 2011). Without invoking an incredible fine tuning, it seems therefore inconceivable to assume that such values were specifically calibrated in these early stages to evolve up to the universe we are observing. This is the *flatness problem*.

The inflaton field

The idea of the inflation precursor Guth (1987) was that instead of having such precise and unstable fined tuned initial conditions, it was more consistent to assume that whatever those initial conditions were, the universe should evolve toward the one we see, as an attractor.

In order to achieve such result, a direct solution is to assume that the comoving horizon was decreasing at the primordial stage of the universe. This allows to define a critical requirement which should be met by inflation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}^{-1} < 0 \; . \tag{2.4}$$

It immediately solves the horizon and flatness problems if the condition 2.4 lasts long enough. Indeed, assuming that this condition is driven by a dominating fluid described by an equation of state parameter $w_f = -1$ (a de Sitter universe), the continuity equation 1.18 implies an exponentially growing scale factor $a(t) \propto e^{Ht}$, or equivalently $a(\tau) \propto -(H\tau)^{-1}$. This has for consequence to push the singularity $a = 0$ toward an initial conformal time $\tau_i \to -\infty$.

On this basis, the idea is to propose a fluid candidate and a mechanism imposing condition 2.4 in a long enough period such that the CMB regions a priory disconnected were in causal contact during the primordial inflation. Quantitatively, observation constrain inflation to last at least $60e$-folds[1] in order to solve both the horizon and the flatness problems.

The proposed inflationary mechanism is based on the release of a phenomenal amount of potential energy in a short period of time from an *inflaton field*, caused by the symmetry breaking between the gravitational and the electro-weak interactions at about $T \sim 10^{15}$ GeV. These beyond standard model particles, forming a scalar

---

[1] The number of e-folds counts the number of times the universe has expanded by a factor e.

field for simplicity[2], is characterised by an equation of state parameter $w \simeq -1$ which, as for Λ, is able to drive a cosmic acceleration ($\ddot{a} > 0$, see equation 1.19).

Be ϕ the inflaton field of spin 0 homogeneously spread in the universe at $T \sim 10^{15}$ GeV and minimally coupled to gravity through the total action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) - \frac{R}{2}\right] \equiv S_\phi + S_{\mathrm{EH}} \ , \tag{2.5}$$

where $V(\phi)$ stands for its potential energy, $g \equiv \det(g_{\mu\nu})$ and taking S_{EH} and S_ϕ as respectively the Einstein-Hilbert action (leading to equations 1.1) and those of the scalar field.

Applying the least action principle $\delta S = 0$ allows to identify the stress-energy tensor of the ϕ-field, and in consequence its equation of state parameter. On the other hand, $\delta S_\phi = 0$ gives its equation of motion. They read

$$w_\phi = \frac{\dot{\phi}^2/2 - V(\phi)}{\dot{\phi}^2/2 + V(\phi)} \ , \qquad \ddot{\phi} + 3H\dot{\phi} + \frac{\mathrm{d}}{\mathrm{d}\phi}V(\phi) = 0 \ , \tag{2.6}$$

interpreted as a rolling ball on an inclined slope, slowed down by the friction term $3H$. As long as the potential energy of the inflaton remains greater than its kinetic energy, such that $w_\phi < -1/3$, inflation occurs. The potential should therefore be well designed (Starobinsky, 1987) to give enough time to the universe to reach a causal equilibrium and solve the horizon problem[3]. When the particle field reaches the potential well $\min\left[V(\phi)\right]$, its kinetic energy is at its maximum and the universe moves into an era called *reheating*. During this period, the kinetic energy decays in a relativistic and hot plasma of radiation and particles of the standard model (*e.g.* *Bassett et al., 2006*). The radiation domination of the universe begins.

Quantum origin of fluctuations

As already mentioned, in addition to solve the horizon and flatness problem at the same time, inflation provides also a possible explanation for initial inhomogeneities. The inflaton field being a quantum field, it is free to fluctuate according to the quantum harmonic oscillator. Thus, the flat metric is expected to be perturbed as

$$\mathrm{d}s^2 = (1 + 2\Phi)\mathrm{d}t^2 - 2aB_i\mathrm{d}x^i\mathrm{d}t - a^2[(1 - 2\Psi)\delta_{ij} + 2E_{ij}]\mathrm{d}x^i\mathrm{d}x^j \ , \tag{2.7}$$

where Φ and Ψ, B_i and E_{ij} are respectively scalar, vectorial and tensorial perturbations[4] of the metric and δ^{ij} is the Kronecker delta. Defining the gauge transformations $t \rightarrow t + \alpha$ and $x^i \rightarrow \delta^{ij}\beta_{,j}$, one can identify the invariant

$$\mathcal{R} = \Psi + \frac{H}{\dot{\bar{\phi}}}\delta\phi \ . \tag{2.8}$$

Here, $\bar{\phi}$ is the averaged background value of the inflaton field and $\mathcal{R}$ represents the scalar curvature fluctuation resulting from the inflaton fluctuation $\delta\phi$ in addition

---

[2] Note however that a vector field can also lead to condition 2.4.

[3] A quasi-flat slope like the Starobinsky potential allows such equilibrium, justifying the common name given to this model: *slow-roll inflation*.

[4] It can be shown that these perturbations evolve independently and do not influence each other.

to the scalar perturbation of the spatial part of the metric. Without entering into details (see for example Baumann, 2011), it can be shown that $\dot{\mathcal{R}} = 0$ when quantum fluctuations leave the horizon. It means that as $\mathcal{H}^{-1}$ decreases, the quantum perturbations accumulate in the same state and by analogy with Einstein-Bose condensate, quantum fluctuations are stretched and frozen, becoming classical (Riotto, 2003).

Finally, the quantum nature of inflation motivates the fact that curvature fluctuations are distributed according to a Gaussian distribution, characterised by its power spectrum (Harrison, 1970; Zeldovich, 1972)

$$P_{\mathcal{R}}(k) = \frac{H^2}{2k^3}\frac{H^2}{\dot{\phi}^2}\bigg|_{k=\mathcal{H}} \tag{2.9}$$

where k are the Fourier modes. The fact that in eq. 2.9 $k^3 P_{\mathcal{R}}(k) = \text{cte}$ implies that the energy spectral distribution of the curvature perturbations is equally spread over the three dimensions, or equivalently that perturbations are scale-independently distributed.

Thus inflation, in addition to solve both the horizon and flatness problems, allows to set a theoretical ground to describe initial fluctuations into the two metric potentials Ψ and Φ (see equation 2.7). Since density perturbations in the matter field are coupled to the Newtonian potential Φ through the Poisson equation (at least on scales much smaller than the horizon), it turns out that those initial inhomogeneities in the metric will induce perturbations in the matter distribution. It is therefore of paramount interest to predict how the matter fluctuations are growing in time, in order to compare it to observations taken at various stages of the universe. This is the subject of the next section.

2.1.2 Evolution of fluctuations in the matter field

During inflation, the size of the comoving Hubble horizon $\mathcal{H}^{-1}$ is decreasing and curvature (or metric) perturbations are forced to get out of it. When inflation ends, the inflaton scalar field decays and only curvature (or potential) fluctuations remain. As already mentioned, this is providing a starting point for matter fluctuations to form and evolve.

Also in chapter 1, we saw that at early time, the universe is dominated by radiation until the matter-radiation equality, happening at $a_{\text{eq}} = \Omega_{r,0}/\Omega_{m,0}$. After this equality, the universe then starts being dominated by matter. As a consequence, we expect the matter fluctuations to behave in a different way depending on which cosmic fluid is dominating the total energy budget of the universe.

First, during radiation domination, matter perturbations δ are growing as the scale factor ($\delta_{\vec{k}} \propto a$ in Fourier space) when they are defined on scales larger than the size of the horizon. Instead, matter fluctuations on scales smaller than the horizon are frozen, *i.e.* they do not grow in time. In turn, after the equality, matter fluctuations evolve as the scale factor independently from their scale (either outside or inside the horizon). To this must be added the information that the comoving horizon starts growing from the end of inflation (while it was decreasing during inflation).

Thus as a reference, the equality time can be translated into a crossing scale $k_{\rm eq}$ corresponding to the Fourier analogue of the size of the horizon at the equality epoch ($a_{\rm eq}$). Large scales, thus small wave modes ($k < k_{\rm eq}$) progressively re-enter the horizon after the equivalence and are not affected by the freezing due to radiation domination. Therefore these modes have grown as the scale factor since the end of inflation. Instead, considering smaller scales ($k > k_{\rm eq}$), modes entering the horizon during the radiation epoch are more and more suppressed, as their scale becomes smaller. Indeed, as short wave modes are entering earlier the horizon, it results that they are more affected by the freezing (Meszaros, 1974).

All of these informations can be encoded in the evolution of matter fluctuations through the transfer function $T(k,a)$, such that the matter perturbation can be expressed in terms of the initial distribution (at the end of inflation) as $\delta_k(a) = T(k,a)\delta_k(a_i)$. The transfer function, that has been widely investigated by many authors (*e.g.* Bardeen et al., 1986; Efstathiou et al., 1992; Eisenstein & Hu, 1998b), can roughly be describe by

$$T(k) \simeq 1 \qquad\qquad k \ll k_{\rm eq} \;, \tag{2.10}$$
$$T(k) \simeq (k_{\rm eq}/k)^2 \ln(k/k_{\rm eq}) \qquad\qquad k \gg k_{\rm eq} \;. \tag{2.11}$$

In practice other effects such as BAO can be encoded in the transfer function, as well as the effect of massive neutrinos which can be efficiently computed using Boltzmann codes like `CLASS`[5] (Blas et al., 2011) or `CAMB`[6] (Lewis et al., 2000; Howlett et al., 2012). Moreover, since the equations which have to be solved are valid only in the linear regime (when perturbations remain small), then during the matter domination when relativistic species are subdominant and on scales well inside the horizon, one can resort to Newtonian dynamics to model the evolution of matter fluctuations, even when perturbations become non-linear. Indeed, from an Eulerian point of view one can require that the change in mass of a volume is given by the flux of matter getting in or out of this volume, namely the mass conservation. In addition, we can apply Newton's law for the dynamics of a volume element. This is equivalent to assume that the matter field must satisfy to the Navier–Stock fluid equations.

Usually in Cosmology, we convert the system in comoving coordinates and in conformal time, while we define perturbations with respect to background quantities. In particular the *matter density contrast* (or *density fluctuations*) field is defined as

$$\delta(\vec{x},t) = \frac{\rho(\vec{x},t)}{\bar{\rho}(t)} - 1 \;. \tag{2.12}$$

In such a coordinate system, the velocity due to the Hubble flow is not explicit and we can define the peculiar velocity field $\vec{v} \equiv a\ \mathrm{d}\vec{x}/\mathrm{d}t$ as

$$\vec{V} = H\vec{r} + \vec{v} \;, \tag{2.13}$$

where $\vec{V}$ represents the true physical velocity (accounting for both the Hubble flow and the peculiar velocity). We thus obtain the system

---

[5] class-code.net

[6] camb.info

$$\frac{\partial \delta}{\partial \tau} + \vec{\nabla}.[(1+\delta)\vec{v}] = 0 \qquad \text{Continuity equation} \tag{2.14}$$

$$\frac{\partial \vec{v}_p}{\partial \tau} + \mathcal{H}\vec{v} + a^{-1}(\vec{v}.\vec{\nabla})\vec{v} = -\vec{\nabla}\phi \qquad \text{Euler equation} \tag{2.15}$$

$$\vec{\nabla}^2\phi = 4\pi G\bar{\rho}a^2\delta \leftrightarrow \vec{\nabla}^2\phi = \frac{3}{2}\Omega_m\mathcal{H}^2\delta \ . \qquad \text{Poisson equation} \tag{2.16}$$

In the above equations, ϕ represents perturbations in gravitational potential, which is related to the matter density perturbation through the Poisson equation. The two last relations can be combined in order to give the non-linear system ruling the evolution of the matter density contrast δ and its associated peculiar velocity field $\vec{v}$

$$\frac{\partial \delta}{\partial \tau} + \theta = -\vec{\nabla}.(\delta\vec{v}) \qquad \text{Continuity equation} \tag{2.17}$$

$$\frac{\partial \theta}{\partial \tau} + \mathcal{H}\theta + \frac{3}{2}\Omega_m\mathcal{H}^2\delta = -\vec{\nabla}.[(\vec{v}.\vec{\nabla})\vec{v}] \qquad \text{Euler + Poisson equations} \tag{2.18}$$

where $\theta \equiv \vec{\nabla}.\vec{v}$ is the divergence of the velocity field.

Solving this system of equations is of great cosmological interest. As pointed out by Wang & Steinhardt (1998) and as shown by equation 2.18, the cosmological parameters such as the matter density $\Omega_{m,0}$ or H_0 are affecting the way the divergence of the velocity field is evolving. Thus through the continuity equation, the evolution of the expansion rate is modifying the way the density contrast is changing in time. That is the reason why clustering probes allow to constrain cosmological parameters.

Due to the non-linear couplings between density and velocity appearing on the right hand side of equations 2.17 and 2.18, the full non-linear solutions are not trivial to compute. However, in order to understand how fluctuations are evolving, the system of equations can be linearised by setting to zero the non-linear terms. This allows to express a differential equation for the density contrast which is

$$\ddot{\delta} + 2H\dot{\delta} - \frac{3}{2}H^2\Omega_m\delta = 0 \ , \tag{2.19}$$

where the dot means partial derivative with respect to the cosmic time t. Analogous to the equation of motion 2.6, equation 2.19 means that derivatives of δ are sourced by the matter density $H^2\Omega_m$ ($\propto \rho_m$), showing that a higher matter density will produce a quick growth of matter fluctuations. On the other hand, the way the universe is expanding tends to slow down the growth of matter fluctuations, as shown by the friction term $2H$. As a result, equation 2.19 shows that the cosmological parameters are indeed modifying the way linear fluctuations are growing.

In addition, equation 2.19 admits a separable solution in space and time, such that $\delta(\vec{x}, t) = D(t)\epsilon(\vec{x})$ (Peebles, 1980a). The spatial part refers to the configuration of the density field at a given time (initial conditions for example), while the *growth factor* $D(t)$ shows how fluctuations are evolving. Being a second order linear differential equation, the general solution for the growth factor $D(t)$ can be built upon linear combinations of two independent solutions, namely the growing and decaying modes D^+ and D^-. It has been shown by Heath (1977) that in a ΛCDM universe (possibly admitting a non zero spatial curvature), when neglecting radiation (deep in the matter domination), the decaying mode is proportional to the expansion rate H. This knowledge allows to resort to the *variation of the constant* method by finding a function of time K such that $D^+(t) = K(t)H(t)$. In practice, during matter domination, the universe is close to an Einstein-de Sitter universe, implying that the decaying mode decreases as $a^{-3/2}$. It is then common practice to neglect it, leading to the expression of the growth factor as a function of the scale factor of the universe

$$D(a) \propto H(a) \int_0^a \frac{\mathrm{d}a'}{[a'H(a')]^3} \,. \tag{2.20}$$

The above expression shows explicitly the dependence of the growth factor $D(a)$ with respect to the expansion rate $H(a)$.

Having a solution in the linear regime for the matter density contrast, the way the matter velocity field is behaving as a function of the scale factor of the universe can also be predicted. Thanks to the linearised continuity equation 2.17, it appears that on large scales, the divergence of the peculiar velocity field is given by the conformal time derivative of the matter density contrast, leading to the relation

$$\theta(x, a) = -aH(a)f(a)\delta(x, a), \tag{2.21}$$

where the function $f(a) \equiv \partial \ln \delta / \partial \ln a$ is the *growth rate of matter fluctuations*. Equation 2.21 allows to understand that on large (linear) scales the velocity field is maximally correlated to the matter density contrast. As a result, by observing how the divergence of the velocity field evolves, one could get some constraints on the growth rate of matter fluctuations $f(a)$. Indeed, several authors paid attention to the cosmological dependence of the growth rate (Peebles, 1980a; Fry, 1985; Lahav et al., 1991), where it has been shown that a interesting way of exhibiting its cosmological dependence is through the growth index γ parametrised such that $f(a) = \Omega_m^\gamma$, then theoretically justified (Steigerwald et al., 2014).

In conclusion, in the linear regime, primordial fluctuations are scale-independently amplified by gravitation, while this effect can be modulated by the cosmological parameters. This is showing that the general expression of matter fluctuations can be written in terms of initial conditions as

$$\delta_k(a) = \delta_k(a_i)T(k, a_m)D(a)/D(a_m), \tag{2.22}$$

where the particular scale factor a_m is chosen to be deep in the matter domination ($a_m \sim 10^{-2}$). This way the transfer function introduced before is taking into account the early evolution (from the end of inflation to the matter domination) of the matter fluctuations, while the growth factor D expresses the linear evolution starting in the matter domination up to present day.

Of course, when decreasing scales down to few Mpc, structure formations can no longer be described linearly. In this regime, peculiar velocities as well as the density contrast can reach values much larger than unity, and the right hand side of equations 2.17 and 2.18 cannot be safely neglected. Despite that several attempt are still made to understand the weakly non-linear regime (Bernardeau et al., 2002), it is necessary to resort to N-body simulations to probe properly the non-linear structures formation.

2.2 The clustering of matter as a cosmological probe

As shown in the previous section, tracking the evolution of matter perturbation or those of its velocity field divergence across cosmic time, allows to constrain cosmological parameters. However observationally, we only have access to our past light cone, meaning that we cannot follow a single perturbation in time. We thus need to rely on a statistical description of the matter fluctuations in the universe, called clustering analysis of the large scale structure of the universe. In the present section, I show that the study of the statistical properties of the matter fluctuations is a powerful cosmological probe. This will motivate the work carried on during these three years of PhD.

2.2.1 The power spectrum

In addition to avoid following single perturbations to measure how fast they are growing in time, adopting a statistical description allows to coherently apply the cosmological principle. But because of the presence of inhomogeneities, the cosmological principle as stated in the previous chapter seems to be violated. Relying on a statistical description of cosmic fields actually allows to avoid this issue by reformulating it as "the universe is statistically invariant by translation and rotation". This means that it is rather the statistical properties and not the universe itself that are satisfying to it.

Thus, cosmic fields are regarded as a stochastic process from which the observed universe represents a single realisation[7]. As a result, the matter density contrast δ can be seen as a random field described by a space and time-dependent distribution function, for which one can define its one-point moments as

$$\langle \delta^n \rangle = \int \mathrm{d}\delta P_\delta(\delta)\delta^n \ , \tag{2.23}$$

where $\langle\rangle$ denotes the ensemble average and $P_\delta(\delta)$ the one-point *Probability Distribution Function* (PDF) of the cosmic field. As previously said, one would expect this PDF to be dependent on both spacial position and time, but once again thanks to the cosmological principle we can guess that the spatial dependence can be removed.

However, the one-point PDF is only partially characterising a stochastic field. In order to fully statistically describe such random field one needs to know its n-point distribution function $P[\delta(\vec{r}_1), ..., \delta(\vec{r}_n)]$, or equivalently its associated n-point moments

$$\langle \delta(\vec{r}_1)...\delta(\vec{r}_n)\rangle = \int \mathrm{d}\delta(\vec{r}_1)...\mathrm{d}\delta(\vec{r}_n) P\left[\delta(\vec{r}_1), ..., \delta(\vec{r}_n)\right] \delta(\vec{r}_1)...\delta(\vec{r}_n) \ , \tag{2.24}$$

which following the *cluster expansion* (Fry, 1984b; Bernardeau et al., 2002), can be expressed as a combination of moments of orders $n' \leq n$. For example the two and three-point moments reads

---

[7] Note that this is also in agreement with the initial Gaussian distribution of curvature fluctuations produced at the end of inflation.

$$\langle\delta(\vec{r}_1)\delta(\vec{r}_2)\rangle = \langle\delta(\vec{r}_1)\rangle\langle\delta(\vec{r}_2)\rangle + \langle\delta(\vec{r}_1)\delta(\vec{r}_2)\rangle_c \tag{2.25}$$

$$\begin{aligned}\langle\delta(\vec{r}_1)\delta(\vec{r}_2)\delta(\vec{r}_3)\rangle &= \langle\delta(\vec{r}_1)\rangle\langle\delta(\vec{r}_2)\rangle\langle\delta(\vec{r}_3)\rangle + \langle\delta(\vec{r}_1)\delta(\vec{r}_2)\rangle_c\langle\delta(\vec{r}_3)\rangle \\ &+ \langle\delta(\vec{r}_1)\delta(\vec{r}_3)\rangle_c\langle\delta(\vec{r}_2)\rangle + \langle\delta(\vec{r}_1)\rangle_c\langle\delta(\vec{r}_2)\delta(\vec{r}_3)\rangle + \langle\delta(\vec{r}_1)\delta(\vec{r}_2)\delta(\vec{r}_3)\rangle_c\end{aligned} \tag{2.26}$$

where the subscript 'c' denotes the cumulent/connected part, *i.e.* an expansion terms that cannot be split into lower orders (we assume that $\langle\delta\rangle = \langle\delta\rangle_c$).

Note however that in the case of Gaussian random fields, the *Wick's theorem* (Bernardeau et al., 2002) implies that moments of order higher than two can be expressed in terms of moments of order one and two. Thus, in such specific Gaussian case, the stochastic field can be entirely characterised by its mean and two-point moment or by its *correlation function* $\xi(\vec{r}_1, \vec{r}_2) = \langle\delta(\vec{r}_1)\delta(\vec{r}_2)\rangle - \langle\delta\rangle^2$. Note that the correlation function is built in such a way that the additional information that it brings is independent from lower order moments (*i.e.* the mean).

Since density contrasts are by definition with null expectation values ($\langle\delta\rangle = 0$, thanks to the ergodic theorem allowing to compute spatial averaging with ensemble averages), it follows that the two-point correlation function is the first moment allowing to describe a stochastic field. Noticing that for a statistically homogeneous and isotropic universe, the two-point correlation function must only depends on the distance separation between two points, *i.e.* $\xi(\vec{r}_1, \vec{r}_2) = \xi(r)$, where $r \equiv |\vec{r}_1 - \vec{r}_2|$.

However, studying correlations does not restrict ourself to configuration space. One can show that the correlation between Fourier modes can be written as

$$\begin{aligned}\left\langle\delta_{\vec{k}}\delta_{\vec{k}'}\right\rangle &= \delta^{\mathcal{D}}(\vec{k} + \vec{k}')\frac{1}{(2\pi)^3}\int \mathrm{d}^3\vec{r}\xi(r)e^{-i\vec{k}.\vec{r}} &(2.27)\\ &\equiv \delta^{\mathcal{D}}(\vec{k} + \vec{k}')P(k)\ , &(2.28)\end{aligned}$$

where $P(k)$ is defined as the Fourier transform of the two-point correlation function, the *power spectrum.* The Dirac delta $\delta^{\mathcal{D}}$ coming from the statistical translation invariance, it shows that the great advantage of Fourier space is that modes at different wavelengths are uncorrelated while this is not the case in configuration space. Note that adding statistical invariance by rotation, one expects the power spectrum to depends only on the modulus of wave modes and not on their orientation.

In general, the fact that two modes at two different wave numbers are uncorrelated does not involves that modes are independent from each other. Indeed, it does not mean that all the possible correlation configurations (3-point, 4-point, etc) are also null, thus their PDF cannot necessarily be written as a product. This would only be true in the Gaussian case for which n-point correlation functions are null above order two, thus showing that the probability distribution of two Fourier modes can be expressed as the product of the distribution of the two modes[8].

As a result, definition 2.28 can be generalised to higher order Fourier space cumulent moments

$$\left\langle\delta_{\vec{k}_1}\delta_{\vec{k}_2}\delta_{\vec{k}_3}\right\rangle_c \equiv \delta^{\mathcal{D}}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)B(\vec{k}_1, \vec{k}_2)\ , \tag{2.29}$$

$$\left\langle\delta_{\vec{k}_1}\delta_{\vec{k}_2}\delta_{\vec{k}_3}\delta_{\vec{k}_4}\right\rangle_c \equiv \delta^{\mathcal{D}}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4)T(\vec{k}_1, \vec{k}_2, \vec{k}_3)\ , \tag{2.30}$$

---

[8] This is the meaning of statistical independence

where B and T stand respectively for the *bi-spectrum* and the *tri-spectrum* of the stochastic field.

Therefore, studying two-point correlations can be done in configuration and in Fourier space. When the second offers the great advantage of uncorrelating wave modes, they actually both suffer from the fact that their sole estimation is cosmology-dependent. Indeed, constructed from a 3-D density contrast field $\delta_{\vec{k}}$ (see eq. 2.28), the estimation imposes to know the comoving position of fluctuations, an information which is not direct from galaxy redshift surveys. Thus estimating the $P(k)$ or the $\xi(r)$ involves the assumption of a *fiducial cosmology* needed for the redshift-to-comobile conversion (eq. 1.15).

This is the reason why it appears attractive to study galaxy clustering in the same space as for the CMB, *i.e.* in the spherical harmonics space. Here the two-point correlation function is called the *angular power spectrum* (usually denoted as C_ℓ) and does not impose to know the comoving separation between fluctuations, but rather their cosmology-independent angular separations (and redshift) from the observer. With the same Fourier space advantage of uncorrelating modes, the harmonic space description is also highly strategic in terms of cross-combination analysis. Being defined on a sphere (although keeping 3-D information), it makes easier the combination of galaxy clustering with other probes also defined on the sphere. In particular, combining the C_ℓ's estimated from galaxy clustering, weak lensing and CMB allows to probe in a joint analysis the whole light cone at once.

The expansion of the density field in harmonic space, as well as the description of some properties of the angular power spectrum can be found in appendix A, but will not be exploited before chapter 5. Furthermore, since both the angular power spectrum and the two-point correlation function are related to the power spectrum, the rest of the discussion will only be focused on the power spectrum and its cosmological dependence.

2.2.2 Cosmological dependence to the power spectrum

In the previous section we saw that the Fourier space density perturbation δ_k can be expressed through its initial configuration, the transfer function and the growth factor (see equation 2.22). Thus this is allowing to express its power spectrum at any epoch as

$$P(k,a) = P_0(k) T^2(k, a_m) \frac{D^2(a)}{D^2(a_m)} \ , \tag{2.31}$$

which forms the only quantity of interest that is necessary to fully characterise the matter as long as it remains in the linear regime (thus remaining a Gaussian field if the primordial curvature distribution is Gaussian).

According to the discussions in the previous section, the primordial power spectrum at the end of inflation can be written as $P_0(k) = A_s(k/k_0)^{n_s}$, where n_s is expected to be close to unity and represents a slight deviation from scale independence previously discussed. Indeed, translated in terms of gravitational potential perturbations, which are linearly linked to curvature perturbations (see relation 2.8

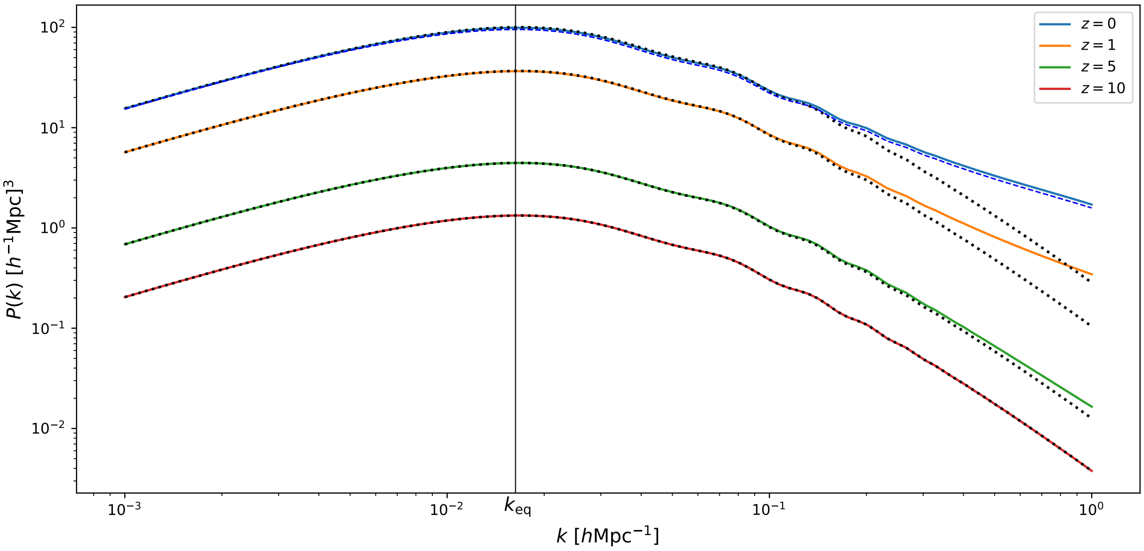


Figure 2.1: Matter power spectrum as predicted by `CLASS` in ΛCDM cosmology for four different redshifts, as a function of the Fourier wave modes. In solid lines are represented the non-linear power spectra (using the `HALOFIT` prescription of Bird et al., 2012; Smith et al., 2003; Takahashi et al., 2012) with their corresponding linear versions (dotted black lines). At $z = 0$ is plotted dashed blue line the non-linear power spectrum when adding to the ΛCDM cosmology massive neutrinos of total mass $M_\nu = 0.16$eV. The vertical line shows the scale k_{eq} of the horizon at the matter-radiation equivalence time.

by taking care of the change of notation), it implies that $k^3 P_\phi(k)$ represents the initial energy spectral distribution. If one assume that no scale is to be preferred at the end of inflation (see eq. 2.9), it means that $k^3 P_\phi(k) = \mathrm{cst}$. Thus, from the Poisson equation 2.16 in Fourier space, *i.e.* $k^2\phi_k = 4\pi G\rho_m a^2 \delta_{k,m}$ we see that $k^4 P_\phi(k) \propto P_\delta(k)$. As a result the initial matter power spectrum should be proportional to k. Moreover, observational constraints from CMB measurements have shown that the scalar index is $n_s \simeq 0.96$ and the scalar amplitude is $\ln(10^{10}A_s) \simeq 3.04$ (Aghanim et al., 2018) at the pivot scale $k_0 = 0.05\mathrm{Mpc}^{-1}$.

In order to illustrate the previous discussion, figure 2.1 displays the shape of the matter power spectrum in a linear and non-linear description, for various redshift, and accounting or not for massive neutrinos in a flat-ΛCDM cosmology. As previously said, the shape of the power spectrum deep in matter domination ($z \gg 1$) is described by two behaviours. For $k < k_{\mathrm{eq}}$, we have a power low according to $P_0 \propto k^{n_s}$, while for $k > k_{\mathrm{eq}}$, the primordial power spectrum P_0 is suppressed with respect to the transfer function $T^2(z, k)$ (see eq. 2.11). In addition the BAO takes the form of wiggles around $k \sim 0.1h/\mathrm{Mpc}$. In the linear case for varying redshift, the power spectrum is rescaled accordingly to the growth factor $D^2(z)$, while the non-linear description additionally has an effect on the amplitude of the correlations at small scales (large k), given by a scale-dependent growth function numerically predicted. Also for $k < k_{\mathrm{eq}}$, the slope of the log-log curve is given by n_s and is not affected by the level of clustering progress. Finally, as depicted in section 1.2.2, adding massive neutrinos tends to slightly damp the clustering at non-linear scales.

Finally on can relate the variance of the matter field (in the linear scales) to the power spectrum. Noticing that the variance is $\sigma^2 \equiv \xi(r=0) = \int \mathrm{d}^3\vec{k}P(\vec{k})$, it follows that without cutoff, the variance is diverging even in the linear regime (when $P(k)$ goes as k^{-2}), it is therefore necessary to introduce a smoothing function with which the matter density field must be convoluted. A common convolution function is the spherical Top-Hat window function which in Fourier space is given by

$$W_{\mathrm{TH}}(kR) = \frac{3}{(kR)^3}\left[\sin(kR) - kR\cos(kR)\right] \ , \tag{2.32}$$

where R is the radius of the sphere on which the matter density field is smooth. In particular, taking $R = 8h^{-1}$Mpc allows to quantify the global amplitude of the matter clustering. This is also allowing to see whether the clustering is linear ($\sigma_8 \ll 1$) or non-linear ($\sigma_8 \sim 1$). Such a *rms* of matter fluctuations in spheres of $8h^{-1}$Mpc can be computed from

$$\sigma_8^2 = \int \mathrm{d}^3\vec{k}P(\vec{k})W_{\mathrm{TH}}(kR_8) = 4\pi \int \mathrm{k}^2\mathrm{d}kP(k)|W_{\mathrm{TH}}(kR_8)|^2 \ . \tag{2.33}$$

In galaxy clustering analysis, it is often used as a cosmological parameter when locally measured ($\sigma_{8,0}$) and is therefore directly related to the scalar amplitude A_s. An other cosmological observable that is sensitive to $\sigma_{8,0}$ is the galaxy-galaxy lensing which allows to estimate the global amplitude of the matter clustering.

In conclusion, interpreting cosmic fields as statistically predictable random fields leads to construct robust and observable cosmological probes to test gravitational models, although it leads to new tensions (see $\sigma_{8,0}$ tensions in section 1.2.1). A great advantage of this kind of development is that common cosmological parameters can be estimated in independents probes (CMB, BAO, SN-Ia, Weak Lensing, Gravitational Waves). In particular using the matter power spectrum, it allows to constrain the inflation parameters A_s, n_s that govern respectively the amplitude of the power spectrum and its slope at $k < k_{\mathrm{eq}}$, then Ω_r, Ω_b, Ω_{cdm} by regarding the position of k_{eq}, the BAO fluctuations and the overall amplitude of the spectrum but also the neutrino mass and H_0 looking at the amplitude of the correlations after k_{eq}.

2.3 Measurements of the cosmological parameters

The main idea in the measurement of the cosmological parameters is first of all to define an observable that can be measured and compared to its theoretical expectation (computed in a given cosmological model). Then one has to vary the values of the cosmological parameters within the chosen cosmological model, in order to find out the values of the cosmological parameters that offers the best description of the observable.

Of course, to any measurement one needs to associate a confidence level, *i.e.* an error bar. As a result, it is important to understand how the uncertainties of the measured observable translate into errors on the parameters estimation. It is also of paramount importance to know how measurements are correlated with each other in order to avoid to bias the error estimation on the parameters of a model.

The particularity of cosmological analysis is that we can observe only a single realisation of our universe, while for example in particle physics it is possible to take many measurements of the same experiment. In fact, when one can redo a measurement as much as necessary, one can infer correlation and dispersion in the data. In the contrary when only as single realisation is at our disposal, we need to be able to predict how measurements of a given observable are correlated and dispersed. This specific need is also necessary when one wants to perform a combined analysis of multiple cosmological probes (observables).

In this section, I will present the standard statistical analysis procedure to estimate the cosmological parameters, using the Bayesian statistics framework. We will see in this context the importance of the *covariance matrix* to constrain the cosmological parameters, leading us to discuss how they are predicted/estimated in the specific case of the power spectrum. This will finally help me to set out the main motivations of the method developed in this document.

2.3.1 The Bayesian statistics

The Bayesian statistical analysis can be used at two different inference levels. The first allows to find out the best values of a set of cosmological parameters within a given theory that reproduces the measurements, called the *best-fit*. The second aims at discriminating cosmological models/theory that are built on different ground basis. For instance the standard model ΛCDM could be compared to modified gravity models such as the ones described in the EFT formalism (see section 1.2.2). More generally, one can also assess whether adding degrees of freedom within a model allows to significantly get a better description of the measurements.

Within the Bayesian approach, one has to keep in mind that it will always be possible to find out a best-fit, even if the underlying cosmological model is known to be unsuitable. Thus, the Bayesian approach involves taking care in the interpretation of the results.

In any case, the analysis is using the fact that the prediction of an observable within a cosmological model is depending upon the free parameters of the model. Such dependence is illustrated in the case of the power spectrum in Fourier space in figure 2.2 using `CLASS` and in harmonic space in figure 2.3 using the `AngPow` software[9] (Campagne et al., 2017; Campagne et al., 2018). In these two figures, the dependency of 9 cosmological parameters on the shape of the correlation function are exhibited. In particular, one can see that A_s and Ω_k (as well as w_0 and w_a in a dark energy context using the CPL parametrisation $w(a) = w_0 + w_a(1 - a)$ (Chevallier & Polarski, 2001; Linder, 2003)) are unequally controlling the overall amplitude of the power spectrum. One can then expect these parameters to be at some extent degenerated in a fitting analysis. When n_s only affect the slope of the correlations for $k < k_{\rm eq}$, an increasing of the h and m_ν parameters respectively improves and damps the clustering amplitude. Also, a reduction of the baryon density tends naturally to suppress the BAO shape, as it would be equivalent to put to zero Ω_r. Finally, a higher $\Omega_{\rm cdm}$ value tends to increase the total matter clustering for $k > k_{\rm eq}$. Note that these plots are only reflecting an overall behaviour and cannot be taken as a reference since it is not always possible to vary only one parameter. This is for instance the case for the energy contents, as they need to verify the Friedmann

---

[9] gitlab.in2p3.fr/campagne/AngPow

equation 1.32.

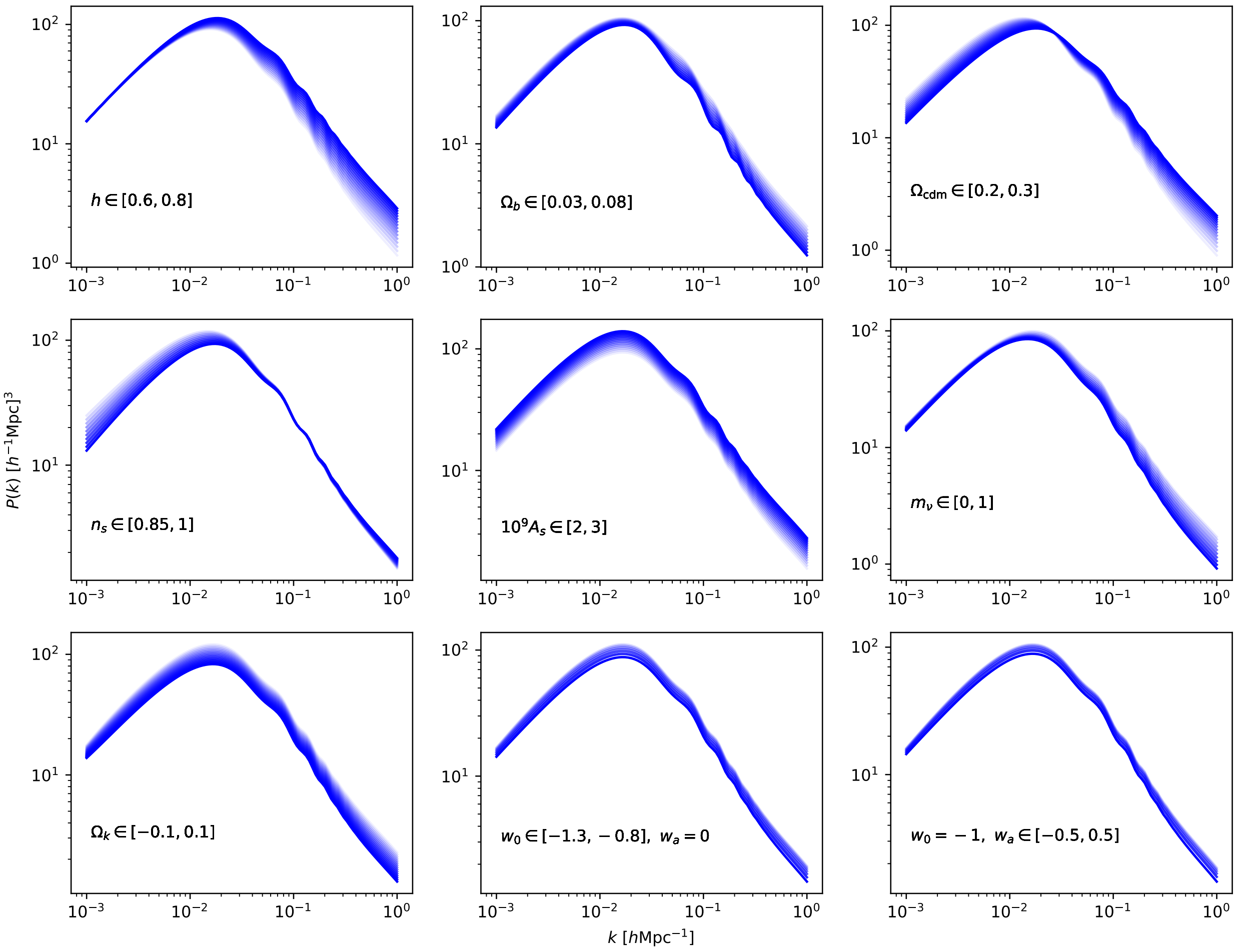


Figure 2.2: Non-linear power spectra as predicted by `CLASS` at $z = 0$ when varying the value of one parameter as referred in each panel while the others are fixed to the ΛCDM values $h = 0.67$, $\Omega_b = 0.05$, $\Omega_{\rm cdm} = 0.28$, $n_s = 0.96$, $A_s = 2.1265 \times 10^{-9}$, $m_\nu = 0$, $\Omega_k = 0$, $w_0 = -1$, $w_a = 0$. The colour become more intense when increasing the value of the parameter in its range of variation. Note that when using $w_a \neq 0$, a dynamical dark energy equation of state parameter is parametrised as $w(a) = w_0 + (1 - a)w_a$ using the CPL parametrisation (Chevallier & Polarski, 2001; Linder, 2003).

As a result, finding the values of the parameters of the model which are in agreement with observations can be seen as estimating a probability that the parameters are lying in a certain range. Thus in the following I detail this probabilistic interpretation of parameter estimation based on the measurements of an observable.

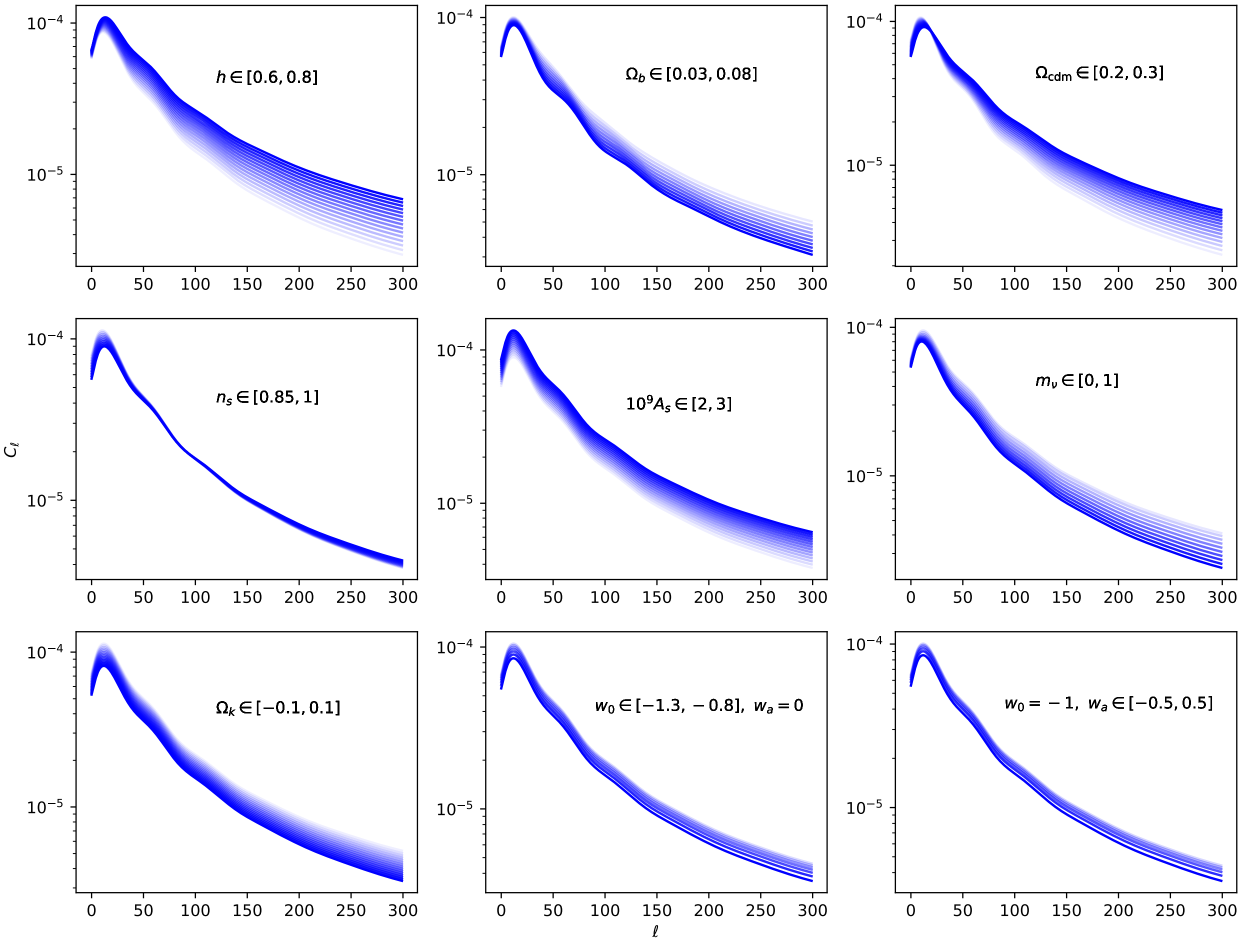


Figure 2.3: Non-linear angular power spectrum as predicted by [`CLASS` + `AngPow`] in a Top-Hat radial window function in the redshift range $z \in [0.25, 0.35]$ when varying the value of one parameter as referred in each panel while the others are fixed to the ΛCDM values $h = 0.67$, $\Omega_b = 0.05$, $\Omega_{\rm cdm} = 0.28$, $n_s = 0.96$, $A_s = 2.1265 \times 10^{-9}$, $m_\nu = 0$, $\Omega_k = 0$, $w_0 = -1$, $w_a = 0$. The colour become more intense when increasing the value of the parameter in its range of variation. Note that when using $w_a \neq 0$, a dynamical dark energy equation of state parameter is parametrised as $w(a) = w_0 + (1-a)w_a$using the CPL parametrisation (Chevallier & Polarski, 2001; Linder, 2003).

Bayes equation

The basis of Bayesian inference is built on the conditional probability relation

$$P(A)P(B|A) = P(B)P(A|B) \ , \tag{2.34}$$

where $P(X)$ is the probability of the event X while $P(X|Y)$ is the probability of the event X given the event Y. When applying this relation in our context, naming

θ the set of cosmological parameters in the framework of a cosmological model M and d the measured data (observable), we get

$$P(\theta|d,M) = \frac{P(d|\theta,M)P(\theta,M)}{P(d,M)} \ . \tag{2.35}$$

The probability $P(\theta|d,M)$ is called the *posterior* and shows the distribution of the values of the parameters given the data d within the model M. This is the probability that represents the estimation of the cosmological parameters. The Bayes equation shows that the posterior can be obtained by computing the *likelihood* $P(d|\theta,M)$, which is the probability of observing the data given the values of the set parameters θ weighted by the *prior* $P(\theta,M)$. The latter is an external knowledge that could be related to external data d'. It is the probability to find a given set of parameters when considering a model (for example setting the neutrino mass value positive, or computing the posterior in a given range of parameters, etc). In order to ensure the left hand-side to be a probability, the numerator on the right is normalised by the *evidence*, computed as

$$P(d,M) = \int \mathrm{d}\theta P(d|\theta,M)P(\theta,M) \ , \tag{2.36}$$

and which does not depends on θ.

Several methods can be designed in order to sample the posterior. Their main difference is the way they are exploring the parameter space. The brute-force consists on setting a n-dimensional grid such that each node is defined by a specific set of n parameters. This way, the output (the posterior) will consist on a spread probability distribution function over a given range of parameters set by the prior. Depending on the number of parameters to be tested and the grid precision, performing such analysis may quickly become highly CPU-demanding and poorly optimised, since probability are most of the time computed in parameter regions that are very unlikely. Especially when the number of parameters in high, other prescriptions has been developed to sample the parameter space in an optimised way. For example, the commonly used *Monte Carlo Markov Chain* (MCMC) sample the posterior distribution by performing a random (Monte Carlo) walk (Markov Chain) through the parameter space.

In the end computing the posterior through a MCMC or a regular grid sampling requires to be able to known how data (or measurements) are distributed given certain values of the cosmological model, *i.e.* the likelihood.

The Likelihood

A straightforward ansatz consists in assuming that data are distributed according to an K-variate Gaussian distribution. Be a set of K measurements $\{\mu_1, \mu_2, ..., \mu_K\}$ which can be stored into a column vector $\vec{\mu}$. The expectation value of each measurement $\vec{\mu}^{\mathrm{th}}$ must be given by theory and depends on the parameters of the model. In addition, in order to fully characterise the K-variate Gaussian, one has to provide the covariance matrix C of the measurements. It consists in a $K \times K$ matrix containing the error of each individual measurement (on the diagonal) and the covariance between them (off-diagonal elements). The covariance can be either due to instrumental limits (pixel resolution, mechanical constraints, electromagnetic spectrum estimation) or from a cosmological origin (discussed below). In practice this is the correlation matrix ρ which shows how much data are correlated, and can be obtained element by elements such that $C_{ij} = \rho_{ij}^2 \sqrt{C_{ii} C_{jj}}$. As a result the likelihood can be formally written as

$$\mathrm{B}(\vec{\mu}) \equiv (2\pi)^{-K/2} |C|^{-1/2} e^{-\frac{\chi^2}{2}} \ , \tag{2.37}$$

where

$$\chi^2 \equiv \left(\vec{\mu} - \vec{\mu}^{\mathrm{th}}\right)^T C^{-1} \left(\vec{\mu} - \vec{\mu}^{\mathrm{th}}\right) \ . \tag{2.38}$$

Since the vector μ^{th} depends on cosmological parameters θ, it is clear that the quantity χ^2 is function of the values of the parameters. In the absence of a prior distribution $P(\theta, M)$, the posterior is given by the Likelihood and it is thus also a function of the parameters. This is the posterior which is telling what is the most probable value of θ and what is its confidence interval. Typically, by integrating the posterior, one can get the constraints on a single parameter by marginalising on the others. This is a meaningful procedure only in the case where the posterior is close to an N-variate Gaussian distribution, where N is the number of parameters. However, one should be aware of the fact that marginalising over a highly non-Gaussian posterior can lead to spurious effects, thus one needs to be cautious in interpreting the marginalised posterior. A region of the parameter space could be excluded by the marginalisation, while before this marginalisation, those very-same values of the parameters could be allowed.

Let me notice furthermore that the likelihood is usually constructed as $\mathcal{L} = e^{-\chi^2/2}$, leaving out the determinant of the covariance matrix (see eq. 2.37) when assuming that it is locally constant around $\min(\chi^2)$. However a perfectly rigorous likelihood construction would re-estimate the covariance matrix at each set of tested parameters[10].

As already mentioned, sampling the posterior may also be done by explicitly imposing some priors to the analysis. Such combination generally takes the form of setting one or several fitted parameters $\theta^i = \theta^i_{\mathrm{prior}} \pm \sigma^i_{\mathrm{prior}}$ constrained by an other experiment, and contributing to the χ^2 by adding to it

$$\chi^2_{\mathrm{prior}} = \left(\frac{\theta^i - \theta^i_{\mathrm{prior}}}{\sigma^i_{\mathrm{prior}}}\right)^2 \tag{2.39}$$

if and only if the two underlying experiments are independents. Obviously, writing eq. 2.39 implies a Gaussian prior. In the context of different forms of Likelihood, it could be sometimes more appropriate to multiply Likelihoods. Furthermore as previously mentioned, setting boundaries to the parameter space to be explored (only restricting to positive mass, upper bound value, etc) is actually a Top-Hat prior but do not requires to explicitly write a χ^2_{prior}.

Finally as previously mentioned, a second level of Bayesian inference can be used to discriminate different cosmological models. Indeed, adding extra parameters to a model lead automatically to a better fit by reducing $\min(\chi^2)$. Starting from

---

[10] As we will see later, continuously re-estimating the covariance matrix for each set of tested parameters is technically impossible.

the basic idea that a good model is a model that minimises the number of free parameters (the Occam's razor), some methods has been developed to penalise the extra parameter models when running the analysis in several cosmological contexts (see Akaike Information Criterion or Bayesian Information Criterion in Akaike, 1974; Schwarz, 1978).

As shown in equation 2.38, the dependence of the χ^2 with respect to the parameters depends on the covariance matrix C, or more precisely on the precision matrix C^{-1}. Playing a fundamental role in cosmological analysis based on Bayesian inference, it is therefore of key interest to predict or evaluate the covariance matrix of a given observable in a reliable way. Note also that if the data vector combines several observable measurements, it requires to know the cross-covariance between these observables.

Given an observable O decomposed in its components o_i, its covariance matrix is formally defined as

$$C_{ij} = \langle o_i o_j \rangle - \langle o_i \rangle \langle o_j \rangle \ . \tag{2.40}$$

In general, it is not always possible for an arbitrary observable, as the correlation functions, to predict it in a fully analytical way (see discussion in section 2.3.2). That is why it is common to resort to simulations of such observables to ensure that overly simplified analytical approximations are not used. In this case, if one can produce many realisations (a number N) of simulated data, the covariance can be estimated as

$$\hat{C}_{ij} = \frac{1}{N}\sum_{s=1}^{N}[o_i^s - \mu_i][o_j^s - \mu_j] \ , \tag{2.41}$$

$$\mu_i = \frac{1}{N}\sum_{s=1}^{N} o_i^s \ . \tag{2.42}$$

However, using such estimator lead to various biases in the cosmological constraint analysis outcomes.

- First, if the covariance matrix is estimated by sampling independent realisations of the data, it can be shown that the estimator 2.41 is biased and can be corrected by switching $N \to N-1$. This is only related to the fact that the mean value of the observable is not the theoretical one, but must be estimated from data as well.

- Once an unbiaised estimator of the covariance matrix is obtained, computing the Likelihood 2.37 involves an estimation of the *precision matrix* $\hat{\Psi} = \hat{C}^{-1}$ whose the properties are described by the inverse Wishart distribution (Wishart, 1928), also biased as $\hat{\Psi} = \nu\,(\nu - K - 1)\,\hat{\Psi}^{\rm UB}$ (known as the Hartlap factor, see Hartlap et al., 2007) where $\nu = N - 1$ and $\hat{\Psi}^{\rm UB}$ standing for the unbiased estimator of the precision matrix.

- A third bias directly concerns the effect of N on the inferred parameters constraints. Indeed following Percival et al. (2014a), errors in the determination of the elements of the covariance matrix propagate to the estimated error on parameters. To partially correct for it, inferred parameters error must be corrected by the factor (taking N_p as the number of parameters to be fitted)

$$m_1 = \frac{1 + B(K - N_p)}{1 + A + B(N_p + 1)} \,, \tag{2.43}$$
$$A = \frac{2}{(N - K - 1)(N - K - 4)} \,, \quad B = \frac{N - K - 2}{(N - K - 1)(N - K - 4)} \,.$$

In conclusion, performing a reliable cosmological parameter inference involves an accurate estimation of the covariance matrix. A statistical noise due to a small N would affect both the n-dimensional ellipses volumes of standard deviations and their position in parameter space, similar to a systematic effect. In the next section, I will show that to some extent, one can predict the covariance matrix of the power spectrum. Once again, I will go through Fourier space, but it is worth mentioning that the following discussion is applicable to the C_ℓ.

2.3.2 The power spectrum covariance matrix

Already introduced earlier in this chapter, the matter power spectrum depends directly on a large set of cosmological parameters. Constraining them as depicted above, requires thus to associate an error[11], or more generally a covariance matrix, to the estimator of the power spectrum. As a result, one must first of all to define a way of estimating the power spectrum of a density field, *i.e.* an estimator.

An estimator of the power spectrum

Let me start first introducing the covariance matrix not directly of the power spectrum, but for one of its byproduct. Indeed, from a galaxy survey, we can extract a subvolume of the universe and the estimated power spectrum is, in consequence, expected to be defined for each 3-D modes $\vec{k}$, given a full grid $P(\vec{k})$. On the other hand, the cosmological principle, from which most cosmological models are based, implies a spherical symmetry that can be used in order to define a 1-D estimator of the power spectrum.

In order to define an estimator for the power spectrum, one can rely on the statistical translation invariance and assume a periodic universe, involving $\langle \delta_{\vec{k}} \delta^*_{\vec{k}} \rangle = k_F^3 P(\vec{k})$ ($k_F = 2\pi/L$ is the fundamental mode in a periodic box of size L), obtained from equation 2.28 in the context of a periodic universe and a real density field. Thus, once a Fourier transform has been applied to compute the density field δ_k, we can estimate the 3-D power spectrum as $\hat{P}(\vec{k}) = |\delta_{\vec{k}}|^2/k_F^3$. This ensures that the expectation value of the estimator converges to the true value of the power spectrum. However, when we add the statistical invariance by rotation (statistical isotropy), one can think of reducing the noise by making a shell average in Fourier space at constant k.

---

[11] Here we are talking about the inherent error/covariance of the cosmological model to be tested. It is not related to any instrumental error.

Since periodic signals have power only at multiples of the fundamental frequency k_F, it is conventionally chosen to define shells in Fourier space of width k_F and centred on modes $|\vec{k}_n| = nk_F$ with $n \in \mathbb{N}^*$. This leads to a discreet sum over the 4π solid angle of each shell of modes that are not independent. Indeed, as $\delta(\vec{x}) \in \mathbb{R}$, it implies for each mode that $\delta_{-\vec{k}} = \delta^*_{\vec{k}}$ and in consequence that $\hat{P}(-\vec{k}) = \hat{P}(\vec{k})$, meaning that not all wave modes provide an independent information. As a result, one can write down the estimator of the shell-average power spectrum $\hat{P}(k)$ as

$$\hat{P}(k_n) = \frac{k_F^3}{M_{k_n}} \sum |\delta_{\vec{k}_n}|^2 \ , \tag{2.44}$$

where the sum is made over the M_{k_n} independent modes within the shell centred around k_n. Note also that the wave mode k is usually obtained by averaging the modes within the shell in the same way, this allows to reduce the effect of the grid (in Fourier space) when the density field is actually not periodic.

Power spectrum covariance matrix

From the definition of the covariance matrix 2.40, one can write the covariance matrix of the power spectrum estimator defined by equation 2.44 as

$$C_{ij} \equiv \left\langle \hat{P}(k_i)\hat{P}(k_j) \right\rangle - \left\langle \hat{P}(k_i) \right\rangle \left\langle \hat{P}(k_j) \right\rangle \ . \tag{2.45}$$

Equation 2.45 can be combined to equation 2.44 to explicitly express the power spectrum covariance (Scoccimarro et al., 1999) as

$$C_{ij} = \frac{\hat{P}(k_i)^2}{M_{k_i}} \delta^K_{ij} + k_F^3 \bar{T}(k_i, k_j) \ . \tag{2.46}$$

The above equation shows that power spectrum covariance matrix can be split into two contributions. The first term is the so-called *Gaussian contribution* that only depends on the amplitude of the power spectrum, weighted by the number of independent available modes per shell. The more Fourier modes fall into shells, the smaller the error will be. In the case of a Gaussian field, it is the only one to contribute and it only affects the diagonal of the matrix (δ^K_{ij} being the Kronecker delta): each Fourier mode being an independent Gaussian random variable, the different i, j shells are consequently uncorrelated. The associated Gaussian variance (located on the diagonal of the covariance matrix) rapidly decreases in the case of a non-linear power spectrum from the mildly linear regime ($k \sim 3 \times 10^{-2}$hMpc^{-1}) where it scales[12] as k^{-4}, affecting to a large extent low wave modes. Thus, in the case of the CMB or galaxy power spectrum analysis in the linear regime ($k \lesssim 10^{-1}$h/Mpc), the Gaussian limit of equation 2.46 appears to be an accurate approximation (Tegmark, 1997a; Bond et al., 1998). However, when trying to maximise the information that can be extracted from a survey, it appears interesting to get into the mildly non-linear regime, in which case it becomes important to take the second term into account.

The second contribution involves the 4-point correlation function in Fourier space, namely the tri-spectrum of the fluctuation field defined in relation 2.30. In

---

[12] In this regime the non-linear matter power spectrum decreases as k^{-1}.

the same way as for equation 2.44, $\bar{T}(k_i, k_j)$ is the cross shell-average (or the averaging in two shells centred in k_i, k_j) version of the tri-spectrum which can be written in the continuous limit as

$$\bar{T}(k_i, k_j) = \int_{k_i} \int_{k_j} \frac{\mathrm{d}^3\vec{k}_1}{V_s(|\vec{k}_i|)} \frac{\mathrm{d}^3\vec{k}_2}{V_s(|\vec{k}_j|)} T(\vec{k}_1, -\vec{k}_1, \vec{k}_2, -\vec{k}_2) \ , \tag{2.47}$$

where V_s is the shell volume in Fourier space. Involving a cumulent greater than two (here of order four) that vanishes in the case of a normal field (due to the Wick's theorem), it is naturally referred as to the *non-Gaussian term* that contributes both to the diagonal and off-diagonal elements of the covariance matrix. Its scaling in k, respectively to the Gaussian term, is more difficult to predict without expanding it using perturbation theory or without resorting to numerical simulations. In Scoccimarro et al. (1999), they report that it starts to dominate around $k \sim 0.2h\mathrm{Mpc}^{-1}$ in the ΛCDM context and at $z = 0$. The relative importance between the Gaussian and non-Gaussian contribution will be discussed in details in chapter 4.

Nevertheless, their relative behaviour as a function of the volumes $V = L^3$ and V_s can be discussed. Indeed, factorising equation 2.46 as

$$C_{ij} = k_F^3 \left[\frac{2\hat{P}(k_i)^2}{V_s(|\vec{k}_i|)} \delta_{ij}^K + \bar{T}(k_i, k_j) \right] \ , \tag{2.48}$$

reveals that both terms are proportional to k_F^3, or inversely proportional to the volume of the survey. Also, while the first term is inversely proportional to the shell volumes in Fourier space V_s, the second is blind to it (V_s only acting in the averaging, see equation 2.47). As a result, minimising the covariance matrix in linear regime can be done by increasing the survey volume V or the size of the shells (lowering the binning). In the non-linear regime, when the tri-spectrum is dominating, we reach a barrier that depends on the clustering properties of the density field.

2.4 Numerical estimation of covariances

We saw that the covariance matrix is composed of two components. The Gaussian contribution can be computed as soon as the power spectrum is known, while the non-Gaussian contribution $\bar{T}$ requires some knowledge about the tri-spectrum. Unfortunately, since the tri-spectrum of the matter field arises from non-linear couplings between long and short wave modes during the gravitational evolution of perturbations, it turns out to be hard to compute using perturbation theory.

The prediction of the tri-spectrum is complicated as it includes, the survey window function, the bias between matter and galaxies[13] and the shot noise (described in more detail in chapter 4) that modify the structure of the covariance matrix. But above all, it should include the effect of peculiar velocities of galaxies (commonly called Redshift-Space Distortions, also introduced in chapter 4) that leads to highly non trivial contributions to the shell average tri-spectrum $\bar{T}$.

Because of these complications, over the last two decades, a formidable effort has been made to develop numerical simulations of the large scale structure of the uni-

---

[13] Galaxy surveys is only able to probe galaxies, a biased tracer to the underlying matter field.

verse and investigate the shape of the covariance matrix, in the non-linear regime where perturbation theory breaks down. estimating it with several output from numerical simulations. This alternative makes possible to faithfully represent the impact of structure formation on the covariance matrix through its unbiased estimator over a number N of simulations

$$\hat{C}_{ij} = \frac{1}{N-1}\sum_{s=1}^{N}[P^s(k_i) - \mu_i][P^s(k_j) - \mu_j] \ , \tag{2.49}$$
$$\mu_i = \frac{1}{N}\sum_{s=1}^{N} P^s(k_i) \ ,$$

where $P^s(k_i)$ is the power spectrum in the s-th realisation (or simulation) evaluated at mode k_i.

N-body simulations are run in spatially flat cosmologies using Newtonian approximation in order to evolve an initial field of particles (defined by their position, velocity and mass). The initial conditions are prescribed by the Zel'dovich approximation (Zel'Dovich, 1970) usually around $z \sim 100$ (deep in the matter domination). In a Lagrangian fluid description, particles are submitted to a displacement field by estimating the gravitational influence of particles surrounding it, and re-evaluated for each time increments. It imposes their successive local displacement up to their final state. Note that for each evolution step, the whole comoving volume is defined by the same cosmic time, calling each of these pictures a *snapshot*. A single catalogue accounting for varying redshift (a *light cone* where the observer is placed at the centre), can then be reconstructed using the procedure of Fosalba et al. (2008).

Since the number of time iterations is important, N-body simulations are computationally expensive. The idea is to find the right balance between setting a large volume to increase the number of available modes on large scales and adopting an acceptable spacial and mass resolution to minimise shot noise and reach non-linear scales.

Investigating the covariance matrix using N-body was initiated in Scoccimarro et al. (1999) and Meiksin & White (1999). When realising that for successive surveys, the observational limit was pushed toward smaller scales, the numerical option seemed like the most effective one to estimate wave mode correlations (*e.g.* Hamilton et al., 2006; Takahashi et al., 2009; Li et al., 2014). The recent development and democratisation of super computers allowing now to produce massive data as the DEUS-PUR (Blot et al., 2015) simulations constituted of 12288 catalogues of 256^3 dark matter particles in a volume of $(656h^{-1}\text{Mpc})^3$ in a flat ΛCDM cosmology or in the QUIJOTE set of 43100 simulations (Villaescusa-Navarro et al., 2020) spanning large volume boxes of $(1000h^{-1}\text{Mpc})^3$ with multiple cosmological models (ΛCDM, massive neutrinos, varying dark energy equation of state parameter) and resolutions (256^3 , 512^3 and 1024^3 dark matter particles).

2.4.1 The N-Body approach : `DEMNUni_cov`

It is therefore opportune to present in more detail a particular numerical experiment on which some of the results of my thesis rely on. Originally developed for testing different probes in the presence of massive neutrinos, the Dark Energy and Massive Neutrino Universe (DEMNUni) experiment is a suit of N-Body simulations (Carbone et al., 2016; Castorina et al., 2015) exploiting the FERMI super computer at CINECA[14].

The code GADGET-3 of Springel (2005) and further developed by Viel et al. (2010) for the implementation of massive neutrinos, has been run for several cosmological scenarios. Among them the ΛCDM model, cosmologies with several massive neutrinos values, and cosmologies with time varying dark energy equation of state with the CPL parametrisation $w(a) = w_0 + w_a(1-a)$ (Chevallier & Polarski, 2001; Linder, 2003). All the DEMNUni suit of simulations have in common the following cosmological parameters:

$$\begin{aligned} \Omega_k &= 0 \ , \\ \Omega_m &= 0.32 \ , \\ \Omega_b &= 0.05 \ , \\ h &= 0.67 \ , \\ n_s &= 0.96 \ , \\ A_s &= 2.1265 \times 10^{-9} \ . \end{aligned}$$

The initial conditions have been set at redshift $z = 99$ using the Zel'dovich approximation (Zel'Dovich, 1970) and the public code described in Zennaro et al. (2017), Zennaro et al. (2018) and Zennaro et al. (2019).

The evolution of cosmic structures is simulated up to redshift $z = 0$, while saving 62 snapshots with equally spaced logarithmic scale factor intervals. Among these snapshots, we restrict our analysis to the five following redshifts[15].

$$z = \{0,\ 0.48551,\ 1.05352,\ 1.45825,\ 2.05053\}.$$

Regarding our study oriented toward the estimation of covariance matrices, only a DEMNUni subset, hereafter called `DEMNUni_cov`, will constitute our reference basis for the estimation of the covariance matrices. This subset is made of two ΛCDM cosmologies, one without massive neutrinos and the other one with massive neutrinos (added at the level of massive particles inside the simulation) of total mass $M_\nu = 0.16$ eV. The main interest of these two suits of simulations is that they have been ran 50 times each with different realisations of the initial conditions. This way they offer the possibility of estimating to some extent covariance matrices.

Each simulation consists in 1024^3 dark matter particles of mass $m_p \simeq 8 \times 10^{10} h^{-1} M_\odot$ in a large cubical volume of size $L = 1000 h^{-1}$Mpc with periodic boundary conditions. For all realisations we assign all particles to a regularly spaced grid consisting in 1024 grid points within the comoving volume. We choose the mass assignment scheme Peace-wise Continuous Spline allowing to minimise the aliasing effect (hereafter described in chapter 3) in Fourier space (Sefusatti et al., 2016). We can finally deduce the dark matter density field which allows to estimate the power spectrum in each realisation at five redshift comoving outputs.

---

[14] cineca.it

[15] for simplification in the following, these five redshift will be referred to the round numbers $z = \{0, 0.5, 1, 1.5, 2\}$.

The estimated averaged power spectrum over 50 realisations together with the evaluated dispersion are represented in figure 2.4.

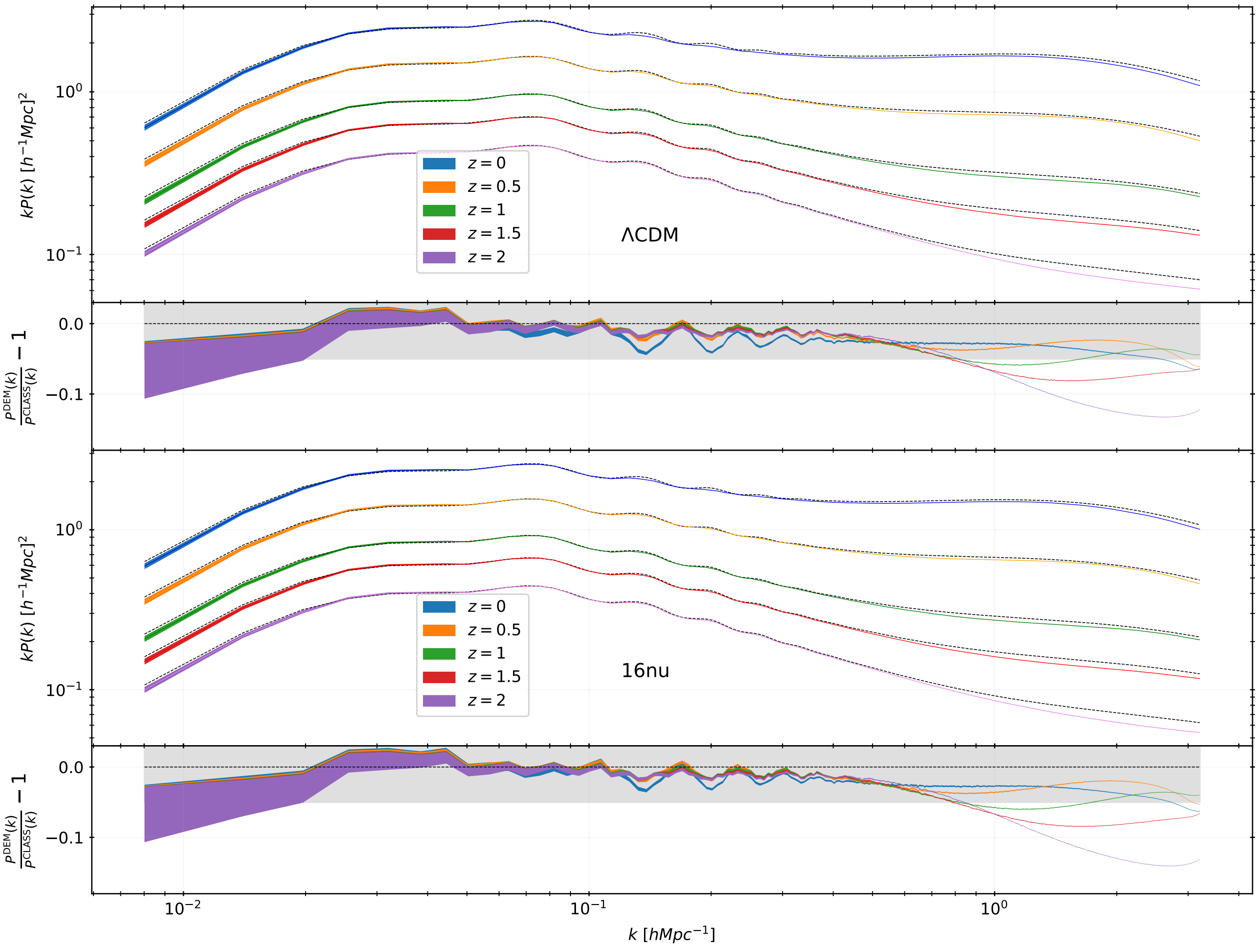


Figure 2.4: Estimated averaged power spectra at redshifts $z = \{0, 0.5, 1, 1.5, 2\}$ compared to `HALOFIT` non-linear prescription and their relative deviation for the two cosmologies with (lower panel) and without (upper panel) massive neutrinos.

We can see in this figure the BAO from $k \sim 0.06$ to $0.3\ h\mathrm{Mpc}^{-1}$ which are actually not well reproduced by the `HALOFIT` non-linear prescription. As expected, as redshift goes to zero the agreement gets worth.

The most striking feature is the clear missing power at the largest scales accessible with the simulations $k \sim 0.01\ h\mathrm{Mpc}^{-1}$ which reveals a 5% lake of power with respect to linear theory. We also see that this is independent from the cosmology and from redshift, indeed we verified that this was already present in the initial conditions of the simulations.

In the non-linear regime, low redshifts are in better agreement with theory than at high redshifts[16]. But overall, in the interesting range probed by redshift surveys ($k < 0.4$) the agreement is at the level of 5% and this is independent from the chosen neutrino mass.

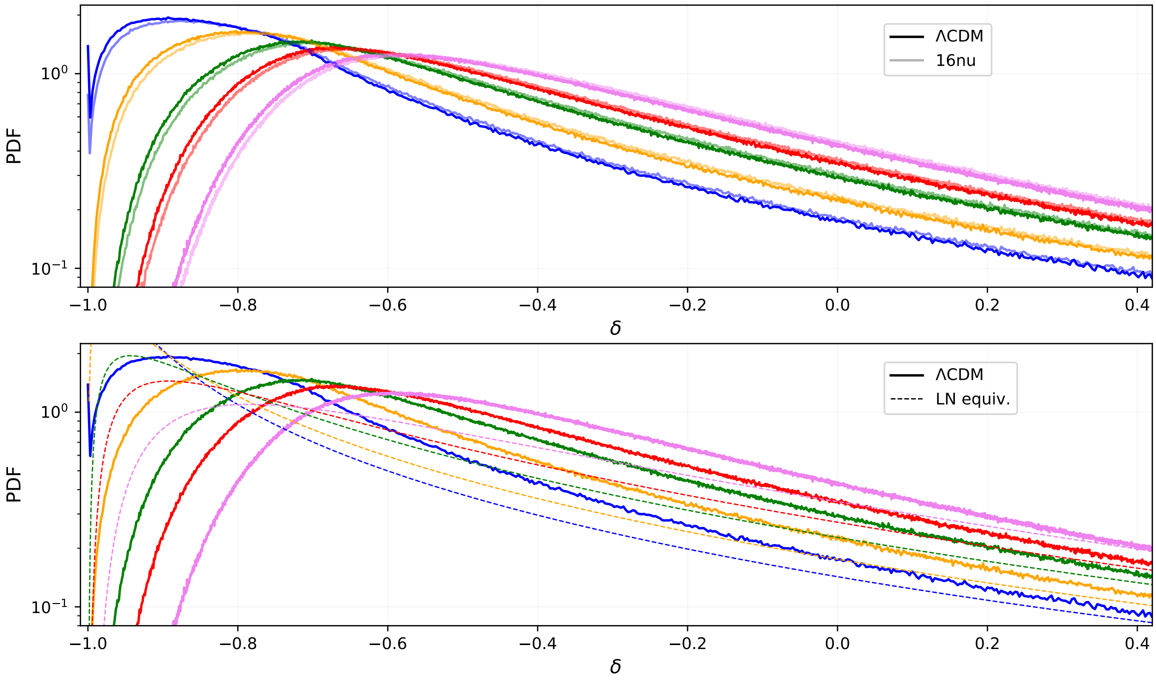


Figure 2.5: Estimated PDF of the dark matter density contrast δ in the first `DEMNUni_cov` simulation. The five redshifts $z = \{0, 0.5, 1, 1.5, 2\}$ are respectively plotted in blue, orange, green, red and violet. *top:* High and low colour density associates respectively the ΛCDM and $M_\nu = 0.16$ eV cosmologies. *bottom:* Only the ΛCDM is represented with solid lines together with the corresponding Log Normal PDF (dashed lines) with the same variance.

We can also estimate the probability density distribution (PDF) of the density contrast $\delta(\vec{x})$ (computed on the same configuration space grid as the power spectrum). This is shown in the upper panel of figure 2.5 where the neat effect of massive neutrinos is to reduce the probability occurrence of extremely empty regions ($\delta < -0.9$). In fact they tend to reduce the asymmetry of the PDF because they slow down the growth of structure, thus the non-linear evolution of the dark matter clustering (see section 1.2.2).

Finally, despite the fact that the density field is smoothed on a scale of about $1h^{-1}$Mpc (due to the mass assignment scheme), approximating its probability density with a Log-Normal provides a very poor agreement with the estimated PDF, even at redshift $z = 2$ (lower panel of figure 2.5). Anticipating the rest of this section, this last remark highly motivates the implementation of Monte Carlo realisations

---

[16] In general, analysis on galaxy surveys in redshift-space are performed up to $k_{\max} \sim 0.2h/\text{Mpc}$ (de Mattia et al., 2020).

without restricting the choice of the PDF to a Log-Normal.

2.4.2 The approximated approach

Nevertheless, N-Body simulations are highly CPU-consuming and several months are needed on super computers to produce a high sample of catalogues ($\gtrsim 1000$ simulations) for a given cosmological model with comparable statistics ($\sim 10^8$ particles) and effective volume ($\sim [1000\mathrm{h}^{-1}\mathrm{Mpc}]^3$) as for future galaxy surveys. Especially since, depending on the precision expected in the estimation of the cosmological parameters, the covariance must be estimated using several thousand simulations[17].

Indeed, as shown in section 2.3.1, noise in the estimated covariance matrix propagates up to the final cosmological parameter constraints (Taylor et al., 2013; Dodelson & Schneider, 2013; Percival et al., 2014b; Sellentin & Heavens, 2015), and only the number of realisations N can reduce it (except if wave modes are grouped in bins). Even if some method has been developed to reduce the number of N-body simulations (Hamilton et al., 2006), the number needed to accurately estimate the covariance matrix remains significantly high and alternative faster methods have been developed.

These faster methods, which do not always offer the same level of precision, can be divided into two main categories : the *internal* and *external* error survey algorithms. The first have been developed to avoid the simulation of several catalogues[18] by estimating the internal error of the survey. Among the most popular ones are the *jackknife* and the *bootstrap* methods, both being based on a re-sampling of the initial distribution.

The jackknife method consists in a sub-sampling technique without replacement. From a catalogue defined by a volume V and a number of particles N, a sub-volume V_{sub} corresponding to N_{sub} particles is removed from it. Estimation of the observable can thus be performed on $N - N_{\mathrm{sub}}$ particles and this process is then repeated N_{JN} times. It is clear that this method is not suited to the Fourier space since the Fourier transform of a sub-volume involves complex convolutions with the underling mask.

The bootstrap method is more adapted to Fourier space because the re-sampling does not concern the volume but the objects in the catalogue. On N_{sub} particles randomly selected (with possible repetition) is estimated the observable ; a process repeated N_{BT} times.

Finally, for both method, the covariance matrix estimated for each re-sampling cannot include the large scale cosmic variance, since they are coming from a single realisation of the universe. But even covariance at small scales is not entirely captured in this process. The sub-sample being not rigorously independents, the estimated covariance is biased (Norberg et al., 2009a; Friedrich et al., 2016; Lacasa & Kunz, 2017) and the covariance between short and long wave modes cannot be traced. A review of the two methods can be found in Norberg et al. (2009b).

In order to reduce the computational cost and keep a high level of accuracy in the covariance matrix estimation, it seems attractive to resort to *approximate methods* that aim at reproducing some specific properties of the survey and fall in

---

[17] For example they used typically $> 10^3$ realisations for the BOSS analysis (Eisenstein et al., 2011).

[18] This is mainly due to a lack of CPU and memory resources in the previous two decades.

the category of external error survey algorithms. They use either approximations in the dynamics or in the statistics of the density fluctuations. These methods are also called semi-analytic to differentiate them from the fully numerical methods such as N-body simulations. Indeed, they use analytic transformations applied on random fields whose effect and outcome are controlled and anticipated. They can be divided into three categories (Lippich et al., 2019):

1. The *predictive* methods are very close to N-body simulations in the sense that they are also evolving an initial particle field, but by customising this evolution using linear perturbation theory. As N-body, no randomness are used (apart from initial conditions), making the field evolution fully deterministic once an initial random realisation has been generated. For example ICE-COLA (Tassev et al., 2013; Izard et al., 2018) approximates the dynamics of the field speeding up each N-body time step increments, PEAK PATCH (Bond & Myers, 1996) and PINOCCHIO (Monaco et al., 2002, 2013) are identifying haloes in their birth stage ($\sim$ initial conditions) and evolve them using Lagrangian Perturbation Theory up to their final state. They are slightly less demanding in terms of CPU and memory resources than N-body codes.

2. The *bias-based* methods are much faster. A matter field is generated at a given redshift using Lagrangian perturbation theory, then a biasing scheme (a calibration on N-body simulations) is applied to obtain haloes. PATCHY (Kitaura et al., 2013, 2015) and HALOGEN (Avila et al., 2015) can be quoted in this category.

3. The *PDF-assumption* methods focus on the implication of the shape of the PDF on the covariance matrix structure. In these fast methods, catalogues of galaxies or haloes are simulated at a given redshift by imposing a shape for the fluctuation field PDF and for the power spectrum. The assumption is that the covariance matrix depends on these two statistical quantities. For instance the Gaussian recipes of Grieb et al. (2017) can only be used in the fully linear regime while the Log-Normal catalogues of Agrawal et al. (2017) can be used up to the mildly non-linear scales. Their great advantage is that they are totally independent from N-body realisations and do not require, in absolute terms for the quoted methods, any calibration on N-Body whatsoever.

The previously quoted algorithms has been compared at the level of the produced covariance matrices of the correlation function in configuration space (Lippich et al., 2019), the power spectrum (Blot et al., 2019) and the bi-spectrum (Colavincenzo et al., 2019) in Fourier space to the one estimated on the MINERVA set of 300 N-body simulations (Grieb et al., 2016). Concerning the power spectrum, the variance (diagonal of the covariance matrix) is well reproduced up to $k = 0.2h/\mathrm{Mpc}$ at 10% level for the two first classes of approximated methods, except for the HALOGEN one that can reach 50% deviations depending on the setting. For the Log-Normal prescription on the other hand, it totally fails to reproduce the variance better than 10% beyond $k = 0.025h/\mathrm{Mpc}$, where the non-Gaussian contributions starts to contribute. A discrepancy justified by the authors of the study, arguing that the Log-Normal approximation was poorly adapted to the PDF of the N-body simulations. That is the reason why we need robust N-body simulations to be used as reference.

2.4.3 Motivations for a new Monte Carlo approach

Following the previous observations, we can draw an overview of the power spectrum covariance matrix prediction/estimation situation. A fully analytic approach was almost impossible to foresee, however in the recent work done by Wadekar & Scoccimarro (2019), they manage to predict in a perturbative way, the power spectrum covariance matrix up to $k \sim 0.7h/\mathrm{Mpc}$ in redshift-space, without sampling noise and for any arbitrary window survey. This remarkable approach that has been validated in Wadekar et al. (2020) does not, however, break the dynamics of the improvement of alternative methods for the estimation of the covariance matrix. There are still many aspects of survey statistics that need to be studied, in particular the wave mode distribution of the power spectrum that may have an impact on the choice of the Likelihood distribution assumption, the covariance matrix of two-point statistics in configuration and the angular power spectrum in harmonics spaces, combinations of cosmological probes of the large scale structures such as the galaxy lensing and the galaxy clustering. It is therefore relevant to investigate methods that are, as much as possible, theoretically independent from N-Body outputs and as quick as bias-based methods.

An interesting possibility stands in the third class of approximated methods, the PDF-assumption algorithms that are in fact Monte Carlo realisations of a density field. They are based on the local transformation of a random field aiming to target a given power spectrum and PDF, that could finally be discretised in particles. The adaptability of such procedure lies in the fact that the nature of the simulated catalogues can be adapted; it can be either dark matter particles, neutrinos, galaxies, haloes, etc. In fact it only depends on which statistical targets are provided as an input.

In addition, switching from a cosmological model to an other can be done as long as a power spectrum and a density PDF can be predicted (or estimated from N-Body simulations). Introducing some observational features such as masks or radial survey window should only improve the time-execution performance by reducing the field domain treatment and without having, a priori, to predict their effects.

Similarly, dealing with complex non-linear effect of peculiar velocities in perturbation theory is reduced here to a phenomenological study of velocity reconstruction models. However, this class of algorithm is underdeveloped since only two kinds of PDF shape have been studied so far in literature. Indeed we saw in section 2.3.2 that the the power spectrum covariance matrix depends on the tri-spectrum of the density field, so the idea is to be able to reproduce at least its overall amplitude.

Given that the 4-points correlation function in configuration space is related to the trispectrum via an inverse Fourier transform

$$\langle\delta_1\delta_2\delta_3\delta_4\rangle_c = \int \mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{k}_3 T(\vec{k}_1,\vec{k}_2,\vec{k}_3)e^{-i(\vec{k}_1.(\vec{x}_1-\vec{x}_4)+\vec{k}_2.(\vec{x}_2-\vec{x}_4)+\vec{k}_3.(\vec{x}_3-\vec{x}_4))} , \quad (2.50)$$

We can deduce the fourth order cumulent moment of the density contrast probability density ($\vec{x}_1 = \vec{x}_2 = \vec{x}_3 = \vec{x}_4$)

$$\left\langle\delta^4(\vec{x})\right\rangle_c = \int \mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{k}_3\ T(\vec{k}_1,\vec{k}_2,\vec{k}_3) . \quad (2.51)$$

Equation 2.51 shows that if the 4-th order cumulent moment is constrained, then the integral of the tri-spectrum over the full Fourier space has the correct amplitude. Since targeting a given one-point PDF allows to match the n-th order cumulent moments $\langle \delta^n(\vec{x}) \rangle_c$, in particular the fourth order 2.51 is automatically matched.

Thus one can think that the shape of the covariance matrix is partially set by the power spectrum and the PDF of the density contrast. Based on these arguments, the essence of this thesis is to develop a method that allows to generate any catalogue of particles that follows some arbitrary power spectrum $P(k)$ but also any arbitrary PDF, a method not yet presented in literature. Once achieved, it will be embedded in a Monte Carlo process that will sample the covariance matrix of any observable.

To conclude this chapter, we discussed the advantages of probing the galaxy clustering in a statistical approach. Much better suited to an observational comparison, theoretical predictions reveal some difficulties to handle non-linearities. Propagating up to the covariance matrix, several methods has been proposed in literature to estimate such matrices. Either too CPU-demanding or non adapted to the regime of interest, I propose to investigate an original method of covariance matrix estimation from a Monte Carlo point of view. The next chapter is therefore oriented to the presentation of a process to generate such non-Gaussian fields.

Chapter 3

Investigating Monte Carlo realisations

The previous chapter brought the advantages of the Monte Carlo process to estimate the power spectrum covariance matrix[1] from several catalogue simulations. In addition to its speed of execution, the benefit of this choice is that it can offer an accurate control on the power spectrum and the probability distribution function (PDF) of density fluctuations thanks to deterministic transformations. By controlling these two quantities, as already discussed, one can expect to generate Monte Carlo realisations which allows the production of reliable covariance matrices for various cosmological observables, such as the power spectrum, the angular power spectrum or the two-point correlation function.

In this chapter, I will first introduce in section 3.1 the basic setting which seems most suitable to generate Monte Carlo realisations of density fluctuations. This will lead me to discuss the relevance of a mathematical tool to carry out such a project, called the Mehler expansion. It will be used to draw the theoretical pipeline allowing to generate a (non-Gaussian) field with some targeted, and arbitrary, PDF and power spectrum.

Finally after addressing the numerical adaptation of the pipeline and the notion of aliasing, I will go through a first validation of the method in the context of Log-Normal PDF. The interest of this choice is that it makes all the calculations entirely analytical, making easier to detect any eventual bias. These numerical aspects will be presented in section 3.2.

3.1 Generating a non-Gaussian field: theoretical pipeline

3.1.1 Basic outline

First and foremost, it is worth emphasising that, both the power spectrum and the density PDF are depending on the cosmological model. As a consequence, one has to be as general as possible and consider any shape for these two quantities in the construction of the Monte Carlo pipeline.

However, as already mentioned before, the difficulty arises from the fact that we want to simulate a non-Gaussian density field. The expression of the covariance matrix eq. 2.46 is showing that scales, or equivalently wave modes in Fourier space

---

[1] In this chapter, each mention to covariance matrix will refer to the one of the power spectrum. I will go back to the angular power spectrum in chapter 5.

for the power spectrum estimator, are correlated through the tri-spectrum. As a result, the field simulation cannot directly result from an independent and local random sampling following a given PDF. Note that even if we knew how to directly generate a correlated field, these correlations are precisely our unknowns that we need to estimate.

On the other hand, this same expression (eq. 2.46) shows that the covariance matrix of a Gaussian set of variables in Fourier space, is diagonal, *i.e.* modes are uncorrelated. In configuration space instead, the inverse Fourier transform consists in a mixing modes that turns scales correlated, even for a Gaussian observable.

Thus, in the Gaussian case, it appears more straightforward to generate a set of measurements or a field in Fourier space than in configuration space. Each mode are independently sampled from a Gaussian distribution and with a variance given by the power spectrum. This rather standard method is presented in appendix B and actually constitutes the basis for the general method that I am going to introduce.

At this point we are able to generate a field in configuration space (applying an inverse Fourier transform) which follows a Gaussian distribution and characterised by an arbitrary power spectrum. In order to generate a non-Gaussian field, the basic idea is to locally apply on this field a non-linear transformation which changes the Gaussian PDF into a non-Gaussian, targeted one. However, as it will be quantified in the following, this process naturally induces a non trivial transformation on the resulting power spectrum.

This issue can be circumvented by choosing a well designed power spectrum for the Gaussian field in such a way that once the non-linear transformation has been applied, we finally obtain the target power spectrum of the non-Gaussian density field. This simple reasoning constitutes the ground base of the proposed Monte Carlo method and the current section is devoted in presenting its technical aspects.

3.1.2 Controlling the one- and two-point statistics

Targeting a PDF model

Producing a density field following a non-Gaussian PDF is fairly simple. A non-linear local transformation $\mathcal{L}$ applied on a Gaussian random field sampled in configuration space automatically modifies its distribution. However, this transformation can be specifically customised to generate a field following a given PDF model, a process widely used in cosmology (see Coles & Barrow, 1987). Since this procedure is the milestone of the proposed Monte Carlo method, let me review it.

Be ν an initial stochastic field following a centred and reduced (standardised) Gaussian distribution $P_\nu(\nu)$ such that $\langle\nu\rangle = 0$ and $\sigma_\nu^2 \equiv \langle\nu^2\rangle_c = 1$ (the index c refers to cumulent moments). Also, be δ the $\mathcal{L}$-mapped non-Gaussian field, playing the role of the density contrast field

$$\delta(\vec{x}) \equiv \mathcal{L}\left[\nu(\vec{x})\right] \ . \tag{3.1}$$

When targeting a specific PDF denoted $P_\delta(\delta)$, $\mathcal{L}$ can be easily found by applying standard probability transformation rules. Indeed, if we consider a monotonic transformation $\mathcal{L}\, :\ ]-\infty,+\infty[\ \rightarrow [-1,+\infty[$, we can ensure the conservation of probabilities

$$P_\delta(\delta)\mathrm{d}\delta = P_\nu(\nu)\mathrm{d}\nu \ . \tag{3.2}$$

The above equation shows that any linear transformation $\mathcal{L}$, *i.e.* $\mathrm{d}\delta/\mathrm{d}\nu = \mathrm{cte}$, would have no impact on the nature of the distribution, apart from re-scaling it.

In order to find the transformation which turns the Gaussian PDF P_ν into the targeted P_δ, one can integrate each side of eq. 3.2 from the respective lower bound domain of definition

$$C_\delta[\delta] = C_\nu[\nu] \tag{3.3}$$

where we define the cumulative distributions as

$$C_X[X] \equiv \int_A^X \mathrm{d}X' \ P_X(X') \ , \tag{3.4}$$

with A standing for the lower bound domain of the variable X, featured by the distribution $P_X(X)$. The positive definite PDFs implies that the cumulative distributions are monotonic. In consequence, eq. 3.3 can always be inverted in terms of reciprocal functions (*i.e.* $F^{-1}[F(x)] = x$). In particular

$$\delta = C_\delta^{-1}[C_\nu(\nu)] \ , \tag{3.5}$$

which can be identified to definition 3.1 in order to extract the field mapping $\mathcal{L}$.

Setting aside cases for which the PDF $P_\delta(\delta)$ is defined by an analytical local transformation, as for instance the Log-Normal one that will be reviewed in the following, the complexity of realistic forms of PDF would make the previous inversion analytically unfeasible. As an example, Klypin et al. (2018) proposed various parametrisations of the PDF or Uhlemann et al. (2016) and Codis et al. (2016) who derived it directly from large deviation theory and spherical infall models. In addition to this, it would be convenient to reproduce a distribution that follows what is directly measured in galaxy survey (or N-body simulations). For these reasons in a realistic case, the inversion should be done numerically.

Targeting a $P(k)$ model

As anticipated in the introduction of the section, transforming a Gaussian field with a local non-linear mapping $\mathcal{L}$ induces non trivial modifications of the input power spectrum of the Gaussian field. In the following I quantify this impact and explain how to take this into account in order to recover the correct target power spectrum after applying the local transform.

We can generalise eq. 3.2 to relate the two-point distribution of the density field $P_\delta(\delta_1, \delta_2)$ to the bi-variate Gaussian distribution $P_\nu(\nu_1, \nu_2)$ as

$$P_\delta(\delta_1, \delta_2, \xi_\delta)\mathrm{d}\delta_1\mathrm{d}\delta_2 = P_\nu(\nu_1, \nu_2, \xi_\nu)\mathrm{d}\nu_1\mathrm{d}\nu_2 \ , \tag{3.6}$$

where $\delta_i \equiv \delta(\vec{x}_i)$, $\nu_i \equiv \nu(\vec{x}_i)$ and $\xi_X \equiv \langle X_1 X_2 \rangle_c$ representing the two-point cumulent moment of order 2 or the correlation function of the X-distribution (see chapter 2). This way combining eqs. 3.6 and 3.1, the two-point correlation functions in configuration space of the two fields are mapped following

$$\xi_\delta \equiv \langle \delta_1 \delta_2 \rangle = \iint \mathrm{d}\nu_1 \mathrm{d}\nu_2 \mathcal{L}(\nu_1)\mathcal{L}(\nu_2) P_\nu(\nu_1, \nu_2, \xi_\nu) \equiv \lambda(\xi_\nu) \ , \tag{3.7}$$

where for standardised Gaussian field in configuration space, the general and non-diagonal covariance matrix reads $C_\nu = \begin{pmatrix} 1 & \xi_\nu \\ \xi_\nu & 1 \end{pmatrix}$, helping to define the bivariate normal distribution (taking $\vec{\nu} \equiv (\nu_1, \nu_2)$ and $|C_\nu| \equiv \det C_\nu$)

$$P_\nu(\nu_1, \nu_2, \xi_\nu) = \frac{1}{2\pi\sqrt{|C_\nu|}} \exp\left\{ -\frac{1}{2} \vec{\nu}^T C_\nu^{-1} \vec{\nu} \right\} \ . \tag{3.8}$$

Apart from the simple case of uncorrelated data $\xi_\nu = 0$ leading to $\xi_\delta = 0$, relation 3.7 shows that the two-point statistics will inevitably be impacted by $\mathcal{L}$. Thus a local mapping for the field implies a local mapping for the two-point correlation function.

Since the power spectrum is the Fourier transform of the two-point correlation function we can rely on the mapping λ in order to predict the input power spectrum of the Gaussian field. Indeed with $|\xi_\nu| < 1$, λ turns out to be continuous and monotonic in this interval. This implies the existence of the reciprocal function λ^{-1}, ensuring eq. 3.7 to be inverted as $\xi_\nu = \lambda^{-1}(\xi_\delta)$.

As a matter of fact, once ξ_δ is known (provided that $P_\delta(\vec{k})$ is a target), being able to compute the two-point correlation function of the Gaussian field ξ_ν is equivalent to find out the specific power spectrum $P_\nu(\vec{k})$ such that once the Gaussian field is transformed under the local transformation $\mathcal{L}$, the resulting δ-field is following the targeted power spectrum $P_\delta(\vec{k})$. Formally, the pipeline reads

$$P_\nu(\vec{k}) = \mathcal{F}\left\{ \lambda^{-1} \mathcal{F}^{-1} \left[P_\delta(\vec{k}) \right] \right\} \ , \tag{3.9}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ stand respectively for the Fourier and inverse Fourier transforms of the 3-D correlation functions. The presented inversion scheme relies on our ability to evaluate the transformation λ of the two-point correlation function, we thus need to compute numerically eq. 3.7.

3.1.3 Mehler expansion and application

An approach appearing more efficient than the evaluation of the 2-D integral 3.7 consists in using the Mehler expansion (Mehler, 1866) to speed numerical computation of any bi-dimensional integral involving a Gaussian distribution. Indeed, the bi-variate Gaussian can be expanded in powers of the two-point correlation function, thus transforming the 2-D integral into a series of 1-D integrals. For more details we refer the reader to appendix D whereas in the following, I focus on the main results.

In this context, one can express the $\mathcal{L}$-mapping eq. 3.5 in terms of the probabilistic Hermite polynomials $\mathcal{H}_{e_n}$ such that

$$\mathcal{L}(\nu) = \sum_{n=0}^{\infty} c_n \mathcal{H}_{e_n}(\nu) \ , \tag{3.10}$$

where the Hermite coefficients c_n can be obtained by the 1-D integration

$$c_n = \frac{1}{n!} \int_{-\infty}^{\infty} \mathrm{d}x \mathcal{L}(x) \mathcal{H}_{e_n}(x) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \ , \tag{3.11}$$

As a result one can express the mapping λ of the two-point correlation functions in terms of the Hermite coefficients as

$$\xi_\delta = \lambda\left(\xi_\nu\right) = \sum_{n=0}^{\infty} n! c_n^2 \xi_\nu^n \ , \tag{3.12}$$

that will be referred in the following as the *Mehler expansion*, although this is formally a Taylor expansion of the two-point correlation function. The series 3.12 converges as long as $|\xi_\nu| < 1$, which fits well with the fact that the Gaussian field has a unitary variance. Moreover, an interesting fact is that when dealing with realistic targeted δ-PDF (see in next chapter), the convergence of this series occurs rapidly.

Thus, the advantage of such formalism is to switch from a 2-D integral 3.7 into a sum of finite 1-D integrals 3.11, easier to tackle. As a result, computing the Hermite expansion of the local mapping allows to find the Taylor expansion of the 2-point correlation function.

In addition to provide an efficient way of computing numerically the relation between the two-point correlations of the density and of the Gaussian field, it offers the possibility of ordering how the powers of the Gaussian field are contributing through the Mehler coefficients c_n.

Note finally that the obtained power spectrum $P_\nu(\vec{k})$ should not be interpreted as physical as its sole role is to be used to generate a Gaussian field that will be transformed into the physical density field $\delta(\vec{x})$.

In the following, after discussing the numerical application of pipeline 3.9, I will go through a concrete case where λ^{-1} can be analytically derived (the Log-Normal case).

3.2 Practical implementation

3.2.1 Aliasing

Let me consider a periodic, cubical comoving volume of size L (and volume $V = L^3$). In addition, I assume that a field can be sampled in it on a regularly spaced grid characterised by the *sampling parameter* N_s, representing the number of point per side (N_s^3 grid-points in the whole box).

We saw in the previous section that generating a non-Gaussian density field in configuration space requires to start from a Gaussian field in Fourier space, thus the Monte Carlo method extensively uses Fast Fourier Transforms (FFT) in order to be efficient. As a result we have to sample the generated field on a regular grid in configuration space, producing in Fourier space a well known effect called *aliasing*.

Basically the FFT of a sampled field is the Fourier transform of the field itself to which replicas of it are added, this produces an extra power on scales close to the sampling scale, also affecting the phases. Then in the following I explain how we deal with aliasing.

Aliasing properties

Let me begin with the general properties satisfied by a periodic field such as the one that will be generated with the Monte Carlo process. This field (either the Gaussian field or the density field) being periodic, one can show that the Fourier transform is discreet. There will be only power at the multiples of the fundamental frequency $k_F = 2\pi/L$ of the box, implying that the lowest non-zero mode is k_F. In addition, we consider a field sampled on a grid with a sampling size $a = L/N_s$ being a multiple of the period.

According to the Nyquist sampling theorem, which states that in order to sample correctly a sine wave we need at least two points per period, it necessarily means that wave mode higher than the Nyquist frequency $k_N = \pi/a$ are meaningless. As a result, it redefines the band-width of physical interest in $\mathcal{I}^+ \equiv [k_F, k_N]$, justifying the reason why FFT algorithms return modes up to k_N. In consequence in the following when considering Fourier grid, spectrum analysis will only be carried out up to the specific wave mode k_N.

Formally speaking, the expected 3-D power spectrum with aliasing contributions can be obtained from the true (theoretical) 1-D power spectrum as

$$\hat{P}(\vec{k}) = \sum_{\vec{n}} P(|\vec{k} - 2\vec{n}k_N|) \ , \tag{3.13}$$

where $\vec{n} = (i, j, k) \in \mathbb{Z}^3$.

The first issue that is brought forward by eq. 3.13 is the fact that aliasing is mixing modes in an ordered way (only with multiples of the considered frequency). This could potentially be a problem when one wants to generate a realisation of a Gaussian field because from eq. 3.13 it seems necessary to generate an infinite number of modes, unfortunately numerically unfeasible.

However thanks to homogeneity (*i.e.* invariance by statistical translation), those modes are uncorrelated, ensuring the Fourier phases to remain uniformly distributed. We can thus generate the Fourier modes by simply computing the aliased power spectrum and generating a Gaussian random variable with the corresponding variance (see eq. B.15).

In addition, a second salient effect that is worth to mention, is the fact that the density field cannot be fully isotopic especially for modes close to the Nyquist frequency, a more in-depth discussion will follow in section 3.2.2.

Interplay between sampling and aliasing

Obviously, the infinite 3-D sum 3.13 cannot rigorously be applied, first by numerical considerations but also because galaxy clustering models are not able to provide a reliable matter power spectrum up to arbitrarily high modes. In particular, galaxy and baryonic physics are expected to have a significant impact in the non-linear regime ($k \gtrsim 1h/\text{Mpc}$), an effect not embedded in numerical predictions. Moreover, even discarting these physical effects and considering matter only, theoretical models are able to predict a reliable power spectrum only up to $k \sim 0.2 - 0.5h/\text{Mpc}$ (see chapter 2).

However, no matter what is the exact value of the power spectrum at non-linear scales, it is expected to decrease as a power law. In that respect higher order terms in the sum 3.13 are less and less important, which means that we can truncate it. Let me define the parameter j that quantifies the order of the truncation.

Definition: *The 'mth alias contribution' corresponds to* $j = m$ *and involves the* $(2m+1)^3$ *first terms of eq. 3.13 from* $\vec{n} = (-m, -m, -m)$ *to* $\vec{n} = (m, m, m)$.

This truncation involves computing the theoretical power spectrum up to a maximum mode k^{th} obtained by taking $\vec{k} = (k_F, k_F, k_F) \sim \vec{0}$ in eq. 3.13, namely $k^{\mathrm{th}} = 2\sqrt{3} j k_N$. For instance for $N_s = 512$ and $j = 3$, the theoretical power spectrum, prior to aliasing, must be provided up to $k^{\mathrm{th}} = 11.14h/\mathrm{Mpc}$ in order to return an aliased power spectrum defined in $\mathcal{I}^+$, up to $k_N = 1.61h/\mathrm{Mpc}$.

In practice for a fixed box volume, two main input numerical settings have an influence on the aliasing contribution: N_s and j. First, an increasing sampling parameter N_s will shift toward higher modes the Nyquist frequency while making the computation more CPU and memory expensive. Typically, the N_s values used for the various analysis are[2] $[256, 512, 1024]$ for which, taking $L = 1000h^{-1}\mathrm{Mpc}$, the Nyquist frequencies are respectively $k_N = [0.80, 1.61, 3.22]h/\mathrm{Mpc}$.

At first sight, given these maximum mode values and as previously mentioned, it might seem sufficient to focus only on $N_s = 256$. However, the needs for a high N_s resolution comes first from the fact that input power spectra could be estimated on simulations or directly from galaxy surveys (that offer an estimation of the power spectrum for $k \gtrsim 1h/\mathrm{Mpc}$). In addition, as it will be the subject of a detailed study in next chapter, the notion of filtering introduced by the grid affects the matching between the true and simulated power spectra. For these reasons it is important to allow a higher resolutions as $N_s = 512$ and $N_s = 1024$.

On the other hand, while fixing L and N_s, the nature of the power spectrum itself can make the aliasing effect sub-dominant or on the contrary dominant. For instance, a linear power spectrum starts to behave as k^{-2} around $0.7h/\mathrm{Mpc}$ an aliasing effect will in this case be much more localised around the Nyquist frequency than in the case of an aliased non-linear power spectrum, characterised by a higher scaling $\propto k^{-1}$. If the resolution N_s is not sufficient, the whole interval $\mathcal{I}^+$ will be impacted, even reaching the fundamental mode k_F. This effect is illustrated in figure 3.1 in the case of a standard ΛCDM cosmology and a box size $L = 1000h^{-1}\mathrm{Mpc}$ in linear and non-linear models for the power spectrum and for various number of sampling points N_s.

For any number of sampling, the figure shows that the aliasing effect is larger for the non-linear power spectrum. In particular for low sampling $N_s = 256$, it multiplies the effect by a factor of three. In addition, even in the case of a $N_s = 1024$ grid and a linear power spectrum the aliasing contribution is still of about 1% at the Nyquist frequency of the lowest resolution grid ($N_s = 256$), showing that aliasing must absolutely be taken into account in the Monte Carlo procedure.

Note that 1-D power spectra are plotted here, even if eq. 3.13 produces 3-D grid power spectra. In order to ensure the comparison, a shell-averaging is performed in shells of width k_F and centred at $|\vec{k}_n| = nk_F$ (see eq. 2.44), both for the power spectra and for the 3-D wave modes.

---

[2] Note that FFT algorithms are optimised for a grid sampling $N_s = 2^p$ with $p \in \mathbb{N}$.

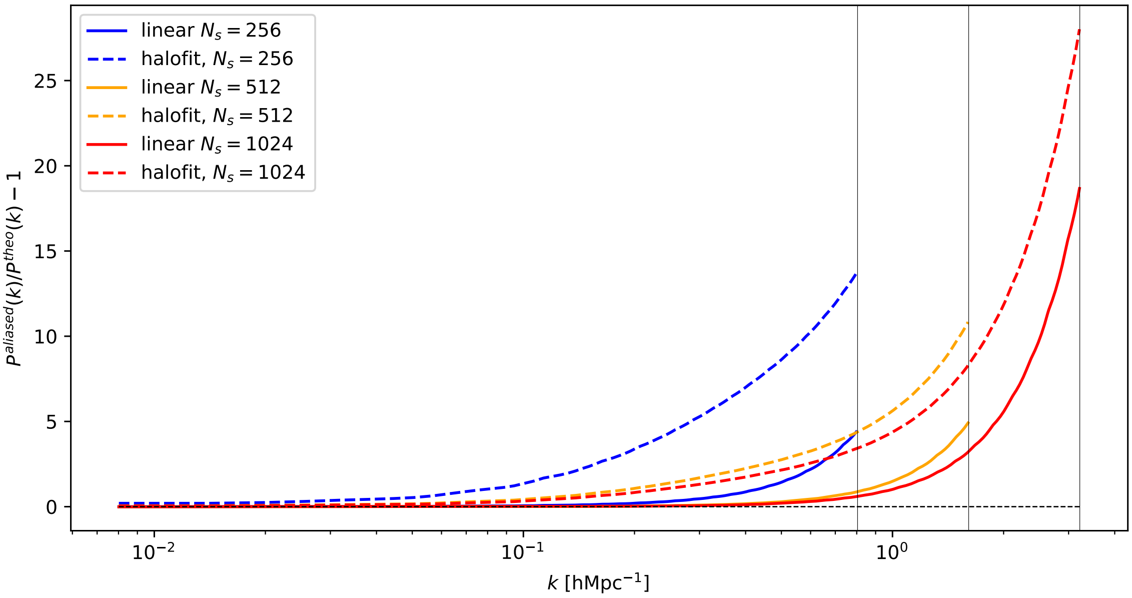


Figure 3.1: Relative deviation between the shell-average aliased and raw matter power spectra using the `CLASS` code in linear and non-linear model (using the `HALOFIT` prescription). Aliasing is artificially added using eq. 3.13 for three different grid precision $N_s = [256, 512, 1024]$. The three solid vertical lines represent the Nyquist frequencies computed at corresponding N_s, respectively $[0.80, 1.61, 3.22] h/\mathrm{Mpc}$.

Incorporating aliasing in the procedure

Coming back to the Monte Carlo method, it is worth mentioning that the Fourier transform is a bijective mapping and so the FFT algorithm, *i.e.*

$$\mathrm{FFT}^{-1}\left[\mathrm{FFT}[\vec{X}]\right] = \vec{X} \ , \tag{3.14}$$

$$\mathrm{FFT}\left[\mathrm{FFT}^{-1}[\vec{Y}_{\vec{k}}]\right] = \vec{Y}_{\vec{k}} \ . \tag{3.15}$$

This is showing that since the Fourier transform of a field in configuration space (sampled on a grid) is aliased then if one wants to create the configuration space field from its Fourier transform by inverse Fourier transforming it, one needs to start from an aliased Fourier transform. Indeed, if this condition is not satisfied then the sampled field in configuration space will be different than the expected one, thus the process would be incoherent.

Technically, the aliasing can be carried out by applying eq. 3.13 on the targeted power spectrum. In the contrary, for an *inconsistent* configuration space (or unaliased Fourier space), the 3-D power spectrum can result from a 3-D interpolation of the targeted power spectrum on the 3-D wave modes $\vec{k}$.

We need to address the question of the influence of the truncation order (parameter j) of the aliasing sum of the power spectrum, keeping in mind that the objective of the method is to reach a high accuracy level of the simulated power spectra up to the Nyquist frequency.

Figure 3.2 helps to give an idea of the impact of the number of alias contributions applied on the input power spectrum for several truncation parameters, keeping in mind that the numerical cost of increasing j goes as $(2j+1)^3$, justifying why I restrict the comparison to $j \leq 6$. For each of the six contribution types, it shows a significant difference in amplitude at high modes with each other, with very slow convergence. This difference in amplitude around the Nyquist frequency is explained by the fact that an increasing j correspond to an increasing k^{th} as well.

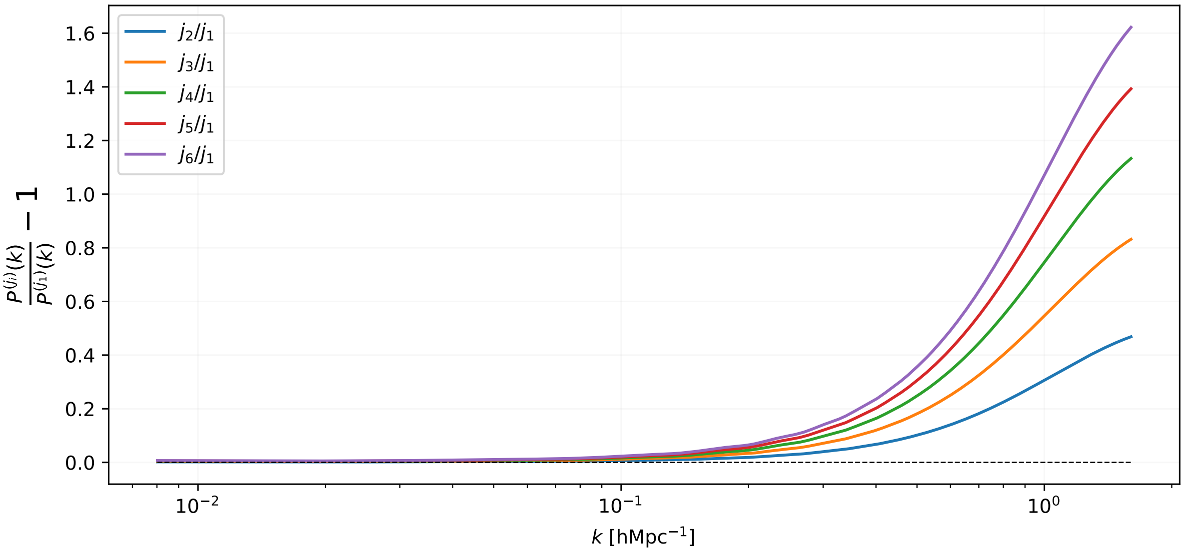


Figure 3.2: Relative deviation between aliased matter power spectra with contributions j from 2 to 6 with respect to $j = 1$ as a function of k. The used sampling is $N_s = 512$ when predicting the linear matter power spectrum.

It will then be of interest to find an optimal parameter j minimising the difference between the simulated power spectrum of the density field and the expected one. Indeed, we need to account for aliasing mainly because we have to generate a Gaussian field in configuration space with a given two-point correlation function in order to ensure that once transformed with the local mapping $\mathcal{L}$ the resulting density field will have the correct expected power spectrum. As a matter of fact if one completely neglect aliasing, then the power spectrum of the Gaussian field would be lower than expected, leading to a mismatch between the power spectrum of the density field and the expected one. Restricting to $j \leq 4$, this is what I am investigating in the next section.

Finally, in order to summary and somehow clarify the different steps involved in the field simulation, figure 3.3 sketches the Monte Carlo pipeline.

3.2.2 Accurate Log-Normal realisations

A typical study case widely explored in literature (*e.g.* Coles & Jones, 1991; Chiang et al., 2013; Greiner & Enßlin, 2015; Agrawal et al., 2017; Xavier et al., 2016a) is to adopt a Log-Normal shape for the targeted δ-PDF. Among its significant advantages, it turned out to be a very realistic approximation of the true distribution measured

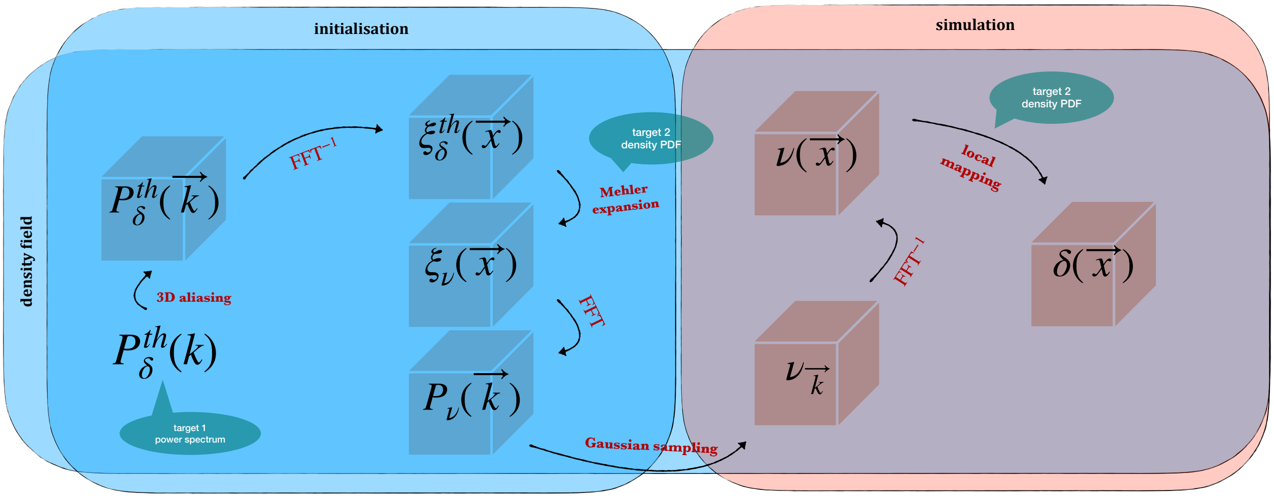


Figure 3.3: Sketch of the Monte Carlo procedure to generate a continuous field on a grid by targeting a given power spectrum and PDF. First of all, we start with the initialisation stage. The user needs to provide the two inputs: the targeted PDF and the 1-D theoretical power spectrum for the density field. To the last is applied eq. 3.13 to get a 3-D (represented as a box) theoretical power spectrum with alias contributions. To it is applied an inverse Fourier transform to get the consistent two-point correlation function ξ_δ in configuration space. Then the Mehler expansion 3.12 allows to know the two-points correlation function to attribute to a Gaussian standardised field ν, given the local mapping $\mathcal{L}$ (see eq. 3.5). The corresponding power spectrum P_ν can then be used to simulate several aliased Gaussian fields $\nu_{\vec{k}}$ in Fourier space using eq. B.15, and in consequence consistent $\nu(\vec{x})$ fields in real space. Finally, applying $\mathcal{L}$ gives the non-Gaussian density contrast field $\delta(\vec{x})$ with controlled PDF and power spectrum.

in galaxy surveys or N-body simulations, for a smoothing window producing $\sigma \lesssim 1$ (linear scales only).

When plotting the histogram of the logarithm of the extra-galactic nebulae[3] number in regular angular patches (see figure 3.4), Hubble (Hubble, 1934) was the first to conjecture the Log-Normal galaxy distribution, even with a really poor statistics of 44000 objects.

The Log-Normal distribution presents some interesting advantages, in particular as it is fully reviewed in appendix E, the mapping between the two-point correlation functions of the Gaussian field ν and the density field δ (see eq. D.15) can be analytically inverted

$$\xi_\nu = \lambda^{-1}(\xi_\delta) = \ln(1 + \xi_\delta) \tag{3.16}$$

when the local non-linear mapping designed to reproduce the Log-Normal PDF reads

$$\delta = \mathcal{L}(\nu) = \exp\left\{\nu\sigma_\phi - \sigma_\phi^2/2\right\} - 1\ , \tag{3.17}$$

---

[3] A controversy in the first half of the XX$^{\text{th}}$ century, called the Great Debate, was to whether interpret extra-galactic nebulae as galaxies or Milky way nebulae.

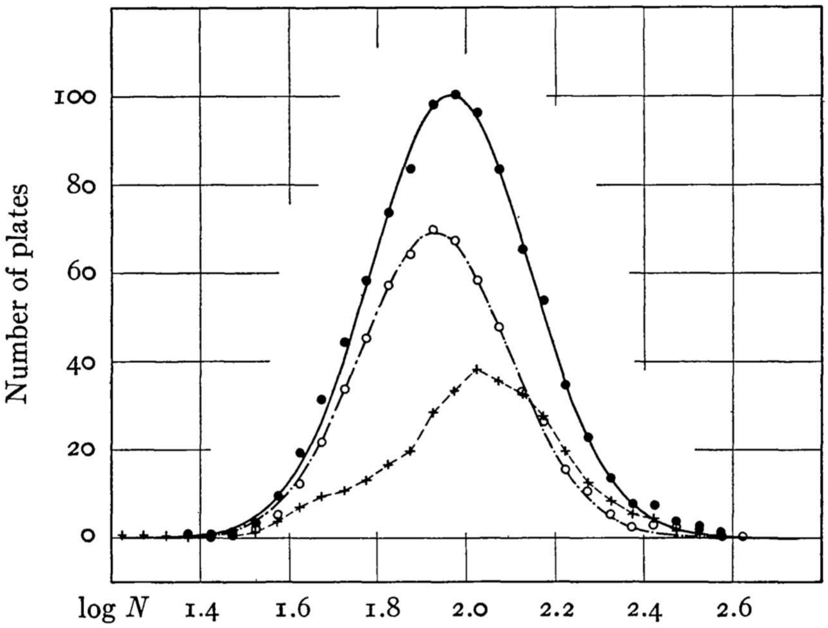


FIG. 8.—Frequency distribution of log N reduced to the galactic poles. Data from Table XIV. Crosses represent the 331 extra-survey fields; circles, the 587 survey fields; disks, the combined data, 918 fields. The smooth curves through the survey fields and the combined data (the two upper curves) are normal error-curves adjusted to the points.

Figure 3.4: credit: Hubble (1934)

with $\nu = \left[\ln(\delta + 1) + \sigma_\phi^2/2\right]/\sigma_\phi$ and $\sigma_\phi^2 = \ln(\sigma_\delta^2 + 1)$. Of course, despite the above analytical results, most of the process remains numerical.

Simulating a Log-Normal PDF

In the following I set a periodic box size $L = 1000h^{-1}$Mpc and a sampling grid $N_s = 256$. In figure 3.5 are featured the various power spectra involved in the Monte Carlo process depicted previously. It shows first the raw power spectra computed with linear (top panel) and non-linear model (bottom panel) together with their aliased version using relation 3.13.

Note the non correspondence between the wave mode intervals of the shell-average spectra with the theoretical ones well defined on $\mathcal{I}^+ = [k_F, k_N]$. This mismatch is due to the fact that shell-averaging is applied not only on the power spectrum but also on the grid norm of the wave modes, as already mentioned.

In addition, the illustration of the non-physical power spectra of the Gaussian fields ν shows for the whole k-range a significant lack of power when compared to the targeted ones. Indeed by construction, the variance of the Gaussian fields $\sigma_\nu^2 = 4\pi \int k^2 \mathrm{d}k P(k) = k_F^3 \sum_i P_\nu(\vec{k}_i)$ (see appendix C) is unitary and finite unlike the variance of the linear and non-linear raw power spectra which are both diverging (when the k-interval is not bounded from above).

Following the method prescription of appendix B, from the Gaussian power spec-

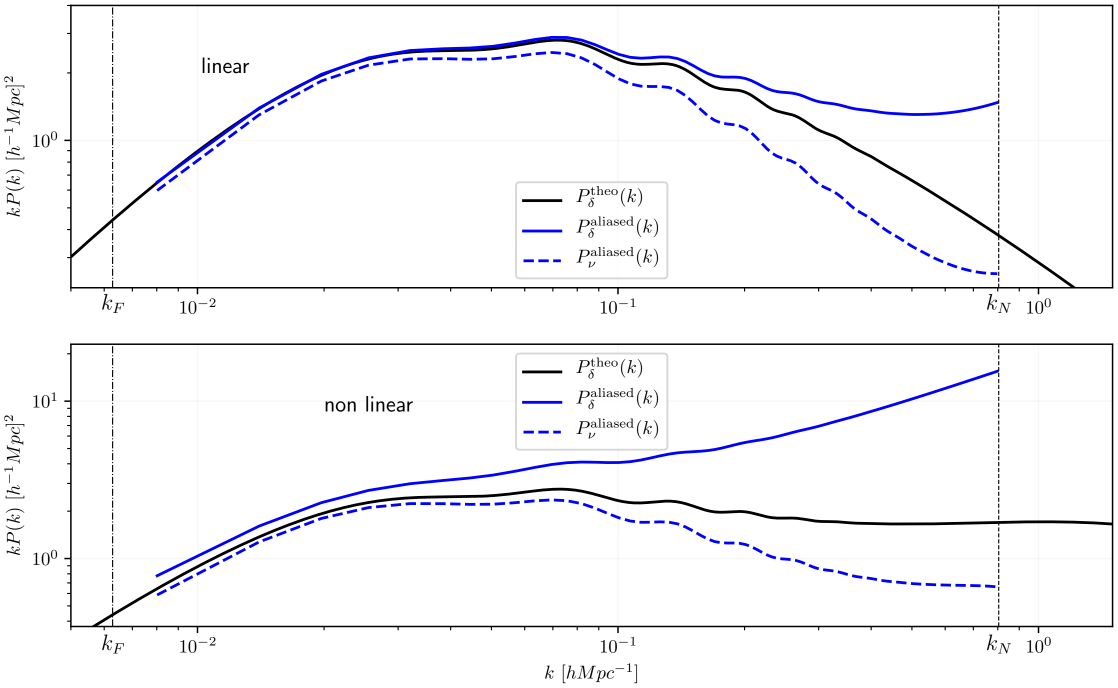


Figure 3.5: Power spectra involved in the Monte Carlo process in the case of linear (top panel) and non-linear (bottom panel) ΛCDM models, for a box size $L = 1000h^{-1}$Mpc and a sampling parameter $N_s = 256$. In solid black lines are represented the theoretical one-dimensional matter power spectra P_δ^{theo} as predicted by `CLASS`. In solid blue and dashed blue lines are the shell-average spectra of respectively the aliased versions of P_δ^{theo} with $j = 1$ and the corresponding power spectra after transformation 3.16. The vertical references represent the fundamental frequency k_F (dash-dotted black lines) and the Nyquist frequency k_N (dashed black lines).

trum can be generated in Fourier space the $\nu_{\vec{k}}$ field using the local mapping eq. 3.17.

In figure 3.6 is represented the measurement of the δ-field PDF in configuration space compared to the expected Log-Normal one (see appendix E)

$$P_\delta(\delta) = \frac{1}{\sqrt{2\pi\sigma_\phi^2}} \mathrm{e}^{-\nu^2/2-\nu\sigma_\phi+\sigma_\phi^2/2}\ . \tag{3.18}$$

The near-matching between the two PDFs, taking the form of a $\sim 2\%$ systematics, is exclusively due to the fact that the measured PDF is made on bins while the expected one is computed at exact δ-values. The ratio is therefore computed for non equal, but close, δ-values. This effect vanishes when increasing the binning.

In conclusion, a Log-Normal PDF can be reproduced with a high-level of reliability in any regime (linear for $\delta \lesssim 1$ and non-linear for $\delta \gg 1$) when using the non linear transformation 3.17.

Distribution of the power spectra

Now concerning the reliability of the simulated power spectra, only the linear case will illustrate the following analysis. Indeed when targeting the non-linear power

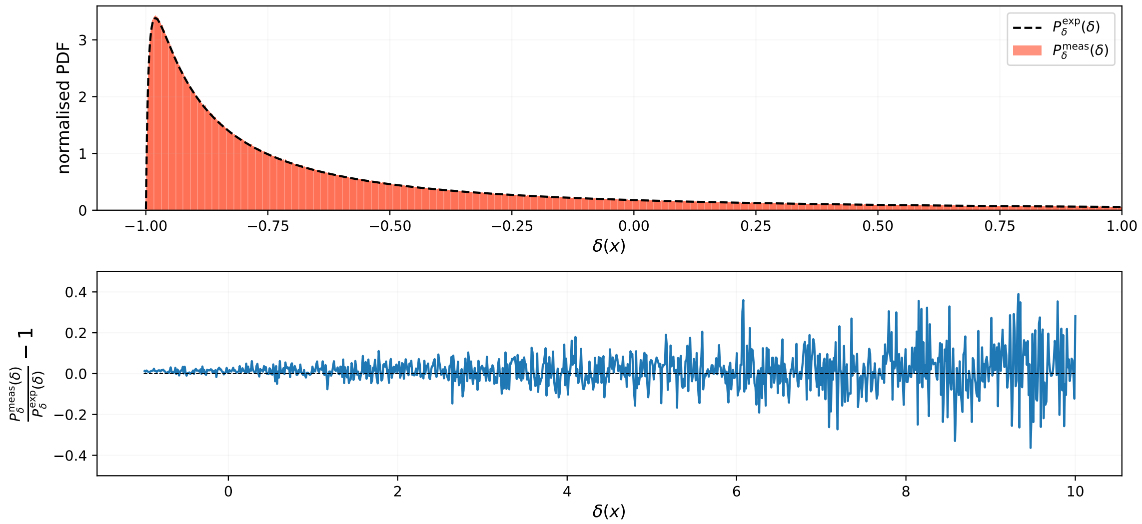


Figure 3.6: *top panel:* Distribution measurement for one realisation of the Log-Normal Monte Carlo field δ (in orange) as compared to its prediction (dashed black line, see eq. 3.18) for the truncated interval $\delta \in [-1, 1]$. *bottom panel:* relative deviation between the two distributions in a wider interval $\delta \in [-1, 10]$.

spectrum, the Gaussian power spectrum $P_\nu(\vec{k})$ obtained from the Mehler expansion 3.17 is produced with negative elements in the vicinity of the Nyquist frequency (in 3-D). This non physical and undesired effect, that seems to be due to the shape of the power spectrum at high modes, prohibit any field generation following the Monte Carlo method, in particular the Gaussian sampling step. This is why in the case of the Log-Normal PDF, only the linear power spectrum is used to produce non-Gaussian fields. This issue will be tackle in the context of a more realistic case in section 4.2.1, in order to allow non-linear power spectrum simulations.

Moreover, before going through the method validation in terms of power spectrum, one has to discuss the way power spectra are averaged when considering several realisation of the density field. In particular, do the shell average and average over realisations commute ? This question is specifically motivated by the observation in figure 3.7 of the non-Gaussian distribution nature of the power spectrum[4] for two grid samplings and 10000 Log-Normal realisations and for varying j.

Regarding $k = k_F$, the distribution is simply a matter of measurement. It is exclusively due to the low number of modes to average at large scales, thus preventing the central limit theorem. In consequence the resulting distribution should follow a Gamma distribution whose the dispersion is described by the *cosmic variance*.

As expected the distribution appears to be almost Gaussian for intermediate mode k_G ($\sim 0.13 h\mathrm{Mpc}^{-1}$) while surprisingly a strong non-Gaussian behaviour is exhibited at the Nyquist frequency no matter what is the value of j. Visually speaking, the distribution shape at k_N can be approximated to a Gamma distribution as well. Note that this is an effect already mentioned in N-body results (see Blot et al., 2015) but whose the origin is difficult to identify. In the present Monte Carlo

---

[4] I'm not referring here to the field non-Gaussianity weighted by the PDF. Instead, I'm talking about the distribution of the power spectrum for each mode k.

procedure instead, the origin of theses non-Gaussianities seems to be rather related to a numerical than a physical origin. However there is no guaranty that the origin of such an effect is similar for the N-Body outcomes.

It can also be notified that this figure does not presents the results for $j = 0$. Indeed when the target power spectrum is simply interpolated on the 3-D wave mode $\vec{k}$ (rather than aliasing it using eq. 3.13) the procedure presented above is aborted because it once again produces some negative elements in $P_\nu(\vec{k})$. This result is a strong argument that aliasing needs to be correctly addressed.

When varying the j-parameter, no significant change can be seen on the distributions for the modes k_F and k_G. On a other hand, it's a distinct response that is observed at k_N, then comforting us in the idea that aliasing induces a spread of the power spectrum distribution, probably induced by subsequent anisotropies.

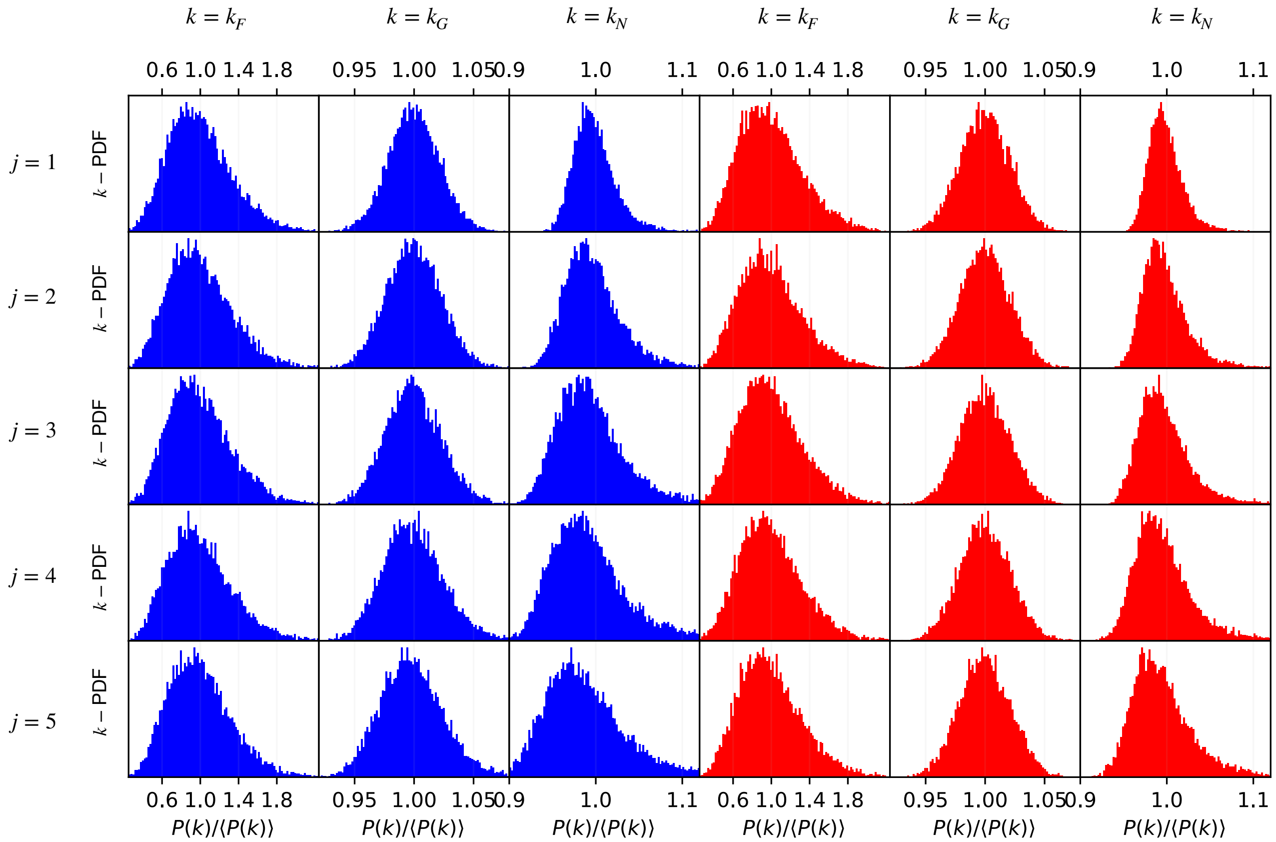


Figure 3.7: Normalised mode distribution of the power spectrum divided by its averaged value over 10000 realisations of Log-Normal fields for two grid samplings $N_s = 256$ (in blue) and $N_s = 512$ (in red). The j-parameters are varying overs rows from $j = 1$ to $j = 5$ while columns look at the distribution for different modes: the fundamental frequency $k_F \sim 0.008h/\mathrm{Mpc}$, an arbitrary intermediate scale $k_G \sim 0.13h/\mathrm{Mpc}$ and the Nyquist frequency $k_N \sim 0.8h/\mathrm{Mpc}$ for $N_s = 256$ and $k_N \sim 1.6h/Mpc$ for $N_s = 512$.

To be more quantitative, figure 3.8 shows the deviation from the Gaussian distribution for the whole range of wave modes through its asymmetry (the skewness S_3) and its spreading (the kurtosis S_4). It can be observed that non-Gaussianity starts to act at higher modes when decreasing the sampling parameter (from $N_s = 512$ to $N_s = 256$) while shifted toward lower modes when j increases. It confirms that at the Nyquist frequency, the relative S_3 ad S_4 values between the two grid sampling does not seem to follow any trend.

On the other hand, the Gamma distribution at small k, as expected, gradually converts into a Gaussian one around $k \sim 5 \times 10^{-2} h/\mathrm{Mpc}$.

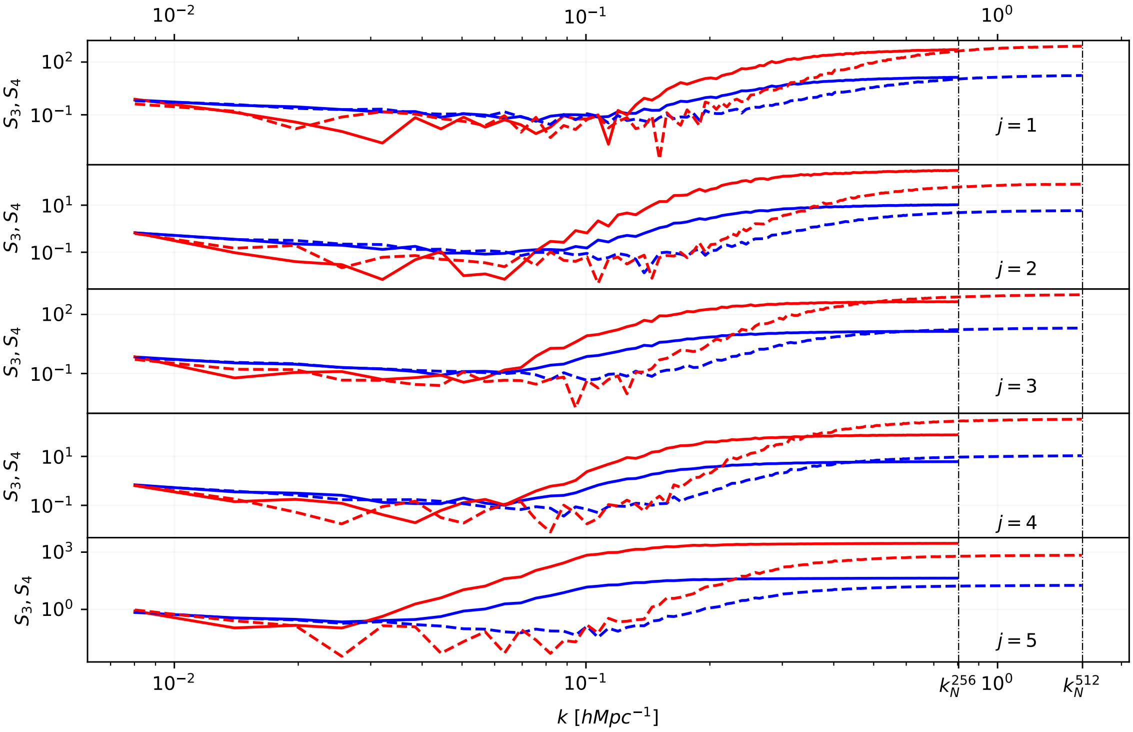


Figure 3.8: Absolute values of the skewness S_3 and the kurtosis S_4 of the shell-average grid power spectra wave mode distribution over 10000 realisations of Log-Normal fields for two grid sampling: $N_s = 256$ (solid lines) and $N_s = 512$ (dashed lines). S_3 is represented in blue while S_4 in red as a function of k up to the Nyquist frequency $k_N \sim 0.8\mathrm{hMpc}^{-1}$ for $N_s = 256$ and $k_N \sim 1.6\mathrm{hMpc}^{-1}$ for $N_s = 512$. The different panels account for different j-parameters.

As a matter of fact, comparing statistically the simulated grid power spectra with their corresponding prediction can be carried out in two different ways.

- The first consists in comparing the average over realisations of the shell-average power spectra with the 1-D targeted power spectrum. However following the above considerations and as represented in the panels 1 and 3 of figure 3.9, the non-Gaussian distribution implies an asymmetry of the corresponding relative deviations.

- This effect is dampened when performing the comparisons in a second, more rigorous way, also represented in the same figure in panels 2 and 4. By definition of the power spectrum estimated in a volume, $P(\vec{k}) \propto \left\langle \delta_{\vec{k}}^2 \right\rangle$, the expectations means that this is an average over realisations of the 3-D power spectra. Thus the second method consists in shell-averaging the averaged 3-D power spectra over realisations.

The two distinct outcomes highlighted in figure 3.9 show that shell-averaging and realisation averaging are not commuting. In the second case (the more rigorous one), it reveals an agreement better than the percent level of the Monte Carlo method for the whole range of wave modes, without unsealing any systematic. As a result, one should keep in mind that aliasing has an effect on the power spectrum estimator when it is shell averaged.

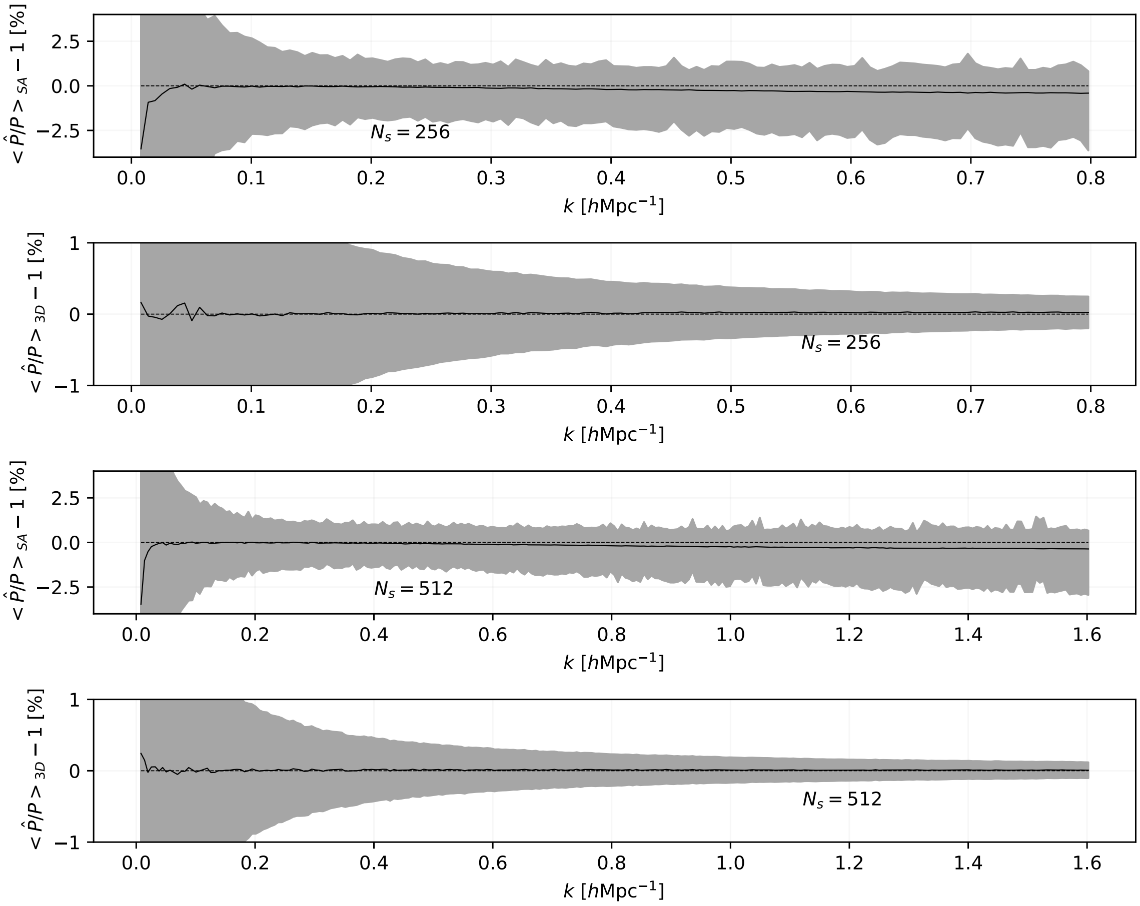


Figure 3.9: Relative deviations in percent with error bars of 10000 simulated power spectra in boxes of length $L = 1000h^{-1}$Mpc, $j = 1$ and for two grid sampling parameters $N_s = [256, 512]$. $<\hat{P}/P>_{SA}$ means that averaging is made on shell-average power spectra while its the other way around for $<\hat{P}/P>_{3D}$: averaging are first made on each 3D modes and the resulting spectrum is shell-averaged. In the first case, the solid black line represents the median of the distribution.

Finally, putting aside the influence of j on non-Gaussianity, figure 3.10 shows the matching level between the reproduced spectra and their targets, when varying this parameter. Even if its effect on the relative deviation between models and measures is lower that 0.1%, for both samplings, the value $j = 2$ provides an excellent compromise between execution time and accuracy. For this reason, in the following when considering the Log-Normal case we will set the aliasing truncation parameter to $j = 2$.

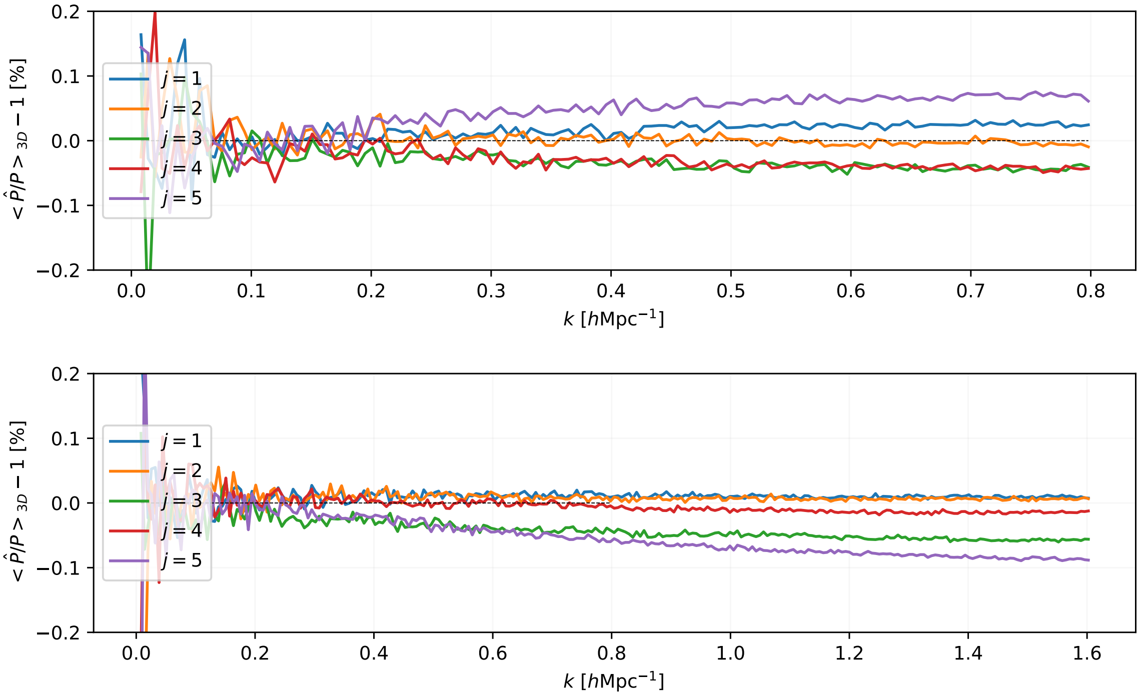


Figure 3.10: Relative deviation in percent of shell average power spectra after 3-D averaging over 10000 Log-Normal fields realisations and the targeted one for two grid samplings $N_s = 256$ (top panel) and $N_s = 512$ (bottom panel) and varying j-parameter.

Covariance matrix

Now that we are confident with the reliability of the simulated power spectra, it could be pertinent to check the statistical behaviour of the produced density fields, especially in terms of the covariance matrix. In particular we can manage to compare the estimated matrix with its corresponding prediction, still in the Log-Normal context.

Indeed, since the local transformation of the Gaussian field ν into δ, as well as the targeted power spectrum, are specified, we can in principle predict the n-point moments of the non-Gaussian field. This can be performed efficiently in a perturbative way by using the decomposition of $\mathcal{L}$ on the basis of Hermite polynomials (see eq. 3.10). Especially, we can focus on the shell-average trispectrum $\bar{T}$ entering the expression of the covariance matrix 2.46.

In appendix C of Baratta et al. (2020) are presented the corresponding derivations for $\bar{T}$, in the context of an arbitrary non-Gaussian field. For the diagonal terms it reads

$$\begin{aligned} \bar{T}(k_i, k_i) \simeq 8c_1^2\left(4c_2^2 + 3c_1c_3\right) P_\nu^3(k_i) \qquad\qquad (3.19)\\ + 24\left(3c_1^2c_3^2 + 4c_1c_2^2c_3 + 12c_1^2c_2c_4\right) P_\nu^2(k_i) P_\nu^{(2)}(k_i) \qquad (3.20)\\ + 144c_1^2c_3^2 P_\nu^{(2)}(0) P_\nu^2(k_i)\ , \qquad\qquad (3.21) \end{aligned}$$

where $P_\nu^{(n)} \equiv \mathcal{F}[\xi_\nu^n]$. The above expression depends explicitly on the Hermite coefficients c_n which can be computed as long as we know the local transform $\mathcal{L}$.

Three kind of non-Gaussian contributions are obtained (3.19), (3.20) and (3.19). Let me label their cumulative contribution through the definitions

- 1-NG : (3.19)
- 2-NG : (3.19)+(3.20)
- 3-NG : (3.19)+(3.20)+(3.21)

in order for each of them to be compared in the Log-Normal case (for which the Hermite coefficients are simply given by $c_n = (1 - \delta_{n0})/n!$, see appendix E) with the Monte Carlo procedure.

Thus in figure 3.11 are compared the cumulative predictions of the diagonal elements of the covariance matrix (adding to it the Gaussian variance) to the one estimated over the 10000 realisations of Monte Carlo fields with Log-Normal PDF.

Error bars on the covariance matrix elements are computed assuming data to be distributed according to a normal distribution. Given that the covariance matrix is estimated by sampling independent realisations of Gaussian distributed data (the power spectra), its properties are described by the Wishart distribution and the error on the covariance matrix can be predicted as

$$V\left[\hat{C}_{ij}\right] = \frac{C_{ij}^2 + C_{ii}C_{jj}}{N-1} \ , \tag{3.22}$$

where N stands for the number of realisations. The prediction eq. 3.22, in the present case, allows to give an idea of the dispersion.

Moreover one can see from figure 3.11 that at least up to $k \sim 0.2h\mathrm{Mpc}^{-1}$, the perturbative prediction and the estimation of the diagonal of the covariance matrix are consistent. Notice also the fact that generating such a high number of realisations (it takes only 3 hours to generate 10000 realisations for $N_s = 512$) allows to get a high signal to noise ratio, enabling the Baryonic Acoustic Oscillations to be clearly seen in the estimated variance of the power spectrum.

On the other hand, it can also be noticed that for high modes, the perturbative prediction didn't converge, as (G+2-NG) and (G+3-NG) still present substantial deviation. Thus, for modes $k > 0.2h/\mathrm{Mpc}$, the comparison can be misleading and no conclusion can be drawn.

To conclude this section, we used the Log-Normal to validate the Monte Carlo method in terms of the statistical targets that must be reproduced. Indeed I have shown that the proposed Monte Carlo process allows to match with a high-level of accuracy the power spectrum and the PDF used as input.

In addition, we verified that the estimated diagonal of the covariance of the power spectrum behaves as expected by theory. Of course, a next step is to compare the covariance matrix estimated with Monte Carlo realisation to the one obtained with N-body simulations. To this end, I have to show how to create catalogues of objects that could be compared with outputs from N-body simulations.

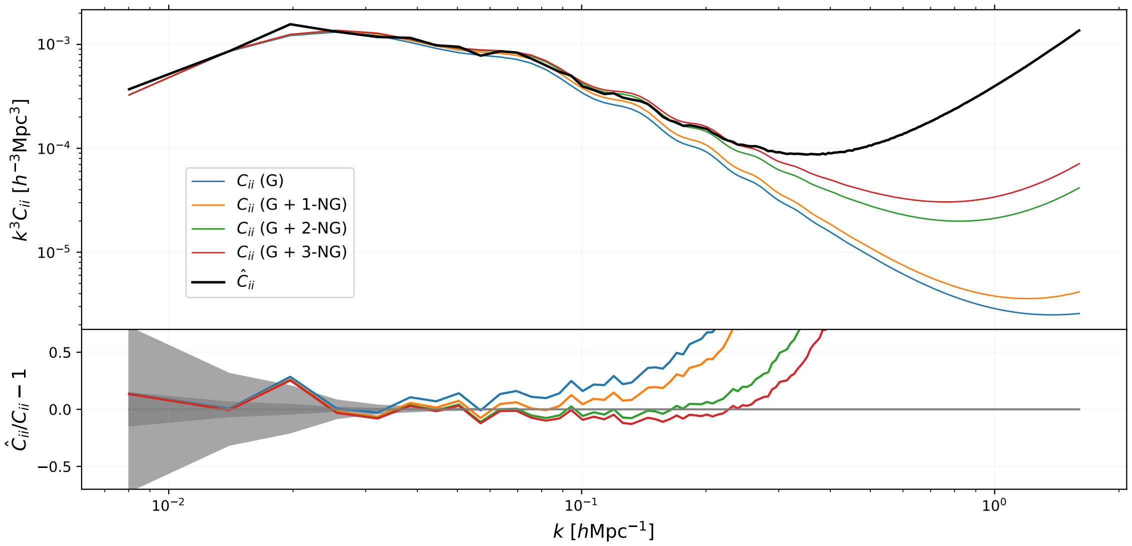


Figure 3.11: *Top:* diagonal of the power spectrum covariance matrix estimated over 10000 Log-Normal field realisations (solid black line) as compared to the different contributions of the perturbative derivation of $\bar{T}$ to which is added the straightforwardly predicted Gaussian contribution in eq. 2.46. The used setting is $j = 2$ and $N_s = 512$. *Bottom:* relative deviation for each level of perturbative contribution with the measured one. The dark and light grey areas represent respectively the 1σ and 5σ Gaussian error limits.

Chapter 4

Monte Carlo catalogues

In the previous chapter, I have shown how to generate a continuous fluctuation field δ sampled on a regular grid with some controlled PDF and power spectrum. In the present chapter I describe how to convert such density field into a catalogue of objects, first in comoving space, then in redshift-space when introducing peculiar velocities. In particular, the creation of a catalogue out of the density field is mainly motivated by the following ideas:

1. Within the Monte Carlo method, we are building the density field on a regular grid that leaves a characteristic non-physical imprint in Fourier space, aliasing. Since at the end of the day we need to produce some reliable power spectra, keeping a grid description does not allow to extract ourself from aliasing. However converting the density field into a point process in real space, as previously said, enables to use some two-point statistics estimator, highly reducing aliasing. Of course, one has to keep in mind that the conversion from density to a catalogue will alter the targeted power spectrum. However, I will show that this change can be accurately predicted and then included at the level of the target power spectrum.

2. The second motivation concerns the redshift-space covariance matrix. In fact, since redshift-space distortions are, by definition, distorting the power spectrum, this will propagate into the covariance matrix. In order to investigate such effect, we need to turn the density field into a discrete set of objects mimicking a galaxy, a halo or a dark matter particle catalogue. The distortions induced by peculiar velocities on the measured position of particles are exact when one wants to change the apparent position of an object given its peculiar velocity. In the contrary, to apply redshift-space distortions to a density field, one needs to expand the Jacobian of the transformation, although it can be singular. As a result, having a catalogue of objects is more suited for redshift-space analysis, partly ensuring to reproduce non-linear effects such as the Fingers-of-God (Jackson, 1972) that comes (partially) from the velocity field dispersion at small scales.

3. Finally, observations will provide galaxy catalogues characterised by some instrumental constraints such as angular masks and radial redshift distribution functions ($n(z)$), designed to be applied on objects. Thus, if one wants to

study the effect of the chosen observational strategy on the covariance matrix of a given observable, it is mandatory to apply the same pipeline to the observed galaxies and to their Monte Carlo realisations.

The first part of the present chapter focuses on presenting some discretisation procedures (section 4.1), or more specifically how to simulate catalogues of objects (galaxies, cold dark matters particles, halos) out of the previously obtained density fields sampled on a regular grid. Then, I will go through the method validation (section 4.2) in comoving space when comparing the estimated power spectrum covariance matrices produced by the Monte Carlo method and the ones coming from N-body simulations. After reviewing how the distortions induced by peculiar velocities on the observed redshifts are affecting the measured power spectrum in section 4.3, I will present how I assign velocities to the objects of the Monte Carlo catalogues in section 4.4. Finally, this method will be validated at the level of the redshift-space power spectrum and its covariance matrix in section 4.5.

4.1 Poisson sampling

The basic idea to turn the density field sampled on a regular grid $\delta(\vec{x})$ into a discreet catalogue of objects is to apply a Poisson sampling (Layzer, 1956) of it. Providing a mean number density ρ_0 expected in a catalogue of volume V and a local density contrast field $\delta(\vec{x})$, one can estimate the local number of objects Λ expected at the same position within a small volume element v as $\Lambda = \rho_0 v(1+\delta)$.

Of course, this expected number Λ is not an integer. However it can be seen as the expectation value that one should see if the sampling was made several times at position $\vec{x}$. Thus the number of observed objects in the small volume v can be drawn from a Poisson distribution with expectation value Λ. In the following, I describe in more details this process leading to a catalogue of objects.

4.1.1 Effect on the density field

From continuous to Poisson field

Before explaining how we produce a catalogue of objects, it is important to clarify the notion of field continuity. The simulated density field, that has been so far generated on a grid, is continuous in the sense that it exists at every spatial positions. However we know it only on the regularly spaced spatial grid corresponding to the Fast Fourier Transform algorithm applied on the generated Fourier fields. As a consequence the sampled field $\delta_{\text{III}}(\vec{x})$ (on the grid) can be formally written as

$$\delta_{\text{III}}(\vec{x}) = \delta(\vec{x}) \text{Ш}_a(\vec{x}), \tag{4.1}$$

where $\delta(\vec{x})$ stands for the continuous and periodic density field and Ш_a is the sampling function representing the spatial grid normalised such that its integral over the period L is unity. In particular, the sampling function can be represented by a Dirac comb of size a corresponding to the spatial grid. The grid size a must be related to the period of the density field L and the number of grid point per side N_s

through $a = L/N_s$. Thus rigorously, what is generated on a grid is a continuous field sampled on specific positions in real (or Fourier) space, but for the sake of simplicity, the subscript Ш will be dropped in the following, as it was the case in chapter 3.

It is clear that due to the sampling grid we don't know the density field in between the grid nodes, while we specifically want to populate this space with points representing galaxies, haloes or dark matter particles. It is therefore necessary to interpolate the density field in order to compute the expected number of objects contained in a given volume that we will call a *cell* or a *voxel*.

As anticipated in the introduction of this section, an average number density of particle ρ_0 has to be arbitrarily set at the beginning of the process. It allows the construction of an interpolated local density in the vicinity of each grid point $\rho_i(\vec{x}) = \rho_0(1+\delta_i(\vec{x}))$ where the index i labels the grid position. The expected number of objects can be evaluated by integrating the local interpolated density around each cell $\Lambda_i = \int \mathrm{d}^3\vec{x}\rho_i(\vec{x})$, where the integration domain is the volume of a cell.

Following the work done by Layzer (1956), we assume the *Local Poisson Process Approximation* claiming that the probability of finding a number of object N in a cell, where the expected number of object is Λ, follows a Poisson distribution

$$P_N = \Lambda_N e^{-\Lambda}/N! \; . \tag{4.2}$$

This way by definition, the expectation value made over many trials will be Λ.

Interpolation and assignation schemes

Given the knowledge of the local number of object to be positioned, in the end in order to create a catalogue we only need to specify an interpolation scheme. The most straightforward of them consists in assuming that the density is constant within a cell. This can be obtained by convolving the sampled density field with a Cartesian Top-Hat window function $W_{\mathrm{TH}}(\vec{x})$ which provides $1/a^3$ within the cell and 0 outside. However, more sophisticated methods can be designed.

In general, one can see the interpolation scheme as choosing a convolution function $W(\vec{x})$ to be applied on the sampled density field. For this reason we adopt the same notations as introduced by Sefusatti et al. (2016) in the case of assignment scheme. Let me clarify these two notions.

- An *assignment scheme* is a recipe which consists in reconstructing a grid from a point-like distribution, for example a density grid from a catalogue of particles, galaxies, etc.
- An *interpolation scheme* is the other way around, it consists in populating with points a distribution from a grid, which is the approach we ought to follow.

A generic Cartesian window function can be expressed as $W(\vec{x}) = a^{-3}W(x/a)W(y/a)W(z/a)$, where W defines the interpolation or the assignment scheme. Basically the order of the window function represents the order of the interpolation scheme. As a matter of examples I show below the various possibilities for the window function. First,

$$W^{(1)}(s) = \left\{ \begin{array}{l} 1 \text{ for } |s| < \frac{1}{2} \\ 0 \text{ otherwise} \end{array} \right. , \tag{4.3}$$

is called Nearest Grid Point (NGP) and corresponds to a 0^{th} order interpolation (Top-Hat). Then,

$$W^{(2)}(s) = \left\{ \begin{array}{l} 1 - |s| \text{ for } |s| < \frac{1}{2} \\ 0 \text{ otherwise} \end{array} \right. , \tag{4.4}$$

is the Cloud In Cell (CIC), corresponding to a linear interpolation (order 1). In addition, two higher order smoothing functions can be reported, the Triangular Shaped Cloud (TSC)

$$W^{(3)}(s) = \left\{ \begin{array}{l} \frac{3}{4} - s^2 \text{ for } |s| < \frac{1}{2} \\ \frac{1}{2}\left(\frac{3}{2} - |s|\right)^2 \text{ for} \frac{1}{2} \leq |s| < \frac{3}{2} \\ 0 \text{ otherwise} \end{array} \right. , \tag{4.5}$$

and the Piece-wise Cubic Spline (PCS)

$$W^{(4)}(s) = \left\{ \begin{array}{l} \frac{1}{6}\left(4 - 6s^2 + 3|s|^3\right) \text{ for } 0 \leq |s| < 1 \\ \frac{1}{6}\left(2 - |s|\right)^3 \text{ for } 1 \leq |s| < 2 \\ 0 \text{ otherwise} \end{array} \right. . \tag{4.6}$$

These four window functions cases are represented in figure 4.1 as a function of the normalised position s with respect to the centre of the cell. The main difference between the various interpolation schemes is the nature of the resulting density field. When the order 0 ($W^{(1)}$) is chosen, we end up with a density field which is not continuous nearby regions that can be described by a very different value of the

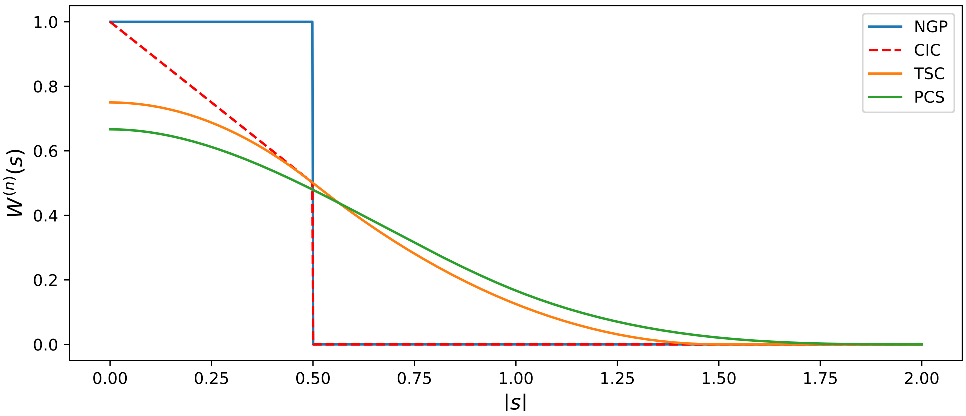


Figure 4.1: Illustration of the four described mass assignment schemes as a function of the normalised position s. $s = 0$ corresponds to the grid node and $|s| > 1/2$ is the area beyond the boundaries of the cell. The orders $n = 1, 2, 3, 4$ stand respectively for the NGP, CIC, TSC and PCS schemes.

density field. In turn, when we choose the linear interpolation of order 1 ($W^{(2)}$), the field cannot be differentiated at each grid point, but at least it is continuous.

In practice, choosing an interpolation scheme too high is not computationally efficient. Indeed, once an interpolation has been provided, it is necessary to generate a random distribution of points following the interpolated density field within each cell. This operation requires to invert the target cumulative distribution inside the cell which is straightforward if the PDF is linear, since its cumulative is a polynomial of order 2. In this case, the cumulative can be analytically inverted while in higher order cases this inversion becomes too demanding.

As a result, I did not try to go beyond linear interpolation, for which all details are described in appendix F. Asking for more computation than the straightforward Top-Hat one (see equations F.8), it turns out to be only two times more CPU-consuming. For this reason, unless specified, I will assume a linear interpolation scheme in the rest of this manuscript.

4.1.2 Effect on the power spectrum

Convolution effects

Since we saw that the choice of the interpolation scheme acts on the density field as a convolution, we can immediately guess that it will also affect the power spectrum, especially on scales comparable to the grid size ($k \sim k_N$). Indeed, the interpolated density contrast within the cell $\tilde{\delta}(\vec{x})$ is obtained by the convolution of the grid sampled density field with the convolution functions $W^{(n)}(\vec{x})$ described above. Thus, in Fourier space, it results in a product

$$\tilde{\delta}_{\vec{k}} = \delta_{\vec{k}} W^{(n)}_{\vec{k}}, \tag{4.7}$$

which allows to express the corresponding power spectrum as

$$\tilde{P}(\vec{k}) = P(\vec{k}) |W^{(n)}(\vec{k})|^2 \; . \tag{4.8}$$

Here, $P(\vec{k})$ is the power spectrum of the sampled density field, namely the aliased power spectrum (see eq. 3.13). This results in an expected power spectrum of the created catalogue as (see eq. 3.13)

$$\hat{P}(\vec{k}) = |W^{(n)}(\vec{k})|^2 \sum_{\vec{m}} P\left(|\vec{k} - 2\vec{m}k_N|\right) + \frac{1/(2\pi)^3}{\rho_0}, \tag{4.9}$$

where the additional constant term on the right corresponds to the shot noise contribution due to the auto-correlation of particles with themselves (see appendix G).

In the various interpolation cases, the window function takes the general form (Sefusatti et al., 2016)

$$W^{(n)}(\vec{k}) = \left[j_0(k_x a/2) j_0(k_y a/2) j_0(k_z a/2)\right]^n , \tag{4.10}$$

where j_0 stands for the spherical Bessel function of order 0 and the index n corresponds to the order of the interpolation scheme. An illustration of the four Fourier transforms of the smoothing functions is displayed in figure 4.2. It shows first that

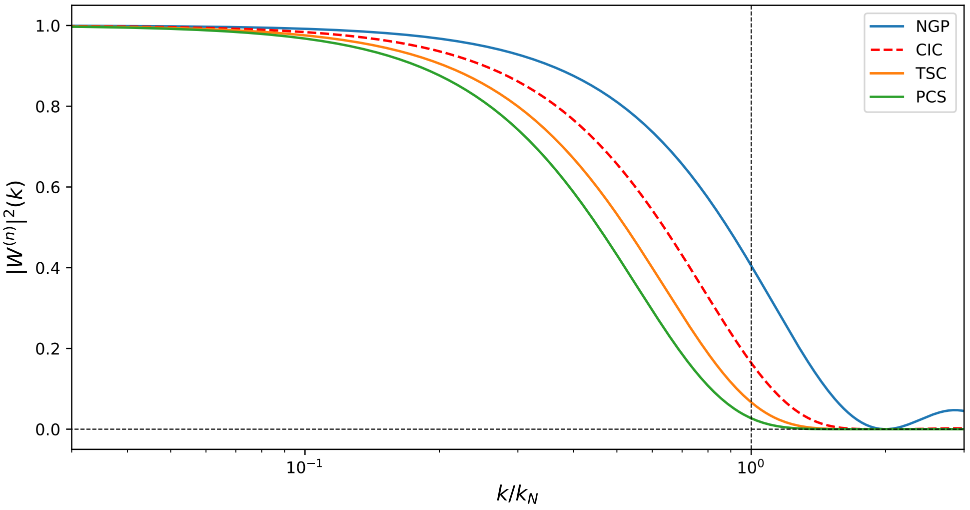


Figure 4.2: Illustration of the first four described smoothing functions as a function of the Fourier modes normalised by the Nyquist frequency $k_N = \pi/a$ with $a = L/N_s$, $L = 1000h^{-1}$Mpc and $N_s = 1024$. The orders $n = 1, 2, 3, 4$ stand respectively for the analogous NGP, CIC, TSC and PCS assignment schemes.

the Fourier transform of the chosen convolution function W is cutting the power on small scales, which is equivalent to smooth the density field on the size of the cells. In addition, increasing the interpolation order implies an increasing dropping power rate as well. Thus the higher is the filtering, the more efficient is the cut-off.

Application on Log-Normal fields

We generate hundred Poissonian Log-Normal fields described by a ΛCDM linear matter power spectrum at $z = 0$, composed of 2×10^7 particles in boxes of size $L = 1200h^{-1}$Mpc and using the grid sampling parameter $N_s = 512$. We estimate the power spectrum of the catalogues with the method described by Sefusatti et al. (2016) employing a particle assignment scheme of order four (PCS, see 4.6) and the interlacing technique to reduce aliasing effects. Notice that these choices are intrinsic to the way we estimate the power spectrum of the distribution of the generated objects and has nothing to do with the way we generate the catalogues.

In figure 4.3, we compare the estimated power spectra for the two interpolation schemes described above (Top-Hat and linear ones). As expected, on small scales, the tri-linear interpolation scheme provides a better fit to the targeted power spectrum that the Top-Hat one thanks to the improvement of continuity at the grid scale. Intuitively, on can expected that the Top-Hat scheme artificially add correlations on these scales, given more amplitude to the power spectrum. This effect is damped with the tri-linear one.

In the same figure are also plotted the expected power spectra computed with eq. 4.9, and corresponding to the two mentioned interpolation schemes. We can see that in both cases we are able to predict with a good accuracy (better than the percent level) the generated power spectra below the Nyquist frequency of the

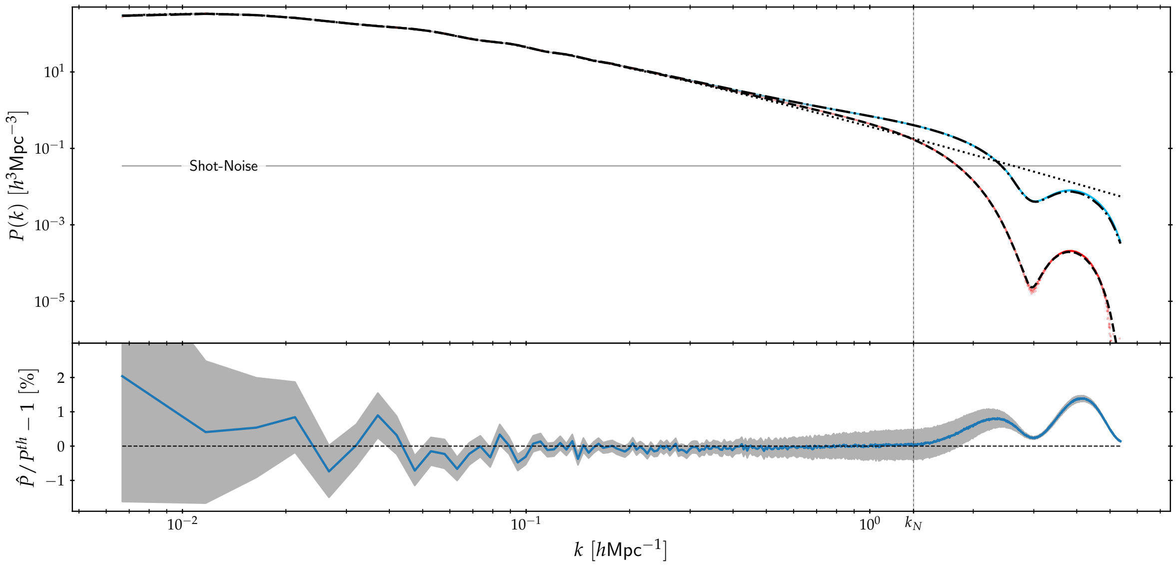


Figure 4.3: *Top:* Measured power spectra averaged over 100 realisations of the Poissonian Log Normal field for the Top-Hat interpolation scheme (blue curve with prediction in dash-dotted black line) and for the linear interpolation scheme (red curve and prediction in dashed line). The shot noise (solid horizontal black line) has been subtracted and its level is about 3.48×10^{-2} $h^3\text{Mpc}^3$. The dotted black line represents the raw theoretical power spectrum computed by `CLASS`. *Bottom:* Relative deviation in percentage between the estimated power spectrum (with shot noise contribution) and the expected prediction for the Top-Hat interpolation scheme. Snapshots are computed for a grid of size $L = 1200$ h^{-1}Mpc and parameter $N_s = 512$. Here comparisons are made well beyond the Nyquist frequency of the created catalogues (vertical line) at $k_N \sim 1.34$ $h\text{Mpc}^{-1}$. The cosmological scenario here is the ΛCDM one at redshift $z = 0$.

initial density field. In addition, we can also explain the peculiar shape of the power spectrum below the grid size when taking into account the interpolation scheme induced by the Poisson process.

In summary, I have presented how to convert a density field into a catalogue of objects. It requires to choose an interpolation scheme which will affect the power spectrum on small scales. Finally, I have shown that despite the fact that each interpolation scheme modifies the power spectrum amplitude in a different way, we can predict it very accurately.

Moreover in the following, a linear interpolation scheme will be exclusively used since, in fact, it ensures the continuity of the density field. Also, the simulated power spectra describe a better matching with the true power spectrum, at least up to the Nyquist frequency.

4.2 Method validation in comoving space

Being able to generate some catalogues of objects for a fixed redshift with controlled PDF and power spectrum, it would be opportune to compare it with a method of catalogue production that can be qualified as a reference concerning the covariance matrix estimation. In particular N-body simulations, as already discussed in chapter 1, can be used for such a reference.

Presenting numerous adapted setting concerning the Monte Carlo – N-body comparison, we chose the `DEMNUni_cov` set of catalogues (introduced in chapter 2) for the whole validation of the Monte Carlo procedure, whether it is in comoving space in the present section, in redshift-space in section 4.5 or for cosmological parameter inference in harmonic space in chapter 5.

Thanks to its large volume ($L^3 = 1000^3 h^{-3}\mathrm{Mpc}^3$) and its thin spatial resolution (1024^3 simulated particles), it will allow a range of study from large scales down to mildly non-linear scales, comparable to future galaxy surveys.

The purpose of this part is therefore to apply (and sometimes adapt) the Monte Carlo method in order to produce clones of the `DEMNUni_cov` simulations, with some specific identical settings, essential for a regular comparison.

4.2.1 Cloning the `DEMNUni_cov` snapshots

Settings

As in the Log-Normal case, the first step consists in choosing the settings and identifying the targets of the Monte Carlo method. Cloning the `DEMNUni_cov` simulations being the whole point, we will first set our simulations on the same box size $L = 1000h^{-1}\mathrm{Mpc}$ in order to reach the same fundamental frequency $k_F = 2\pi/L$.

Moreover, in the interest of chapter 5 where cosmological parameter inference will be addressed, we will focus exclusively on the standard cosmology including massive neutrinos, that I will call "16nu". The clones will be generated at the same comoving outputs, namely redshifts $z = [0, 0.5, 1, 1.5, 2]$[1], for which the statistical targets will be the power spectra as measured and averaged over the 50 `DEMNUni_cov` (see figure 2.4) realisations and the estimated PDF over one simulation, for instance the `DEMNUni_cov01` PDF[2] shown in figure 2.5.

The grid will be set to a sampling parameter $N_s = 1024$, ensuring both a time efficiency and a pretty high spatial resolution. Finally, as fixed in chapter 3, the aliasing parameter will take the value $j = 2$.

In the following, I explain in detail some key points which need to be addressed in order to make sure that that the Monte Carlo is able to mimic both the power spectrum and the PDF.

Reproducing the PDF

In the case of a general PDF, whose the shape is not parametric, the process needs to be adapted compared to the Log-Normal case described before. Indeed, the targeted power spectrum put a strong constraint on the variance of the density field σ_δ^2. Thus, when the PDF is analytical (*e.g.* eq. 3.18), it is straightforward to find the parameters shaping the PDF which allows to match the variance.

---

[1] Actually this is a short-cut, since the true redshift at which the Monte Carlo catalogues are simulated are the same as for `DEMNUni_cov` and are given in chapter 2.

[2] The high signal-to-noise and the high level of PDF reproductibility from one N-body simulation to an other (at the same redshift) allows to pick only one PDF of reference.

In the contrary, when the PDF is obtained from a measurement, the matching of the variance coming from the PDF and the variance of the target power spectrum is not immediately possible. If this condition is not met, the resulting Gaussian variance won't have a unitary variance and the whole process would go wrong. In fact, if the variance of the Gaussian field is not the expected one, then the local transform won't convert it into the target PDF and the power spectrum will also be different from the expected one. This issue arises because unlike the power spectrum that can be deconvolved from the smoothing (or mass assignment scheme) used to estimate it, no deconvolution method exists for the PDF estimator.

A possible way to deal with this, is to enforce the variance coming from the power spectrum to match the one of the PDF. The idea is to filter out short wave lengths in order to reduce the total power. This implies to set a filtering shape w.

In doing so, an illustration of the application of $\mathcal{L}$ on a Gaussian field consisting in 10^9 particles is shown in figure 4.4 and admits a perfect reproducibility of the δ-PDF on the whole field interval $\delta \in [-1, 3052]$ of `DEMNUni_cov01`.

Adapting the targeted power spectrum

In this part, I review more in detail how the choice of the filtering mentioned above is affecting the Monte Carlo process.

As explained before, the variance estimated on the two targets (PDF and power spectrum) must be the same. The target power spectrum (which is by definition deconvolved : $P^{\rm dec}(k)$) must be filtered in such a way that its variance[3] matches the one given by the PDF

$$\left\langle \delta^2 \right\rangle_{\rm PDF} \equiv 4\pi \int \mathrm{d}k k^2 P^{\rm dec}(k) F(k) \ . \tag{4.11}$$

This filtering can be performed using any type of low-pass filter, for example

$$F(k) = \exp\left[-(kR)^i\right] \ , \tag{4.12}$$

where R is the filtering radius, and i the *filtering shape parameter*. Note that for $i = 2$, eq. 4.12 reduces to a Gaussian filtering. As illustrated in figure 4.5, the higher is i, the sharper will be the filtering.

It would be appealing to chose $i \gg 1$ in order to keep untouched the power spectrum up to a maximal mode, but as it will be shown further, an other effect occurs within such limit.

In order to find out the power spectrum of the Gaussian field P_ν, the local application $\mathcal{L}$ can be decomposed on the basis of Hermite polynomials where the corresponding coefficients c_n are computed using eq. 3.11. Noticing that

$$c_0 = \int_{-1}^{\infty} \delta \mathcal{P}_\delta(\delta) \mathrm{d}\delta = \langle \delta \rangle \ , \tag{4.13}$$

we should precise that, as shown in the top panel of figure 4.6, the value of c_0 in the cloning procedure is not rigorously null. This is due to the binning effect of the

---

[3] The variance from the power spectrum can be computed using that $\sigma^2 = \xi(0) = \int \mathrm{d}^3\vec{k} P(\vec{k}) = 4\pi \int \mathrm{d}k k^2 P(k)$ for an isotropic field.

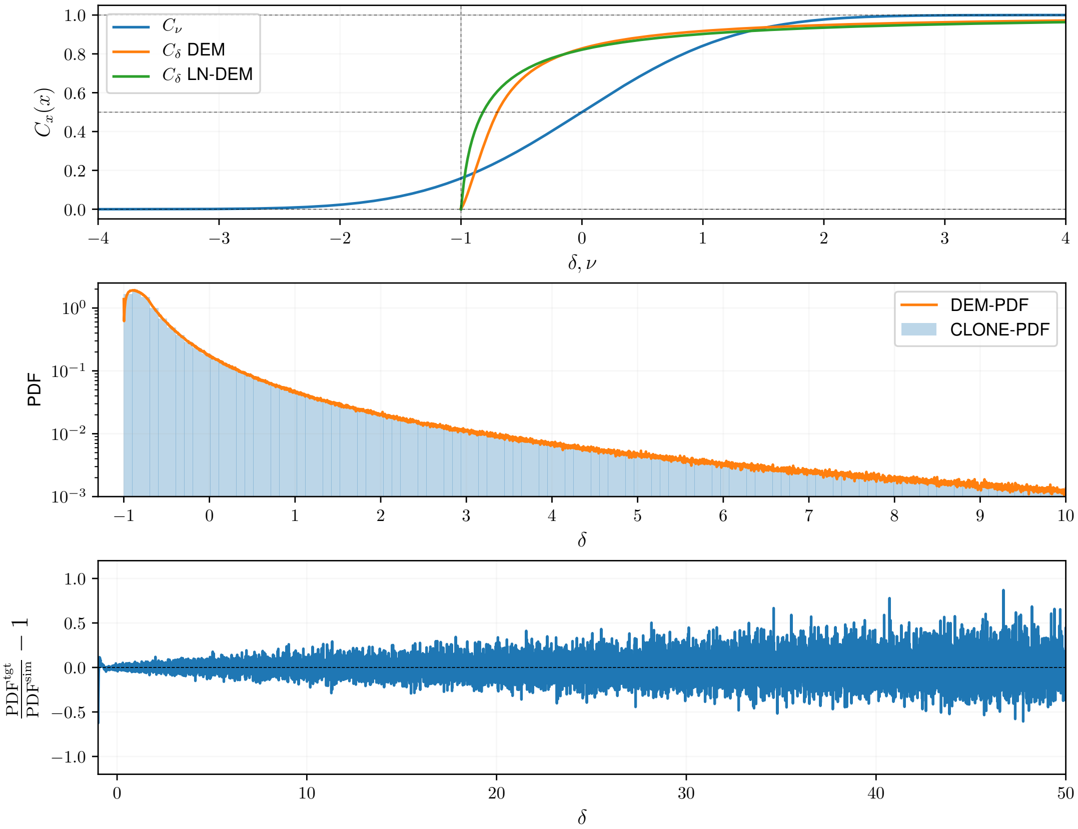


Figure 4.4: *top panel :* normalised cumulatives of the Gaussian (C_ν) and non-Gaussian (C_δ) distributed fields involved in the Monte Carlo procedure described in chapter 3. The `DEMNUni_cov01` grid-PDF at $z = 0$ in 16nu cosmology is used for the orange solid line and compared to the equivalent (same variance) Log-Normal PDF case in green solid line. *centre panel :* normalised histogram (in blue) of the non-Gaussian field δ after the numerical local mapping (eq. 3.5) on a Gaussian field ν consisting in 10^9 particles and targeting the `DEMNUni_cov01` PDF (in orange). *bottom panel :* relative deviation between the simulated ($\mathrm{PDF}^{\mathrm{sim}}$) and expected ($\mathrm{PDF}^{\mathrm{tgt}}$) δ distributions. All panels are presented in truncated intervals of the total one $\delta \in [-1, 3052]$.

PDF when computing the integral 4.13. However, as shown in figure 4.6 the value of c_0 remains small, the main effect of a non zero c_0 would be to change the total number of object in the catalogues.

The other coefficients c_n of the Hermite expansion allowing to match the probability density of the `DEMNUni_cov` are shown in figure 4.6. One can see that only the first five order are above 0.1 and the convergence is very fast, this is showing that the expansion can be safely stopped at order 20 ($\sim 10^{-12}$). As a result, we can compute the transformation $\lambda^{-1} : \xi_\delta \to \xi_\nu$ displayed in the bottom panel of figure 4.6. Note that $\xi_\delta(\vec{r} = \vec{0})$ must be mapped from $\xi_\nu(\vec{r} = \vec{0}) = 1$ (the variance of the Gaussian field is defined as unitary but of course not those of δ) and that, as expected, the relation crosses $\xi_\nu = \xi_\delta = 0$.

This being done, λ^{-1} can be applied on the inverse Fourier transform of the aliased targeted power spectrum in order to find out the input power spectrum P_ν

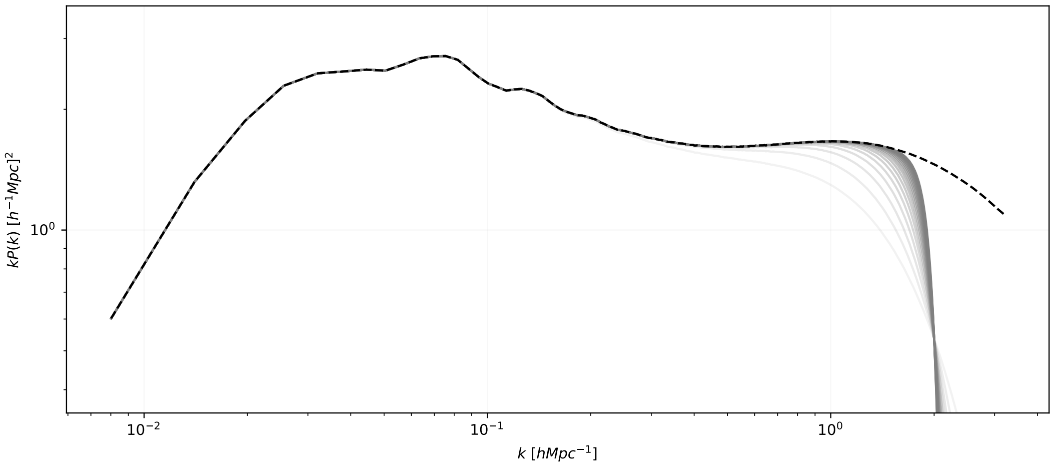


Figure 4.5: Different low-pass filters of type 4.12 applied on the averaged power spectrum of `DEMNUni_cov` (dashed black line). The increasing grey colour density means an increasing filtering power parameter $i \in \mathbb{N}$ within $i \in [2, 20]$. In this example the filtering radius is set to $R = 0.5h^{-1}$Mpc.

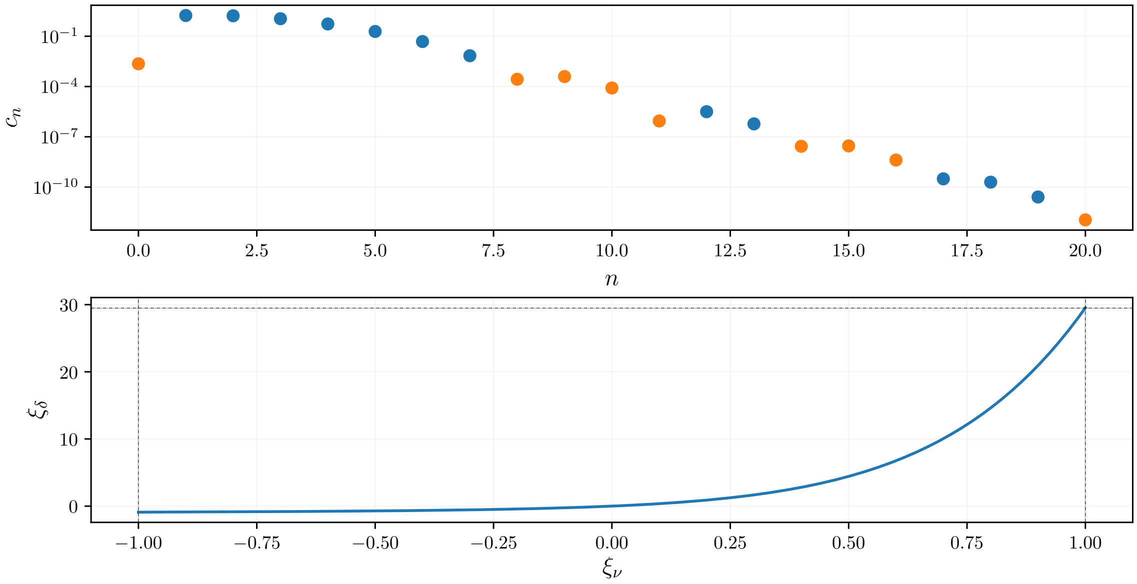


Figure 4.6: *top panel :* Hermite coefficients (eq. 3.11) of the local transformation mapping the Gaussian field ν to the field δ characterised by the `DEMNUni_cov` PDF. In blue dots are plotted the positive values while in orange are the absolute values of the negative ones. The horizontal dashed black line represents the mean value of the targeted PDF and expected to correspond to the c_0 coefficient. *bottom panel :* relation between the two 2-points correlation functions of the fields ν and δ following eq. 3.12. The horizontal dashed black line represents the value of the targeted PDF variance and expected to correspond to $\xi_\delta(r = 0)$.

of the centred reduced Gaussian field. We recall that the input power spectrum P_ν is computed on the same Fourier grid that will be used to generate the Gaussian field in Fourier space.

As explained before, in order to match the variance of the PDF, the power spectrum has to be filtered. Thus, it seems irrelevant to artificially alias the target power spectrum, since high modes are suppressed around the Nyquist frequency. Nevertheless, we keep the alias contributions ($j = 2$) in order to keep coherency between the various generated fields.

As already mentioned in section 3.2.2, the Monte Carlo procedure is not rigorously adapted to any shape of the power spectrum, particularly the one of the non-linear matter power spectrum, producing negative values in the 3-D power spectrum $P_\nu(\vec{k})$, even when the aliasing is taken into account. In our specific cloning case with a high resolution grid $N_s = 1024$, corresponding to a small filtering power at highly non-linear modes when compared to $N_s = 512$ or 256, about $\sim 60\%$ of negative elements are produced. They are mainly located in the Fourier volume above the Nyquist frequency (which represents nearly 50% of the total Fourier volume). The number of negatives is reduced when using $N_s = 512$ ($\sim 10\%$) and disappears for $N_s = 256$. But choosing such poor grid setting would mechanically reduces the range of useful simulated modes.

As a result, the variance $\sum_i P_\nu(\vec{k}_i)k_F^3$ estimated from the power spectrum of the Gaussian field is actually well unitary, as required by the procedure. However these negative elements prohibit the simulation of the Gaussian field $\nu_{\vec{k}}$. Indeed, the power spectrum represents the variance of the real and imaginary part of the Fourier modes thus a negative variance does not make any sense (see eq. B.15).

To address this restriction, two corrective methods can be suggested. The first consists in setting to zero all negative values. Let me call it the *clipping method*. The second takes their absolute value in order to counterbalance negative in the shell averaging and thus recover some lost power, called the *absolute method*. As a direct consequence, the variance computed on these new 3-D power spectra automatically deviates from unity with $\sigma_\nu > 1$. In order to recover the procedure condition $\sigma_\nu = 1$, an additional 3-D filtering (such as 4.12) can be applied on it to match the unit variance. Obviously, this filtering has the direct and undesired consequence of reducing the maximum mode for which the power spectrum is simulated. I will quantify such wave mode deterioration in the following.

Moreover even if high modes are filtered, the aliasing structure[4] is conserved during this process. Although this method is not ideal, it enables non-linear modes simulation to be still performed. Before determining up to which typical mode the method can reliably simulate a power spectrum, let met discuss about which corrective method is best suited. Considering a filtering shape parameter $i = 2$ in these two corrective methods, figure 4.7 shows all involved shell-average power spectra. In particular the prediction (using eq. 3.12) of the estimated grid power spectrum of the non-Gaussian field $P_\delta^{\mathrm{clip}}(k)$ in the clipping procedure and $P_\delta^{\mathrm{abs}}(k)$ in the absolute procedure.

Few differences can be noticed between the two methods, except that the clipping one seems to fit the true power spectrum up to slightly higher mode than the absolute one. For this reason, this method should be adopted instead of the other one[5].

---

[4] Here I am referring to anisotropic structure induced by relation 3.13 and a possible responsible for non-Gaussian wave modes distribution of the power spectrum discussed in section 3.2.2.

[5] Remark: a different choice was used in the rest of the manuscript, those of the absolute method.

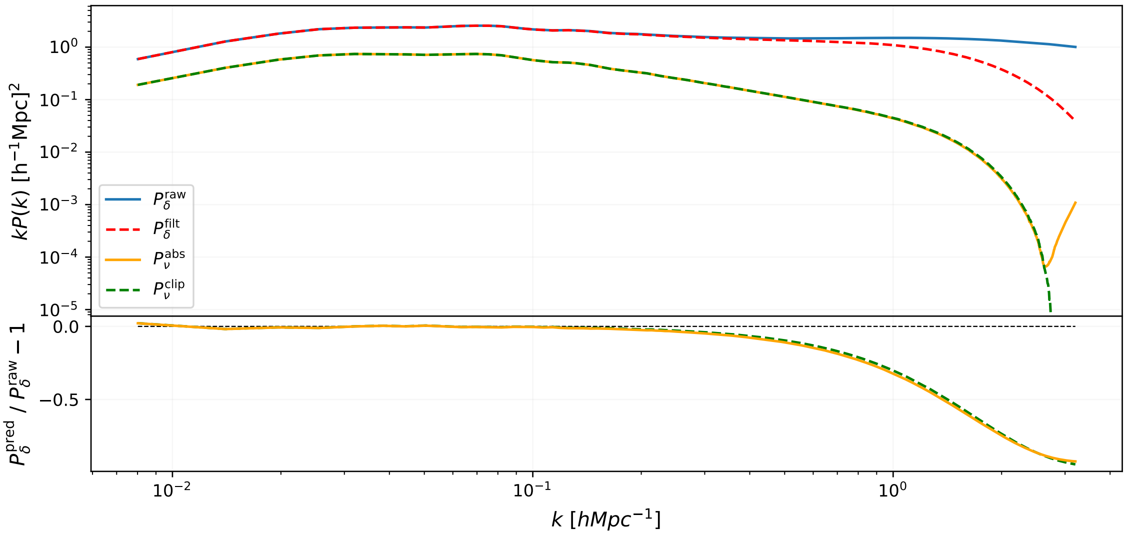


Figure 4.7: *top panel :* shell-average power spectra involved in the two types of corrective methods. The blue solid line is the raw power spectrum estimated on the 50 `DEMNUni_cov` simulations (alias-free, deconvolved and averaged) and its smoothed version in red dashed line, defining the targeted power spectrum. The orange solid line and green dashed line are the power spectra of the Gaussian field ν in the absolute and clipping method cases, respectively. *bottom panel :* relative deviation between the predictions of simulated power spectra using eq. 3.12 in the two corrective method, with the raw power spectrum. The same colour-code is used.

In brief, this adaptation of the Monte Carlo procedure has a cost. It involves the application of a filtering on a power spectrum already smoothed at the grid scale. It should not be forgotten that the field will be effectively filtered once again when discretising it, following eq. 4.9.

In order to avoid ending up simulating only the linear scales, it turns out to be fundamental to play with the various involved filtering shape parameters. First, instead of using a Gaussian filter on the deconvolved power spectra (those displayed in figure 2.4), I now relax the filtering shape parameters in eq. 4.12 and call it i_1. Also, let me call i_2 the free filtering power parameters in charge of the $P_\nu(\vec{k})$ filtering. In other words, an optimal combination on these two parameters i_1 and i_2 must be found in order to maximise the interval of matching modes. In this respect, figure 4.8 shows the deviation between the predicted $P_\delta(k)$ (after i_1 and i_2 application) and the true one by successive scanning of the two parameters; an exercise performed for the five redshifts in the 16nu cosmology (equivalent results are obtained in the ΛCDM case).

After inspecting figure 4.8, one can say that the optimal combination allowing to match a wider wave mode range is changing with redshift. It seems to be related to the overall amplitude level of the power spectrum (only depending on redshift), constituting another argument to pretend that the Monte Carlo method must be

---

This choice was motivated by reasons that are now obsolete, but will have no consequences. It is important to mention that the DC mode $P_\nu(\vec{k} = 0)$ that is not rigorously zero is nevertheless clipped, indeed this is imposing that the mean of the Gaussian field is null.

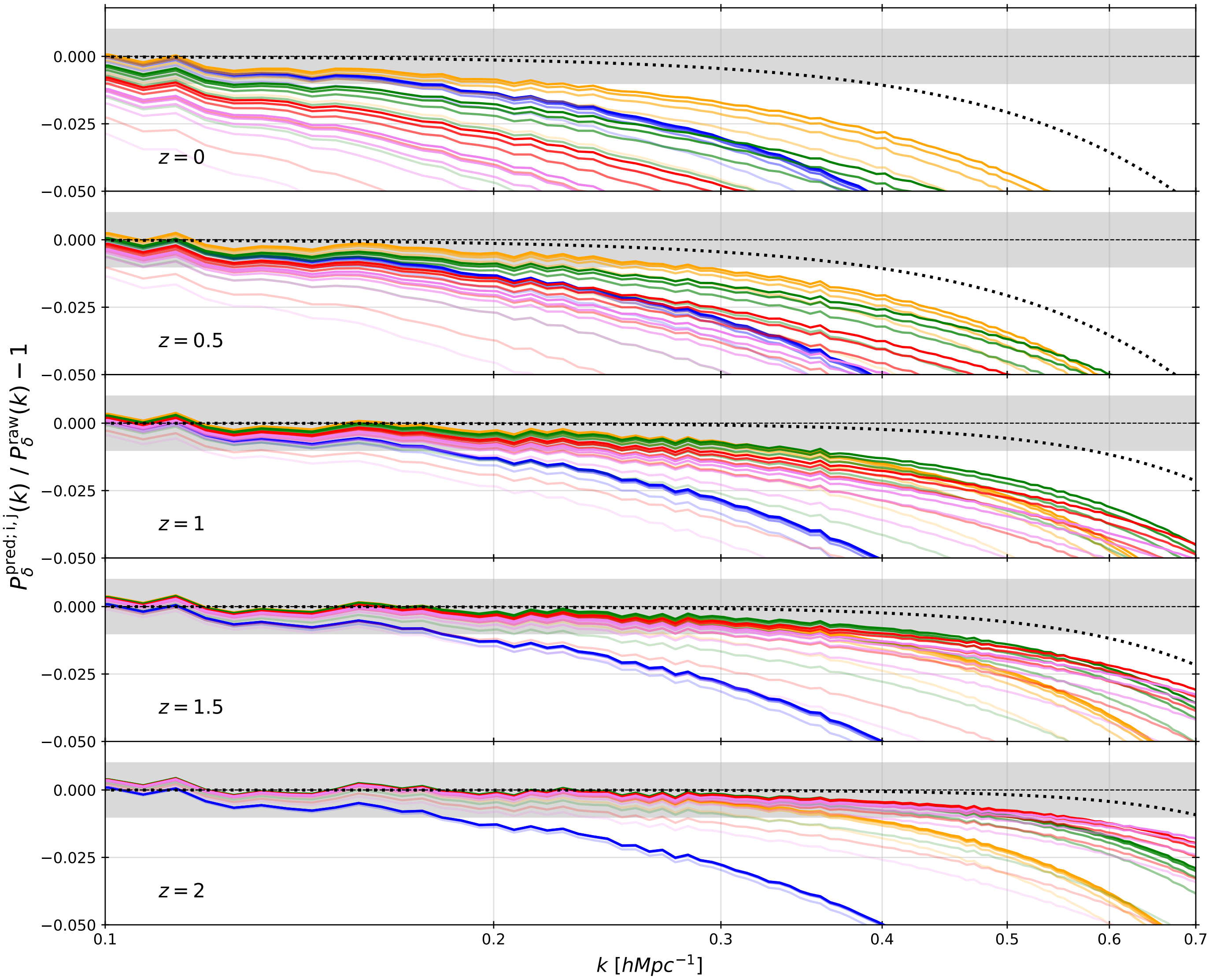


Figure 4.8: Relative deviation between the predicted $P_\delta(k)$ (after the two filterings) and the corresponding averaged power spectra estimated from the 50 `DEMNUni_cov` simulations at redshifts $z \simeq [0, 0.5, 1, 1.5, 2]$ in 16nu cosmology. The predicted spectra are computed for all combinations of parameters $i_1 = [2, 3, 4, 5, 6]$ and $i_2 = [2, 4, 6, 8, 10]$. Each colour represents a certain value of i_1, respectively blue, orange, green, red and violet. An increasing intensity of colour means an increasing i_2 value. The dotted black line is the relative deviation between the same raw power spectrum and the filtered one (before corrective methods) taking the optimal i_1, respectively 3, 3, 4, 4, and 5. The grey area shows the 1% limit while is displayed the truncated interval of interest $k \in [0.1, 0.7]h/\mathrm{Mpc}$.

adapted according to the target power spectrum. While the i_1 parameter cannot be pushed to the limit of sharp filtering $i_1 \gg 1$ without loosing significant range, it is a more predictable effect obtained for i_2. Indeed, for a given i_1, it seems more profitable to set $i_2 \gg 1$.

Moreover, in the same figure each dotted black line shows the predicted power spectrum that one would obtain in the case where no negative modes are produced, showing the impact of the corrective method. Thus, for the five successive redshifts, the optimal power filtering shape parameters will be chosen as $i_1 = [3, 3, 4, 4, 5]$ and i_2 to the arbitrary high number $i_2 = 18$, allowing to reach on the grid $k \sim 0.24h/\mathrm{Mpc}$ at $z = 0$ up to $k \sim 0.58h/\mathrm{Mpc}$ at $z = 2$ if the percent accuracy of the simulated power spectra is targeted.

In conclusion, the cloning procedure starts from a given targeted PDF, smoothed on a certain scale. For the procedure to be coherent, the power spectrum must be filtered (using i_1) in such a way that its variance matches the one of the PDF.

Because of the limits of the Monte Carlo method, the power spectrum of the Gaussian field is produced with negative values. I then proposed a corrective method that first clip to zero the negatives and filter (using i_2) the extra power to reach a unitary variance. A phenomenological analysis finally allowed me to find the optimal filtering power parameter combination to simulate fields up to mildly non-linear scales.

4.2.2 Comparative analysis

Power spectrum multipoles

All the ingredient are now put together to simulate clones (up to a certain maximal mode discussed in the following) of the `DEMNUni_cov` simulations. The goal of this part is to compare the simulated power spectra from the Monte Carlo to the ones estimated on `DEMNUni_cov`.

Using the tri-linear interpolation scheme (see section 4.1) when populating particles in voxels, we simulated 1000 catalogues/snapshots for the five successive redshifts in 16nu cosmology. Each of these catalogues is composed of $N_p = 10^8$ particles distributed in a box of volume of $(1000h^{-1}\mathrm{Mpc})^3$, corresponding to a particle number density of $\rho_0 = 0.1h^3\mathrm{Mpc}^{-3}$. About ~ 25 min of CPU-time on 50 processes of the Dark Energy Centre[6] are necessary to generate one catalogue.

We estimate the multipols of the power spectrum, namely the monopole $P^{(0)}(k)$, the quadrupole $P^{(2)}(k)$ and the hexadecapole $P^{(4)}(k)$ defined as

$$P^{(n)}(k) = \frac{2n+1}{2} \int_{-1}^{1} \mathrm{d}\mu_k P(k, \mu_k) \mathcal{P}_n(\mu_k) \ , \tag{4.14}$$

where the $\mathcal{P}_n(x)$ are the Legendre polynomials of order n, $\mu_k \equiv \cos(\theta_{\vec{k}})$ and $\theta_{\vec{k}}$ is the angle between the line-of-sight and $\vec{k}$. Here, the line-of-sight is chosen along the direction z of the comoving volume. The monopole is the usual shell-average power spectrum with which we have dealt up to now, while the quadrupole and

---

[6] The Dark Energy Centre (DEC) is composed of 1624 threads spread over 29 nodes, 28 of them being 2.4 GHz processors.

hexadecapole quantify the anisotropy of the 3-D power spectrum. Despite the fact that usually this kind of anisotropy is usually produced by redshift-space distortions (see next section), however aliasing can affect the isotropy of the power spectrum.

We compare the estimated multipols (after removing shot noise contributions) to the one estimated on the `DEMNUni_cov` in figure 4.9 and 4.10. As expected, it shows

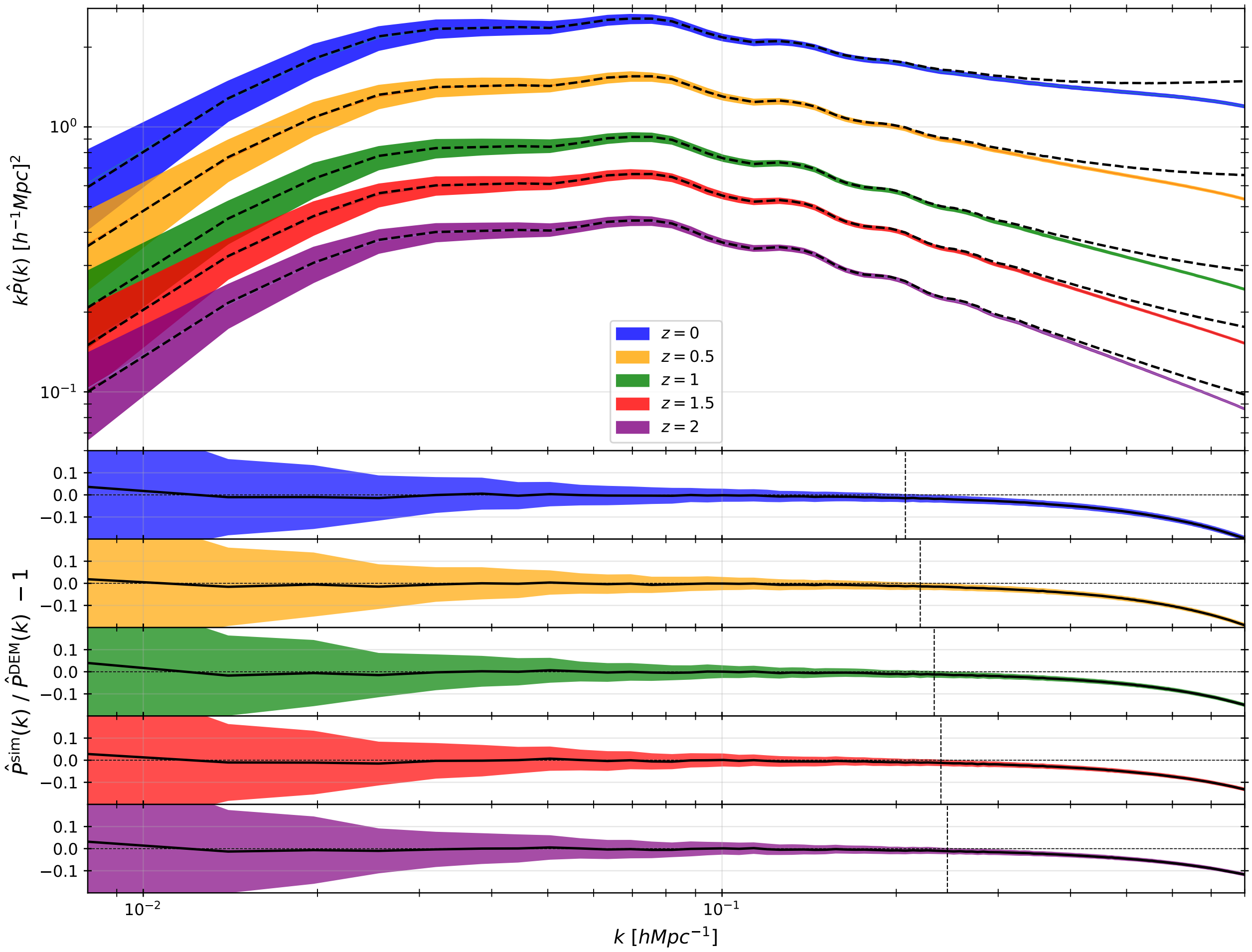


Figure 4.9: *panel 1:* estimated and averaged monopoles over 1000 realisations of Poisson catalogues in 16nu cosmology, each consisting in 10^8 particles with error bars represented in colour. Overlapping dashed black lines represents the averaged power spectra over 50 `DEMNUni_cov` simulations per redshift. The shot-noise contribution has been removed from all spectra. *panel 2,3,4,5,6 :* their relative deviation shown in black solid line with error bars following the same colour-code. The vertical dashed lines indicates the wave mode limit for which the respective error bars are crossing zero. These corresponding values for the successive redshifts are respectively $k_{\rm max} \sim [0.21, 0.22, 0.23, 0.24, 0.25] h/{\rm Mpc}$.

for the monopole a smaller range of reliable simulated wave modes when compared to its grid equivalent in figure 4.8. Indeed, as already discussed, the Poisson sampling gives rise to a smoothing of the cosmic field that visually starts to dominate around $k \sim 0.2 h{\rm Mpc}^{-1}$. This smoothing has for direct consequence to reduce the redshift dependence of the maximum well simulated Fourier mode (in the 1-σ limit) around

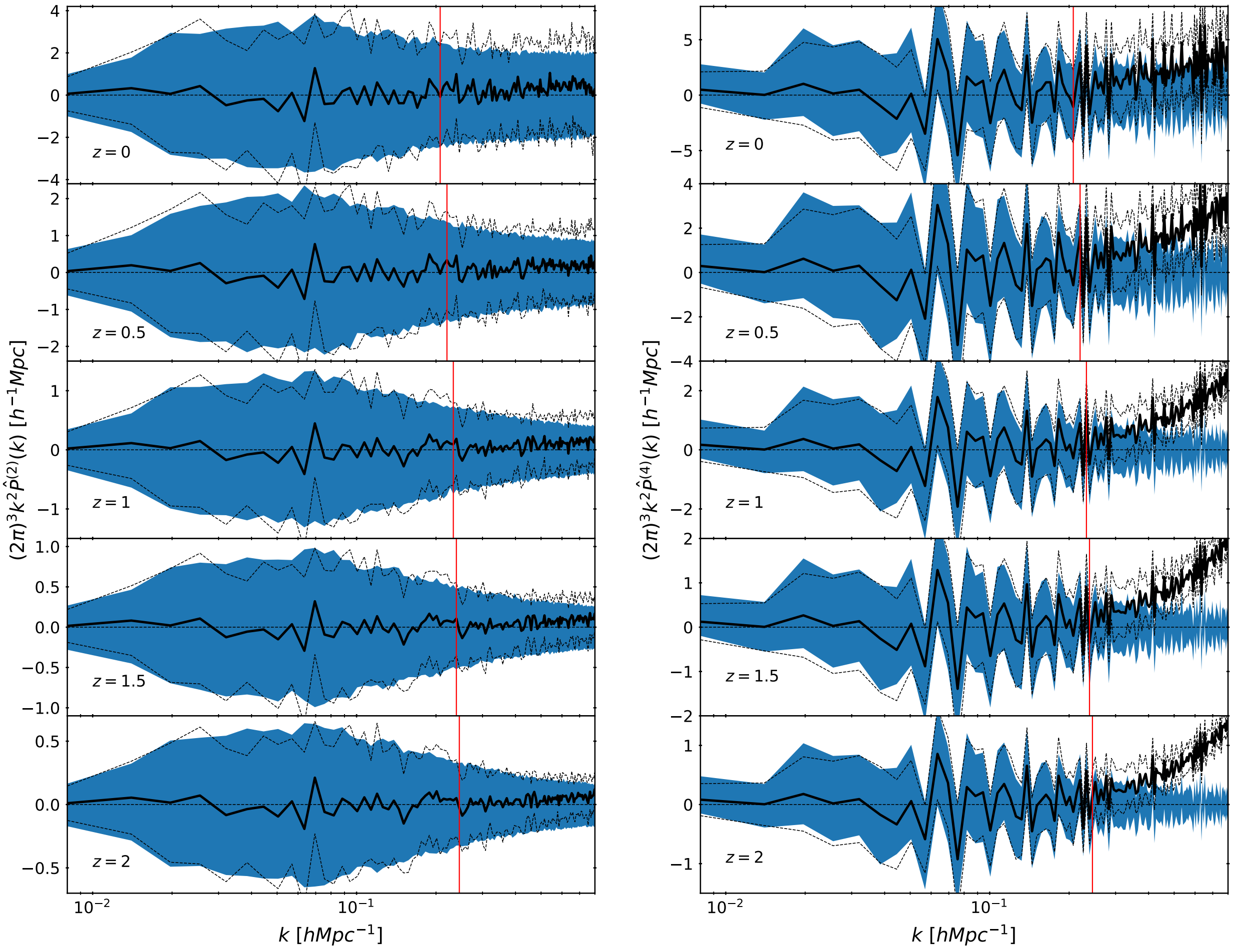


Figure 4.10: Estimated and averaged over 1000 simulations of the quadrupoles (left column) and hexadecapoles (right column) with error bars in blue for the five simulated redshifts as compared to the ones estimated on `DEMNUni_cov` in black solid lines with error bar in dashed black lines. The vertical references in red show for each redshift the maximal mode up to which the monopole is well reproduced in the $1-\sigma$ limit (see figure 2.4), *i.e* $k_{\max} \sim [0.21, 0.22, 0.23, 0.24, 0.25]\mathrm{hMpc}^{-1}$.

$k \sim 0.22h/\mathrm{Mpc}$, and being more constraining for high-z than for low-z. In practice, one could think of taking this drop of power at the level of the input Gaussian field but I leave this for future improvements.

Furthermore, regarding the quadrupole and the hexadecapole (as well as their dispersion), they match those estimated on the `DEMNUni_cov`, at least up to the maximum k for which the monopole is well reproduced. It can be noticed that the Monte Carlo method does not produce for the whole range of frequency any quadrupole or hexadecapole different from zero. This is showing that even the aliasing does not affect the isotropy of the generated density field. On the other hand in the `DEMNUni_cov` case, $P^{(2)}$ and $P^{(4)}$ starts to deviate from zeros respectively around $k \sim 0.6h/\mathrm{Mpc}$ and $k \sim 0.25h/\mathrm{Mpc}$. In particular a clear non zero hexadecapole is measured and this effect is amplified when increasing the redshift. The origin of this anisotropy is not clear however one could think of an effect coming from the initial positions of the particles of the simulation. Indeed, when starting a N-body simulation, particles are placed on a grid and are slightly displaced to generate the initial power spectrum. As a result, it might happened that this initial grid distribution might still be imprinted across time. The fact that the effect increases with redshift is in agreement with this hypothesis. Finally, one could be surprised by the fact that the fluctuations of the hexadecapole at $k < 0.2h\mathrm{Mpc}^{-1}$ are the same in the `DEMNUni_cov` and in the Monte Carlo, this is due to the Fourier grid which is the same for the two (we tested to put the expected 3D power spectrum on the same grid and found that it also follows the same pattern).

Covariance matrix estimator for the Monte Carlo

The next step would be to compare the covariance matrix estimated (eq. 2.49) with the Monte Carlo method to the one obtained from the `DEMNUni_cov` simulations. However, in doing so without caution one would conclude that the covariance matrix obtained from the Monte Carlo is systematically higher. This is a spurious effect due to the fact that the $k = 0$ mode of the density field obtained from the Monte Carlo is not 0. Indeed, in the previous section I stressed that the $k = 0$ mode of the Gaussian power spectrum is likely to be negative while we impose it to be null in order to make sure that the Gaussian field has a zero mean (the whole process relies on this). This is producing an undesired non null $k = 0$ mode in the power spectrum of the density field which introduce a correlation between the volume averaged density $\bar{\delta}$ with the power spectrum at all k modes. Thus, the power spectrum at different wave mode being correlated with a hidden variable ($\bar{\delta}$), the covariance appears larger than expected. Let me explain how to quantify and correct from this effect.

A Monte Carlo realisation following my process will be characterised by a finite total number of object X, which will vary from a realisation to an other. In principle, the variance should be given by the Poisson distribution as $\sigma_X^2 = N_p^2$, where $N_p \equiv \langle X \rangle$ is the expectation value of the number of objects in the catalogues. Thus, the expectation value of the number of objects can be obtain by multiplying the volume L^3 by the expected number density ρ_0: $N_p = \rho_0 L^3$.

When computing the cross correlation $C_{XP_k} \equiv \left\langle (X - N_p)\left(P(\vec{k}) - \langle P(\vec{k})\rangle\right)\right\rangle$ between the number of object and the amplitude of the power spectrum at a given wave mode k, one can show that it is given by

$$C_{XP_k} = N_p k_F^3 \bar{B}(0, k). \tag{4.15}$$

In eq. 4.15, $\bar{B}(k_1, k_2)$ stands for the shell average bi-spectrum, a quantity in principle expected to be null when taken at any k_1 or k_2 being 0. This is given by the fact that the mean number density is supposed to be known, thus the mean density contrast is by definition zero and no power appears at any wave-modes $\vec{k} = \vec{0}$.

As anticipated, if a negative value appears at the $\vec{k} = \vec{0}$ mode, then it is enforced to be zero. It results that the $\vec{k} = \vec{0}$ mode of the density power spectrum is expected to be larger than 0. The first effect of this is to provide an extra variance to the Poisson one, such that $\sigma_X^2 = N_p^2(1 + k_F^3 P_\delta(\vec{k} = \vec{0}))$.

The second effect is to induce a non negligible shell average bi-spectrum $\bar{B}(0, k)$. Indeed, at leading order one can express $B(0, k) \simeq 2c_2c_1^2P^2(k)$ with respect to the Hermite coefficients c_1 and c_2. This is showing that an extra correlation is introduced between the total number of object (or the mean density) and the power spectrum. After having tried to predict precisely the correlation coefficient (using the Mehler expansion of the bi-spectrum) in order to remove it from the computed covariance, we found that the following empirical formula works well and allows to efficiently recover the correct the covariance

$$P_{\text{deb}}(k) = \left[\frac{N_p}{X}\right]^2 (\hat{P}(k) - P_{\text{SN}}), \tag{4.16}$$

where $P_{\text{SN}} = (1/(2\pi))/\rho_0$ is the shot noise contribution to the power spectrum. The overall correction is just taking into account that the global effect of having more or less particles is to resize the amplitude of the power spectrum. At the end of the day, the shot noise corrected power spectrum that will be used for covariance matrix estimation will therefore be $P_{\text{deb}}(k)$.

Covariance matrix comparison

Once the estimated power spectra unbiased using relation 4.16, the whole covariance matrix can be estimated over the 1000 simulations per redshifts, using its estimator eq. 2.49. In figure 4.11 are shown the diagonal of the covariance (the variance), both in the `DEMNUni_cov` and Monte Carlo cases. The figure shows that for the five redshifts, all terms are matching, at least up to the maximal modes at which the monopole is well reproduced.

In addition, one can use the decomposition in Gaussian and non-Gaussian contributions (see eq. 2.46) to compare directly the tri-spectrum contribution between the Monte Carlo and the `DEMNUni_cov`. Indeed, the Gaussian contribution can be computed from the average power spectrum over the realisations (see figure 4.9 and figure 2.4), whereas the tri-spectrum contribution can be obtained subtracting the Gaussian contribution from the total variance. In figure 4.11, it can be observed that in the linear regime ($k < 0.1$h/Mpc), the amplitude of the trispectrum term is about one order of magnitude lower than the Gaussian variance. At the red vertical limit of $z = 0$, the trispectrum contributes about half of the total variance and starts dominating for $k \gtrsim 0.3h$/Mpc. Finally, one can see that the tri-spectrum of the same shell ($\bar{T}(k_i, k_i)$) is well reproduced by the Monte Carlo even at higher wave modes than the maximum wave at which the monopole is well reproduced.

The comparisons can be extended to the off-diagonal elements of the covariance matrix where solely the non-Gaussian contribution coming from the tri-spectrum $\bar{T}(k_i, k_j)$ is expected to have an influence. When respectively plotting terms up to $k^{\text{max}} \sim 0.3h$/Mpc and $k^{\text{max}} \sim 0.8h$/Mpc, figures 4.12 and 4.13 are featuring two types of comparison between the N-body and the Monte Carlo methods. The left panels shows the estimated covariances terms for the two experiments, re-normalised by the Gaussian error eq. 3.22 of the covariance elements of `DEMNUni_cov`. This quantity is an indicator of the compatibility level of the covariance with zero, *i.e.* where the underlying field is incompatible with a Gaussian description.

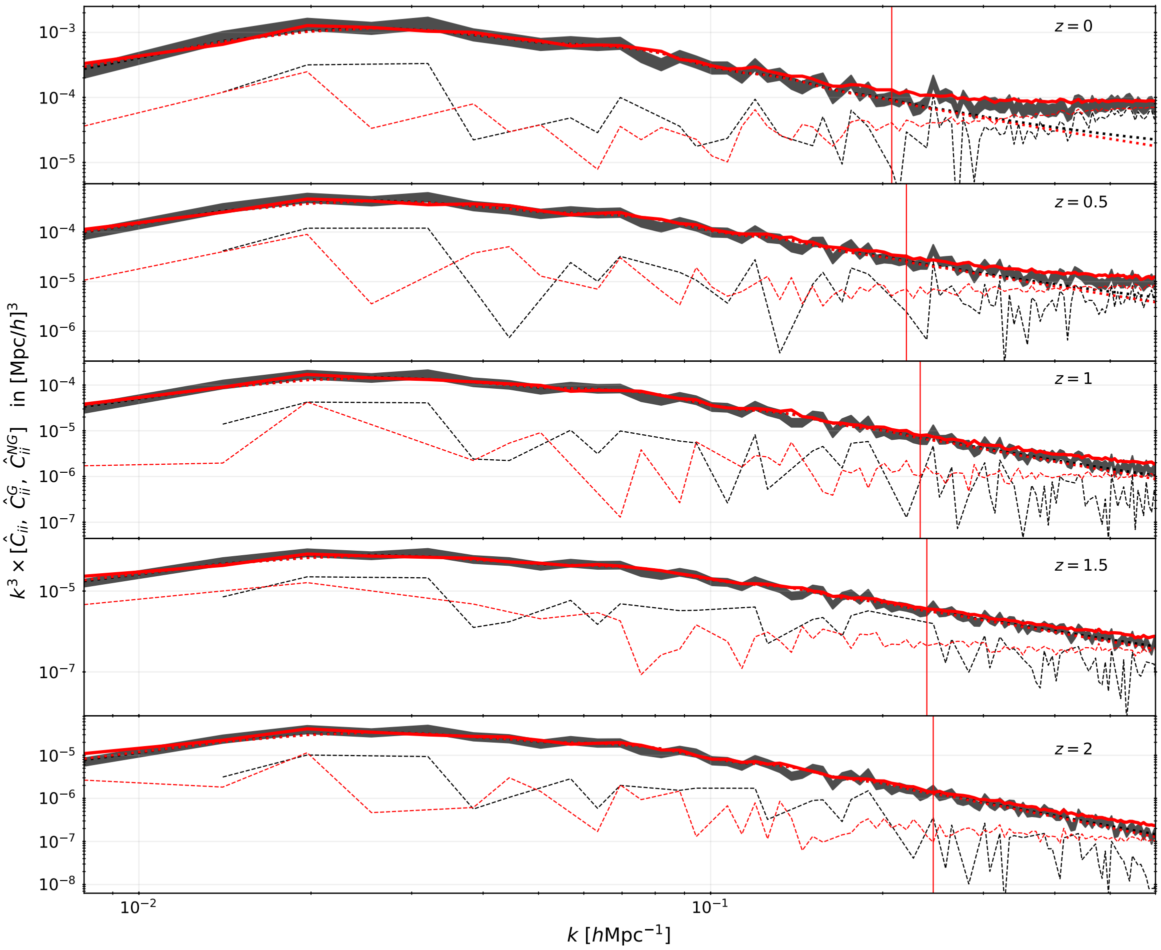


Figure 4.11: Estimated contributions of the diagonal of the covariance matrix from 1000 realisations of Monte Carlo catalogues in configuration space as compared to the ones from `DEMNUni_cov` and for the five redshifts. The estimated diagonal $\hat{C}_{ii}$ accounting for all contributions is represented in solid red line for the Monte Carlo and compared to those of the N-body in black with Gaussian error bars $2C_{ii}^2/(N-1)$ (see eq. 3.22) where $N = 50$. Using the same colour-code, in dotted lines are displayed the estimated Gaussian contribution $\hat{C}_{ii}^{\mathrm{G}}$ (first term of 2.46) estimated from the average power spectra. The estimated trispectrum contributions $\hat{C}_{ii}^{\mathrm{NG}}$ are drawn in dashed lines and are coming from the residual between the total and the Gaussian variance (only their positive values are represented). The vertical references in red shows, for each redshift, the maximal mode up to which the monopole is well reproduced in the $1-\sigma$ limit (see figure 4.9).

Note that the Gaussian errors are computed using $N = 50$ for both experiment, even if the Monte Carlo covariance terms are estimated using 1000 realisations. This choice allows to compare the amplitude of each element while appreciating the relative dispersion due to the difference of statistics. We see that the Monte Carlo method is able to reproduce the same trend observed in the `DEMNUni_cov`. In figure 4.13, I show that the same result hold even if we consider a much higher scale range

($k < 0.8h\mathrm{Mpc}^{-1}$).

The right panel of both figure 4.12 and 4.13 however represents the residual between the covariance matrices estimated from the 1000 Monte Carlo and the 50 `DEMNUni_cov` divided by the error on this difference. We see that even on the extended scale range (figure 4.13), the deviations between the two are not statistically significant. Thus allowing to say that with only 50 realisations of the `DEMNUni_cov`, we cannot spot an incompatibility of the Monte Carlo at all the tested redshifts between 0 and 2.

In figure 4.14, one can also appreciate the visual comparison between the correlation matrices defined as

$$\mathrm{r}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \,. \tag{4.17}$$

It first shows that mode mixing produces essentially positive correlations, migrating from high modes down to low modes for decreasing redshift, progressively shrinking the range of the linear regime. Such behaviour can be related to the crossing migration between the Gaussian and non-Gaussian variance terms discussed in figure 4.11. Note finally that the apparent extra correlation generated by the Monte Carlo method in the lower left panel (at $z = 0$) are those quantified in the top right panel of figure 4.12. However, as shown before, this is not a significant effect.

In turn, we can see clearly the smoothness of the correlation matrix coming from the Monte Carlo with respect to the one coming from the `DEMNUni_cov`. Of course, this is only due to the fact that we have 50 `DEMNUni_cov` while we generated 1000 Monte Carlo realisation. This is feasible, because it took nearly the same time and CPU resources to generate the 1000 realisations and to run a single N-body simulation.

In conclusion, we have checked statistical quantities related to one-point distribution and particle pairs (but also four-points) in Fourier space, ensuring a successful correspondence between the Monte Carlo procedure to generate snapshots of particles and those obtained from N-body simulations. Note however an extra correlation of order $1-\sigma$ produced at low redshift for modes $k \gtrsim 0.2h/\mathrm{Mpc}$. This can be due to the fact that the power spectrum is filtered (using i_1) in a different way of the smoothing scale at which the PDF is estimated. Moreover, the PDF is estimated on a grid built from a PCS assignment scheme (equivalent to a Gaussian window at first order) while the adopted interpolation scheme is a linear window. Of course more work is needed to confirm that there is indeed a discrepancy (more N-body simulation need to be ran).

It is worth noticing the fact that, in the present comoving space case, it would have been easier to estimate the covariance matrix directly on the density field before Poisson sampling. This way we would have enlarged the range of well simulated modes without shot noise, all of this in much shorter computation times. Indeed as high frequency are cut, aliasing reconstruction 3.13 is not in first approximation distinguishable from a simple 3D $\vec{k}$-grid interpolation of a 1D targeted power spectrum. In other words, apart from its anisotropical structure, aliasing is of minimal interest.

However, since in the next section I will detail how to implement in the Monte Carlo the velocity field in order to reproduce redshift-space distortions of the ap-

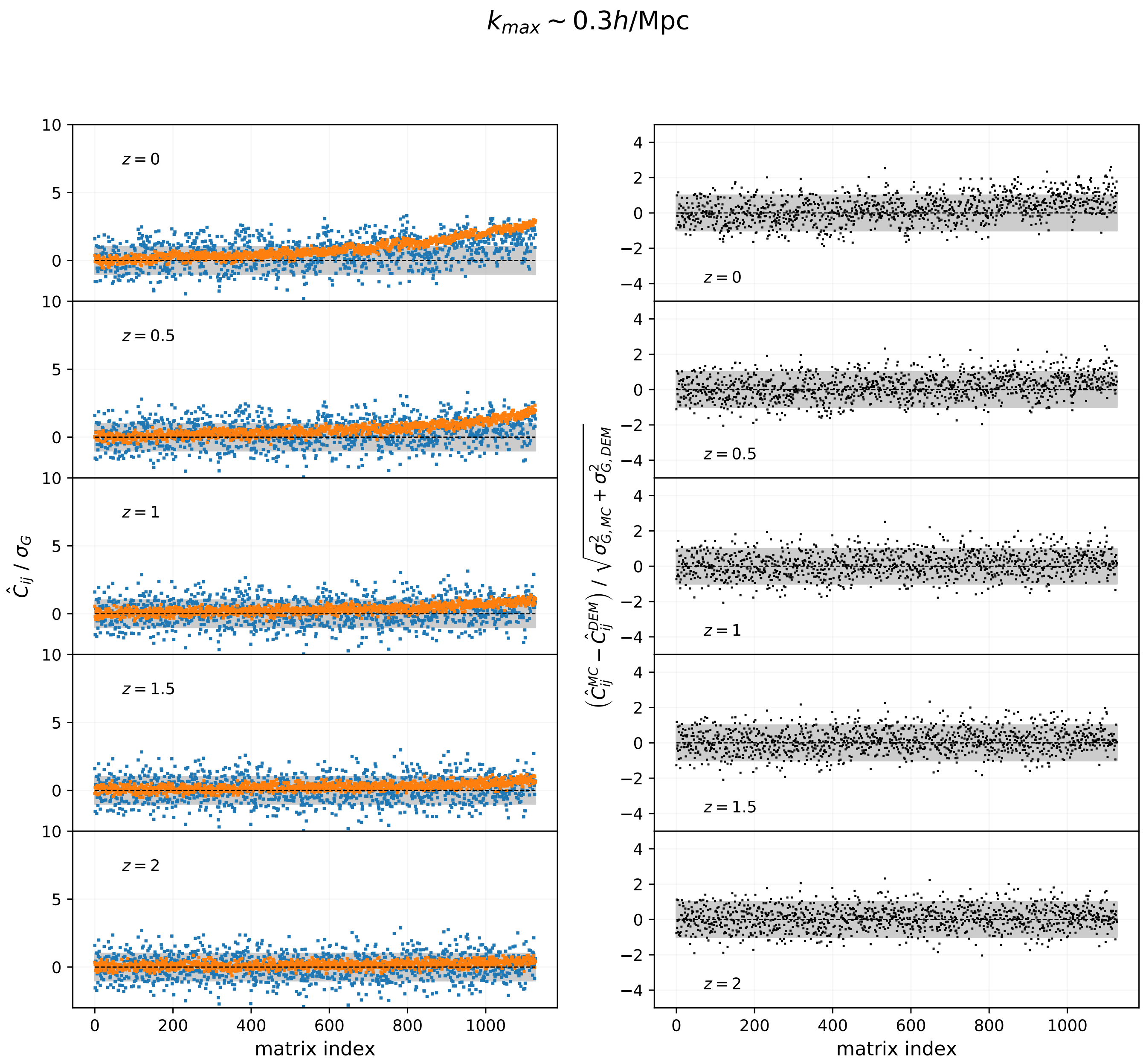


Figure 4.12: *left panels:* estimated off-diagonal elements of the covariance matrix for the `DEMNUni_cov` (blue dots) and the Monte Carlo (orange dots), normalised by their respective Gaussian error estimated using the averaged values of the simulated power spectra for the five redshifts. The elements are ordered column by column of the lower half of the (48×48) matrices by referring to the matrix index in order to display covariances up to $k \sim 0.3h/\mathrm{Mpc}$. Note that the diagonal elements are not included. right panels residual of the two matrices using the same ordering and normalised by the combined Gaussian errors (black dots). Note that Gaussian errors are computed using $N = 50$ for both experiments even if the Monte Carlo covariance is estimated over 1000 realisations. In all panels, grey areas represent the $1 - \sigma$ limit.

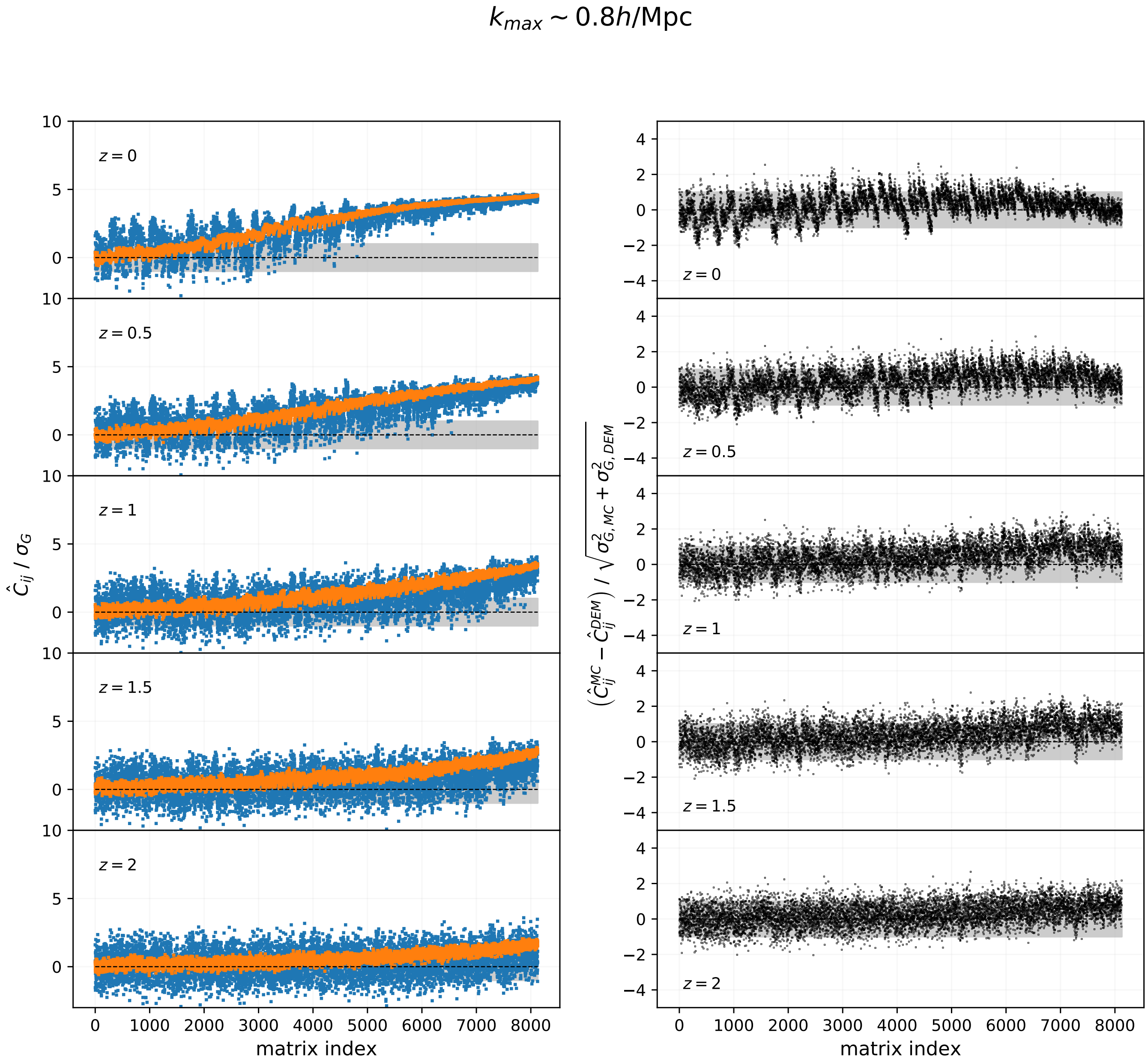


Figure 4.13: *left panels:* estimated off-diagonal elements of the covariance matrix for the `DEMNUni_cov` (blue dots) and the Monte Carlo (orange dots), normalised by their respective Gaussian error estimated using the averaged values of the simulated power spectra for the five redshifts. The elements are ordered column by column of the lower half of the (128×128) matrices by referring to the matrix index in order to display covariances up to $k \sim 0.8h/\text{Mpc}$. Note that the diagonal elements are not included. right panels residual of the two matrices using the same ordering and normalised by the combined Gaussian errors (black dots). Note that Gaussian errors are computed using $N = 50$ for both experiments even if the Monte Carlo covariance is estimated over 1000 realisations. In all panels, grey areas represent the $1 - \sigma$ limit.

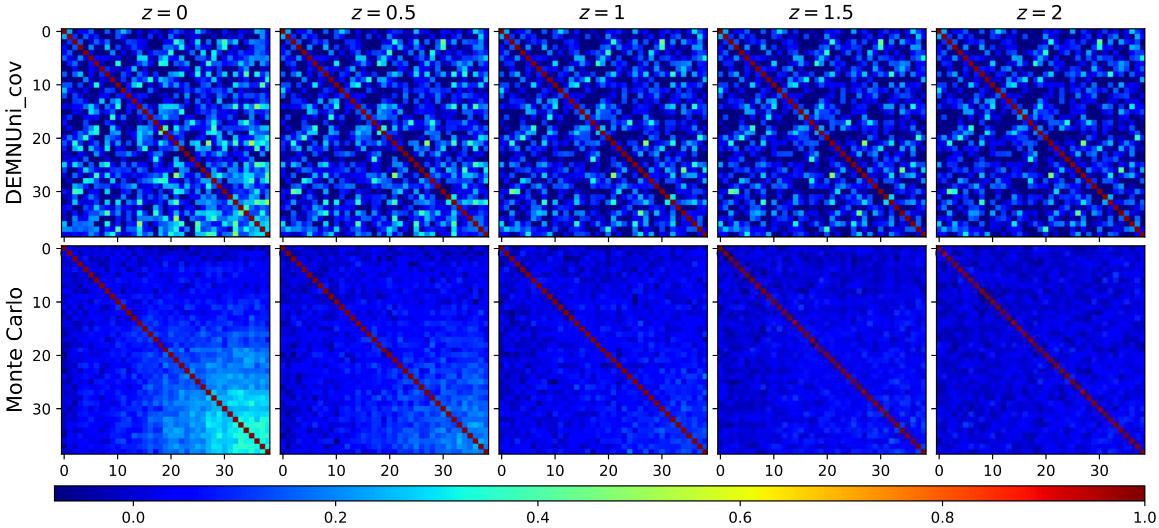


Figure 4.14: Estimated correlation matrices Corr_{ij} in the `DEMNUni_cov` (top panels) and Monte Carlo (bottom panel) cases up to $k \sim 0.25$ h/Mpc, (i.e (39×39) sub-matrices) for the five redshifts.

parent positions of objects, and given the fact that point-like distributions are more suited to apply redshift-space distortions, I preferred to stick to the same setting in comoving space, *i.e.* applying the Poisson sampling.

4.3 Redshift-Space Distortions

So far, catalogues of objects have been generated in comoving space, a coordinate system nevertheless not directly accessible trough spectroscopic galaxy surveys. In fact, each observed galaxy is described in an other coordinate system defined by a set of three numbers (z, θ, ϕ). The redshift z is the radial coordinate while θ and ϕ track the angular position on the sky. The measured redshift (or redshift-space) results not only from the Hubble flow mentioned in chapter 1 but also from the classical Doppler effect due to the projection of the peculiar velocity of an object on the line-of-sight. As a result, the redshift-space is polluted by peculiar velocities of galaxies with respect to the observer frame, resulting in a misinterpretation of distances. Thus, statistical quantities introduced in previous chapter are also affected by *Redshift-Space Distortion* (RSD).

4.3.1 Overview

In the same way as structure formation introduced in chapter 2, two distinct regimes can be described. In the linear regime ($\delta \ll 1$), if velocities were randomly distributed, no statistical effect could be expected (at least on two-points statistics, see Davis & Peebles, 1983) since the distance shift is small when compared to the considered large scales. However, on large scales the matter (or galaxy) velocities are rather coherent (not random). Indeed, small patches of matter are converging toward over-dense regions. On one side, this is changing the apparent amplitude of the clustering and on the other side, this is breaking its statistical isotropy. It turns out that on large scales and along the line-of-sight the clustering is enhanced, this is the so-called *Kaiser effect* (Kaiser, 1987).

In the non-linear regime where $\delta \gg 1$, over-dense regions are likely to be virialized, thus random motions are dominating and distribute positions in an uncorrelated way. This tends to dilute the clustering amplitude along the line of sight.

In both regimes, since the distribution of galaxies (or matter) is systematically perturbed, it follows that all observables constructed from the contrast density field are going to be affected as well. Thus, the angular averaged power spectrum is higher for low modes and appears suppressed for high modes with respect to its true amplitude. These features are summarised in figure 4.15, where the lower panel shows the ratio between the RSD monopole power spectrum and the real space power spectrum. We can see that on large scale ($k < 0.1$) this ratio converges to the expected Kaiser boost ($1 + 2/3f(z) + f^2(z)/5$, where f is the growth rate).

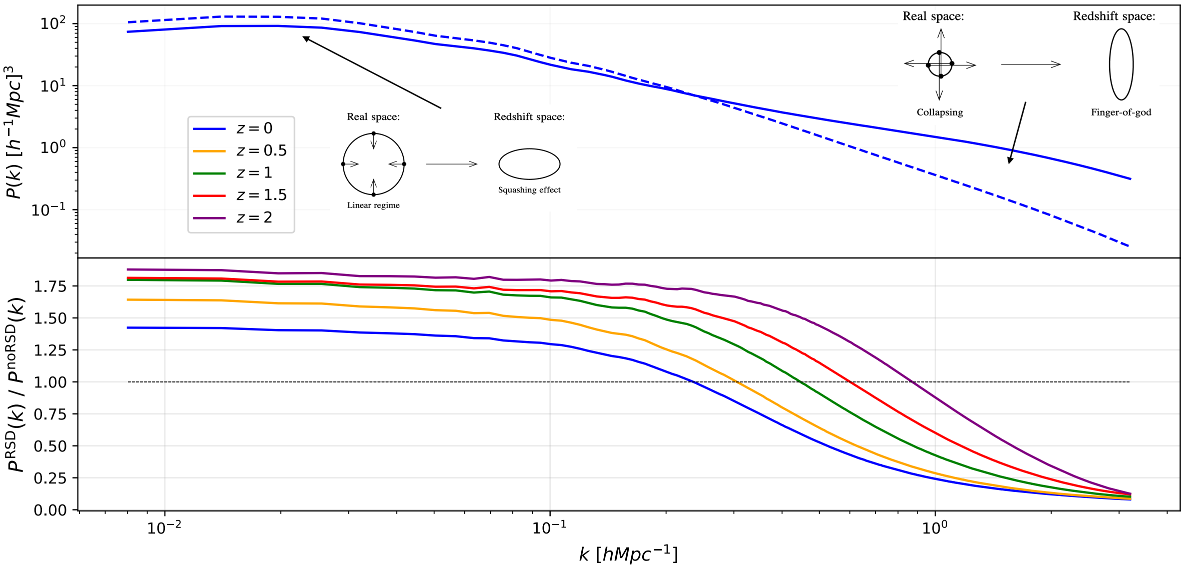


Figure 4.15: *upper panel:* Averaged over 50 `DEMNUni_cov` realisations of the power spectrum in comoving space (solid blue line) and in redshift-space (dashed blue line) at $z = 0$ in 16nu cosmology. *bottom panel:* Ratio between redshift and comoving spaces power spectra for the five studied redshifts. Sketch illustrating distortion of volumes from comoving to redshift-space are picked from Hamilton (1997): in linear regime, apparent volumes are squashed along the line-of-sight. In this case structures appear smaller and lead to interpret the clustering as enhanced. In the non-linear regime, volume are elongated along the line-of-sight, called the Finger-of-god effect. In this regime, clustering appears under-estimated resulting in a power suppression.

In the following I will present models in different regimes that quantifies RSD effects on the power spectrum.

4.3.2 Models in the linear regime

In order to introduce the original linear model for RSD, called the Kaiser model (Kaiser, 1987), it is necessary to explicit the coordinate transformation between real and redshift-space.

Be $\vec{v}$ the peculiar velocity of a particles/galaxies with respect to the rest frame, in comoving space (real space). It induces a non negligible supplementary spectral shift Δz (due to Doppler effect) on the observed redshift z_o following

$$z_o = z_c + \Delta z \ , \tag{4.18}$$

where z_c is the redshift resulting from the cosmic expansion. Thus, in redshift-space, the observed distance s of the object will be distorted with respect to its true comoving position r. Indeed, combining eq. 1.15 with eq. 4.18 one can express the apparent distances as

$$s = \frac{c}{H_0} \int_0^{z_o} \frac{\mathrm{d}z'}{E(z')} = \frac{c}{H_0} \left[\int_0^{z_c} \frac{\mathrm{d}z'}{E(z')} + \int_{z_c}^{z_o} \frac{\mathrm{d}z'}{E(z')} \right] \ , \tag{4.19}$$

where $E(z)$ is the normalised expansion rate $E(z) \equiv H(z)/H_0$. The first term of eq. 4.19 represents the true comoving distance r while the second one can be simplified assuming that z_o is close enough to z_c, resulting in

$$s = r + \frac{c}{H_0 E(z_c)} \Delta z \ . \tag{4.20}$$

Moreover, Δz can be obtained by rescaling the redshift solely induced by velocity by the scale factor at emitting time, *i.e.* $\Delta z = z_v / a_c$. Assuming $|\vec{v}| \ll c$, this contribution can be related to the non-relativistic Doppler–Fizeau effect eq. 1.14 as $z_v = v^{\|}/c$ where $v^{\|} = \vec{v} \cdot \frac{\vec{r}}{|\vec{r}|} \equiv \vec{v} \cdot \hat{e}_{\|}$ is the projection of the peculiar velocity along the line-of-sight. Thus, without lake of generality, eq. 4.20 becomes

$$\vec{s} = \vec{r} + v^{\|} \frac{(1+z_c)}{H_0 E(z_c)} \hat{e}_{\|} \ . \tag{4.21}$$

From now on, I will define the scaled velocity field $\vec{u}$ such that $\vec{u} \equiv -\vec{v}(1+z)/\left[f(z) H_0 E(z)\right]$, where f is the growth rate of structures. This change of variable allows to write down eq. 4.21 in a simpler form

$$\vec{s} = \vec{r} - f(z)(\vec{u}.\hat{e}_{\|})\hat{e}_{\|} \ . \tag{4.22}$$

Note that in the following I will drop the subscript $z_c \to z$, but still referring to the same quantity. Since one wants to quantify the effect on the power spectrum, one has to propagate the change of coordinates induced by peculiar velocities $(r_{\|}, \theta, \phi) \to (r_{\|} - f(z) u_{\|}, \theta, \phi)$ up to the density contrast δ. Let me define the scaled velocity projected along the line-of-sight as $u_{\|} \equiv \vec{u}.\hat{e}_{\|}$. First we can assume the local mass conservation between the two coordinate systems

$$\rho^s(\vec{s}) \mathrm{d}^3\vec{s} = \rho(\vec{r}) \mathrm{d}^3\vec{r} \ . \tag{4.23}$$

Then, one can use the Jacobian of the transformation to map the volumes between the two coordinate systems $\mathrm{d}^3\vec{s} = |J|\mathrm{d}^3\vec{r}$. Assuming, that the Jacobian is not singular (no shell crossing), from the mass conservation eq. 4.23 and the continuity equation 2.17 one can obtain the system

$$\delta^s(\vec{s}) = \frac{\delta(\vec{r}) + f(z)\partial_\parallel u_\parallel(\vec{r})}{1 - f(z)\partial_\parallel u_\parallel(\vec{r})} \ , \tag{4.24}$$

$$\dot{\delta}(\vec{r}, z) = f(z)H(z)\vec{\nabla} \cdot \left([1 + \delta(\vec{r}, z)]\vec{u}(\vec{r}, z)\right) \ . \tag{4.25}$$

In the Kaiser model, several strong but sensible approximations are made. Only $\delta \ll 1$ are considered leading to linearise the system, on top of neglecting velocity gradient, $\partial_\parallel u_\parallel(\vec{r}) \ll 1$ and assuming a potential velocity flow ($\vec{\nabla} \times \vec{u} = 0$). The system can be further simplified assuming the *plane parallel approximation*, postulating that velocity distances are much smaller than the distance separation between us and the object (*i.e.* $u_\parallel \ll r_\parallel$) and that the line-of-sight is not changing according to the position (which is clearly not true). In Fourier space, the system reduces to

$$\delta^s_{\vec{k}} = \delta_{\vec{k}} + \mu_k^2 f(z)\delta_{\vec{k}} \ , \tag{4.26}$$

where I recall the definitions $\mu_k \equiv \cos(\theta_{\vec{k}})$ and $\theta_{\vec{k}}$ being the angle between the line-of-sight and the considered wave vector $\vec{k}$. It is now easy to predict the distorted power spectrum in redshift-space in the Kaiser limit

$$P^s_K(k, \mu_k) = \left(1 + f(z)\mu_k^2\right)^2 P(k) \ . \tag{4.27}$$

The above equation shows, as explained before, that the power spectrum acquires an angular dependence with respect to the line-of-sight but also that along the line-of-sight ($\mu_k = 1$) the power is enhanced.

In practice, the anisotropy of the clustering can be encoded with a Legendre expansion in terms of the Legendre polynomials $\mathcal{P}_n(\mu_k)$ as $P^s(k, \mu_k) = \sum_{n=0}^{\infty} P^{s,(n)}(k)\mathcal{P}_n(\mu_k)$. In the Kaiser limit the series stops at order 4 and all odd contributions are null for symmetry reasons

$$P^{s,(0)}_K(k) = \left[1 + \frac{2}{3}f + \frac{1}{5}f^2\right] P(k) \ , \tag{4.28}$$

$$P^{s,(2)}_K(k) = \left[\frac{4}{3}f + \frac{4}{7}f^2\right] P(k) \ , \tag{4.29}$$

$$P^{s,(4)}_K(k) = \left[\frac{8}{35}f^2\right] P(k) \ . \tag{4.30}$$

A more accurate description, avoiding the issue of having a possible singular Jacobian which is likely to happen if we don't restrict ourselves to the linear regime, is the *streaming model* (Scoccimarro, 2004). It starts from the mass conservation eq. 4.24 and the plane parallel approximation, yielding the exact power spectrum

$$\delta^{\mathcal{D}}(\vec{k}) + P^s(\vec{k}) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3\vec{r} e^{-i\vec{k}\cdot\vec{r}} \left\langle e^{ifk_z\Delta u_z} \left[1 + \delta^r(\vec{x})\right] \left[1 + \delta^r(\vec{x}')\right] \right\rangle \ , \tag{4.31}$$

where $\vec{r} = \vec{x} - \vec{x}'$ and $\Delta u_z = u_z(\vec{x}') - u_z(\vec{x})$ when going back to Cartesian coordinates, with the line-of-sight aligned with the z-coordinate.

Such an expression turns out to be equivalent to define a moment generating function for the moments of the pairwise velocity distribution along the line-of-sight

$$\mathcal{M}(\lambda, \vec{r}) \equiv \frac{\left\langle e^{\lambda \Delta u_z} \left[1 + \delta^r(\vec{x})\right] \left[1 + \delta^r(\vec{x}')\right] \right\rangle}{\left\langle \left[1 + \delta^r(\vec{x})\right] \left[1 + \delta^r(\vec{x}')\right] \right\rangle} \ , \tag{4.32}$$

which allows to recover eq. 4.31 when setting $\lambda = ifk_z$. Thus knowing all the n-moments of the pairwise velocity distribution $\left[\partial \mathcal{M}^n / \partial^n \lambda\right](\lambda = 0)$ should be enough to describe exactly the RSD, both in linear and non-linear regime.

Expanding the moment generating function at order 4, Scoccimarro (2004) proposed a model which reads

$$P^s_S(k) = P_{\delta\delta}(k) + 2f\mu_k^2 P_{\delta\theta}(k) + f^2 \mu_k^4 P_{\theta\theta}(k) \ , \tag{4.33}$$

where θ refers to the divergence of the scaled velocity field $\theta \equiv \vec{\nabla} \cdot \vec{u}$ and where the cross-power spectra are defined as

$$\left\langle a_{\vec{k}} b_{\vec{k}'} \right\rangle = \delta^{\mathcal{D}}(\vec{k} + \vec{k}') P_{ab}(\vec{k}). \tag{4.34}$$

Note that eq. 4.33 is equivalent to eq. 4.27 when taking $\delta = \theta$ (fully linear regime), which is a byproduct of the linearisation of eq. 4.25.

In practice, expression 4.33 is not directly related to observations. Indeed, from galaxy surveys, properties of the matter field δ can only be inferred from biased tracers (see chapter 2). On large scales, one can relate the galaxy density contrast to the matter density contrast through a linear mapping called the linear bias b. Further assuming that the galaxies are following the velocity flow of matter, then equations 4.27 or 4.33 can be recast as

$$P^s_K(k, \mu_k) = \left[(b\sigma_8)(z) + (f\sigma_8)(z)\mu_k^2\right]^2 P_{\delta\delta}(k, z = 0) \ , \tag{4.35}$$

$$P^s_S(k, \mu_k) = (b\sigma_8)^2 P_{\delta\delta}(k, z = 0) + 2(b\sigma_8)(f\sigma_8)\mu_k^2 P_{\delta\theta}(k) + (f\sigma_8)^2 \mu_k^4 P_{\theta\theta}(k, z = 0) \ , \tag{4.36}$$

where $\sigma_8(z)$ is the normalisation of the power spectrum at redshift z. The above equations show that $\sigma_8(z)$ and the linear bias b are completely degenerate, while the angular dependence of the redshift galaxy space power spectrum can be used to constrain the combination $f\sigma_8$ as already anticipated in chapter 2.

4.3.3 Models in the non-linear regime

In order to model the progressive suppression of correlation power at smaller scales (see *Fingers-of-God* in figure 4.15) and extract as much information as possible from galaxy surveys, some methods has been proposed, as the *dispersion models* (Peacock & Dodds, 1994). They consist in convolving (multiplying in Fourier space) the linear model previously introduced with a damping term D as

$$P^s(k, \mu_k) = P^s(k, \mu_k) D(k, \mu_k, \sigma_{\mathrm{FoG}}) \tag{4.37}$$

where σ_{FoG} is a scale independent velocity dispersion. In practice, the dispersion σ_{FoG} is left as a free parameter in galaxy surveys. Several empirical models for the Damping term has been proposed in literature. Among them the Gaussian and the Lorentzian filtering

$$D_G(k, \mu_k, \sigma_{\mathrm{FoG}}) = \exp\left[-(k\mu_k\sigma_{\mathrm{FoG}})^2\right] \ , \tag{4.38}$$

$$D_L(k, \mu_k, \sigma_{\mathrm{FoG}}) = \left[1 + (k\mu_k\sigma_{\mathrm{FoG}})^2\right]^{-1} \ . \tag{4.39}$$

However, despite the variety of damping functions that can be used in this model, the fact that the pairwise dispersion σ_{FoG} remains a global constant leads to a difficult link with the physical property of the velocity field.

Lastly, the main issue with this non-linear description is that it does not take into account possible couplings between short and long wave modes through the velocity field. As a result, in the next section I adopt a different approach for the treatment of redshift-space distortions. Instead of applying a specific model to the power spectrum we want to work at the level of the velocity field, in order assign velocities to the objects in the Monte Carlo catalogues.

4.4 Sampling peculiar velocities

In principle, applying the Monte Carlo pipeline detailed in chapter 3 to generate a continuous velocity field, as long as a power spectrum and a PDF are provided, is perfectly feasible. However, several subtle points must be addressed. The first is related to the constraint that both the velocity and the density fields must be correlated to verify the continuity equation 2.17. This implies that the grid velocity field cannot be simulated in an independent way from the grid of the density field δ. In addition, once the velocity grid is obtained, one needs to associate a velocity to a particle lying at a certain position in between grid nodes.

In this section, I extend the pipeline in the same way as for the density field; first discussing how to generate the velocity field on a grid and then explaining how to associate velocities to particles.

4.4.1 Extending the Monte Carlo procedure to velocity field

In the following, for simplification I will assume that the studied velocity fields are deriving from a scalar potential ϕ_v, or equivalently curl-free fields. Even if it is not true in virialized structures, this approximation is still good enough on non-linear scales (Scoccimarro, 2004). Also, discussions will be based on the different results derived in appendix H concerning the byproducts of this assumption, in particular the relations between the velocity field and its divergence (and their power spectra). However I will recall in this sections the main results.

Identifying the targets

As previously explained in the case of the density field, the first step of the task is to identify the targets that we want to match at the end of the procedure. In this present exercise as I am going to discuss, both theoretical and numerical constraints make the identification of which statistical quantities will be targeted, difficult.

On top of being correlated to the density field, it turns out that the three components of the velocity field are also correlated with each other, as shown in eq. H.18

$$P_{u_i u_j} = \frac{k_i k_j}{k^4} P_{\theta\theta} \ , \tag{4.40}$$

where i and j label the vector components in Cartesian coordinates. In the mean time, since we assume that the velocity field is curl-free, these internal correlations are fully deductible from the power spectrum of a single scalar field which could be the scalar potential ϕ_u or the divergence of the scaled velocity field $\theta \equiv \vec{\nabla} \cdot \vec{u}$. Thus an approach that is likely to alleviate the difficulties is to apply the generation process introduced in chapter 3 to the divergence of the velocity field.

The power spectrum of the divergence of the velocity field $P_{\theta\theta}$ can be deduced from the linear power spectrum of matter fluctuations $P_{\delta\delta}$ using fitting prescriptions as I will discuss later. Unfortunately I was not aware of any theoretical modelling of the velocity divergence PDF[7]. Also I preferred to concentrate on the velocity filed itself which is the one of main interest when considering redshift-space distortions. In consequence the targeted PDF must be obtained directly from N-body simulations, once again relying on the `DEMNUni_cov` measurements for our tests.

However estimating the PDF of either the velocity components $\vec{u}$ or θ is not feasible for any grid sampling, in particular the one that I used so far, $N_s = 1024$. Indeed the corresponding estimation procedure is 1) construct a velocity grid by averaging each velocity component falling in voxels 2) estimate its divergence in Fourier space using eq. H.5 : $\theta_{\vec{k}} = i\vec{k} \cdot \vec{u}_{\vec{k}}$ and 3) go back in configuration space to estimate the θ-PDF or the $\vec{u}$-PDF. But the main problem in doing so is that for high grid resolution, the number of voxel which does not contain particles turns out to be large[8]. In these voxels, determining the average velocity of particles is thus prohibited and applying assignment schemes as listed in section 4.1 in order to increase the particle influence domain would not be sufficient to remove all indeterminations. As a consequence, these indetermination are preventing us to perform any Fourier transform, stopping us at point 2).

A naive idea would be to clip to zero each indetermination as it would seem realistic to pretend that a null velocity, whatever if the voxel belongs to an over- or under-density, has no statistical effect on the divergence. I checked this hypothesis by first setting $N_s = 512$, which counts no indetermination, and estimate the θ-PDF. Then this measure has been compared to the case when some indeterminations appears in voids. To do that, the most straightforward method is to decrease the

---

[7] In fact, I found out afterwards that Bernardeau et al. (1997) proposed a parametrisation of the smoothed θ-PDF. Although its form does not allow the analytical determination of the local mapping of the fields eq. 3.5 or for the correlation functions eq. 3.12, the numerical treatment of such a PDF model is totally allowed in this pipeline.

[8] These empty cells reach about $\sim 70\%$ for the `DEMNUni_cov` density.

particle density of the whole snapshot, while fixing N_s. In the end, when arbitrarily clipping to zero the indeterminations, the PDF appears non-trivially distorted, rejecting the hypothesis.

An other dead-end idea is a deconvolution/convolution procedure. Keeping a grid size that does not produce any indetermination ($N_s^{(1)}$), the obtained divergence field in Fourier space $\theta_{\vec{k}}^{\mathrm{conv}}$ is expected to be smoothed by a spherical/cubic Top-Hat window already defined in eq. 2.32 as

$$W_{\mathrm{TH}}(kR_{(1)}) = 3\frac{j_1(kR_{(1)})}{kR_{(1)}} \ , \tag{4.41}$$

where j_1 is the spherical Bessel function of order 1 and $R_{(1)} \equiv L/N_s^{(1)}$ the smoothing radius. One can think about deconvolving it applying $W_{\mathrm{TH}}^{-1}(kR_{(1)})$ then reconvolving it on the targeted grid smoothing using $W_{\mathrm{TH}}(kR_{(2)})$ of radius $R_{(2)} = L/N_s^{(2)}$ such that

$$\theta_{\vec{k}}^{\mathrm{rec}} = \left[\theta_{\vec{k}}^{\mathrm{conv}} W_{\mathrm{TH}}^{-1}(kR_{(1)})\right] W_{\mathrm{TH}}(kR_{(2)}) \ . \tag{4.42}$$

Although this operation is correct, no go back in configuration space is permitted. Indeed, convolution/deconvolution makes sense concerning Fourier amplitudes but is is non trivial at the level of Fourier phases. According to the central limit theorem, an inverse Fourier transform $\mathcal{F}^{-1}\left[\theta_{\vec{k}}^{\mathrm{rec}}\right]$ would produce a simple Gaussian field. Then estimating a velocity field in the same way as we estimate the density field is not possible and more sophisticated methods are required (Bernardeau & van de Weygaert, 1996; Pueblas & Scoccimarro, 2009b).

Thus due to these constraints, it turns out to be more convenient to work with velocity field targets, still through its divergence θ to ensure velocity components correlations (see eq. 4.40). The advantage of this choice is that the indeterminations have no impact on the PDF of each velocity field component, apart obviously from the noise (simply by removing the zeros when estimating the PDF).

Now concerning the targeted velocity power spectrum, one faces once again the issue of the indeterminations, as it requires to switch to Fourier space. Fortunately, in literature we can find fitting prescriptions to predict $P_{\theta\theta} = f\left(P_{\delta\delta}\right)$. In particular, one can use the fitting functions proposed by Bel et al. (2019) that gives

$$P_{\theta\theta}^{\mathrm{lin}}(k) = f^2(k) P_{\mathrm{cb}}^{\mathrm{lin}}(k) \ , \tag{4.43}$$

$$P_{\theta\theta}(k) = P_{\theta\theta}^{\mathrm{lin}}(k) \ \exp\left(-a_1 k - a_2 k^2 - a_3 k^3\right) \ , \tag{4.44}$$

where $P_{\mathrm{cb}}^{\mathrm{lin}}(k)$ stands for the linear cold dark matter plus baryons power spectrum (numerically predicted by Boltzmann codes) which gives the linear velocity divergence power spectrum $P_{\theta\theta}^{\mathrm{lin}}$. In turns eq. 4.44 provides its non-linear version using the fitting coefficients

$$a_1 = -0.817 + 3.198\sigma_{8,m} \ , \tag{4.45}$$

$$a_2 = +0.877 - 4.191\sigma_{8,m} \ , \tag{4.46}$$

$$a_3 = -1.199 + 4.629\sigma_{8,m} \ , \tag{4.47}$$

where $\sigma_{8,m}$ must be predicted for the total matter (eventually accounting for massive neutrinos). Finally, one can invoke eq. H.14, namely

$$P_{v_i v_i}(k) = \frac{1}{3k^2} P_{\theta\theta}(k) \ , \tag{4.48}$$

mapping the theoretical and targeted velocity power spectrum.

As for the density field, the variance estimated from the input power spectrum must match the variance estimated from the PDF. However, surprisingly, the velocity power spectrum eq. 4.48 provides a variance lower than the one estimated from the grid velocity PDF. This might be due to the way I computed the velocity on the grid by simply averaging the velocity of particles inside each cell. Indeed, it has been shown in literature (Bernardeau & van de Weygaert, 1996; Bernardeau et al., 1997; Hahn et al., 2015; Pueblas & Scoccimarro, 2009a) that this estimator does not allow to correctly trace the velocity field but rather the density weighted velocity. Instead, a more rigorous velocity grid reconstruction from catalogue should use a Delaunay tessellation algorithm. As a cross check, I used the data of Bel et al. (2019) where they estimated the velocity field on a 512^3 grid with such a reconstruction scheme. In this case the variance is much smaller and becomes compatible with the power spectrum. However, I learned these notions only during the writing of this manuscript. Thus, the rest of the discussion does not mention them.

In order to circumvent this problem (PDF variance grater that power spectrum variance), one can assume a shape for the velocity PDF with a free parameter related to the variance of the velocity field, this way once we provide a velocity power spectrum, then we can compute the velocity PDF with the corresponding variance.

After having studied the shape of the velocity PDF coming from the `DEMNUni_cov` simulations, I finally choose to model the velocity PDF shape following the inverse hyperbolic cosine distribution

$$P_v(v) = \frac{A}{\pi \cosh(Av)} \ , \tag{4.49}$$

$$A = \frac{\pi}{2\sqrt{\sigma_v^2}} \ , \tag{4.50}$$

where A is the normalisation factor both ensuring $\int \mathrm{d}x P_v(x) = 1$ and the matching of the variance σ_v^2 constrained by eq. 4.48. In figure 4.16 is displayed the PDF of the 1024 velocity grid (in green) and the corresponding inverse hyperbolic cosine PDF (blue line) obtained by matching the velocity variance. It appears that they do not match, actually the measured PDF of the velocity grid is closer to a Gaussian (represented with red line). However, I also represented on the same figure, the raw velocity distribution of particles in the simulation (hereafter referred to as one-point velocity distribution) and the later matches better the inverse hyperbolic cosine distribution. As a result, I propose to use eq. 4.49 to model the grid-PDF. This choice is motivated by the fact that, when addressing velocity assignment in next section, it will provide a PDF closer to the true one than the Gaussian model plotted in figure 4.16.

In conclusion, I didn't find any possibility of targeting the velocity-divergence properties. First because its grid PDF, to my knowledge, was not predictable and

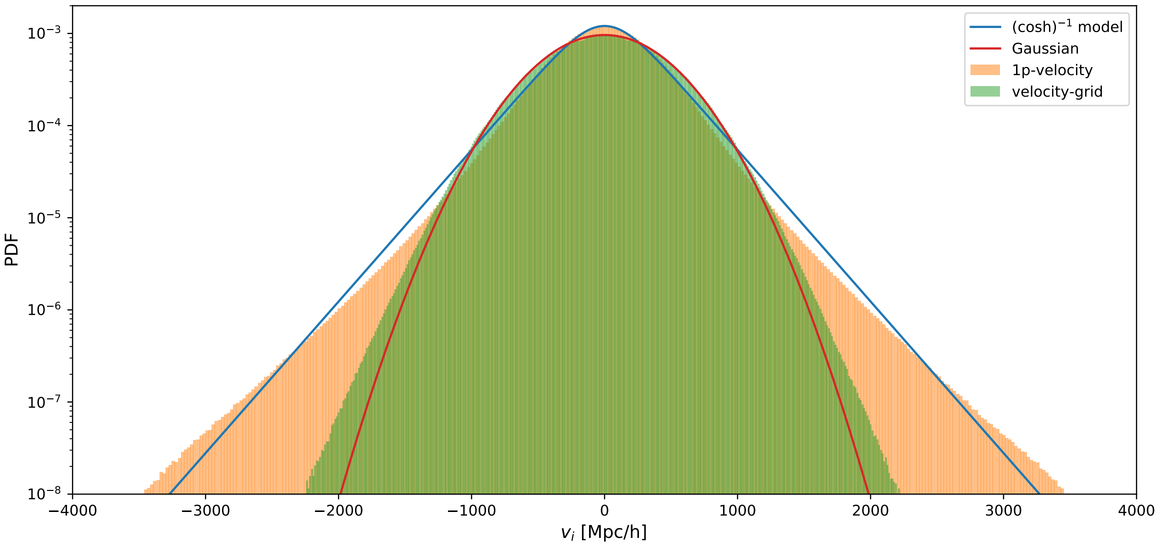


Figure 4.16: Various velocities PDF (normalised) involved in the discussion. In green is represented the grid-PDF (after removing the indeterminations) of the velocity field using a Top-Hat assignment scheme and in orange the one-point velocity distribution. In red and blue solid lines are represented the normal distribution and the model eq. 4.49, both of same variance as the velocity-grid PDF one.

secondly because its estimation on `DEMNUni_cov` for an arbitrary sampling parameter N_s is extremely unstable. I then bypassed this issue by rather choosing velocity properties as target (also more suited to the redshift-space distortions treatment). Therefore one has to keep in mind that even for the validation steps in the next section, both targets are theoretically modelled independently from N-body simulation, or at least not extracted from simulations.

In the following, I will specify how concretely the Monte Carlo procedure can be adapted to play at the same time with velocity component correlation and velocity-density relation.

Extending the Monte Carlo procedure

As for the density field, the identification of targeted PDF and power spectrum allows to recover the conditions presented in chapter 3. We first compute, using the prescription 3.5, the transformation $\mathcal{L}_\mu$ locally mapping a standardised Gaussian field $\mu(\vec{x})$ to the non-Gaussian velocity component field of power spectrum $P_{v_iv_i}(k)$ (see eq. 4.48) and with the PDF 4.49. The key part is to find out the specific power spectrum $P_\mu(k)$ for the sampled Gaussian field such that $P_{v_iv_i}(k)$ is correctly mapped. This can, once again, be done resorting to the Mehler expansion 3.12.

A major difference with the density field is that we are not anymore dealing with the simulation of a scalar field but rather with a vector field whose components are correlated (see eq. 4.40). As a result, independently simulating each velocity component is not conceivable. Likewise, using the same random seed for all components, leading to $P_{v_iv_j}(k) = P_{v_iv_i}(k)$, would be in contradiction with eq. H.15. In turns, each component must also be correlated to the local density field in order to make sure that RSD effects are correctly emulated (specially the Kaiser boost).

All of these constraints can be treated at once by resorting to the realisation of a single scalar field, the divergence of the velocity field. By computing the divergence of μ in Fourier space, noted Θ, one can also know its power spectrum $P_\Theta(k)$ using 4.48. Thus simulating this scalar Gaussian field[9] in Fourier space $\Theta_{\vec{k}}$, the resulting vectorial velocity field (see eq. H.12) $\vec{\mu}_{\vec{k}}$ is still expected (in configuration space) to be Gaussian distributed, standardised, and internally correlated. This ensures the mapping $\mathcal{L}_\mu$, apply on its real-space equivalent, to get the non-Gaussian velocity field $\vec{v}(\vec{x})$.

Establishing the velocity-density correlation can simply be achieved using the same Fourier phases for the random sampling of $\Theta_{\vec{k}}$ as the one used for the Gaussian density field $\nu_{\vec{k}}$. In doing so, one can expect that under-densities will locally correspond to small values for θ, thus of coherent in-fall. Instead, to over-densities will be associated a high velocity divergence. Note that this point only allows to recover the Kaiser effect with high accuracy, but with such a method, one can not yet expect to reproduce the Finger-of-God effect. As for the density field, a sketch summarising all the steps of this pipeline is presented in figure 4.17.

An ambiguity brought up by such a method concerns aliasing. In the density field case, no transformation is applied on aliased fields in Fourier space. The only two transformations $\mathcal{L}$ and λ^{-1} operated in comoving space, where alias contribution was expected to be suppressed. In this extended method instead, four transformations are accounted for: two in real space (for two-points correlation function and velocity fields) and two in aliased Fourier space (computing $P_\Theta(k)$ and transforming $\Theta_{\vec{k}}$ into $\vec{\mu}_{\vec{k}}$. In doing so there is no guaranty that the aliasing is properly conserved. However, this is without mentioning that the amplitude of the velocity power spectrum, combining eqs. 4.44 and 4.48, is widely suppressed with respect to the density power spectrum. In short, we don't expect the aliasing to be important in our working setting, $N_s = 1024$.

Thus this method can be used to generate a continuous grid of velocity, characterised by well controlled statistical quantities. In the following, I'm going to tackle the attribution of velocity to particles, making sure to simulate non-linear effects as introduced in section 4.3.

4.4.2 Modelling peculiar velocity assignment

Gaussian prescription

With such a method then, each velocity grid component is conceived to correspond to a single local density value. But when several particles can populate a single voxel using interpolation schemes (see section 4.1), only one velocity vector is associated with them. Hence it's not straightforward to make a correspondence between the two fields.

---

[9] It must be by construction a Gaussian field, without nevertheless the guarantee that it remains reduced. This fact is not a problem since unit variance is required when locally transforming the field to target a PDF.

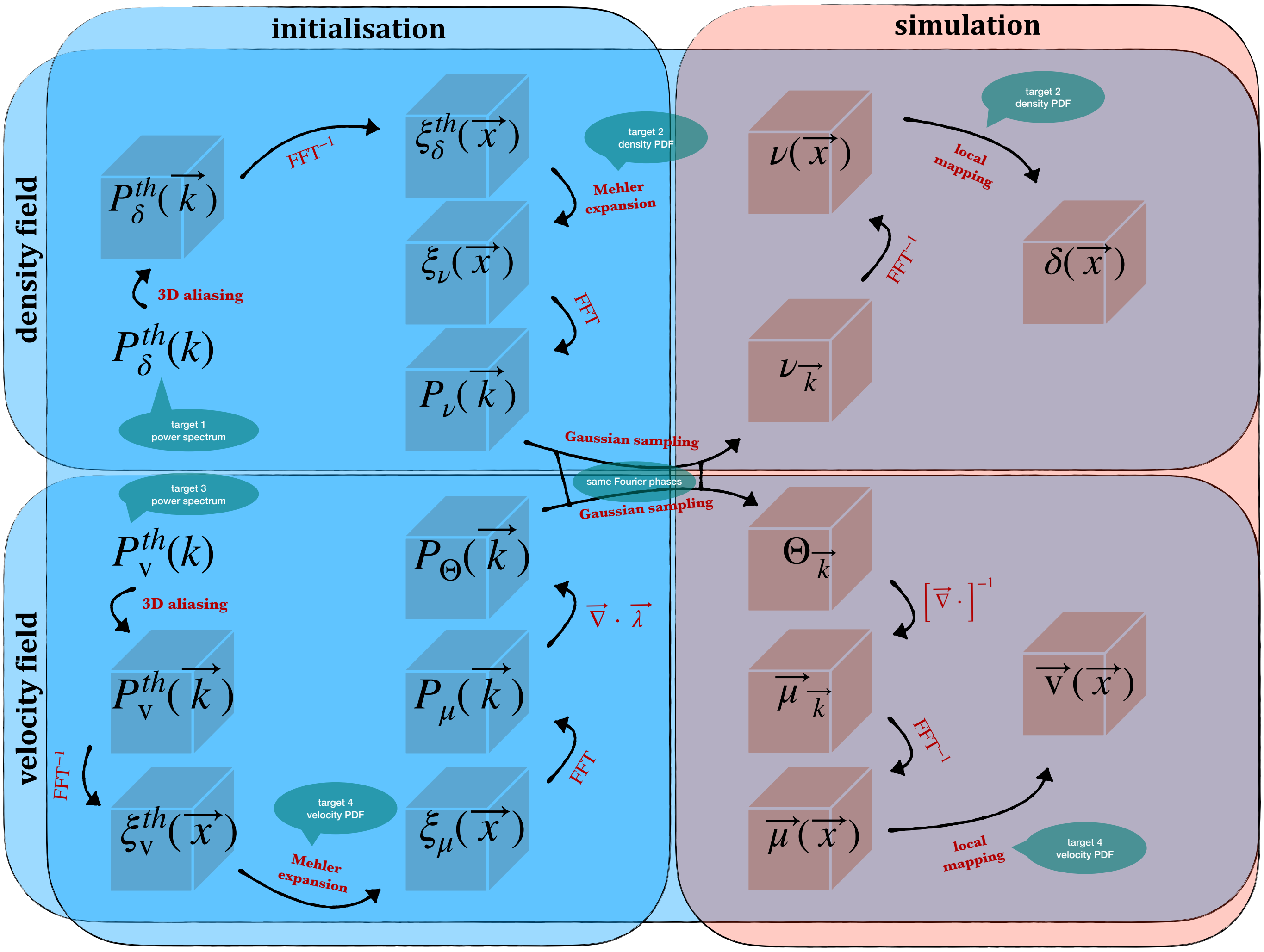


Figure 4.17: Sketch of the Monte Carlo procedure to generate correlated density and velocity fields on a grid by targeting a given power spectrum and PDF. Concerning the density field, we refer the reader to figure 3.3. For the velocity field, the first step of the initialisation procedure is to identify the statistical targets: the velocity field power spectrum eq. 4.44 (here noted P_V^{th}) and PDF eq. 4.49. Whereas the procedure allows to find the power spectrum of the Gaussian field $P_\mu(\vec{k})$ in the same way as for the Gaussian field ν, this extended pipeline rather deals with the velocity divergence of μ, using eq. 4.48, here noted $P_\Theta(\vec{k})$. This initialisation is done and the simulation procedure starts. Using the same Fourier phases as for the generation of $\nu_{\vec{k}}$, the Gaussian field $\Theta_{\vec{k}}$ is transformed into $\mu_{\vec{k}}$ using eq. H.12. Finally, switching to real space, one obtains the grid velocity field $\vec{v}(\vec{x})$.

A basic idea would be to attribute the same peculiar velocity at each particle populating a voxel, equivalent to a Top-Hat interpolation scheme. In doing so, the velocity field at the typical scale of the grid is free from dispersion, yet essential to simulate RSD at non-linear scales. Also this argument holds in the tri-linear interpolation scheme that would not fix the dispersion issue.

Rather it is necessary to be guided by the main features of the RSD at small scales. The Finger-of-God indeed inherit from the random nature of peculiar velocity characterised by a dispersion. However the great defect of the dispersion model (see section 4.3) is that it assumes that there is no relation between the local density and the velocity dispersion. This is at odd with what is observed in dark matter haloes identified in N-body simulations; their velocity dispersion depends on their mass thus on the local density contrast. Following this simple argument, it is possible to reproduce the velocity dispersion density contrast relation extracted from an N-body simulation in order to assign velocities to particles of our Monte Carlo.

From the velocity field $\vec{v}(\vec{x})$ in comoving space, let us rather re-sample the grid at the particle positions using a tri-linear grid interpolation, and doing this for each velocity component. Let me call this non-regularly spaced grid $\vec{w}(\vec{x}_i)$, where i labels each particles. This allows, in the same spirit as the density field, to improve the field continuity at the grid scale and to push toward higher modes the accuracy of the power spectrum in redshift-space. As previously mentioned, $\vec{w}$ cannot be directly used to attribute peculiar velocities. Thus in order to introduce randomisation, one can make the choice that peculiar velocities $\vec{V}_p$ are attributed following a normal distribution

$$\mathcal{P}^i_{\vec{V}_p} \equiv \mathcal{N}\left[\vec{w}(\vec{x}_i), \Sigma^2(\rho_i)\right] \ . \tag{4.51}$$

In eq. 4.51, $\Sigma^2(\rho_i)$ represents the velocity variance as a function of the local and tri-linearly interpolated density value at particle coordinates, which from now on is a free function of the local density ρ_i.

Modelling the $\Sigma^2(\rho)$ function

Fortunately having at our disposal a large sample of N-body simulations, one can get an insight on the shape of $\Sigma^2(\delta + 1)$ by estimating it[10] on the basis of the `DEMNUni_cov` snapshots.

By fixing the grid sampling to $N_s = 1024$, one can run an estimation of the velocity variance as a function of the local density contrast value over the 50 simulations in the two cosmologies 16nu and ΛCDM and for the five redshifts $z = [0, 0.5, 1, 1.5, 2]$. The combination of these results are presented in figure 4.18. Note that for small δ, their estimation result from a Poisson distribution rather than the underlying continuous density contrast field. A bias is therefore expected when a low number of particles is counted in cells (small δ). Thus, rather than trying to correct it, one can simply focus the discussion exclusively on regions with a high number count (large δ). Moreover, given the fact that we want to reproduce the redshift-space distortions in the non-linear regime, it makes sense to focus the dispersion-density relation in high density regions (which are supposed to dominated the non-linear signal).

In figure 4.18, it appears that the local variance as a function of the local density $1 + \delta$ can be well approximated by a simple power law

$$\Sigma^2(\rho) \equiv \beta\rho^\alpha \ . \tag{4.52}$$

---

[10] One would prefer to work with $\delta + 1$ rather than ρ in order not to be depend on the choice of the averaged density $\bar{\rho}$.

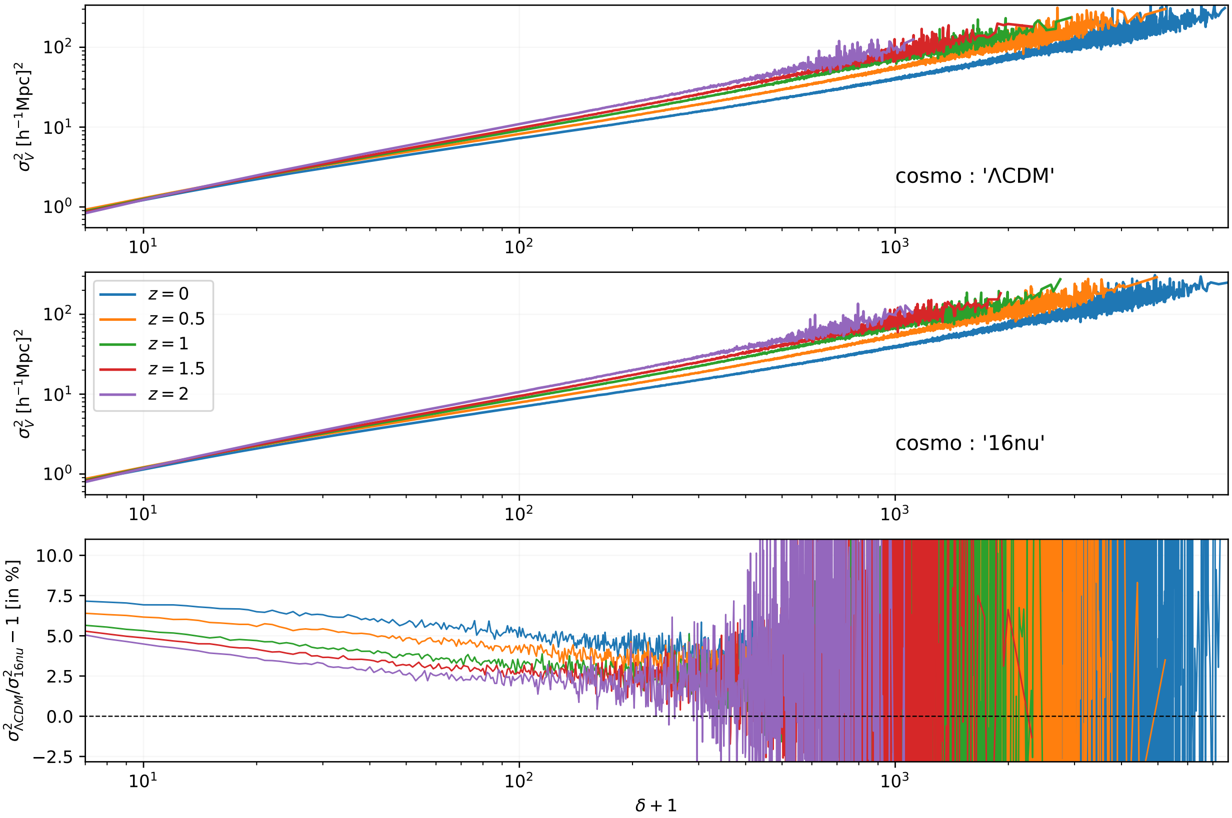


Figure 4.18: *upper and centre panels:* merged velocity components variance as a function of local density contrast for the `DEMNUni_cov` set of 50 simulation and for the five redshifts $z = [0, 0.5, 1, 1.5, 2]$ in the case of the standard cosmology with and without massive neutrinos *bottom panel:* relative deviation in percent, redshift per redshift, between the two cosmologies. The three panels are using the same colour-code.

These new free parameters α and β, on top of depending on the redshift, are also sensitive to the cosmological model. In particular, adding massive neutrinos is damping the variance, a result coherent with the fact that neutrinos tends to reduce the clustering amplitude. Note also that the exponent α of the power law is related to dynamics inside haloes, indeed we checked that if one use haloes to estimate the above relation then the same value for α is measured (but with a different normalisation β). Thus, in the following I will consider the exponent α as fixed by measurements in the simulation while I will allow β to vary.

One-point velocity distribution as a new target

Note that when going from the velocity grid to a peculiar velocity description, there is no reason for the one-point velocity variance σ_V^2 to match the one targeted on the grid. That is the reason why I leave the normalisation β of the dispersion-density relation as being free to vary. Thus, it can be set in such a way that σ_V^2 matches the variance of the velocity of the particles in the N-Body simulation. In table 4.1 are

	σ in ΛCDM	σ in 16nu	α	β
$z = 0$	3.9873	3.8331	~ 0.37	~ 48
$z = 0.5$	3.6853	3.5488	~ 0.40	~ 45
$z = 1$	3.2197	3.1063	~ 0.43	~ 41
$z = 1.5$	2.9332	2.8337	~ 0.45	~ 39
$z = 2$	2.6021	2.5188	~ 0.46	~ 39

Table 4.1: *rms* (in Mpc/h) of the one-point velocity distribution as measured from particles in the `DEMNUni_cov01` for the two cosmologies. Estimation of α and β inferred from figure 4.18.

listed the estimated variances for `DEMNUni_cov01` in each cosmology and redshift configuration, as well as the estimated values of α and β. Once again, it appears that introducing massive neutrinos reduces the matter velocity dispersion, for the same reason discussed previously. This effect is maximised for low redshift ($\sim 4\%$).

Note that matching the one-point velocity variance allows to provide a physical description of the amplitude parameter β. Thus, let me show how to predict how the one-point variance that we will obtain from the Monte Carlo is impacted by the prescription 4.51 when going from a grid velocity to a particle description. Deriving this modification can be done by considering the vector components V_c of $\vec{V}_p$, w_c of $\vec{w}$ and the locally interpolated value ρ. The corresponding tri-variate probability distribution function reads

$$\mathcal{P}(V_c, w_c, \rho) = \mathcal{P}_{\vec{V}_p}\mathcal{P}(w_c, \rho) \ , \tag{4.53}$$

where $\mathcal{P}_{\vec{V}_p} \equiv \mathcal{P}(V_c|w_c, \rho)$ is defined in eq. 4.51. Also in its discrete form, σ_V^2 can be obtained using the expectation value

$$\sigma_V^2 = \frac{1}{N}\sum_{i=1}^{N} V_i^2 \ , \tag{4.54}$$

where one would prefer to work with its continuous limit in order to exploit eqs. 4.51 and 4.53. In such representation, one has to take care that the number of generated particles depends on the local density. As a matter of fact the n^{th}-order moments of the one-point velocity distribution reads $\sigma_V^n = \langle \rho V_c^n \rangle$ and in the case we are interested in[11],

$$\sigma_V^2 = \langle \rho V_c^2 \rangle = \iiint \mathrm{d}V_c \mathrm{d}w_c \mathrm{d}\rho \ \rho V_c^2 \mathcal{P}(V_c, w_c, \rho) \ . \tag{4.55}$$

Replacing eqs. 4.51 and 4.53 in this expression and using that the moment of order 2 of the Gaussian distribution V_c reduced to the sum of its variance and the squared of its expectation value, it simplifies in

$$\sigma_V^2 = \iint \mathrm{d}w_c \mathrm{d}\rho \ \mathcal{P}(w_c, \rho)\rho \left[w_c^2 + \Sigma^2(\rho)\right] = \sigma_w^2 + \sigma_\Sigma^2 \ . \tag{4.56}$$

---

[11] Note here that ρ must be taken as adimensional, then is exceptionally defined as $\rho \equiv \rho/\bar{\rho}$ where $\bar{\rho}$ is arbitrary.

In the above expression the first term $\sigma_w^2 \equiv \int \mathrm{d}w_c \mathrm{d}\rho \; \rho w_c^2 \mathcal{P}(w_c, \rho)$ is the variance of the interpolated velocity field that can be estimated independently from the choice of the function $\Sigma^2(\rho)$, while this is not the case for the second marginalised term, defined as $\sigma_\Sigma^2 \equiv \int \mathrm{d}\rho\rho\Sigma^2(\rho)$. Then coming back to our simple model 4.52, the β parameter can be obtained using eq. 4.56 as

$$\beta = \left(\sigma_V^2 - \sigma_w^2\right) / \sigma_\alpha^2 \,, \tag{4.57}$$

where $\sigma_\alpha^2 \equiv \langle \rho^{\alpha+1} \rangle = \frac{1}{N} \sum_i \rho_i^\alpha$ is directly estimated on the interpolated density field at particle coordinates.

To conclude, I have reviewed all velocity field properties and transformations ensuring the RSD effect to be simulated in redshift-space, both in linear and non-linear regime. Since the presented pipeline was designed to match the `DEMNUni_cov` outputs (apart from the velocity PDF), the way is paved for a legal method validation in the next section.

4.5 Method validation in redshift-space

4.5.1 On the distorted power spectrum

Accordingly to the overall pipeline presented in the previous and in the present chapters, we generate 1000 catalogues for the five redshifts consisting once again in $N_p = 10^8$ particles distributed in a box of volume of $(1000h^{-1}\text{Mpc})^3$ or equivalently $\rho_0 = 0.1h^3\text{Mpc}^{-3}$ in 16nu cosmology. Now attributing peculiar velocities to particles, about ~ 40 min of CPU-time on 50 processes of the Dark Energy Centre are necessary to generate one catalogue.

As in section 4.2 for the comoving space, this present section aims at comparing the different estimated statistical quantities between the Monte Carlo process and the `DEMNUni_cov` simulations in redshift-space. Starting with the estimated monopoles, figure 4.19 shows the averaged of the estimated power spectra in the 1000 realisations with error bars for the whole set of redshifts. With high accuracy and without detecting any systematic, the simulated monopoles are perfectly fitting the N-body ones, at least up the maximal wave modes for which the monopoles in comoving space were reliable (see figure 4.9).

It is also interesting to look at the ratio between the redshift-space monopole and the real space one, this is what is shown in figure 4.20. One can see that the effect of redshift-space distortions on the monopole is reproduced in the Monte Carlo at percent level for higher non-linear modes ($k \sim 0.5h/\text{Mpc}$) than the comparison between the monopoles of the Monte Carlo and the N-body in comoving and redshift-space. Indeed, due to the various filtering on power spectra previously introduced, their tight sharpness (see i_1 and i_2) around these typical wave modes could be the reason of the deterioration of the ratio $P^{\text{RSD}}(k)/P^{\text{noRSD}}(k)$. Thus testing the validity of the RSD model beyond $k = 0.5h/\text{Mpc}$ will be the object of further studies.

However, the left and centre panels of figure 4.20 features an interesting physical process, related to the migration of the crossing between the power spectra in both spaces. Due to the enhancement of clustering and the formation of structures as time passes, virialisation at non-linear scales are more and more dominant, progressively

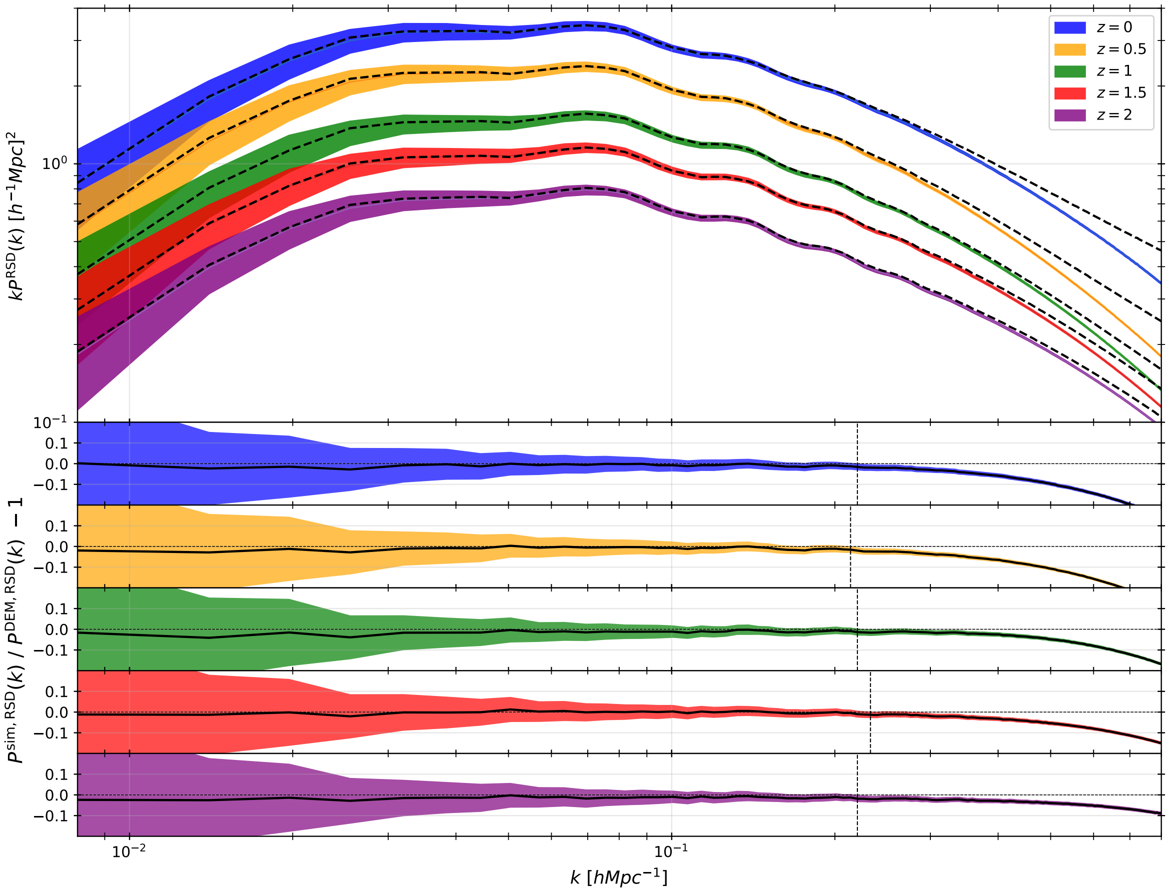


Figure 4.19: *panel 1:* measured averaged power spectra over 1000 realisations of Poisson catalogues in 16nu cosmology, each of 10^8 particles with error bars represented in colour. Overlapping dashed black lines represents the averaged power spectra over the 50 `DEMNUni_cov` simulations per redshift. *panel 2,3,4,5,6 :* their relative deviation showing in black solid line with error bars following the same colour-code. Vertical dashed lines indicates the wave modes for which error bar cross zero. The corresponding values for redshifts $z = [0, 0.5, 1, 1.5, 2]$ are respectively $k_{\max} \sim [0.22, 0.21, 0.22, 0.23, 0.22]h/\mathrm{Mpc}$.

shrinking the linear domain. It results in an increase in domination of the Finger-of-God over the power spectrum definition domain.

In addition, we can see in the centre panel that the amplitude of the Kaiser effect is reduced for decreasing redshift. This is an effect that was in fact expected, as evidenced by the eq. 4.28, where it should be mentioned that f is a decreasing function of time (since the cosmological constant Λ damps the rate at which clustering operates). Notably, deep in the matter domination, the ratio between the two monopoles can be predicted to be $(1 + 2/3 + 1/5) \approx 1.9$.

Let us concentrate now on the study of the simulated quadrupoles and hexadecapoles, figure 4.21 shows the anisotropy induced by RSD. Note that both of them are allowed to be negative, in case of the quadrupole, the change of sign can be related to the change of regime, the point for which the random motions become

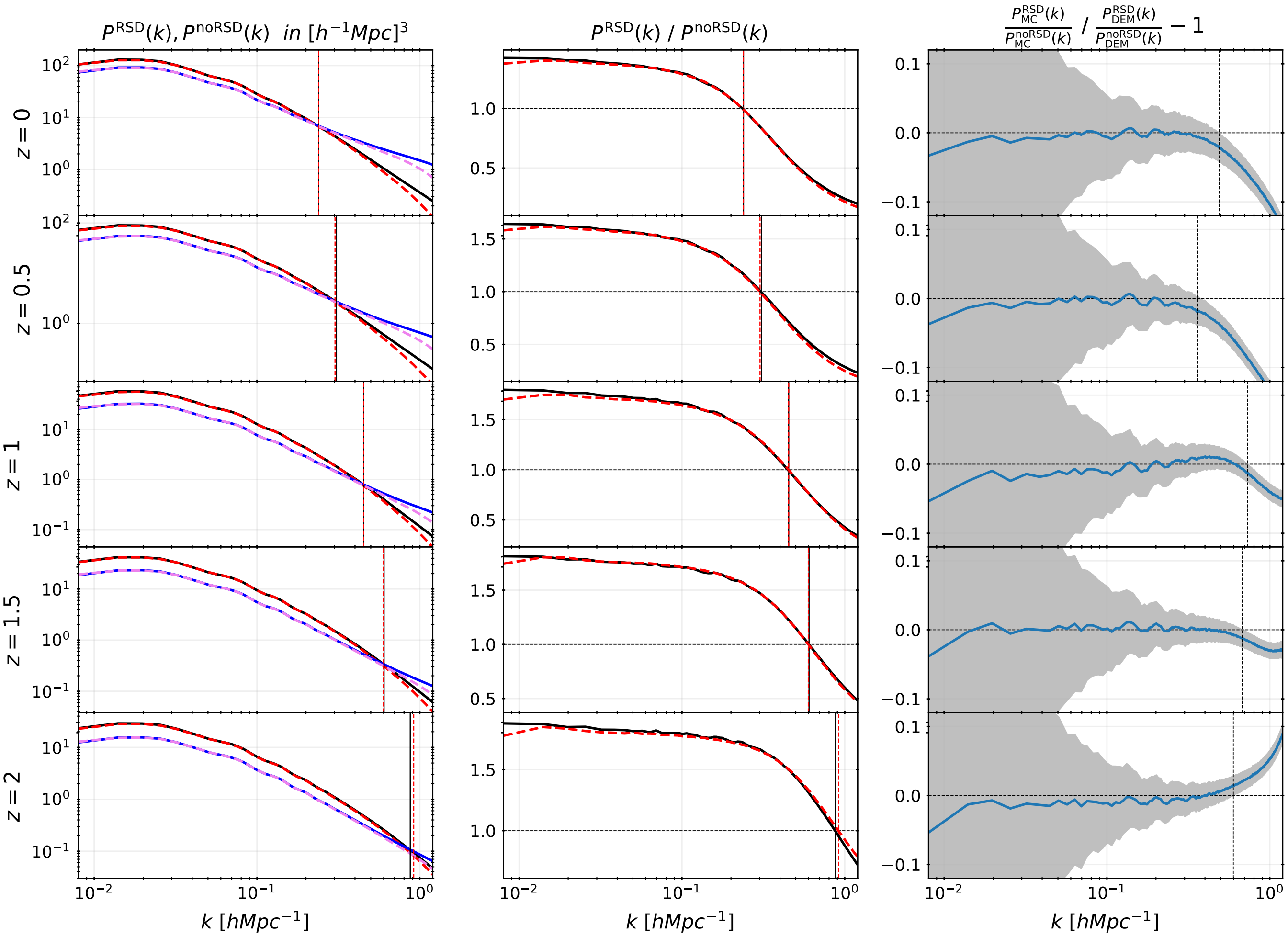


Figure 4.20: *Left column:* averaged power spectra in comoving space $P^{\mathrm{noRSD}}(k)$ for the Monte Carlo method (pink dashed lines) and `DEMNUni_cov` (blue solid lines) and in redshift-space $P^{\mathrm{RSD}}(k)$ for Monte Carlo (red dashed lines) and `DEMNUni_cov` (black solid lines) for the five redshifts. The successive vertical lines correspond to the crossing wave modes k_{cr} such that $P^{\mathrm{noRSD}}(k_{cr}) = P^{\mathrm{RSD}}(k_{cr})$ in black dashed lines and blue dashed lines for the N-body and for the Monte Carlo, respectively. They values are for increasing redshifts $k_{cr} \sim [0.23, 0.30, 0.45, 0.60, 0.87]h/\mathrm{Mpc}$. *Centre column:* ratio between monopoles with and without RSD contribution in black solid lines for `DEMNUni_cov` and in red dashed lines for the Monte Carlo. The vertical references are the same as for the first column. *Right column:* ratio between the previous quantity in blue solid line with error bars in grey (accounting for estimated variance for `DEMNUni_cov` and the Monte Carlo in both redshift and comobile spaces). The vertical references in dashed black lines corresponds to the successive modes for which the RSD model becomes incompatible at $1-\sigma$. They are for increasing redshifts $k_{1\sigma} \sim [0.49, 0.36, 0.72, 0.68, 0.60]h/\mathrm{Mpc}$.

dominant. As for the simulated redshift-space multipoles, $P^{(2)}$ and $P^{(4)}$ turn out to be just as well simulated, up to wave modes higher than the ones for which the monopole in comoving space has been considered as reliably reproduced.

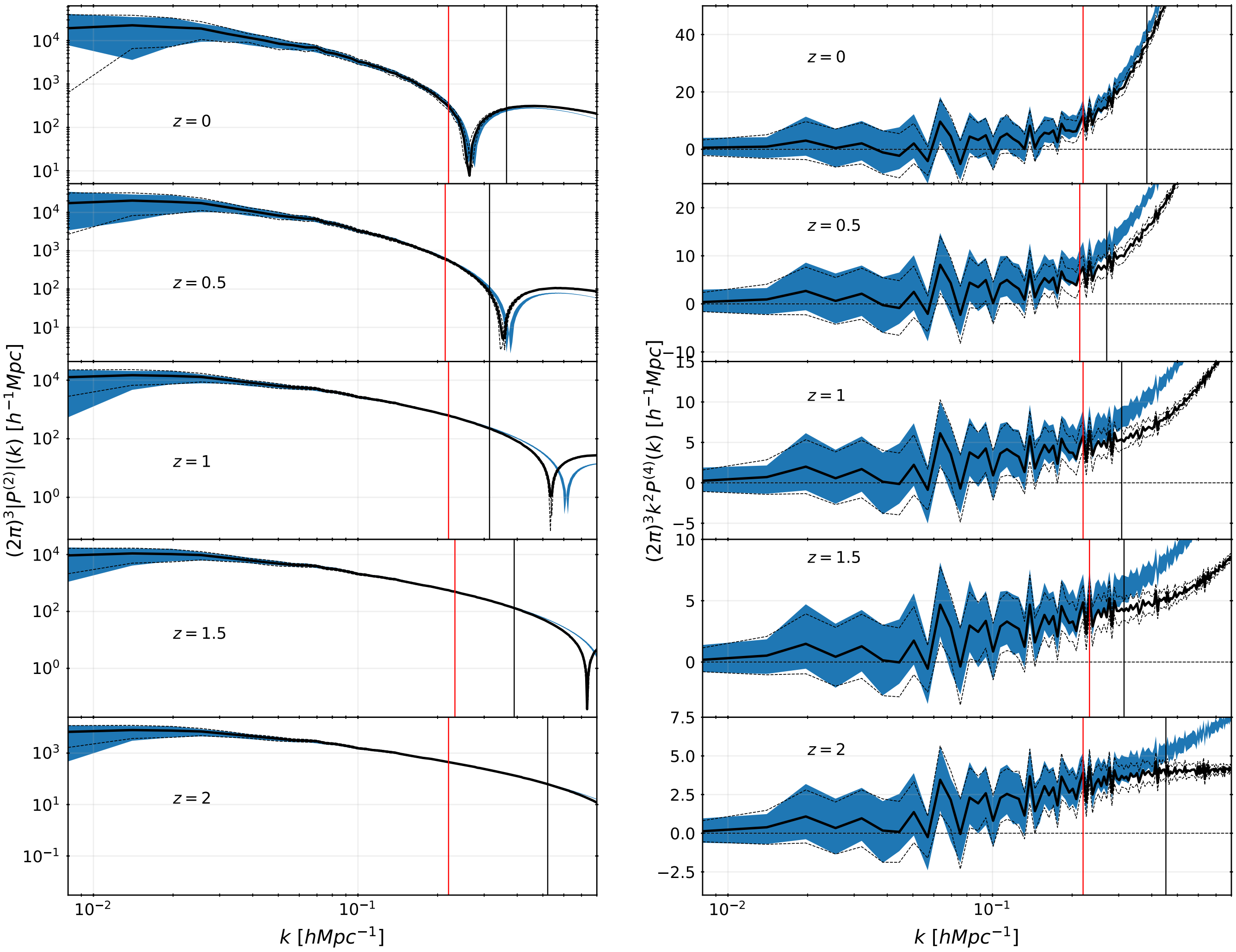


Figure 4.21: Absolute values of the averaged quadrupoles (left column) and averaged hexadecapoles (right column) for the `DEMNUni_cov` in solid black lines (error bars in dashed black lines) and for the Monte Carlo with error bars in blue for the five redshifts. The red vertical references stand for the maximum wave mode value for which the monopole in redshift-space matches the N-body one (at $1-\sigma$) $k_{1\sigma}^{\mathrm{M,RSD}} \sim [0.22, 0.21, 0.22, 0.23, 0.22]h/\mathrm{Mpc}$, while the black vertical solid line is the equivalent but for the quadrupole and the hexadecapole. In the case of the quadrupole : $k_{1\sigma}^{\mathrm{Q,RSD}} \sim [0.36, 0.31, 0.31, 0.39, 0.52]h/\mathrm{Mpc}$, and for the hexadecapole : $k_{1\sigma}^{\mathrm{H,RSD}} \sim [0.38, 0.27, 0.31, 0.31, 0.45]h/\mathrm{Mpc}$.

The hexadecapole becomes also more relevant in the non-linear regime, a feature well reproduced in the Monte Carlo. However, for both the quadrupole and hexadecapole we can see a deterioration of the quality of the Monte Carlo when going from redshift 0 to redshift 2. This is at odd compared with the fact that the power spectrum itself is better reproduced at higher redshift. I am currently investigating the origin of this effect, it seems that the balanced between the velocity on the grid and the dispersion is not optimal when the redshift is increasing. One possible explanation for that could be the choice of an inverse hyperbolic cosine for the PDF of the grid velocity. Indeed, we saw that it is in reality closer to a Gaussian, thus it might be that this choice could affect the highest redshift for which this model is not accurate enough. Actually the first preliminary results on that are going in the right direction, taking a target Gaussian PDF for the velocity grid seems to alleviate the problem. Nevertheless, more work is needed to properly address this issue.

4.5.2 Redshift-space covariance matrix

To conclude this chapter, let me go through the estimated covariance matrices of the simulated monopoles, once again debiased using eq. 4.16. First focusing on its diagonal, where both Gaussian and non-Gaussian signature are contributing, figure 4.22 provides very similar result as in figure 4.11 for comoving space. Once again, it shows a good accuracy for the matching between the variance estimated from N-body simulation and the one from the Monte Carlo realisations, even beyond the maximal modes for which the monopole is well reproduced. In addition, one can notice that compared to comoving (real) space that the contribution of the tri-spectrum to the total variance is higher in redshift-space in the Monte Carlo. This is actually worsening when increasing the redshift thus I argue that this might be related to the same effect pointed out before regarding the quadrupole and hexadecapole.

The study can be extended to the off-diagonal terms where that time, only the tri-spectrum contribution has an effect. When setting the maximal mode to $k \sim 0.3h/\mathrm{Mpc}$ in figure 4.23 and $k \sim 0.5h/\mathrm{Mpc}$ in figure 4.24, one can see a good matching level between the two methods, even for modes beyond the reproduction domain of the monopole. When comparing to the results in comoving space figures 4.12, no significant degradation can be detected.

However, one can remark that at low modes and for most of the simulated redshifts, extra correlations with large modes about $1-\sigma$ appears. This effect can be easily visualised in figure 4.25 for high z. Once again, such effect requires further studies to be investigated.

In Figure 4.25 I compare the correlation matrix of the monopole RSD power spectrum of the Monte Carlo (bottom) to the one of the `DEMNUni_cov`. It is clear that with 1000 realisation the correlation matrix appears smoother, but we can also detect what I consider to be a spurious correlation (top right of each correlation matrix) which seems to indicate that the largest scales and smallest scale are getting more and more correlated when increasing the redshift. This is really contradictory with the fact that the field is more and more linear thus reducing the coupling between density and velocity. This is also something which could be related to the choice of the inverse hyperbolic cosine distribution for the distribution of the velocities on the grid. Thus this is a possibility I want to explore soon.

In conclusion, switching from a continuous grid to a discreet set of particles has been done by means of an interpolation schemes, whose the convolution effect on the resulting power spectrum can be successfully predicted. To each of these

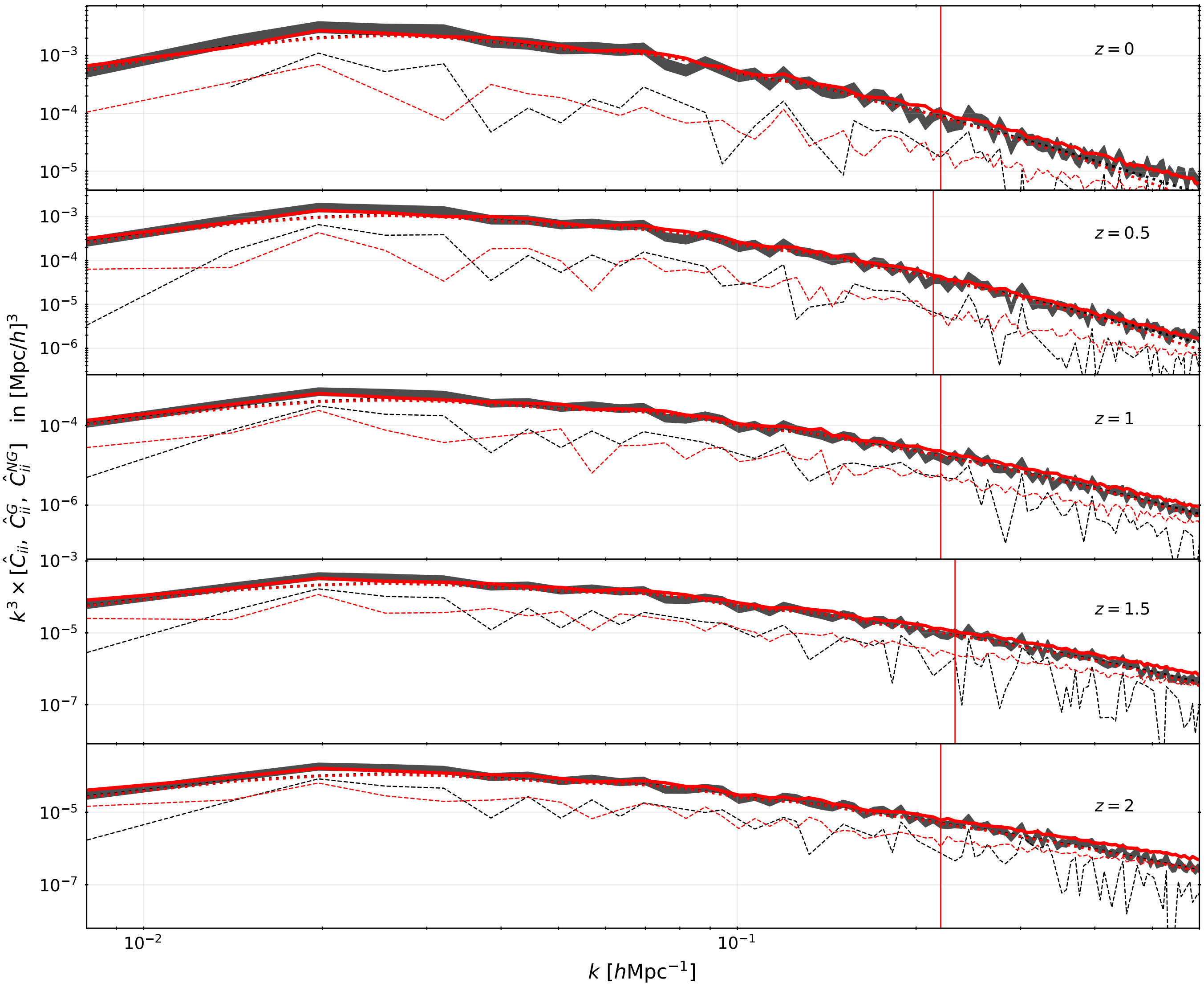


Figure 4.22: Estimated contributions of the diagonal of the covariance matrix estimated over 1000 realisations of catalogues composed of 10^8 particles with 16nu cosmology in redshift-space as compared to the ones for `DEMNUni_cov` for five redshifts. The diagonal C_{ii} accounting for all contributions is represented in solid red line and compared to those of the N-body in black with Gaussian error bars $2C_{ii}^2/(N-1)$ where $N = 50$. In dotted lines are displayed the Gaussian contribution C_{ii}^G (first term of eq. 2.46) estimated from the averaged power spectra in red (Monte Carlo) and in black (`DEMNUni_cov`). The trispectrum contribution C_{ii}^{NG} are drawn in dashed black (`DEMNUni_cov`) and dashed red (Monte Carlo) lines; they are coming from the residual between the total variance and the Gaussian variance and only their positive values are represented. The vertical reference in red shows for each redshift the maximal mode up to which the monopole is well reproduced in the $1-\sigma$ limit, $k_{1\sigma}^{\mathrm{M,RSD}} \sim [0.22, 0.21, 0.22, 0.23, 0.22]h/\mathrm{Mpc}$ (see figure 4.9).

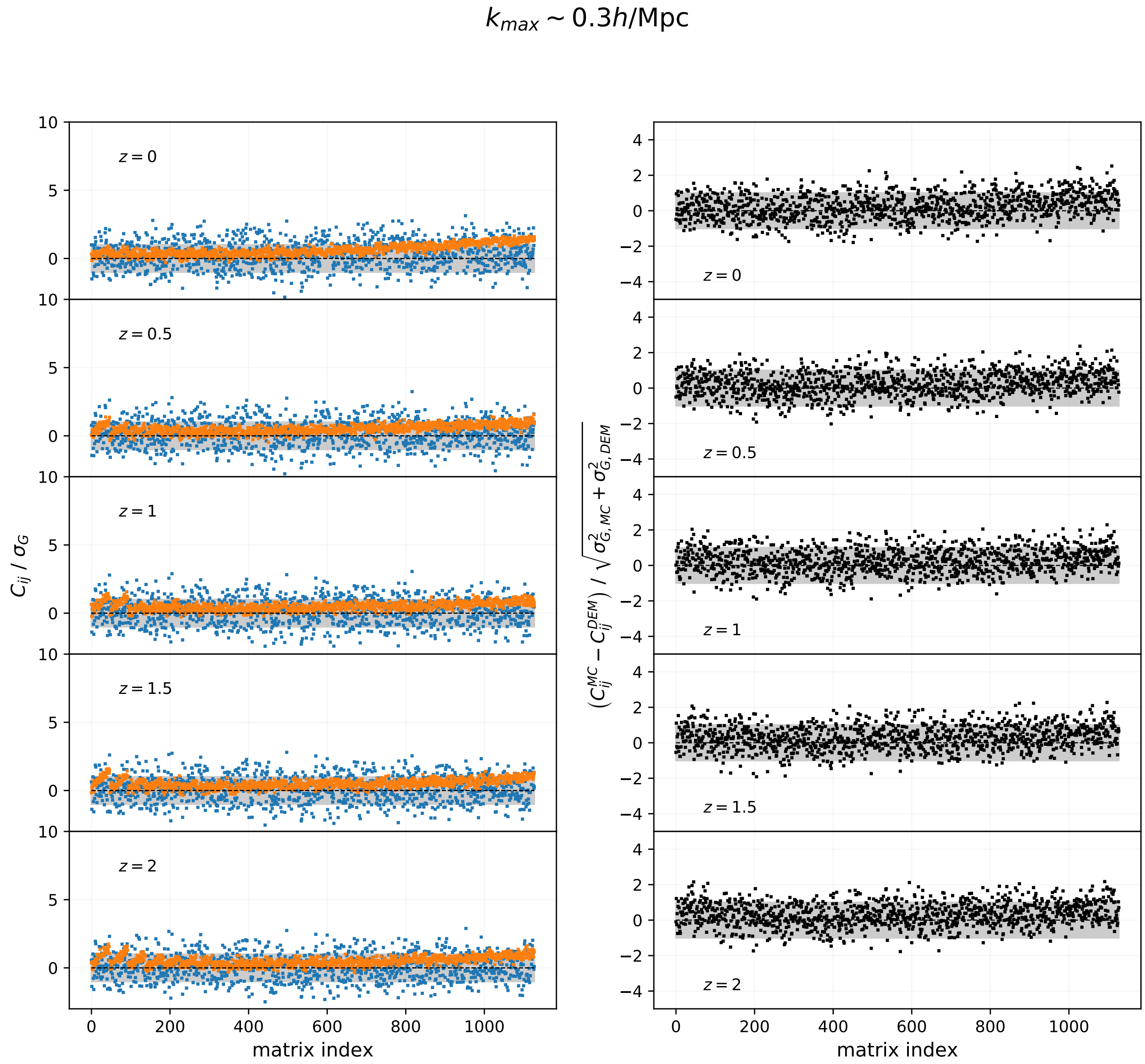


Figure 4.23: *left panels:* estimated off-diagonal elements of the covariance matrix for the `DEMNUni_cov` (blue dots) and the Monte Carlo (orange dots), normalised by their respective Gaussian error estimated using the averaged values of the simulated power spectra for the five redshifts. The elements are ordered column by column of the lower half of the (48×48) matrices by referring to the matrix index in order to display covariances up to $k \sim 0.3h/\mathrm{Mpc}$. Note that the diagonal elements are not included. *right panels:* residual of the two matrices using the same ordering and normalised by the combined Gaussian errors (black dots). Note that Gaussian errors are computed using $N = 50$ for both experiments even if the Monte Carlo covariance is estimated over 1000 realisations. In all panels, grey areas represent the $1 - \sigma$ limit.

particle, a modelisation of the peculiar velocity attribution has been proposed. This allowed to recover in redshift-space the right distortions affecting the monopole,

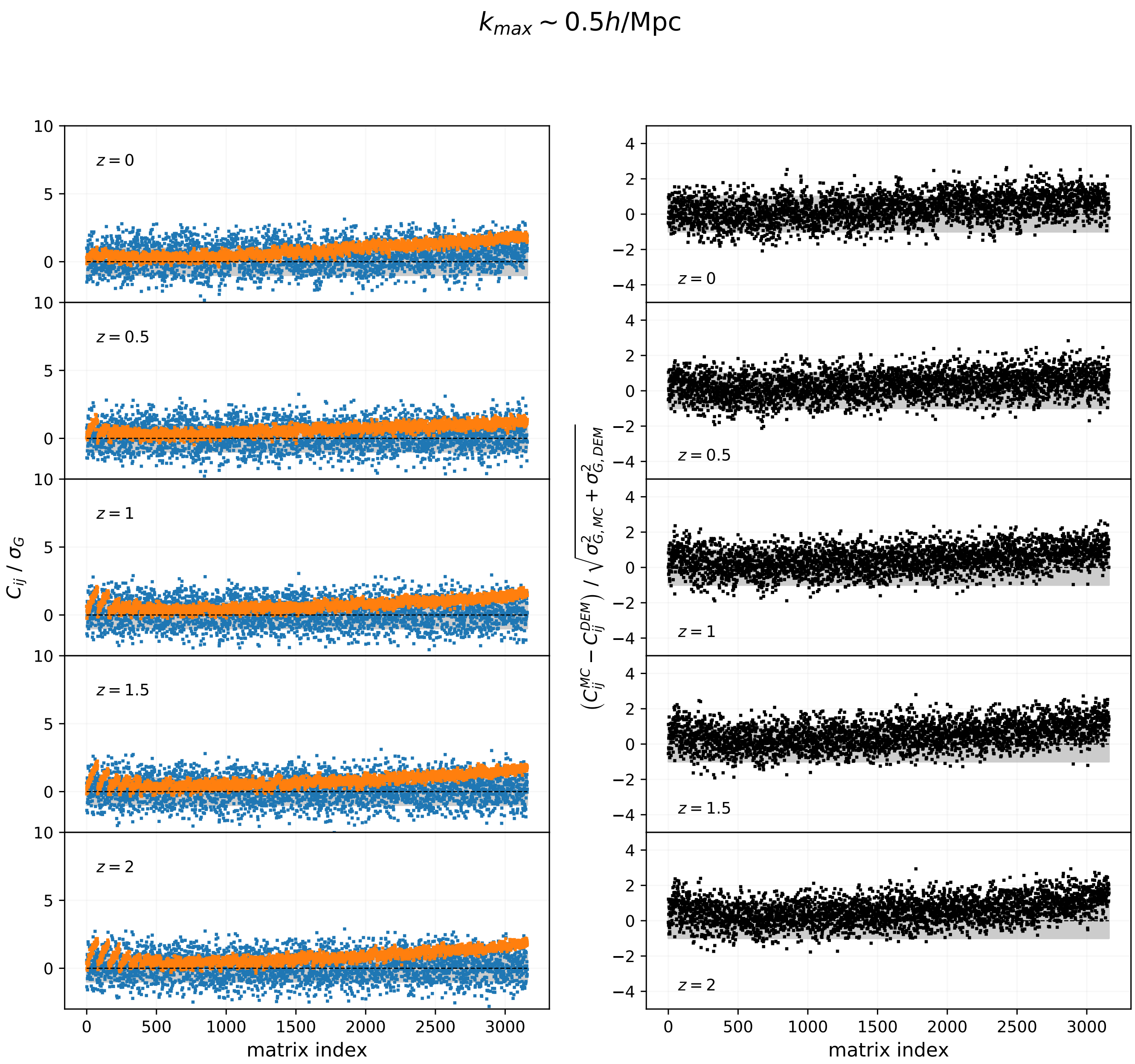


Figure 4.24: *left panels:* estimated off-diagonal elements of the covariance matrix for the `DEMNUni_cov` (blue dots) and the Monte Carlo (orange dots), normalised by their respective Gaussian error estimated using the averaged values of the simulated power spectra for the five redshifts. The elements are ordered column by column of the lower half of the (80×80) matrices by referring to the matrix index in order to display covariances up to $k \sim 0.5h/\mathrm{Mpc}$. Note that the diagonal elements are not included. *right panels:* residual of the two matrices using the same ordering and normalised by the combined Gaussian errors (black dots). Note that Gaussian errors are computed using $N = 50$ for both experiments even if the Monte Carlo covariance is estimated over 1000 realisations. In all panels, grey areas represent the $1 - \sigma$ limit.

both in linear regime with the Kaiser effect and in the non-linear one with the Fingers-of-God. Then, anisotropies in redshift-space has been quantified through

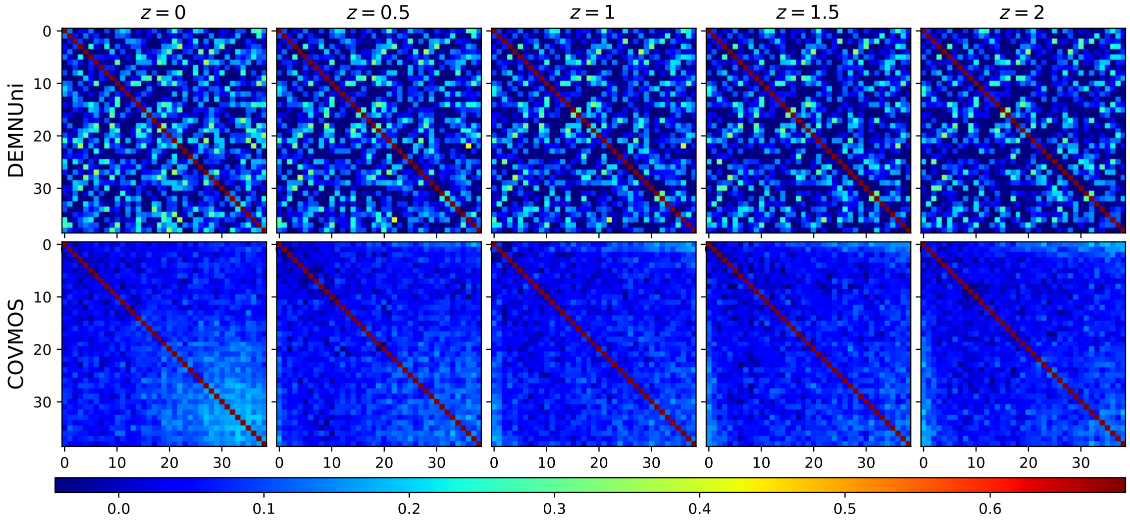


Figure 4.25: Correlation matrices in redshift-space in the `DEMNUni_cov` (left panels) and Monte Carlo (right panel) cases up to $k \sim 0.25$ h/Mpc, *i.e.* (39×39) sub matrices for the five redshifts.

the quadrupole and the hexadecapole, showing a high-level of reliability (1%) of the method on two-points statistics at least up to $k \sim 0.2h/\text{Mpc}$. Finally, the power spectrum covariance matrix has been investigated in redshift and comoving spaces, producing in the $\sim 1\sigma$ limit reliable results for $k < 0.8h/\text{Mpc}$ in comoving space and $k < 0.5h/\text{Mpc}$ in redshift-space.

In the next chapter, I will present an application of the Monte Carlo method to an other galaxy clustering observable; the angular power spectrum C_ℓ.

Chapter 5

Monte Carlo application in harmonic space

So far, the proposed approximated method for covariance matrix estimation has been exclusively exploited in Fourier space. Applying it in the case of two-point correlation statistics in configuration space would have been totally feasible without specific adjustment of the method, and a further study could be dedicated to it. In this chapter, I show an application of the Monte Carlo method to a rather less investigated clustering probe, namely the angular matter power spectrum C_ℓ.

As already introduced in chapter 2 and in appendix A, it consists in expanding the density field in spherical harmonics and computing the power associated to each spherical harmonics. The great advantage of this observable is that only the measurement of the three coordinates (z, θ, ϕ) is necessary to estimate it. This way, it can be considered as a true cosmological observable in the sense that its estimation does not imply to suppose any cosmological model to convert measured redshift in comoving distances, as done in the case of the power spectrum or the two-point correlation function (Bonvin & Durrer, 2011; Montanari & Durrer, 2012; Asorey et al., 2012). In addition, this observable is defined on a sphere, simplifying its combinations with other cosmological probes such as lensing (Cai & Bernstein, 2012; Gaztañaga et al., 2012), CMB or/and Hα intensity mapping.

In the first section, I will recall the main properties of the C_ℓ that will be used in this chapter as its theoretical form, its estimator and its associated variance. Then in section 5.2, I will detail three techniques allowing the treatment of the C_ℓ taking the form of shell reconstruction recipes starting from the Monte Carlo outputs. The *shell method* and the *cell method* purpose will be to reconstruct a light cone with varying redshift, while the *snapshell* method will be design to reshape a single snapshot contained in a box into a shell. Finally in section 5.3, I will show a direct application on the estimation of cosmological parameters.

5.1 Some properties of the C_ℓ

The angular power spectrum C_ℓ can be related to the power spectrum $P(k)$ through the following integral

$$C_\ell(z_1, z_2) = (4\pi)^2 \int \mathrm{d}z\mathrm{d}z' W_1(z) W_2(z') \int_0^\infty k^2 \mathrm{d}k P(k) j_\ell[kr(z)] j_\ell[kr(z')] \ , \tag{5.1}$$

where j_ℓ are the spherical Bessel functions and $W_i(z)$ is a radial window function. In practice, the radial window function is used to project the density field in a thick shell over the sky. For any window function shape (Top-Hat, Gaussian, etc, see appendix A), two descriptions can be met. The first is studying the auto-correlation in the (continuous or discrete) field defined in a single shell, equivalent to take $W_1 = W_2$ in eq. 5.1. The second description can both study auto- and cross-correlations in the field whether the window functions W_1 and W_2 are overlapping or not.

In practice, the numerical evaluation of eq. 5.1 is not simple due to the oscillating behaviour of the spherical Bessel functions. For this purpose, we use the `AngPow` software[1] (Campagne et al., 2017; Campagne et al., 2018), which is fully optimised for this task, especially for linear spectra. Indeed it consists in providing the linear power spectrum at $z = 0$, then the algorithm evolves the power spectrum using the linear growth factor such that $P(k, z) = D^2(z)P(k, z = 0)$.

The obtained angular power spectrum C_ℓ may then be compared to the Monte Carlo simulations. To this purpose, it necessary to estimate it from the Monte Carlo realisations. We can simply project the objects of the catalogue on the sky by weighting them according to the radial window function. Thus we can count them in pixels in order to obtain a spherical density map, and compute the spherical power-spectrum $\hat{C}_\ell$ thanks to the `Healpix` software[2](Zonca et al., 2019; Górski et al., 2005) using the parameter nside $= 2^{11}$.

The shot noise contribution in the Top-Hat case for the auto-C_ℓ case is classically $4\pi/N$, where N is the total number of particles falling in the window. Instead, the cross-C_ℓ of two non-overlapping thick shells will naturally be free from shot noise. We refer the reader to appendix I for more detailed information about the shot noise.

In practice, once the harmonic expansion of the density field has been computed, an estimator of the $\hat{C}_\ell$ reads

$$\hat{C}_\ell \equiv \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} \tilde{\delta}^m_{1,\ell} \tilde{\delta}^{m\star}_{2,\ell} \ , \tag{5.2}$$

where the $\tilde{\delta}^m_{i,\ell}$ is the harmonic expansion of the density contrast field. Recalling the definition of the covariance matrix

$$\mathcal{C}_{\ell\ell'} \equiv \left\langle \hat{C}_\ell \hat{C}_{\ell'} \right\rangle - \left\langle \hat{C}_\ell \right\rangle \left\langle \hat{C}_{\ell'} \right\rangle , \tag{5.3}$$

one can express the it for the estimator 5.2. In the auto-C_ℓ case, we immediately see that the first term of equation 5.3 will let a 4-point moment appear, which can be expanded (Fry, 1984a,c) in terms of cumulent moments (or connected expectation values). It follows that it takes the general form

$$\mathcal{C}_{\ell\ell'} = \frac{2C_\ell^2}{(2\ell + 1)} \delta^K_{\ell\ell'} + \bar{T}_{\ell\ell'}, \tag{5.4}$$

---

[1] gitlab.in2p3.fr/campagne/AngPow

[2] healpix.sourceforge.net

where $\bar{T}_{\ell\ell'}$ accounts for non-Gaussian contributions. Instead, when the underlying $\tilde{\delta}_\ell^m$ is a Gaussian field, we can see that the covariance matrix is diagonal. One can notice that expression 5.4 behaves as the one in Fourier space (eq. 2.46). In addition, the Gaussian contribution in the cross-C_ℓ case takes the following form

$$\mathcal{C}_{\ell\ell'} = \frac{C_{\ell,1}C_{\ell,2} + XC_\ell^2}{\sqrt{(2\ell+1)(2\ell'+1)}}\delta^K_{\ell\ell'}, \tag{5.5}$$

where the $C_{\ell,i}$ are the auto-C_ℓ in each shell, while XC_ℓ is the cross-C_ℓ between the two shells. In particular, eq. 5.5 shows that even if the cross-C_ℓ are not sensitive to the shot noise, its corresponding covariance matrix is. Note that the derivation of the covariances are described in appendix J.

In the following, I show how to adapt the Monte Carlo procedure to the study of the C_ℓ.

5.2 Reconstruction methods for C_ℓ estimation

Provided that the Monte Carlo method described in this manuscript allows to generate realisations of a periodic comoving volume (a snapshot) at a constant redshift, one needs to build the past light cone in order to simulate observations. This is also required for the estimation of the angular power spectrum which needs to be done directly on the light cone (see eq. 5.1).

Thus from a comoving output we need to choose where the observer takes place in order to reconstruct its past light cone. This is natural for galaxy surveys but requires to adapt the Monte Carlo output. Also the total size of the comoving boxes is fixed by the period L. However we might want to build the light cone up to some redshift which radial comoving distance could be larger than the size of the comoving output. This apparent issue can be tackled by taking advantage of the periodicity of the Monte Carlo universes. Indeed as they don't have boundaries, we can therefore reconstruct the light cone up to an arbitrary redshift.

However, the cross-correlation (or cross-C_ℓ) between density maps taken at two different redshifts needs to be reproduced, thus we cannot generate many independent comoving outputs at various redshifts in order to reconstruct the past light cone. We detail two ways of proceeding in the following subsections.

5.2.1 Light cone reconstruction : the *shell method*

The first approach is close to the one proposed by Fosalba et al. (2008) in the case of N-body simulations. Adapted for arbitrary nature of power spectrum (linear or not), it consists on gluing a series of simulated comoving volumes, each taken at constant time z, in order to reconstruct the past light cone shell by shell. The main difference with the Monte Carlo method arises from the fact that a N-body simulation is evolving a density field, thus on can naturally take various outputs at given cosmic times (or redshift). Instead in the Monte Carlo case as stressed before, we need to generate an output for each redshift without loosing correlation between them. It is therefore impossible to generate independent realisations at various cosmic times.

In practice, we first select a redshift interval Δz labelled by $z_{\rm min}$ and $z_{\rm max}$ defining the light cone catalogue and a number $N_{\rm shl}$ of shells within it. For each of these shells at intermediate redshift, a point-like distribution is generated in a comoving volume at constant cosmic time z. However we take care of keeping track of the cross-correlation between modes larger than the size of the shell by starting with the same Gaussian field (of power spectrum given by pipeline 3.9), namely with the same numerical random seed, for all the simulated shells.

Obviously, we perform the Poisson sampling only for the grid-cells contributing to the considered redshift shell to optimise computational resources. In addition, we keep only the objects belonging to the comoving volume spanned by the redshift shell, defined by

$$\mathrm{SHL}_i \in [R(z_i - \mathrm{d}z/2), R(z_i + \mathrm{d}z/2)] \ , \tag{5.6}$$

where z_i corresponds to the redshift of the comoving volume labelled by $i \in [1, N_{\rm shl}]$, $\mathrm{d}z = \Delta z / N_{\rm shl}$, and $R(z)$ the radial comoving distance. The sketch in figure 5.1 allows to clarify and illustrate this first reconstruction method.

The light cone method is designed to cover 4π steradians of the sky and any mask or radial selection function can be applied to it, either truncating the shells, or by adapting the local density. In our study case, the focus will be on generating a point-like distribution of constant radial density in redshift. However, the method was originally designed to generate a constant density in comoving space, details of the transformations allowing such feature can be found in appendix K.

Finally, a non negligible effect of the choice of the number of shells $N_{\rm shl}$ used to build the light cone on the angular power spectrum is expected. This will be discussed in section 5.2.3. The higher the number of shells $N_{\rm shl}$, the smoother the temporal continuity from one shell to another, but at the cost of a higher computational time. It is therefore necessary to consider how to choose the number of shells to be simulated. To do so, we will calibrate the method on the basis of a second reconstruction method, called the cell method which is described in the following.

5.2.2 Light cone reconstruction : the *cell method*

The second method is faster and is mainly intended for calibrating the first. Rather than simulating many redshift shells, a single redshift z_0 is selected at the middle of the radial comoving space spanned by the light cone: $z_0 = (z_{\rm min} + z_{\rm max})/2$. The corresponding Gaussian field is then generated in a comoving volume on a grid at $z = z_0$. At this level we need to include some evolution in the radial direction from the point of view of an observer located at the centre of the box. To do so, there are two possibilities.

1. We can simply think of rescaling the Gaussian field at a comoving radial distance $x(z)$ (from the observer) with the corresponding growth factor $D(z)$ (see section 2.20), which rules the evolution of linear matter perturbations. Once achieved, the local transformation $\mathcal{L}$ is applied to obtain the fluctuation field, on which a Poisson sampling is finally performed. By rescaling the Gaussian field ν rather than the cosmological one δ, it is clear that on large scales the power spectrum of the density field will follow the expected evolution

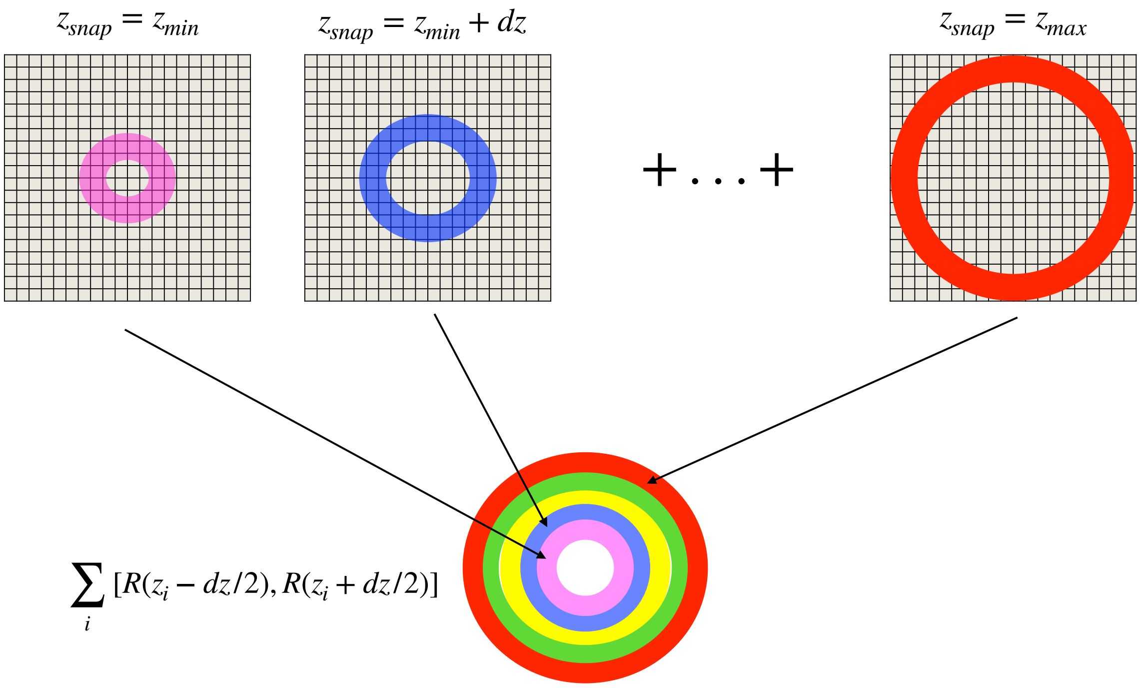


Figure 5.1: Scheme of the shell method: first an interval $[z_{\rm min}, z_{\rm max}]$ is chosen by the user to define the catalogue. $N_{\rm shl}$ snapshots at intermediate redshifts are simulated and the observer is placed at the centre of each box. Shells that correspond to the comoving volume of the redshift intervals 5.6 are selected for each simulated snapshots. Finally, all shells are glued to reconstruct the light cone.

in $D^2(z)$. However, the small scales will be affected in a non-trivial way leading to a modification of the shape of the power spectrum.

2. Rescaling the density contrast field δ rather than the Gaussian one so that the evolution of the density field will rigorously follow the growth factor $D(z)$. For the specific Log-Normal case this would read (see appendix E)

$$\delta_{\rm LN}(\vec{x}, z) = e^{\nu(\vec{x},z) - D^2(z)\sigma^2(0)/2} - 1 \ , \tag{5.7}$$

where the application of the growth factor aims at rescaling the targeted variance of the non-Gaussian field.

This second option is particularly well suited when generating a density field following a linear evolution. Ensuring a perfect redshift-continuity (whose precision is given by the size of the voxels), no bias is expected in the local power spectrum and in the whole angular power spectrum. However the first option, although not rigorously exact, can allow for the fast computation of spectra evolution for more complex cases as when the growth factor $D(k, z)$ depends on the wave number k.

5.2.3 Comparison between light cone reconstruction methods

In this section are compared and discussed, in the linear regime, the shell method and the two cell methods in the specific case of the Log-Normal density field.

We note that as in the figure 4.3, the angular power spectrum is computed on scales smaller than the grid size. On a spherical basis, the equivalent of the Nyquist mode is obtained using $l_N \sim R[z_{\rm mean}]k_N$ (see appendix A), where $z_{\rm mean}$ is the averaged redshift of the particles composing the catalogue.

In figure 5.2 are compared the estimated ΛCDM angular power spectrum in the $z \in [0.2, 0.3]$ redshift range to the predicted one with the shell method described in section 5.2.2 using $N_{\rm shl} = 250$, for one thousand generated light cones. In the lower panel of the same figure we display the relative difference between the two, showing that the agreement is better than the percent level. In order to quantify the impact of the choice of the number of shells set in the shell method, we run a test comparison between the cell method and a various number of shells $N_{\rm shl}$ in the shell method. We note that since we are using a power spectrum that is evolving linearly across cosmic time, it can be expected that the shell method converges to the cell method if the density field is rescaled with the linear growing mode $D(z)$, as described in section 5.2.2.

Figure 5.3 shows the outcome of this analysis, indicating that in the considered redshift range, the shell method is indeed converging to the cell method below the percent level as long as the number of shells is greater than 200 (for reaching the 1% accuracy level in the ℓ-interval displayed).

Finally, we make a comparison within the cell method, rescaling the Gaussian field instead of rescaling the density field. This way we can quantify the deviation when assuming that the Gaussian field evolves linearly as compared to the case when it is the density field that is evolving linearly. In figure 5.4, we show the relative deviation in the two cases with respect to the expected theoretical power spectrum. We can see that the deviation, despite being systematic, remains small (around the percent level).

Therefore, considering these two cell methods, and as stated in the previous section, only the one offering better results will be recommended: the rescaling of the density field (top panel).

Estimating the covariance matrix from $N = 10000$ realisations of angular power spectra, figure 5.5 represents first its diagonal terms with the errors computed using once again the approximation that multipoles are Gaussian distributed following

$$V[C_{ij}] = \frac{C_{ij}^2 + C_{ii}C_{jj}}{N-1} \; . \tag{5.8}$$

Since in the Gaussian case the relative error expected on the diagonal of the covariance matrix elements is given by $\sqrt{2/(N-1)}$, the interest of using such a large

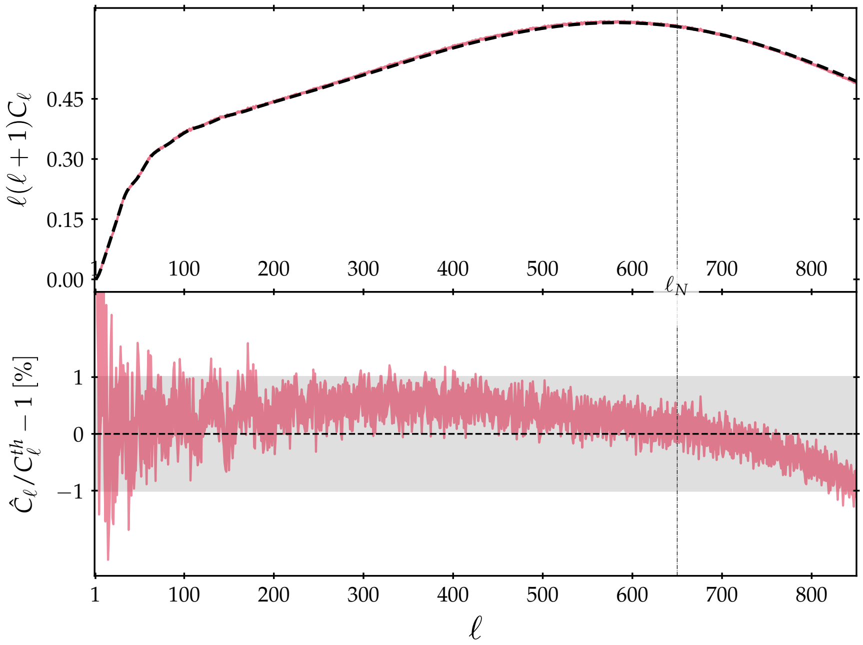


Figure 5.2: *Top panel:* one thousand averaged C_ℓ values for simulated light cones using the shell method with error bars (red curve) and corresponding prediction (dashed black curve). We simulate here a light cone between redshifts $[z_{\min}, z_{\max}] = [0.2, 0.3]$ in a sampling $N_s = 512$ and a number of shells $N_{\rm shl} = 250$ to ensure a sufficient level of continuity in the density field. The spherical Nyquist mode is situated around $\ell_N \sim 650$ and represented by the vertical reference. *Bottom panel:* relative deviation in percent of the averaged predicted C_ℓ values with error bars in red.

number of realisations is that we expect a 1.4% precision on the estimation of the diagonal of the covariance matrix and an absolute precision on the correlation coefficients $r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$ of roughly 0.02. In the bottom panel of figure 5.5 is shown the relative deviation between the Gaussian prediction and the measured variance of the angular power spectrum, indicating that the maximum of deviation is about 45% at $\ell \sim 600$. It appears that deviations from Gaussianity remain small compared to the deviation obtained for power spectrum covariance matrix, which was about two orders of magnitude bigger (see section 3.2.2).

In addition, in figure 5.6 is displayed some off-diagonal covariance elements with their error bars. Despite some fluctuations it is consistent with zero, indicating that the covariance matrix is close to be diagonal, as expected in the Gaussian case (at least for the 300 first elements of the matrix by counting them following the description in the caption).

In order to make sure that this is indeed the case, in figure 5.7 are drawn the correlation coefficients $r_{ij} = C_{ij}/\sqrt{C_{ii}C_{jj}}$; suggesting that the matrix is close to diagonal only considering $\ell < 200$. It therefore confirms that projecting a thick redshift shell onto the sky tends to turn the density field more Gaussian. This is compatible with what it would naively be expected from the central limit theorem

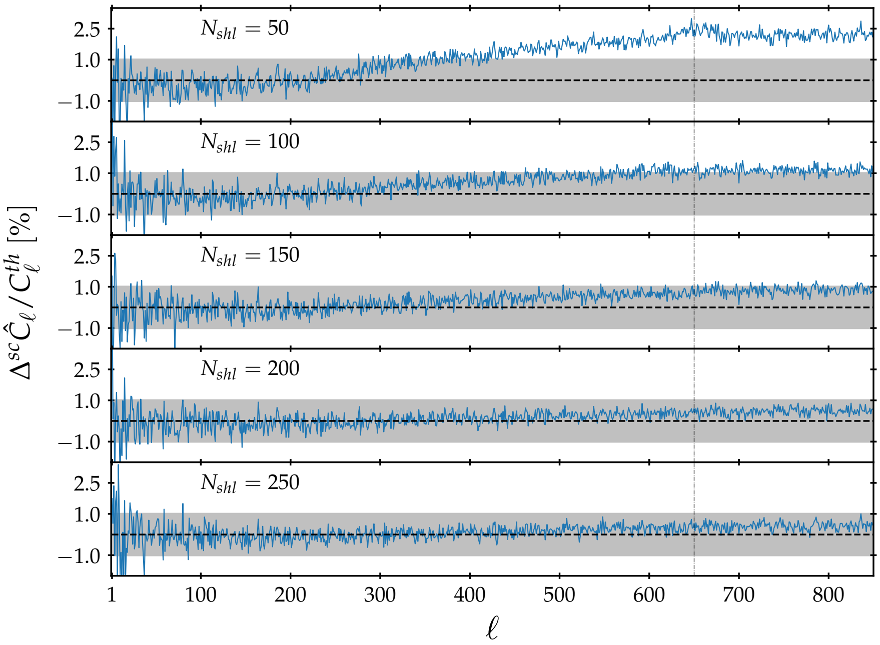


Figure 5.3: Relative difference in percent between shell method and cell method (rescaling δ) for varying numbers of shells. The spherical Nyquist mode is situated around $l_N \sim 650$ and represented by the vertical reference.

since the projection is made by summing over many values of a non-Gaussian field with some weights (see equation 5.2). It appears that the resulting distribution should tend to a Gaussian as the volume of the projection increases. However, for large ℓ-values, a significant amount of correlation is measured, typically of order 10%, reaching 30% at $\ell \sim 600$.

In conclusion, I have shown one fundamental extension of the Monte Carlo pipeline, the construction of a light cone using the shell method which has been validated using the cell method. For further studies, it would be opportune to apply the light cone reconstruction to non-linear power spectrum.

5.2.4 C_ℓ estimation on snapshot : the *snapshell* method

As already mentioned, the method validation in harmonic space will use once again the `DEMNUni_cov` simulations. However, we have access to five snapshots, thus we cannot reconstruct a light cone out of the 50 `DEMNUni_cov` simulations. Indeed the five snapshots being separated by a distant time interval ($z_{\rm snap} = [0, 0.5, 1, 1.5, 2]$), using the shell method would be a too crude approximation as shown in the top panels of figure 5.3.

The most straightforward way of proceeding is to compute the angular power spectrum on thin shells centred at the comoving distance corresponding to the red-

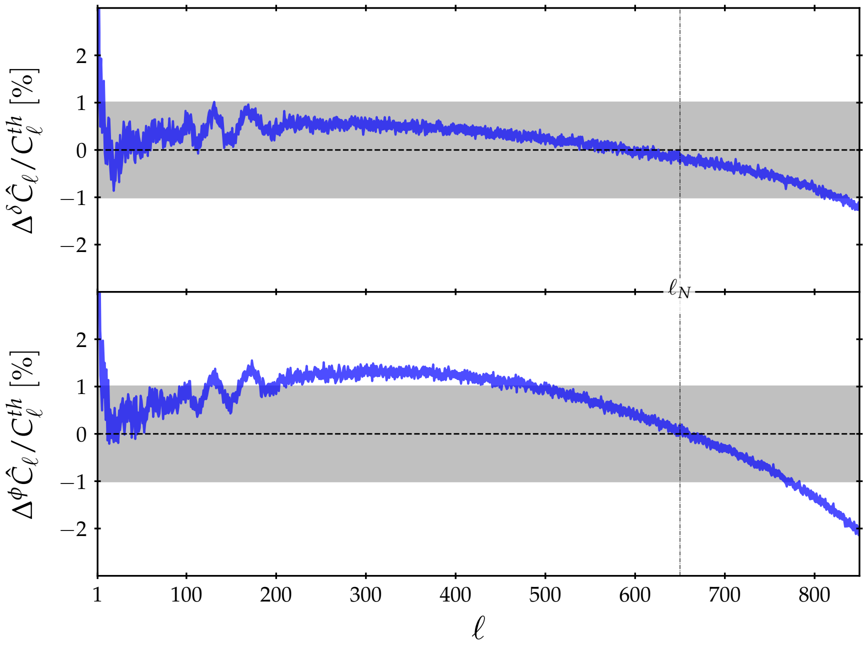


Figure 5.4: Relative deviation in percent with error bars for 10000 averaged realisations of C_ℓ values in the context of the cell method. In the top panel, the density field (non-Gaussian) is rescaled using linear growth function, while in the bottom panel the Gaussian field following the Gaussian power spectrum is rescaled. The spherical Nyquist mode is situated around $l_N \sim 650$ and is represented by the vertical reference.

shift of each comoving output. Of course, this is feasible at all redshifts except the $z = 0$ snapshot for which this is not making sense.

Moreover, an other problem is that the size of the simulation might be smaller than the radial comoving distance associated to each snapshot (to each redshift). Indeed, the `DEMNUni_cov` simulation has a size of $L = 1h^{-1}$Gpc while the lowest useful redshift $z = 0.5$ corresponds to a radial comoving distance $R_{\text{snap}}(z = 0.5) = 1312h^{-1}$Mpc. Thus we need to apply the periodic boundary conditions in order to reconstruct properly the shell.

We also need to fix a criteria to set up the thickness of the shell, for simplicity we require the shell to have the same volume as the comoving output, this way we can take advantage of the whole simulation volume and reach the same shot noise level (the total number of particles is conserved). We are thus able to build a continuous 4π steradians shell of radius $R(z)$ and of thickness e out of each snapshot.

In practice, we need to find the expression of the thickness of the shell $e(L)$ as a function of the size L of the snapshot. By imposing that the volume of each shell is equal to the total volume of the simulation we get

$$L^3 = \frac{4}{3}\pi \left[\left(R_{\text{snap}} + \frac{e}{2} \right)^3 - \left(R_{\text{snap}} - \frac{e}{2} \right)^3 \right] . \tag{5.9}$$

The above equation is a polynomial of order 3 in the thickness e and can be inverted

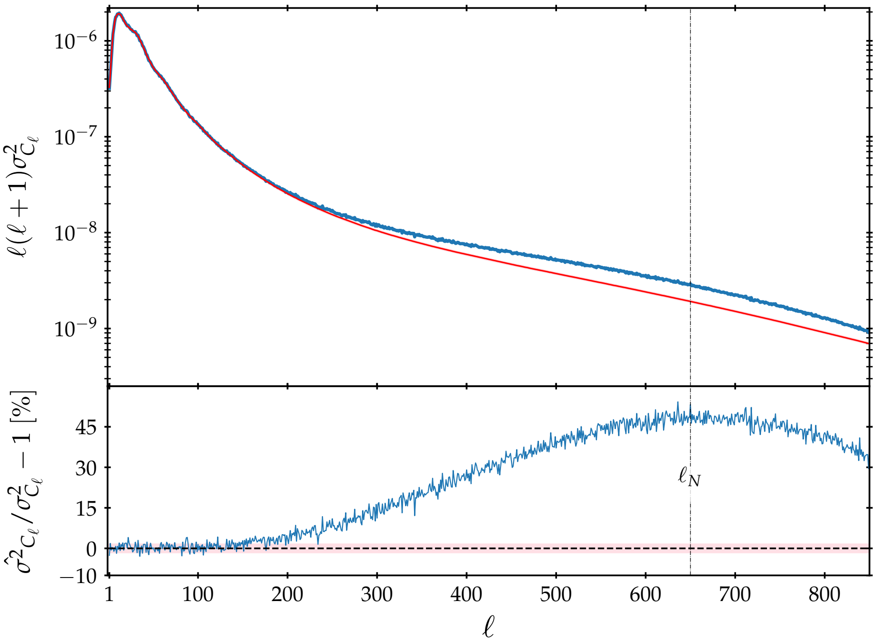


Figure 5.5: *Top:* measured diagonal of the covariance matrix (blue curve) over $N = 10000$ realisations of different light cones. The red curve represents the associated prediction in the case of a Gaussian field with errors computed using eq. 5.8. Here we keep the shot noise effect in the measures and include it in the prediction. The spherical Nyquist mode is situated around $l_N \sim 650$. *Bottom:* relative difference in percent following the same colour-coding.

to express

$$e = \alpha - 4\frac{R_{\text{snap}}}{\alpha} \ , \tag{5.10}$$

where $\alpha = \left[\left(S + \sqrt{256R^6_{\text{snap}} + S^2}\right)/2\right]^{1/3}$ and $S = 3L^3/\pi$.

At this point, we can on one hand estimate the angular power spectrum in each 50 `DEMNUni_cov` simulation and for the 4 redshifts, and on the other hand estimate it in the corresponding Monte Carlo realisations of the `DEMNUni_cov`.

In figure 5.8 are shown the spectra of the Monte Carlo realisations compared to the spectra of the `DEMNUni_cov` for various redshifts. One striking feature of figure 5.8 is the fact that the mean of the Monte Carlo realisations is precisely following the fluctuations of `DEMNUni_cov`. Indeed, as in chapter 4, I used the averaged power spectrum of the `DEMNUni_cov` as target for the Monte Carlo simulations. In addition, I generated many more Monte Carlo realisation (1000 per redshift), that is the reason why residual fluctuations due to the average made over 50 realisations of the `DEMNUni_cov` are reproduced by the Monte Carlo realisations.

Figure 5.8 allows to assess up to which multipole ℓ the Monte Carlo is able to reproduce the N-body simulation. We define the maximum ℓ as the multipole ℓ at which the difference between the two is equal to the 1-sigma error on the

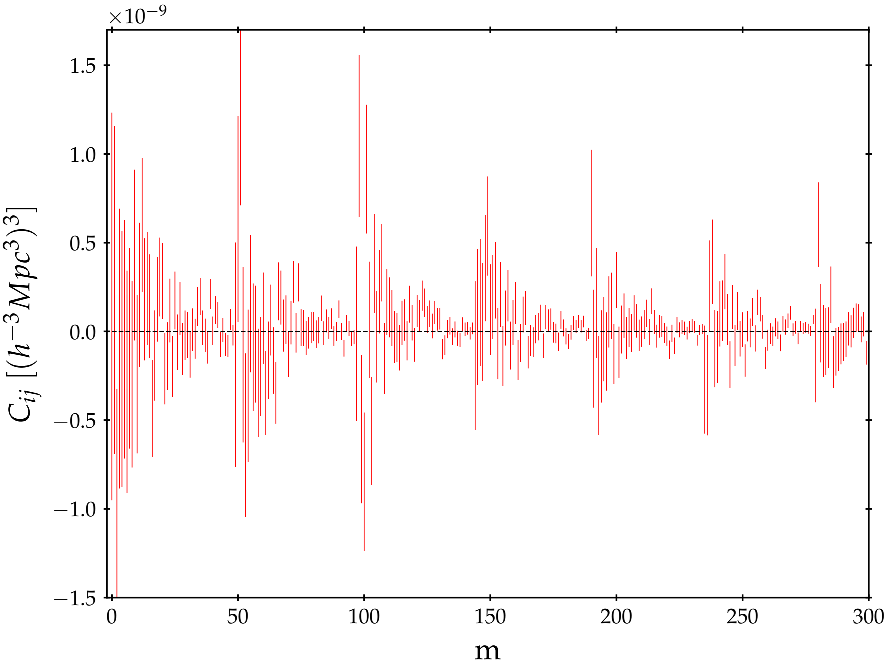


Figure 5.6: First 300 elements measured for the off-diagonal part of the covariance matrix over $N = 10000$ realisations of light cone with Gaussian errors computed using eq. 5.8. The elements are labelled by the index m and are ordered column by column in the lower half of the matrix without passing by the diagonal, i.e. $C_{ij,\ i>j}$.

estimation of the angular spectrum (in a single realisation). This way we know the multipoles that should not be used if one desires to perform a cosmological parameter inference. This limit is represented with vertical dotted line in the lower panels, obtaining for each redshift $\ell \sim [573, 1073, 1308, 1682]$. As a comparison, the typical maximum wave number $k_{\rm max}$ taken into account in the power spectrum analysis is $k_{\rm max} \sim 0.2h/{\rm Mpc}$. In order to ease the comparison, this wave number can be roughly converted into a maximum $\ell_{\rm max}$ for each redshift, corresponding to $\ell_{\rm max} = [256, 474, 591, 722]$. This is ensuring that we would perform the cosmological parameter estimation in a range of multipoles for which the angular spectrum is very well reproduced by the Monte Carlo (this limit is shown in vertical solid line on figure 5.8). Finally, I compared the prediction of the angular spectrum when using the `HALOFIT` prescription for predicting the non-linear matter power spectrum. We can see that the predictions (in dotted black lines) are in very good agreement with the measurements, at least up to the multipoles corresponding to $k \sim 0.2h/{\rm Mpc}$.

When applying the same debiasing as in the relation 4.16 in the case of the C_ℓ simulated by the Monte Carlo method, figure 5.9 shows the estimated diagonal of the covariance matrix for the four studied redshifts, both for the `DEMNUni_cov` and for the Monte Carlo outcomes. As for the C_ℓ in figure 5.8, it shows a good agreement between the two variances, even beyond the multipoles corresponding to $k \sim 0.2h/{\rm Mpc}$. In particular, it shows a high reduction of the covariance noise thanks to the 1000 simulated C_ℓ. In addition is plotted the Gaussian prediction

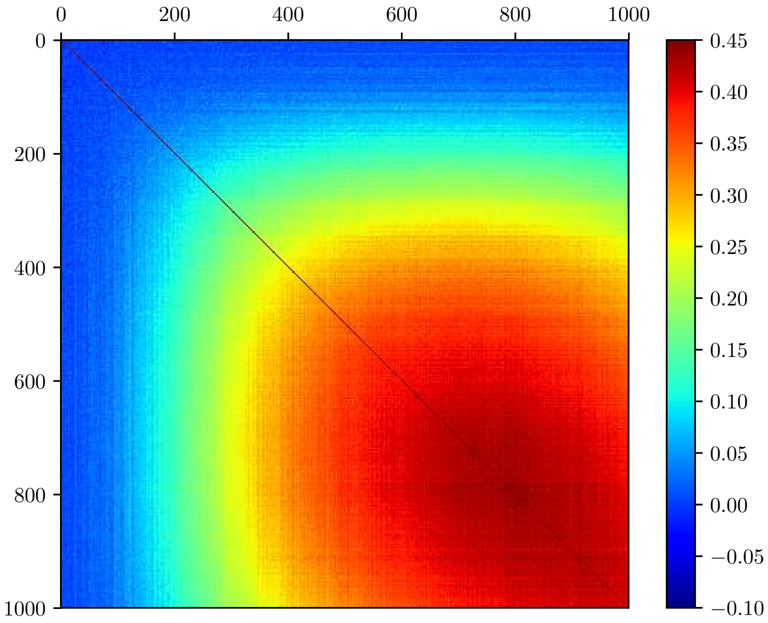


Figure 5.7: Correlation matrix for 10000 realisations of C_ℓ in a simulated universe between redshifts 0.2 and 0.3 and a sampling $N_s = 512$. The $(\ell \times \ell) = (1000 \times 1000)$ correlation matrix is represented here.

(first term of eq. 5.4) computed using the mean value of the C_ℓ estimated on the Monte Carlo realisations, clearly showing a Gaussianisation of the covariance when comparing to the its Fourier equivalent (see figure 4.11), where at least one order of magnitude differentiated the Gaussian to the non-Gaussian variances at $k \sim 0.2h/\text{Mpc}$. By recalling that the non-Gaussian part constitutes one of the big issue of covariance prediction, it appears that probing the C_ℓ (whose the non-Gaussian contribution is very law) rather than the $P(k)$ (where the non-Gaussian contribution is significant) provides therefore one another great advantage.

Note that if one wants to compare the power spectrum to the angular power spectrum, it is clear that the angular power spectrum is losing the radial information, while the power spectrum capture the full 3-D clustering. However, in angular analysis one can partly access to the radial information by cross-correlating shells at two different redshifts, thus measuring the cross-angular power spectrum XC_ℓ. This can be adapted to snapshots by separating in two sub-shells the shell obtained for the C_ℓ estimation. In doing so as already discussed, no shot noise will contribute to the estimated cross-angular spectrum, as opposed to the C_ℓ case. On the other hand, the corresponding covariance matrix is obtained using eq. 5.5. But as it depends on the amplitude of the respective C_ℓ of the two shells, the variance is still affected by shot noise which will naturally be increased. Thus the division of the original shell in two halves makes the two resulting auto-C_ℓ of higher amplitude due to 1) the fact that more shot noise is expected than in the initial shell (the number of particles is divided by two) and 2) reducing the radial window reduces the study

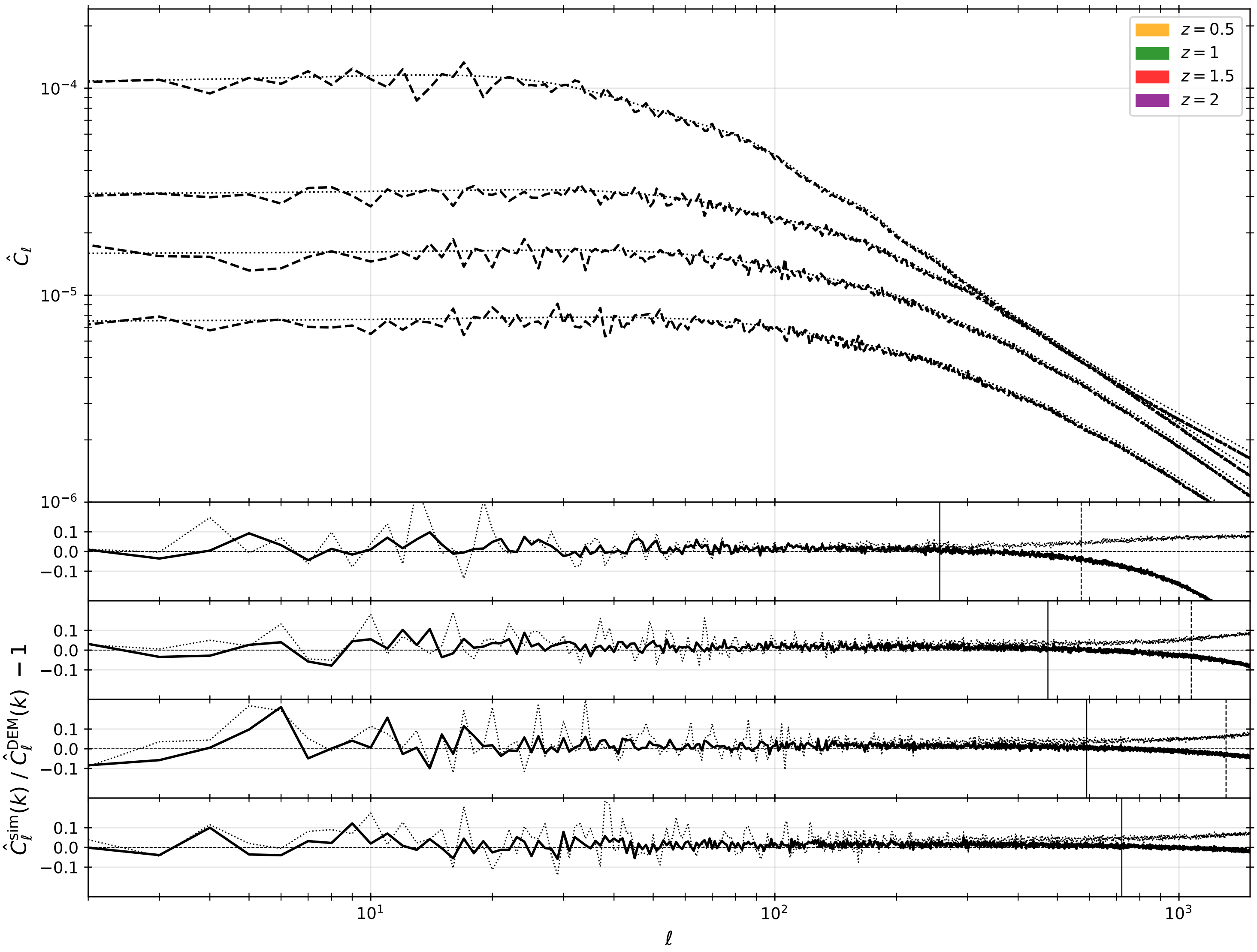


Figure 5.8: *panel 1:* estimated and averaged auto-C_ℓ with error-bar in colour for 1000 realisations of Monte Carlo catalogues. Overlapping dashed black lines are representing the averaged measures made on `DEMNUni_cov` while the dotted black lines are standing for the predicted C_ℓ using `HALOFIT`. *panels 2,3,4,5:* Relative deviation of the predictions (dotted black lines) and the Monte Carlo outputs (solid black lines with error-bars) with those of the N-body, using the same colour-coding. The vertical dashed lines shows the modes for which the Monte Carlo and N-body experiments become incompatible at more than $1-\sigma$, respectively $\ell \sim [573, 1073, 1308, 1682]$. The vertical dashed lines represent the multipole limits for which the parameter estimation is performed : $\ell_{\rm max} = [256, 474, 591, 722]$ that correspond to $k_{\rm max} \sim 0.2h/{\rm Mpc}$.

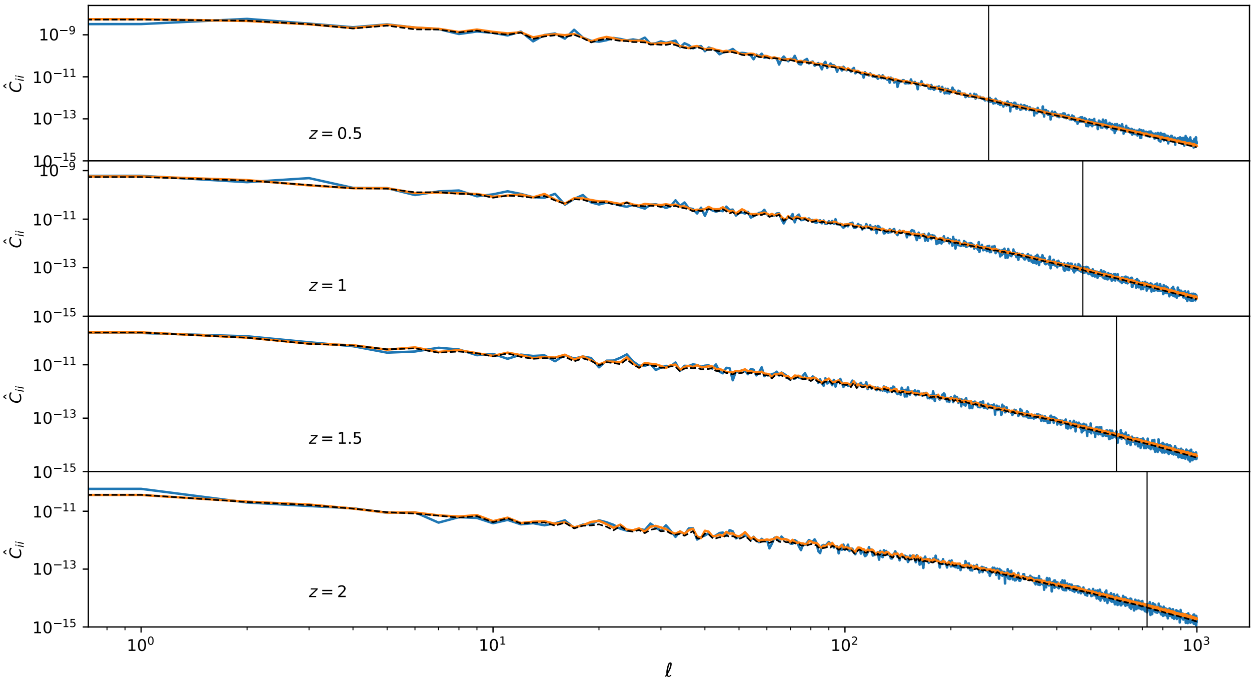


Figure 5.9: Estimated diagonal of the covariance matrix over 1000 realisations of C_ℓ using the Monte Carlo method (solid orange line) as compared to the one estimated over the 50 realisation of C_ℓ from `DEMNUni_cov` (solid blue line). In dotted black lines are shown the corresponding Gaussian prediction (see eq. 5.4) computed on the averaged Monte Carlo C_ℓ. The vertical lines represents, for the four redshifts, the multipoles corresponding to $k = 0.2h/\mathrm{Mpc}$.

to highly correlated particles (going toward the Dirac limit of thin shell). Thus, the thinner are the shells the higher will be the cosmological signal-to-noise but the higher will be the Poisson noise. It is then interesting to find the shell thickness which maximises the total signal-to-noise (cosmic variance and Poisson noise).

Before, we need to set the distance between the two shells. In the case of the Monte Carlo realisations, we start on a grid of finite size which basically set the scale below which the clustering cannot be reproduced. As a result, it is of poor interest to try to cross-correlate two thin shells which are separated by less than the resolution of the Monte Carlo. Anyway, here the resolution is of the order of the h^{-1}Mpc which is still better than the redshift errors that are achieved by galaxy surveys. As a result, we need to set a separation in between the two shells at least by the size of a Monte Carlo voxel to discard correlations between particles that are coming from the same cell. Even if voxel are not uniformly populated in tri-linear interpolation scheme (see section 4.1), such condition may lead to a lack of correlation for all multipoles.

Once the distance is set, we can optimise the choice of the thickness, in figure 5.10, I apply successive factor γ on the two shells volumes separated by the maximal size of a Monte Carlo cell ($\sqrt{3}L/N_s$), in order to find out the optimal volume such that the

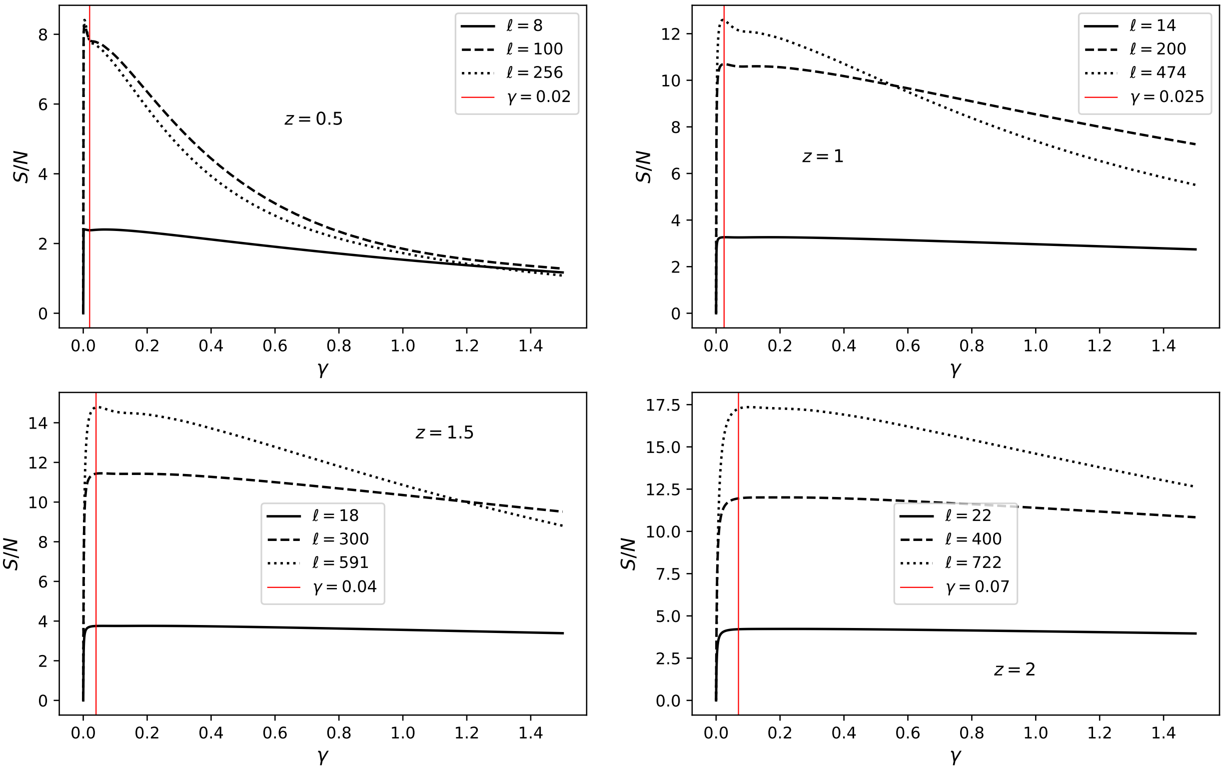


Figure 5.10: Signal-to-noise ratio for several shell volumes such that $V'_{\rm shl} = \gamma V_{\rm shl}$. The signal is the predicted cross-C_ℓ between two volume-flexible shells separated by one Monte Carlo cell $s = \sqrt{3}L/N_s$, while the noise is the square root of the predicted Gaussian variance 5.5. This quantity is given for three multipoles $\ell(z) = [\ell_F(z), \ell_{int}(z), \ell_{max}(z)]$ where $\ell_{max} = R(z)k_{max}$ setting $k_{max} = 0.2$h/Mpc. Here $\ell_{int}(z)$ is arbitrarily set roughly at half-way between ℓ_{min} and ℓ_{max}. The red vertical references shows the chosen γ that allows to reach a high S/N while minimising the shell volume to be reconstituted.

signal-to-noise ratio (S/N) is maximised. This exercise is performed theoretically with `AngPow` and assuming a Gaussian covariance matrix.

It first shows a similar behaviour between all redshifts. In the cross-C_ℓ case (recalling that the shot noise is computed from the `DEMNUni_cov` density), small volumes provide a better signal-to-noise ratio, leading to some non-intuitive situation where the volume must be truncated to improve the analysis quality. S/N reaches a first peak (more pronounced at low z) where the Dirac limit of thin shells is obtained. As the S/N quantity is described by the same behaviour from one ℓ to another, I will set γ arbitrarily such that the whole ℓ-range is roughly maximised (apart from $z = 0.5$ where the peak is dangerously close to $\gamma = 0$). Thus I will set for the four consecutive redshifts the values $\gamma = [0.02, 0.025, 0.04, 0.07]$, corresponding to shell-thicknesses of $\sim [0.48, 0.17, 0.18, 0.21]h^{-1}$Mpc.

Note that shrinking the volume using γ in the cross-C_ℓ case allows to reduce the computational cost[3], since not the whole box is involved in the reconstruction method. In the same spirit as for the shell method, the Poisson sampling and the particle coordinate assignment is only performed in the voxels which participate in the shell reconstruction. It turns out that only 12% of the cells contribute at redshift $z = 0.5$, and increases when reconstructing larger shells : 32% at $z = 1$, 43% at $z = 1.5$ and 53% at $z = 2$. It reduces both memory and CPU-time in the same way.

Figure 5.11 features the cross-C_ℓ outcomes following the previous procedure. Here, the Monte Carlo, the N-body and their corresponding prediction are all compatible (at $1 - \sigma$) up to multipoles slightly lower than the mode of reference $k \sim 0.2h/\mathrm{Mpc}$, namely $k_{\mathrm{max}} \sim 0.15h/\mathrm{Mpc}$, which corresponds to the multipoles $\ell_{\mathrm{max}} = [192, 356, 443, 542]$.

Also in the same spirit of figure 5.9, figure 5.12 is showing the diagonal of the covariance matrix in the cross-C_ℓ case. Here, even if the level of matching between the Monte Carlo and `DEMNUni_cov` is still high, a slight deterioration with respect to the auto-C_ℓ can be noticed. However, on the basis of the diagonal of the covariance, the cross-correlation seems very well reproduced up to $k \sim 0.15h/\mathrm{Mpc}$. This is the reason why in the next section for parameters estimation, the cross-C_ℓ analysis will be carried out up to $\ell_{\mathrm{max}} = [192, 356, 443, 542]$, corresponding to $k_{\mathrm{max}} \sim 0.15h/\mathrm{Mpc}$.

Concerning the Gaussian prediction following eq. 5.5, it would have been totally feasible to predict it using the averaged XC_ℓ and C_ℓ for the two shells. However I did not estimated the auto-C_ℓ (I didn't need it for the following) and I rather plotted a theoretical prediction whose details will be discussed in the next section. Anyway, it shows also in the cross-C_ℓ case a non-Gaussian contribution highly subdominant.

5.3 Application on cosmological parameter estimation

The purpose of this section is to apply the Monte Carlo procedure in the concrete case of cosmological parameter estimation (see section 2.3.1) using the angular power spectrum. For the present test I will use the angular spectra estimated in 4 realisations (one for each redshift) of the `DEMNUni_cov` as observed data. Those data will be analysed with various covariance matrices: the Gaussian covariance, the diagonal of the `DEMNUni_cov` covariance, the diagonal of the Monte Carlo covariance and the full Monte Carlo covariance matrix. For simplicity, I restrict the analysis of the auto and cross-C_ℓ to comoving space, without implementing redshift-space distortions.

5.3.1 Setting

In the auto-C_ℓ case, the same setting as in the method validation in Fourier space will be kept for the Monte Carlo: *i.e.* $N = 1000$ simulated catalogues in ΛCDM cosmology with massive neutrinos ("16nu"), with parameters $N_s = 1024$, $L = 1000h^{-1}\mathrm{Mpc}$, $\rho = 0.1h^3\mathrm{Mpc}^{-3}$ and a tri-linear interpolation scheme for particle coordinates assignment. In the cross-C_ℓ case, the optimal volumes (see γ factors above) was

---

[3] Simulating one snapshot with the same density as for `DEMNUni_cov` $\rho = [(1024/1000)h/\mathrm{Mpc}]^3$ weights $\sim$ 25GB while $\sim$ 1.7 CPU-hour on 56 threads are necessary.

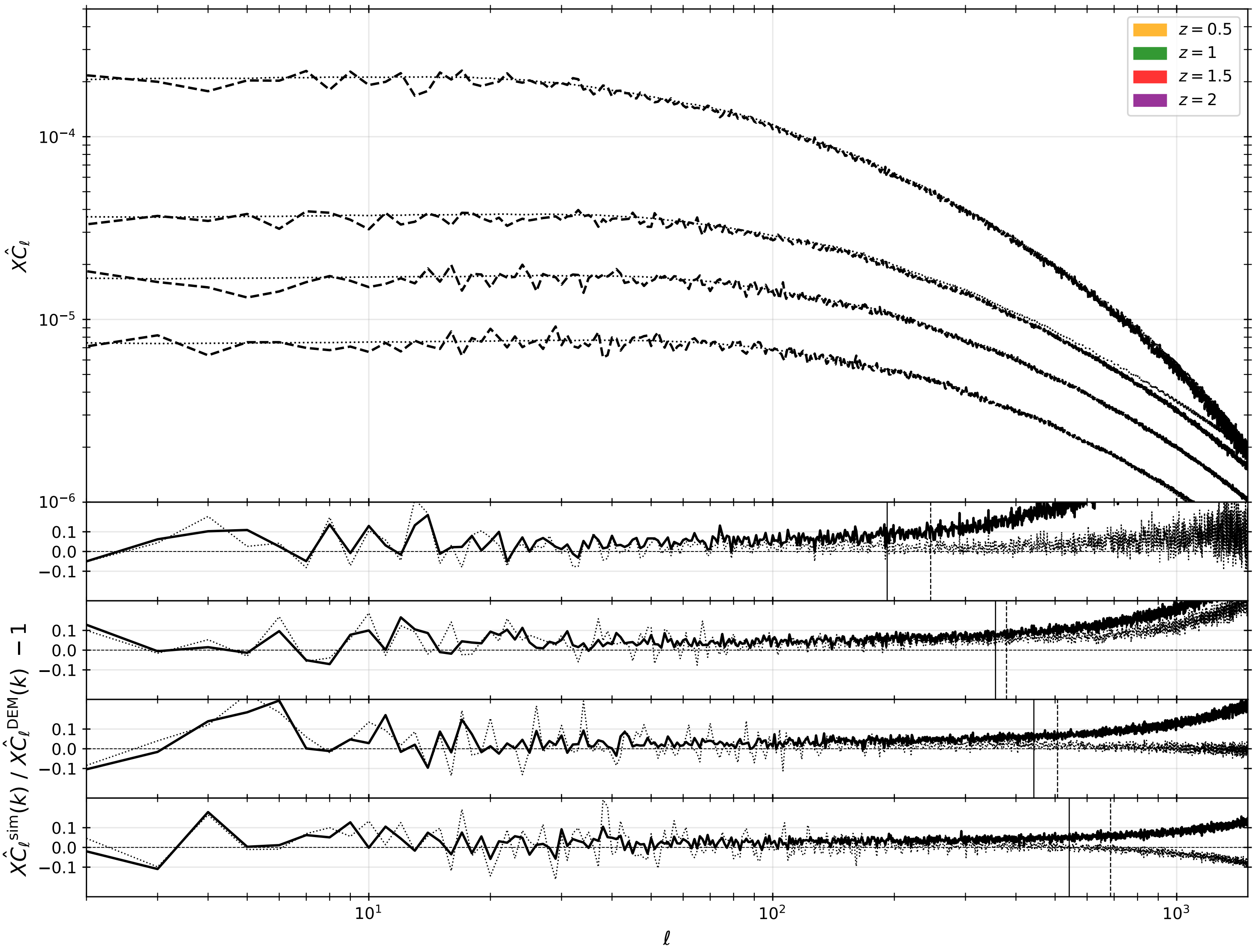


Figure 5.11: *panel 1:* estimated and averaged cross-C_ℓ with error-bar in colour for the Monte Carlo method. Overlapping dashed black lines are representing the averaged measures made on `DEMNUni_cov` while the dotted black lines are standing for the predicted cross-C_ℓ using `HALOFIT`. *panels 2,3,4,5:* Relative deviation of the predictions (dotted black lines) and the Monte Carlo outputs (solid black lines with error-bars) with those of the N-body, using the same colour-coding. The vertical dashed lines shows the modes for which the Monte Carlo and N-body experiments become incompatible at more than $1-\sigma$, respectively $\ell \sim [246, 379, 507, 686]$. The vertical dashed lines represent the multipole limits for which the parameter estimation is performed : $\ell_{\max} = [192, 356, 443, 542]$ that correspond to $k_{\max} \sim 0.15h/\mathrm{Mpc}$.

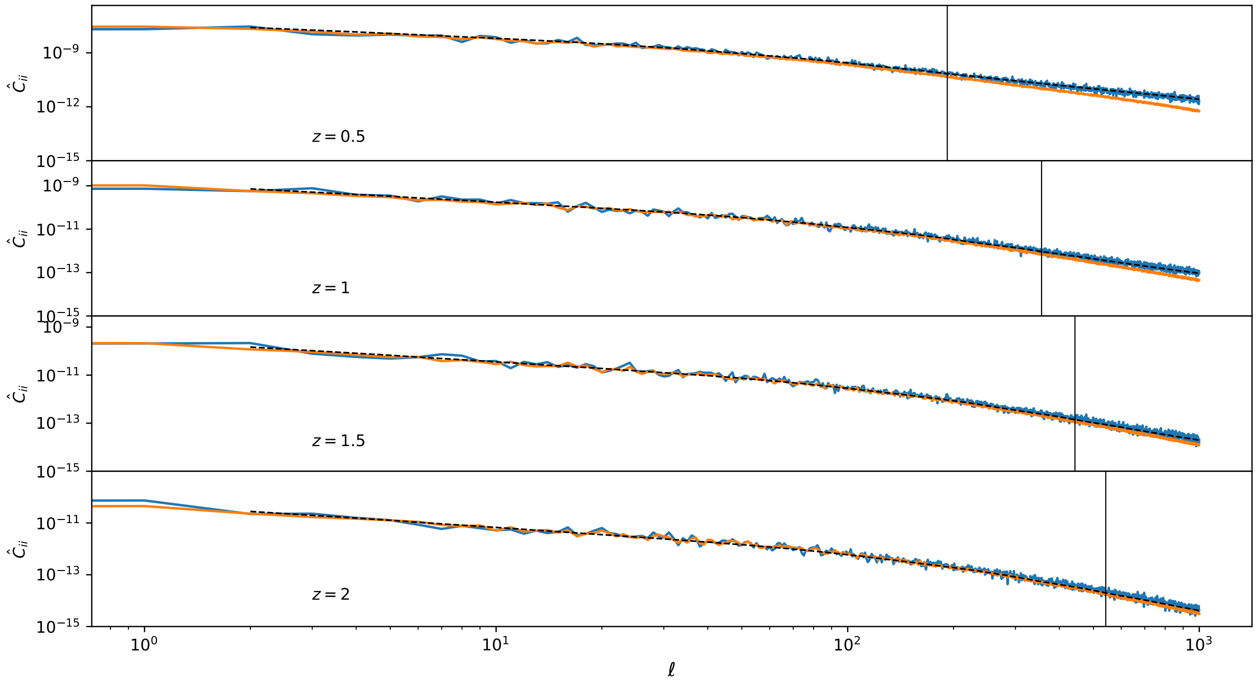


Figure 5.12: Estimated diagonal of the covariance matrix over 1000 realisations of cross-C_ℓ using the Monte Carlo method (solid orange line) as compared to the one estimated over the 50 realisation of the cross-C_ℓ from `DEMNUni_cov` (solid blue line). In dotted black lines are shown the corresponding Gaussian prediction (see eq. 5.5) theoretically computed. The vertical lines represents, for the four redshifts, the multipoles corresponding to $k = 0.15h/\mathrm{Mpc}$.

determined in the `DEMNUni_cov` density case. That is why for a fair comparison, we set the Monte Carlo density also at $\rho = [(1024/1000)h/\mathrm{Mpc}]^3$, applying the same volume truncation. Apart from this density modification, the other parameters are set to the same values as in the auto-C_ℓ case.

Data

The idea is to fit one `DEMNUni_cov` simulation per redshift. But since in the N-body context, snapshots are representing the same simulation at different redshifts, one expects for example that the `DEMNUni_cov` of the same number (`01,02,...,50`) at $z = 0.5, 1, 1.5$ and 2 are not realistically correlated. One can circumvent the problem by picking one `DEMNUni_cov` simulation per redshift that are coming from different initial conditions. Therefore in the following I will take `DEMNUni_cov10` as data for $z = 0.5$, `DEMNUni_cov20` for $z = 1$, `DEMNUni_cov30` for $z = 1.5$ and `DEMNUni_cov40` for $z = 2$. This way, it ensures that the four elements (four redshifts) composing the total data vector are not correlated.

Also given that we construct the shells with the snapshell reconstruction method and even if periodic boundary conditions were set, it seems ambiguous to use the smallest multipols corresponding to the size of the initial box. Indeed this might introduce more correlation than expected through the fact that particles are repeated periodically. That is the reason why, in the analyses that will follow, this aspect will be considered by taking $\ell < \ell_F$ out, where $\ell_F \sim R(z_{\rm snap})k_F$. For the four consecutive redshifts, the minimum multipoles that will be considered will therefore be set to $\ell_F = [8, 14, 18, 22]$.

Covariances

Since data are not correlated between redshifts, the covariance matrix can be constructed in this case as block diagonal, where each block corresponds to the predicted or estimated covariance matrix per redshift, as it has been done so far. As anticipated, four types of covariance matrices will be tested:

1. Using the Gaussian variance eq. 5.4 (first term) for the auto-C_ℓ and eq. 5.5 for the cross-C_ℓ computed using the predictions of the auto- and cross-C_ℓ on the fiducial cosmology (see the `DEMNUni_cov` setting in section 2.4.1). Note that the shot noise contribution is added to the amplitude of the auto-C_ℓ only in eq. 5.5. Indeed, there is no need to add shot noise in eq. 5.4 since it has been removed from the measurements.

2. Using the diagonal only of the covariance matrix estimated over the $(50 - N_z)$ `DEMNUni_cov` simulations. Here N_z is the number of measurements that will be peak for the data fitting ($N_z = 4$) and that cannot be used for the covariance matrix estimation to avoid correlation between data and error. Also there is a reason why I'm only focusing of the diagonal only. As reported in chapter 2, taking the full covariance matrix will led to a biased estimation of the precision matrix $\hat{\Psi}_{ij} = bC_{ij}^{-1}$ where $b = \nu\,(\nu - K - 1)^{-1}$ is the Hartlap factor, $\nu = N - 1$ and K the number of binned modes that are fitted. This theoretical bias is correct as long as it is positive, *i.e.* $K < N + 2$. Going up to an $\ell \sim (50 - N_z)$ corresponding to $k \sim 1.5 \times 10^{-2} h/\mathrm{Mpc}$ at $z = 2$ offers only limited interest in such an analysis where we need to go up to $\ell \sim 700$ at $z = 2$ for reaching $k \sim 0.2 h/\mathrm{Mpc}$. A re-binning or eventually picking only $K = N + 1$ multipoles in the wanted ℓ-range would be conceivable, but since off-diagonal terms are very close to zero in harmonic space, one may prefer to use the diagonal only of the covariance matrix providing a different and advantageous precision matrix bias

   $$\hat{\Psi} = \frac{N-3}{N-1} C^{-1} \,, \tag{5.11}$$

   that do not limit the maximal number of modes to be took. In the same way as the Hartlap factor, obtaining such a bias parameter 5.11 only invokes that the covariance matrix elements follow a Wishart distribution (given for a Gaussian likelihood).

3. For the method validation between `DEMNUni_cov` and the Monte Carlo, the parameter estimation will also be performed with the diagonal of the covariance matrix estimated on $(50 - N_z)$ Monte Carlo realisations. Note here that the

N_z in $(50 - N_z)$ is optional as the data (coming from `DEMNUni_cov`) are in any case not correlated to the errors (estimated from the Monte Carlo). However as previously, the precision matrix must be de-biased in the same way as for the power spectrum in eq. 5.11.

4. Using the full covariance matrix estimated from the Monte Carlo simulations, and whose the precision matrix can be unbiased, that time using the Hartlap factor (see section 2.3.1). In this case, the large number of realisation N allows us to not be constrained by the Hartlap factor condition.

Prediction

Now that data to fit and their associated errors have been defined, the way the prediction of the auto- and cross-C_ℓ will be conducted must be discussed.

In particular, it is worth pointing out that the `AngPow` software is specifically suited to perform eq. 5.1, *i.e.* integrating over the redshift and not over comoving distances. In the context of a true light cone the integration must, indeed, done over redshift. However, the shells with which we are currently dealing with are not light cones, since they are coming from constant-time snapshots. In practice, the `AngPow` algorithm asks the user to provide the linear matter power spectrum $P^{\rm lin}(k, z = 0)$ at redshift $z = 0$ in order to evolve using the mode-independent linear growth factor $D(z)$ proposed in Eisenstein (1997) (equivalent to the one derived in chapter 2), such that $P(k, z) = D^2(z)P^{\rm lin}(k, z = 0)$. Thus it does not support the prediction of the angular power spectrum in the non-linear regime, whereas we want to use the `HALOFIT` prescription in our fitting process. However, since we are dealing with thin shell, one can neglect the redshift dependence D across the shell. Thus, rather than providing the `CLASS` non-linear power spectrum at $z = 0$, one can provide it at the central redshift of the shell but divided by the linear growth factor $\tilde{P}(k, z = 0) \equiv P^{\rm nonlin}(k, z_{\rm snap})/D^2(z_{\rm snap})$. This way, the software will evolve this spectrum linearly up to the shells situated at $R(z_{\rm snap})$, where the thickness allows to approximate $z = z_{\rm snap}$ for the whole shell.

However, in order to assess the impact of the growth factor across the shell as well as the redshift-to-comoving distance in the Bessel term $j_\ell(kr(z))$, we decided to develop a prediction code of auto- and cross-C_ℓ specifically suited for this exercise, respecting all of the above conditions: non-linear power spectrum, growth factor and redshift-to-comoving distance constant throughout the shell.

5.3.2 Preliminary results

In order to sample the posterior distribution introduced in chapter 2, we run some Monte Carlo Markov Chain (MCMC) using the `Monte Python`[4] software (Audren et al., 2013) with adaptative Metropolis algorithm (`superupdate` parameter set to 20).

In this fitting, four parameters are free to vary

---

[4] brinckmann.github.io/montepython_public

$$
\begin{aligned}
&\Omega_{\rm b}h^2 \equiv \omega_{\rm b} \in [5\times 10^{-3}, 3.9\times 10^{-2}] \ , \\
&\Omega_{\rm cdm}h^2 \equiv \omega_{\rm cdm} \in [4.5\times 10^{-3}, 3.6\times 10^{-1}] \ , \\
&h \in [0.3, 1.5] \ , \\
&\sum m_\nu \in [0, 1]\mathrm{eV} \ ,
\end{aligned}
$$

a setting that corresponds to flat priors.

Note that the following results are preliminary. In figures 5.13 and 5.14 are plotted the outputs of the MCMC, namely the marginalised posterior 2-D ellipses representing the one and two-standard deviations (as well as the 1-D histograms) of the posterior distributions. In these panels, we can see that the estimated parameters are incompatible ($> 2\sigma$) with the fiducial parameters (true values) $\omega_{\rm b} = 2.2\times 10^{-2}$, $\omega_{\rm cdm} = 1.2\times 10^{-1}$, $h = 0.67$ and $\sum m_\nu = 0.16$ eV → $m_{ncdm} = 0.053$ eV. Moreover, this incompatibility between the two kinds of data (auto and cross-C_ℓ) appears incoherent : the whole set of parameters are under-estimated with respect to the true cosmology in the auto-C_ℓ case, while it is the opposite behaviour in the cross-C_ℓ case.

Putting aside this mismatch, it appears that the three diagonal covariances are providing the same level of error on parameters, with a quasi-similar constraints when looking at the 1-D posterior distributions. This result is not surprising since, as we have seen in figures 5.9 and 5.12, the three variances present in overall the same amplitude (Gaussian, diagonal of `DEMNUni_cov` and diagonal of Monte Carlo). However, a different behaviour can be observed with the ellipses produced by the full shape of the Monte Carlo covariance matrices. Indeed, the ellipses are all enlarged and shifted, showing that the off-diagonal elements (as well as the noise in the covariance) can affect both the best fit and the cosmological parameter constraints. This result on the other side is a bit more surprising since one would expect that the Gaussianisation of the covariance previously discussed, leading to small off-diagonal terms, would have no impact on the MCMC outcomes. As a result, it seems important to take into account the correlation coefficients.

Moreover, discussing the reliability of the best fit often use the reduced χ^2 defined as $\chi^2_r = \chi^2/n_{\rm dof}$ where $n_{\rm dof} = K - 1 - N_p$. Here K is the length of the data vector and N_p the number of fitted parameters. Following a rule of thumb, a good fit would provide $\chi^2_r \sim 1$, while $\chi^2_r \ll 1$ signifies that error bars are over-estimated and $\chi^2_r \gg 1$ that either the model is not a good fit (need for example more free parameters) or uncertainties are under estimated (or the analysis is subject to systematic errors). Table 5.1 gives the minχ^2 (or $\chi^2_{\rm BF}$) and the χ^2 computed at the true cosmological parameters values ($\chi^2_{\rm FID}$). These numbers should be compared with the number of multipoles K in the total data vector. As we start from $\ell_F = [8, 14, 18, 22]$ and finish at $\ell_{\max} = [256, 474, 591, 722]$ for the auto-C_ℓ case and $\ell_{\max} = [192, 356, 443, 542]$ for the cross-C_ℓ case, this is therefore $K = 1985$ multipoles fitted in the auto-C_ℓ case and $K = 1475$ in the cross-C_ℓ case.

In particular, the third row of table 5.1 gives the χ^2_r values computed at the best fits. It shows first that whatever the used covariance type, the computed χ^2_r are systematically greater than 1, pointing toward either non adapted covariances, or a systematic effect in the model prediction. The first option seems discarded, as

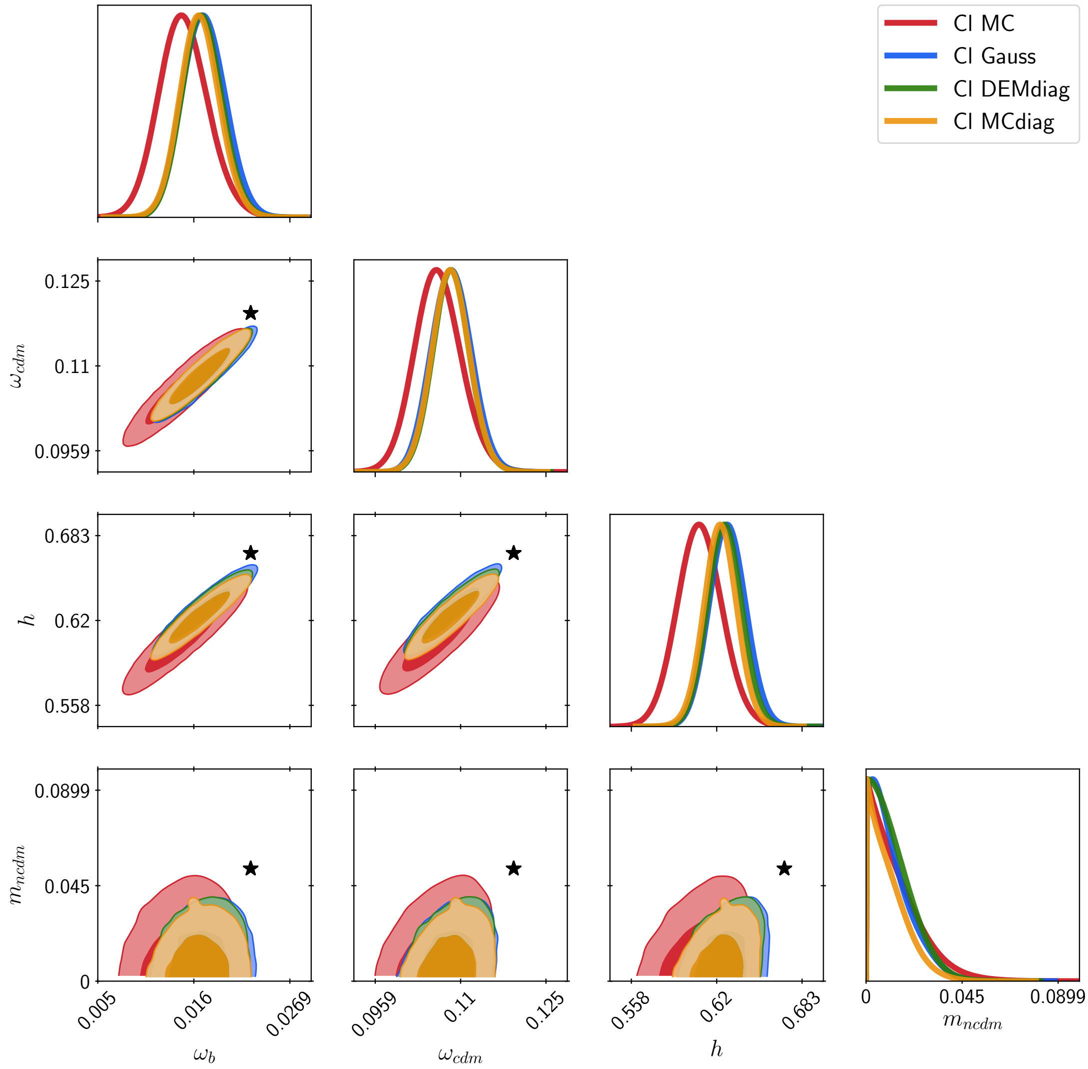


Figure 5.13: Triangular plot representing the 2-D ellipses of one and two standard deviations and 1-D histogram of the posteriors for the four fitted cosmological parameters. Four covariances have been used in the fitting procedure of the measured C_ℓ from the `DEMNUni_cov` data: the Gaussian covariance (Cl Gauss), the diagonal of the `DEMNUni_cov` covariance estimated from 46 simulations (Cl DEMdiag), the diagonal of the Monte Carlo covariance estimated from 46 simulations (Cl MCdiag) and the full Monte Carlo covariance estimated from 1000 realisations (Cl MC). The star labels represent the true cosmological parameters values.

a wide sample of matrix has been used (sometimes directly estimated on the same kind of data), while the second seems more likely to bias the analysis.

In order to try to understand these results, figure 5.15 displays the various outcomes of the analysis: the predictions of the auto and cross-C_ℓ at the best fit and at the true cosmology (the fiducial cosmology) as compared to the fitted data with error bars computed using the predicted Gaussian variance (see the first contribution of eq. 5.4). It shows a very close overlapping between the fiducial and best fit predictions, both matching very well the data. In particular, the deviations between the two predictions is maximised at $z = 2$ for the auto-C_ℓ and at $z = 0.5$ for the cross-C_ℓ, without nevertheless detecting a clear mismatch.

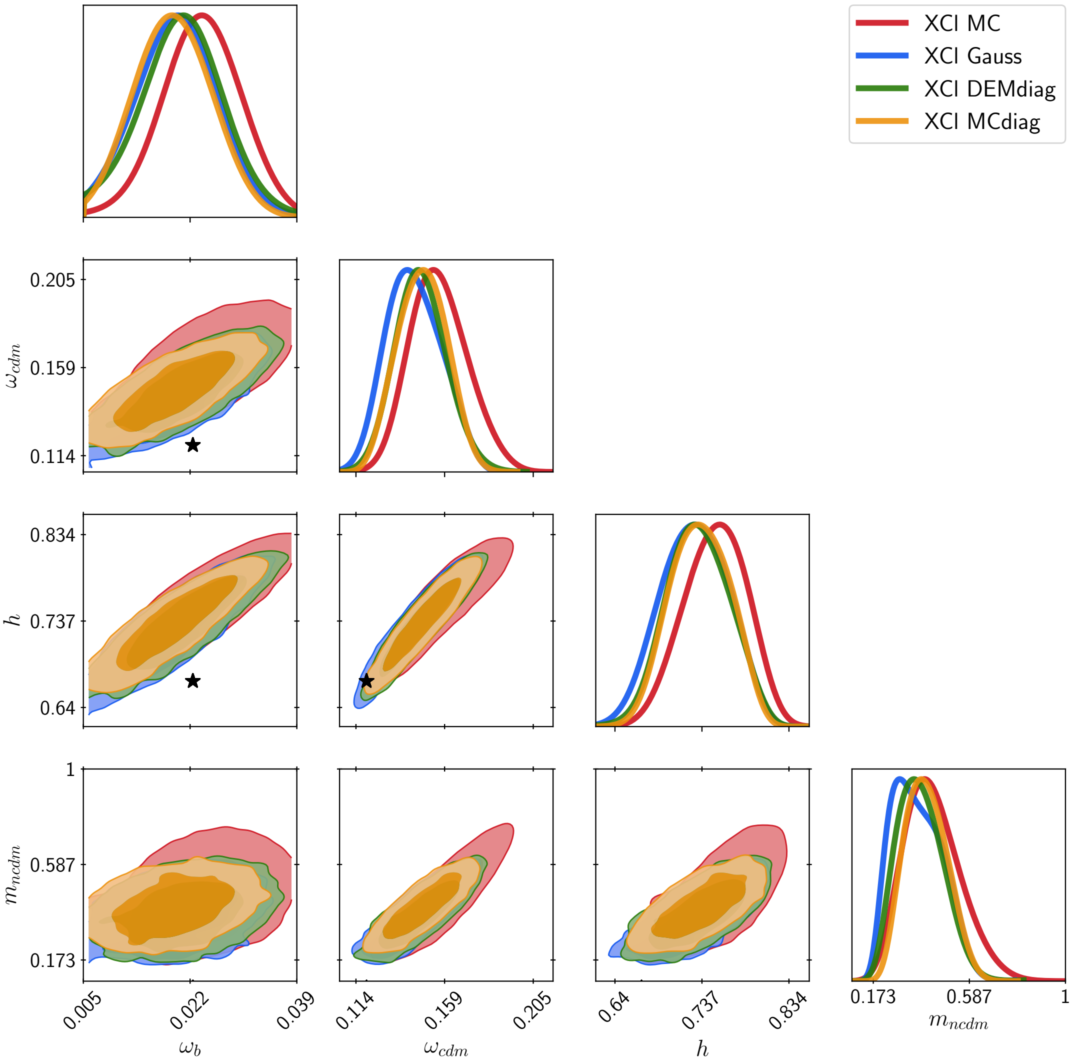


Figure 5.14: Triangular plot representing the 2-D ellipses of one and two standard deviations and 1-D histogram of the posteriors for the four fitted cosmological parameters. Four covariances have been used in the fitting procedure of the measured cross-C_ℓ from the `DEMNUni_cov` data: the Gaussian covariance (XCl Gauss), the diagonal of the `DEMNUni_cov` covariance estimated from 46 simulations (XCl DEMdiag), the diagonal of the Monte Carlo covariance estimated from 46 simulations (XCl MCdiag) and the full Monte Carlo covariance estimated from 1000 realisations (XCl MC). The star labels represent the true cosmological parameters values.

This problem is currently under investigation. It should be noticed that running the analysis by switching for the Fourier power spectrum, every choice of covariance and redshift leads to a fit compatible with the true cosmology. Thus, it seems that something goes wrong when performing the analysis in harmonic space.

In conclusion, I have presented in this chapter an extension of the Monte Carlo for the study of the angular power spectrum. First I went through the shell method (calibrated by the cell method), which is the most rigorous and adapted to observation since it constitutes a light cone reconstruction. It provided a better than

data-type	C_ℓ				XC_ℓ			
cov-type	Gauss	DEMdiag	MCdiag	MC	Gauss	DEMdiag	MCdiag	MC
$\chi^2_{\rm BF}$	2215	2364	2255	2225	1637	1683	1961	1849
$\chi^2_{\rm FID}$	2408	2576	2466	2318	1674	1721	2013	1869
χ^2_r	1.11	1.19	1.13	1.12	1.11	1.14	1.33	1.25

Table 5.1: Values of the χ^2 at the best fit (BF), at the fiducial (FID) set of cosmological parameters and those of the χ^2_r computed at the best-fit.

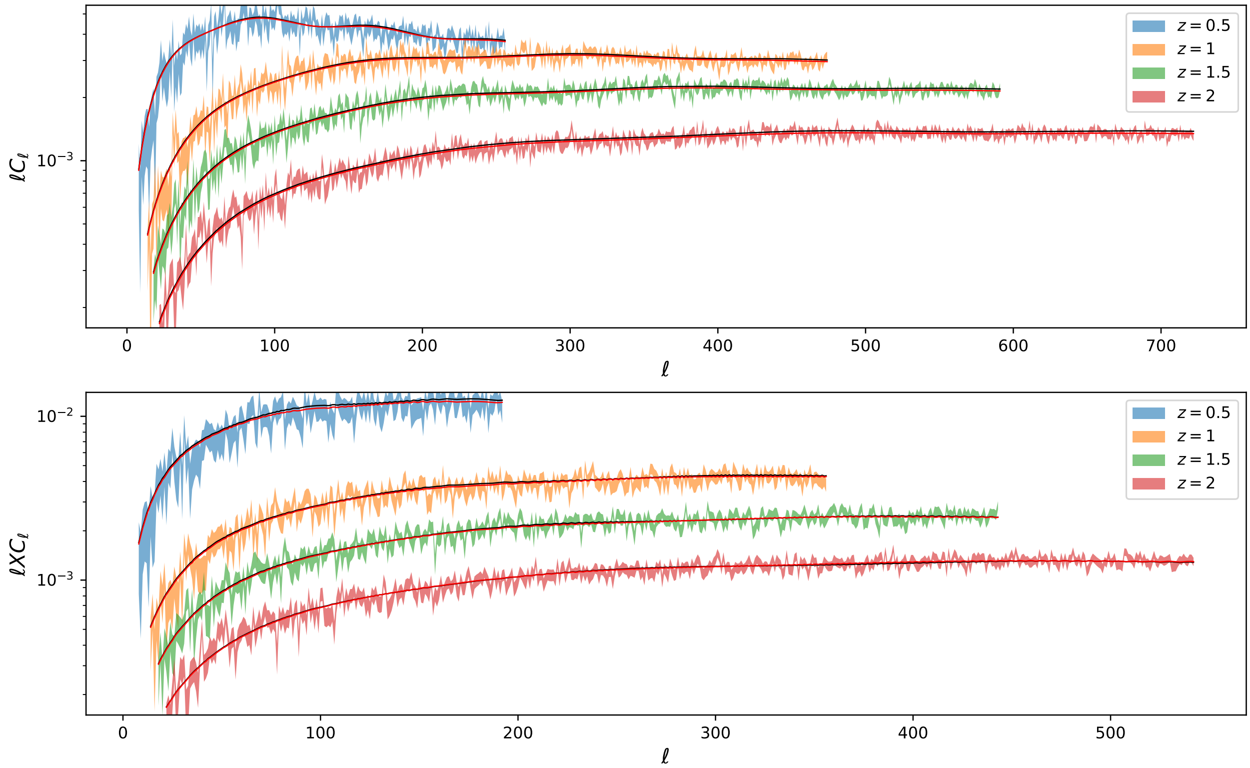


Figure 5.15: *top panel:* Comparison between fitted data C_ℓ with Gaussian error bars in colour for the four redshifts with the corresponding predictions at the best fit (overlapping red solid lines) and at the fiducial cosmology (overlapping black solid lines). *bottom panel:* Same thing but with the cross-C_ℓ.

the percent accuracy between simulation and prediction in the case of linear power spectrum. The third reconstruction method, although showing accurate angular statistics when comparing the Monte Carlo to the `DEMNUni_cov` outcomes, was nevertheless providing non-conclusive results concerning the cosmological parameters estimation. Thus further investigations are required to understand the weakness of such analysis pipeline. In particular, it would appears less biased to run a parameter inference on the N-Body and Monte Carlo outcomes when applying on both of them a light cone reconstruction using the shell method. To do so, more than four snapshot (four redshifts) per `DEMNUni_cov` would be necessary.

Conclusions and perspectives

Modern cosmology and future data analysis for large galaxy surveys are asking for the development of robust statistical tools and accurate estimation of parameter errors. This thesis has focused on providing a method for computing the covariance matrix for two-point statistics for future galaxy clustering surveys based on a fast Monte Carlo method. To do so, I went through the presentation of this new approach to generate catalogues of dark matter particles, haloes or galaxies (depending on the target power spectrum and probability density distribution).

In first place, I described the general pipeline and the theoretical tools needed to simulate any continuous non-Gaussian field characterised by its power spectrum and probability density function. I underlined the necessity of taking care of the aliasing (which is often neglected) in the proposed process. In addition, I have shown that, when it comes to populate the density field with objects (here we considered dark matter particles), the usual Top-Hat scheme is not the only possibility. I proposed then a more sophisticated scheme allowing to keep a smooth density field. In addition, I showed that the discretisation procedure introduces a low pass filtering of the power spectrum that can be accurately predicted, thus allowing to keep a perfect control on the expected power spectrum.

The reliability of the produced power spectra in comoving space has been tested in two working cases. First, for the analytical Log-Normal PDF specifically suited for the detection of any bias, where the percent accuracy has been reached (even below the resolution of the initial grid). Then, I investigated a more realistic case by cloning a N-body simulation. This second case allowed to point out specific issues, such as the apparition of negative values in the input power spectrum of the Gaussian field, even if we take care of the aliasing. Indeed, it seems that the shape and amplitude of the target power spectrum can affect the procedure allowing to compute the input power spectrum. As a result, I proposed several corrections in order to bypass this issue, that was presented in details. As a consequence, the Monte Carlo was validated for the power spectrum at least up to $k \sim 0.2h/\mathrm{Mpc}$ but still with the possibility of predicting the corrections.

I show that we achieve the objective to build an accurate covariance matrix of the power spectrum (and of the angular power spectrum). The covariance matrices obtained from this Monte Carlo has been compared to those estimated on N-body, and has been shown to be in agreement within 95% confidence level up to $k \sim 0.8h/\mathrm{Mpc}$.

I developed then an original method allowing first to create a velocity field with a given PDF correlated to the density and additionally, a way of assigning peculiar velocities to the generated catalogues of objects. The interest of the method is physically motivated by the fact that the velocity dispersion within haloes is higher in more massive structures. This simple model allows to successfully reproduce the distortions induced by peculiar velocities on the power spectrum in redshift-space, both in the linear regime with the Kaiser effect and in the non-linear regime with the Fingers-of-God. In that case, we show that the covariance matrix were reliable up to $k \sim 0.5h/\mathrm{Mpc}$ in the $2-\sigma$ limit. However, when going to higher redshift, the agreement was degraded. More investigation are needed on the possibility that this could depend on the choice for the velocity PDF. After having discussed of the observational advantages to deal with the two-point correlation function in harmonic space, the angular power spectrum C_ℓ, I updated the Monte Carlo method for various light cone reconstruction recipes. For the validation of this method, I still resorted to the Log-Normal reference case, showing a percent accuracy at the level of the simulated C_ℓ. As previously, a test has been performed with more realistic settings, once again relying on N-body comparison. I show that the angular spectra and their covariance matrix where successively reproduced up to the same scales.

Finally I showed an application of the Monte Carlo method on the estimation of cosmological parameters, using a simulated cosmology with massive neutrinos. However, the auto and cross angular power spectra do not allow to recover the values of the cosmological parameters corresponding to the simulation. The problem seems not due to the Monte Carlo, since the same pipeline has been tested on cosmological parameter using the power spectrum, and is successfully working. Despite the tests that have been performed, more work is then required to be the least possible dependent on N-body simulations in the production of covariance matrices. Thus, using cosmology-dependent parametrisations of one-point density and velocity distributions, as well as velocity dispersion vs local density relation, need to be investigated.

Moreover, it is worth emphasising that such method is already mature enough to test a wide scope of applications. For example, a direct application is to assess how the noise in the estimation of the covariance matrix elements is propagating to the cosmological parameter constraints. Yet an other possibility is to study the impact of the distribution of an observable (the likelihood) in a cosmological inference assuming a Gaussian likelihood when it is not. For two-point correlation functions, one can introduce instrumental constraints as angular mask, radial selection function, galaxy bias etc... to do some galaxy survey covariance matrix forecasts, in particular accounting for super sample covariance matrix.

The ultimate goal of such a project is anyway to estimate the cross-covariance between correlated observables. The present setting allows the treatment of correlated monopoles, quadrupoles, hexadecapoles for the power spectrum, but may also produce correlated two-point correlation functions in Fourier, configuration and harmonic spaces. In particular, the eventual advantage of $C_\ell \times P(k)$ combination should be investigated in more details.

Finally, of great interest, such method can be developed to allow the production of correlated C_ℓ data between galaxy clustering and galaxy lensing, a strategic combination to measure accurately the cosmological parameters in future galaxy surveys.

Appendix A

The angular power spectrum / C_ℓ

This appendix introduces the basics of the spherical harmonics space, and how we can define a two-point correlation function in it.

The harmonics vectorial space

For any isotropic cosmic field $\delta(\vec{x})$, some advantages can be found describing it in the scalar spherical harmonics space. The basis elements of degree ℓ and of dimension $2\ell+1$ are defined thanks to the *associated Legendre polynomials of first kind* $P_\ell^m(x)$, with the convention

$$Y_l^m(\theta,\phi) \equiv N_\ell^m P_\ell^m[\cos(\theta)]e^{im\phi} \ , \tag{A.1}$$

where $\ell \geq 0$, $-\ell < m < \ell$ are two integers, $\theta \in [0,\pi]$ the colatitude, $\phi \in [0,2\pi[$ the longitude and

$$N_\ell^m = \sqrt{\frac{2\ell+1}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} \ . \tag{A.2}$$

Furthermore, $P_\ell^m(x)$ are related to the usual *Legendre polynomials* $P_n(x)$ over the interval $x \in [-1,1]$ as

$$P_\ell^m(x) = (-1)^m(1-x^2)^{\frac{m}{2}}\frac{\mathrm{d}^m}{\mathrm{d}x^m}P_\ell(x) \ , \tag{A.3}$$

that verifies the orthogonal property

$$\int_{-1}^{1}\mathrm{d}x P_\ell(x)P_{\ell'}(x) = 0 \ , \ \text{for } \ell \neq \ell' \ . \tag{A.4}$$

Propagating this relation up to the Y_ℓ^m definition, the spherical harmonics space inherit the inner product

$$\int_{\theta=0}^{\pi}\int_{\phi=0}^{2\pi}\mathrm{d}\Omega Y_\ell^m(\theta,\phi)Y_\ell^m(\theta,\phi)^* = \delta^K_{\ell\ell'}\delta^K_{mm'} \ , \tag{A.5}$$

where $\mathrm{d}\Omega = \sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$, δ^K the Kronecker delta (which should not be confused with the cosmic field $\delta(\vec{x})$) and the subscript $*$ standing for the complex conjugate. Any real space $\delta(\vec{x})$ field can be expanded on such basis following the separation of radial and angular variables

$$\delta(\vec{x}) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} \delta_\ell^m(r) Y_\ell^m(\theta, \phi) \ , \tag{A.6}$$

whose the radial coefficients are obtained computing

$$\delta_\ell^m(r) = \int_{\mathcal{S}} \mathrm{d}^2\Omega \ \delta(r, \theta, \phi) Y_\ell^m(\theta, \phi)^* \ . \tag{A.7}$$

Equivalently, the δ_ℓ^m can be obtained directly from the Fourier counterpart $\delta_{\vec{k}}$ using

$$\delta_\ell^m(r) = 4\pi i^\ell \int \mathrm{d}^3\vec{k} \ \delta_{\vec{k}} Y_\ell^m(\theta_{\vec{k}}, \phi_{\vec{k}})^* j_\ell(kr) \tag{A.8}$$

where the j_l are the spherical Bessel functions. To obtain this relation, A.7 has been developed using the inverse Fourier transform of $\delta(\vec{x})$ and by expansion of the plane wave on the spherical harmonics basis (Cahill, 2019)

$$e^{i\vec{k}.\vec{x}} = 4\pi \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} i^\ell j_\ell(kr) Y_\ell^m(\theta_{\vec{k}}, \phi_{\vec{k}})^* Y_\ell^m(\theta, \phi) \ . \tag{A.9}$$

In this relation, $\theta_{\vec{k}}$ and $\phi_{\vec{k}}$ are the angular coordinates of the wave mode completing the radial one $k = |\vec{k}|$.

Weighting functions

Let's define a *weighing radial function* $W_j(r)$ of type j that allows to project on the sphere the cosmic field δ. Also called *Kernel* or *window function*, j can typically refers to

- a Dirac window of infinitely small thickness $W_D = N\delta^{\mathcal{D}}(r - r_a)$ where N is a normalisation coefficient, $\delta^{\mathcal{D}}$ the Dirac delta function and r_a the selected radius. It means that a single redshift is investigated
- a Top-Hat window $W_{\mathrm{TH}} = NH[r_a + r]H[r_b - r]$ where $H(r)$ is the Heaviside step functions defined by $H(x \geq 0) = 1$ or zero either and r_a and r_b the two window radial borders. It means that a continuous range of redshift is equitably investigated
- a Gaussian window $W_G = N\exp\left[-\frac{1}{2}(r - r_a)^2/\sigma_a^2\right]$ where r_a and σ_a are the two parameters of the normal distribution. It means that a continuous range of redshift is investigated with varying influence

Moreover, being related to a probability, they must verify the normalisation

$$\int_0^\infty \mathrm{d}r \ W_j(r) = 1 \ . \tag{A.10}$$

When a Poisson distribution is considered, the radial weighing function is translated to each particles of comoving radius r_i valuing them with the quantity $w_i = w(r = r_i)$. Using it as a projection function, it reads

$$\tilde{\delta}_j(\theta,\phi) \equiv \int_0^\infty \mathrm{d}r\ W_j(r)\delta(r,\theta,\phi)\ . \tag{A.11}$$

Additionally, $W(z)$ can be defined in redshift-space following the same prescription. In this case, the projection function is defined as

$$\tilde{\delta}_j(\theta,\phi) \equiv \int_0^\infty \mathrm{d}z\ W_j(z)\delta(r,\theta,\phi)\ . \tag{A.12}$$

C_ℓ definition

Be $\tilde{\delta}_1^A$ and $\tilde{\delta}_2^B$ two projected fields. Note that the nature of the fields A and B can describe two different tracers/probes. In the same way, there is no requirement that $W_1 = W_2$. Using A.7, they take the notation $\tilde{\delta}_\ell^{Am}(r_1)$ and $\tilde{\delta}_\ell^{Bm}(r_2)$. Let's define the angular cross power spectrum in harmonics space as

$$\tilde{C}_\ell^{AB}(r_1,r_2) \equiv \left\langle \tilde{\delta}_\ell^{Am}(r_1)\tilde{\delta}_\ell^{Bm*}(r_2)\right\rangle\ , \tag{A.13}$$

where we adopt a similar nomenclature as for the Legendre expansions of the Fourier power spectrum; C_0 is called the monopole, C_1 the dipole, C_2 the quadruple and so on. Developing the definition A.13, it yields

$$\begin{aligned}\tilde{C}_\ell^{AB}(r_1,r_2) = (4\pi)^2 \int \mathrm{d}r_1\mathrm{d}r_2 W_1(r_1)W_2(r_2) \int \mathrm{d}k k^2 P^{AB}(k) j_\ell(kr_1) j_\ell(kr_2) \\ \int \mathrm{d}\Omega\ \sin(\theta_{\vec{k}}) Y_\ell^m(\theta_{\vec{k}},\phi_{\vec{k}})^* Y_\ell^m(\theta_{-\vec{k}},\phi_{-\vec{k}})\ ,\end{aligned} \tag{A.14}$$

where $P^{AB}(k) = \left\langle \delta_{\vec{k}}^A \delta_{\vec{k}'}^{B*}\right\rangle \delta^D(\vec{k}+\vec{k}')$. In spherical coordinates, the latter constrain allows to write the system of equations

$$\begin{cases} -k\ \sin(\theta_{\vec{k}})\cos(\phi_{\vec{k}}) = k\sin(\theta_{\vec{k}'})\cos(\phi_{\vec{k}'}) \\ -k\ \sin(\theta_{\vec{k}})\sin(\phi_{\vec{k}}) = k\sin(\theta_{\vec{k}'})\sin(\phi_{\vec{k}'}) \\ -k\ \cos(\theta_{\vec{k}}) = k\cos(\theta_{\vec{k}'}) \end{cases} \tag{A.15}$$

whose the simplest allowed solution that we will use is

$$\theta_{\vec{k}'} = \theta_{\vec{k}} + \pi\ , \phi_{\vec{k}'} = \pi - \phi_{\vec{k}}\ . \tag{A.16}$$

In this case A.1 becomes

$$Y_\ell^m(\theta_{\vec{k}'},\phi_{\vec{k}'}) = N_\ell^m(-1)^\ell P_\ell^m\left[\cos(\theta_{\vec{k}})\right] e^{-im\phi_{\vec{k}}}\ , \tag{A.17}$$

where we use the parity relation of the associated Legendre polynomials $P_\ell^m[-x] = (-1)^{\ell+m}P_\ell^m[x]$. Writing $Y_\ell^{-m}(\theta_{\vec{k}'},\phi_{\vec{k}'})$ using the previous relation, it can be identify to its counterpart $Y_\ell^m(\theta_{\vec{k}},\phi_{\vec{k}})$, producing

$$Y_\ell^m(\theta_{-\vec{k}},\phi_{-\vec{k}}) = (-1)^{\ell-m} Y_\ell^{-m}(\theta_{\vec{k}},\phi_{\vec{k}})\ . \tag{A.18}$$

Coming back to relation A.14, the angular integral vanishes and only remains the usual definition of the angular power spectrum

$$\tilde{C}_\ell^{AB}(r_1, r_2) = (4\pi)^2 \int \mathrm{d}r_1 \mathrm{d}r_2 W_1(r_1) W_2(r_2) \int \mathrm{d}k k^2 P^{AB}(k, z) j_\ell(kr_1) j_\ell(kr_2) \ . \tag{A.19}$$

Switching the radial integration to redshift, this relation can be obtained in a similar way, *i.e.*

$$\tilde{C}_\ell^{AB}(z_1, z_2) = (4\pi)^2 \int \mathrm{d}z_1 \mathrm{d}z_2 W_1(z_1) W_2(z_2) \int \mathrm{d}k k^2 P^{AB}(k, z) j_\ell(kr(z_1)) j_\ell(kr(z_2)) \ , \tag{A.20}$$

from which can be extracted the angular power spectrum between two shells located at radial comoving positions r_1 and r_2, introducing the linear growth function $D(z)$ in eq. 2.20 (thus only adapted to a linear power spectrum)

$$\begin{aligned} C_\ell^{AB}(r_1, r_2) &\equiv \left\langle \delta^{A}{}_{\ell}^{m}(r_1) \delta^{B}{}_{\ell}^{m*}(r_2) \right\rangle \\ &= \int \mathrm{d}k k^2 D(z_1) D(z_2) P^{AB}(k, z=0) j_\ell(kr(z_1)) j_\ell(kr(z_2)) \ . \end{aligned} \tag{A.21}$$

Multipoles and Fourier modes

As notified in Di Dio et al. (2014), the angular correspondence Θ_ℓ (in real/configuration space) with the multipoles ℓ is given by $\Theta_\ell \sim 2\pi/\ell$ for $\ell \gg 1$. Thus for an observer at a comoving position $R(z)$ of the shell, it will interpret a distance d on the sphere as an angle $\Theta_d = \tan(d/R(z)) \sim d/R(z)$ for $R(z) \gg d$. Since Θ_ℓ and Θ_d must represent the same angle, it can be simply written that, for large multipoles ℓ,

$$\ell \sim kR(z) \ . \tag{A.22}$$

Nevertheless, one has to keep in mind that this relation is a rough approximation. For instance, picking a given multipole, relation A.19 implies that an integration is performed on the whole k-range weighted by the Bessel functions.

Appendix B

Sampling a Gaussian field with a given $P(k)$

This appendix gives the general method allowing to sample a random Gaussian field on a grid following an input power spectrum.

Let's start from the definition of the power spectrum $P(\vec{k})$

$$\langle \delta_{\vec{k}} \delta_{\vec{k}'} \rangle = \delta^{\mathcal{D}}(\vec{k}+\vec{k}')P(\vec{k}) \ , \tag{B.1}$$

where $\delta_{\vec{k}}$ stands for the Fourier transform of the fluctuation field $\delta(\vec{x})$ and $\delta^{\mathcal{D}}$ the Dirac function. Reducing the framework of the sampled cosmic fields to a cubic periodic box of volume $V = L^3$ actually simplifies definition B.1. Indeed, the Fourier transform of a periodic function is discrete and has power only at wave numbers which are multiples of the fundamental frequency of the box

$$k_F = \frac{2\pi}{L} \ , \tag{B.2}$$

representing the largest scale contained in the volume. As a consequence, the Fourier and inverse Fourier transforms linking $\delta(\vec{x})$ to $\delta_{\vec{k}_n}$ reads, for $\vec{k}_n = \vec{n} k_F$, $\vec{n} \in \mathbb{Z}^3$,

$$\delta_{\vec{k}_n} = (2\pi)^{-3} \int_V \mathrm{d}^3\vec{x} \ e^{-i\vec{k}_n.\vec{x}} \delta(\vec{x}) \ , \qquad \delta(\vec{x}) = k_F^3 \sum_{\vec{k}_n} e^{i\vec{k}_n.\vec{x}} \delta_{\vec{k}_n} \ . \tag{B.3}$$

Using these results, the discretised power spectrum simplifies in

$$\langle \delta_{\vec{k}_n} \delta_{\vec{k}'_n} \rangle = \delta^K_{\vec{n},-\vec{n}'} \frac{P(\vec{k}_n)}{k_F^3} \ , \tag{B.4}$$

where δ^K is the Kronecker symbol. To simplify further expression, the index n is drop but remains implicit.

Also, $\delta(\vec{x})$ being a real cosmic field, it involves that $\delta_{-\vec{k}} = \delta^*_{\vec{k}}$. Taking $\vec{k}' = -\vec{k}$ yields the expression allowing an estimation of the power spectrum as

$$P(\vec{k}) = k_F^3 \langle |\delta_{\vec{k}}|^2 \rangle \ . \tag{B.5}$$

From now on, lets adopt the notation ν as a reference to Gaussian fields. Splitting $\nu_{\vec{k}}$ in two independent real and imaginary parts, i.e. $\nu_{\vec{k}} = \alpha_{\vec{k}} + i\beta_{\vec{k}}$, the two cases

$\vec{k}' = -\vec{k}$ and $\vec{k}' = \vec{k}$ give respectively

$$\langle |\nu_{\vec{k}}|^2 \rangle = \langle \alpha_{\vec{k}}^2 \rangle + \langle \beta_{\vec{k}}^2 \rangle = P(\vec{k}) k_F^{-3} \tag{B.6}$$

$$\langle \nu_{\vec{k}}^2 \rangle = \langle \alpha_{\vec{k}}^2 \rangle - \langle \beta_{\vec{k}}^2 \rangle + 2i \langle \alpha_{\vec{k}} \rangle \langle \beta_{\vec{k}} \rangle = 0 \ . \tag{B.7}$$

While the imaginary part of B.7 vanishes, involving

$$\langle \alpha_{\vec{k}} \rangle = \langle \beta_{\vec{k}} \rangle = 0 \ , \tag{B.8}$$

$$\langle \alpha_{\vec{k}}^2 \rangle = \langle \beta_{\vec{k}}^2 \rangle = P(\vec{k}) k_F^{-3}/2 \ , \tag{B.9}$$

the two independent variables can be split into

$$\alpha_{\vec{k}} = \lambda_1 \sigma_{\vec{k}} \ , \tag{B.10}$$

$$\beta_{\vec{k}} = \lambda_2 \sigma_{\vec{k}} \ , \tag{B.11}$$

where λ_1 and λ_2 are two random numbers following a standardised normal distribution $\mathcal{N}(0,1)$ as

$$\langle \lambda_1 \rangle = \langle \lambda_2 \rangle = 0 \ , \tag{B.12}$$

$$\langle \lambda_1^2 \rangle = \langle \lambda_2^2 \rangle = 1 \ , \tag{B.13}$$

$$\langle \alpha_{\vec{k}}^2 \rangle = \langle \beta_{\vec{k}}^2 \rangle = \sigma_{\vec{k}}^2 \ . \tag{B.14}$$

This procedure is actually poorly optimised. As it does not contains the property $\nu_{\vec{k}} = \nu_{-\vec{k}}$, this property must be enforced by hand.

A more efficient method rather consists in generating only one set of random numbers. Applying a Fourier transform on a real grid of random numbers following a standardised normal distribution gives a grid of complex numbers constituted of random phases and Gaussian distributed real and imaginary parts. Moreover, as the Fourier transform is applied on a real field, the condition $\nu_{\vec{k}} = \nu_{-\vec{k}}$ is already completed. It finally reads

$$\nu_{\vec{k}} = \mathrm{FFT}\left[\vec{\mathcal{N}}\,(0,1)\right] \sqrt{\frac{N_s^3}{k_F^3} \frac{P(\vec{k})}{k_F^3 \sum_{\vec{k}} P(\vec{k})}} \ , \tag{B.15}$$

where FFT is the Fast Fourier Transform algorithm applied on a real grid of sampling number N_s^3. Note that the quantity $k_F^3 \sum_{\vec{k}} P(\vec{k})$ is equal to the variance of the field in real space (see Parseval's theorem in appendix C).

Appendix C

Parseval's theorem applied to cosmic fields

Be $u(\vec{x})$ a centered random field, sampled on a 3-D grid of sampling number N_s^3 such that $\langle u(\vec{x})\rangle = 0$. The Parseval's theorem states

$$\int \mathrm{d}^3\vec{x}u^2(\vec{x}) = (2\pi)^3 \int \mathrm{d}^3\vec{k}|u_{\vec{k}}|^2 \ . \tag{C.1}$$

Discretizing these continuous integrals on grid elements reads

$$\left(\frac{L}{N_s}\right)^3 \sum_{i=0}^{N_s^3-1} u_i^2(\vec{x}) = (2\pi)^3 k_F^3 \sum_{i=0}^{N_s^3-1} |u_{i,k}|^2 \ , \tag{C.2}$$

or equivalently

$$\frac{1}{N_s^3} \sum_{i=0}^{N_s^3-1} u_i^2(\vec{x}) = k_F^3 \sum_{i=0}^{N_s^3-1} k_F^3 |u_{i,k}|^2 \ . \tag{C.3}$$

The left hand side is nothing else that the variance of the field in real space while the right hand side can be identified using the grid power spectrum B.5

$$\sigma_u^2 \equiv \left\langle u_i^2(\vec{x})\right\rangle = k_F^3 \sum_{i=0}^{N_s^3-1} P(k_i) \ . \tag{C.4}$$

Appendix D

The Mehler equation applied to correlation functions

In this appendix, I present how to adapt the Mehler equation in order to simplify the computation of the integral 3.7.

On the basis of the Mehler formula is the Hermite's differential equation system (Mehler, 1866)

$$\begin{aligned} y'' - 2xy' + 2ny &= 0 \\ z'' + (2n + 1 - x^2)z &= 0 \ , \end{aligned} \tag{D.1}$$

where $x \in \mathbb{R}$, $n \in \mathbb{N}^+$,

$$z = e^{-x^2/2}\mathcal{H}_n(x) \ , \tag{D.2}$$

$$y = \mathcal{H}_n(x) = (-1)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2} \ . \tag{D.3}$$

Here, y is the basis elements of the *physicist Hermite polynomials* in a n-dimensional space of identity element $\mathcal{H}_0(x) = 1$ for the inner product

$$\int_{-\infty}^{\infty} \mathrm{d}x \mathcal{H}_n(x)\mathcal{H}_m(x)e^{-x^2} = \sqrt{\pi}2^n n! \delta^K_{nm} \ , \tag{D.4}$$

where δ^K_{nm} stands for the Kronecker symbol.
The system D.1 can be solved using the *parabolic cylinder functions* (following Bateman, 1953) yielding four general solutions

$$\sum_{n=0}^{\infty} \mathcal{H}_n(x)\frac{z^n}{n!} = e^{2xz - z^2} \ , \tag{D.5}$$

$$\sum_{n=0}^{\infty} (-1)^n \mathcal{H}_{2n}(x)\frac{z^{2n}}{(2n)!} = e^{z^2}\cos(\sqrt{2}xz) \ , \tag{D.6}$$

$$\sum_{n=0}^{\infty} (-1)^n \mathcal{H}_{2n+1}(x)\frac{z^{2n+1}}{(2n+1)!} = e^{z^2}\sin(\sqrt{2}xz) \ , \tag{D.7}$$

$$\sum_{n=0}^{\infty} \frac{(z/2)^n}{n!}\mathcal{H}_n(x)\mathcal{H}_n(y) = \frac{1}{\sqrt{1-z^2}}\exp\left\{\frac{2xyz - (x^2+y^2)z^2}{1-z^2}\right\} \ . \tag{D.8}$$

Getting closer to the bivariate Gaussian distribution, the last solution called the *Mehler formula*, turns out to be valuable in the simplification of equation 3.7. Indeed, switching to the usual variables $x \to \nu_1/\sqrt{2}$, $y \to \nu_2/\sqrt{2}$, $z \to \xi_\nu$ and defining a similar vectorial space of *probabilistic Hermite polynomials* basis

$$\mathcal{H}_{e_n}(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2} \ , \tag{D.9}$$

with inner product

$$\int_{-\infty}^{\infty} \mathrm{d}x \mathcal{H}_{e_n}(x) \mathcal{H}_{e_m}(x) e^{-x^2/2} = \sqrt{2\pi} n! \delta^K_{nm} \ , \tag{D.10}$$

it can be shown that

$$\frac{1}{\sqrt{1-\xi_\nu^2}} \exp\left\{ \frac{2\nu_1\nu_2\xi_\nu - (\nu_1^2+\nu_2^2)\xi_\nu^2}{2(1-\xi_\nu^2)} \right\} = \sum_{n=0}^{\infty} \frac{\xi_\nu^n}{n!} \mathcal{H}_{e_n}(\nu_1) \mathcal{H}_{e_n}(\nu_2) \ . \tag{D.11}$$

Here the left hand side can be identified to the ratio $P_\nu(\nu_1, \nu_2, \xi_\nu) / \left[P_\nu(\nu_1) P_\nu(\nu_2)\right]$ allowing to finally write the *extended Mehler formula* (Kibble, 1945)

$$P_\nu(\nu_1, \nu_2, \xi_\nu) = P_\nu(\nu_1) P_\nu(\nu_2) \sum_{n=0}^{\infty} \frac{\xi_\nu^n}{n!} \mathcal{H}_{e_n}(\nu_1) \mathcal{H}_{e_n}(\nu_2) \ . \tag{D.12}$$

It gets a lot more enlightening when decomposing the $\mathcal{L}$-mapping function (see eq. 3.5) in a series of probabilistic Hermite polynomials

$$\mathcal{L}(\nu) = \sum_{n=0}^{\infty} c_n \mathcal{H}_{e_n}(\nu) \ , \tag{D.13}$$

where obtaining the Hermite coefficients c_n can be done by multiplying both sides of eq. D.13 by $e^{-x^2/2}\mathcal{H}_{e_m}(\nu)$ and then by the use of the orthogonal property D.10, *i.e.*

$$c_n = \frac{1}{n!} \int_{\infty}^{\infty} \mathrm{d}x \mathcal{L}(x) \mathcal{H}_{e_n}(x) \mathcal{N}(x) \ , \tag{D.14}$$

where $\mathcal{N}(x)$ denotes the standardised normal distribution. Combining equations D.12, 3.7 and D.13 allows to reformulate relation λ linking the two two-point correlation functions into

$$\xi_\delta = \lambda\left(\xi_\nu\right) = \sum_{n=0}^{\infty} n! c_n^2 \xi_\nu^n \ . \tag{D.15}$$

Appendix E

Mehler equation in the Log-Normal case

In this appendix I review the main applications of the Mehler equation D.15 in the case of a Log-Normal PDF.

The Log-Normal transformation of a Gaussian field $\nu(\vec{x})$, that initially reads $\delta \equiv \mathcal{L}(\nu) = e^{\nu} - 1$ would not produce an expected cosmological field $\delta \in]-1, \infty[$. Indeed, following the local conservation of probability using this transformation (see equation 3.2), the targeted field would have a distribution $P_\delta(\delta) = e^{-\nu} P_\nu(\nu)$ that has for undesired effect to shift $\langle \delta \rangle < 0$. Instead, the transformation must come from another field that is neither centered nor reduced. Lets redefine $\nu(\vec{x})$ as a field $\phi(\vec{x})$ following

$$\nu = (\phi - \mu_\phi)/\sigma_\phi \ , \tag{E.1}$$

where μ_ϕ and σ_ϕ are respectively the mean value and the *rms* of the ϕ-field. The transformation reads in this case

$$\delta \equiv \mathcal{L}(\nu) = \mathrm{e}^{\phi(\nu)} - 1 \ . \tag{E.2}$$

These two first moments must be well calibrated to reproduce the density contrast field once the non-linear mapping is applied. They read

$$\langle \delta^n \rangle = \int_{-1}^{\infty} \mathrm{d}\delta \left[\mathrm{e}^{\phi} - 1 \right]^n P_\delta(\delta) \ , \tag{E.3}$$

that can be explored using the Newton's binomial

$$(x+y)^n = \sum_{j=0}^{n} \binom{n}{j} x^j y^{n-j} \ . \tag{E.4}$$

Once each terms judiciously identified, it turns out that is can be written as related to the specific expectation value $\langle \mathrm{e}^{j\phi} \rangle$

$$\langle \delta^n \rangle = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} \int \mathrm{d}\phi \mathrm{e}^{j\phi} P_\phi(\phi) = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} \langle \mathrm{e}^{j\phi} \rangle \ . \tag{E.5}$$

For all random variable X can be associated a moment generating function $\mathcal{M}(t)$ such that the moments $\langle X^n \rangle$ are derived from it

$$\left. \frac{\mathrm{d}^n \mathcal{M}(t)}{\mathrm{d}t^n} \right|_{t=0} \equiv \langle X^n \rangle \ . \tag{E.6}$$

Equivalently, using that $\langle X^n \rangle \equiv \int \mathrm{d}X X^n P_X(X)$ and restricting X to a Gaussian sampled scalar fields[1], the moment generating function is written

$$\mathcal{M}(t) = \sum_{n=0}^{\infty} \frac{\langle X^n \rangle}{n!} t^n = \left\langle \mathrm{e}^{Xt} \right\rangle = \exp\left[\mu_X t + \sigma_X^2 t^2/2\right] \ . \tag{E.7}$$

Applying these results on ϕ, the moments can finally be obtained in their simplest form

$$\langle \delta^n \rangle = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} \exp\left[\mu_\phi j + \sigma_\phi^2 j^2/2\right] \ . \tag{E.8}$$

Some constraints can be applied on the two first moments of equation E.8:

- for $n = 1$, $\langle \delta \rangle = 0$ is equivalent to $\mu_\phi = -\sigma_\phi^2/2$,
- for $n = 2$, we get the constraint $\sigma_\phi^2 = \ln(\sigma_\delta^2 + 1)$.

ϕ is therefore totally determined by the input power spectrum as well as for ν. The Log-Normal transformation equation E.2 reads finally

$$\delta \equiv \mathcal{L}(\nu) = \mathrm{e}^{\nu\sigma_\phi - \sigma_\phi^2/2} - 1 \ , \tag{E.9}$$

that correspond to a well shaped Log-Normal distribution

$$P_\delta(\delta) = \frac{1}{\sqrt{2\pi\sigma_\phi^2}} \mathrm{e}^{-\nu^2/2 - \nu\sigma_\phi + \sigma_\phi^2/2} \ , \tag{E.10}$$

where $\nu = \left[\ln(\delta + 1) + \sigma_\phi^2/2\right]/\sigma_\phi$.

Assuming $\sigma_\phi^2 \sim \sigma_\nu^2 = 1$, the inversion of eq. D.15 can be straightforward. Indeed patching equation E.9 into equation D.14 gives

$$c_n = \frac{1}{n!} \int_{-\infty}^{\infty} \mathrm{d}\nu \left(\mathrm{e}^{\nu - \frac{1}{2}} - 1\right) P_\nu(\nu) \mathcal{H}_{e_n}(\nu) \tag{E.11}$$

$$= \frac{1}{n!} \left(\int_{-\infty}^{\infty} \mathrm{d}\nu P_\nu(\nu - 1) \mathcal{H}_{e_n}(\nu) - \int_{-\infty}^{\infty} \mathrm{d}\nu P_\nu(\nu) \mathcal{H}_{e_n}(\nu) \right) \ . \tag{E.12}$$

The first contribution can be solved by the change of variable $\nu \to \nu + 1$ and using the identity $\mathcal{H}_{e_n}(x+1) = \sum_{k=0}^{n} \binom{n}{k} \mathcal{H}_{n-k}(x)$. Then multiplying the two integrands by the identity element $\mathcal{H}_{e_0}(x)$, orthogonality (see equation D.10) is obtained for all $n \neq k$. It reads

$$c_n = \frac{1}{n!} \left[1 - \delta_{n0}\right] \ . \tag{E.13}$$

Combining equation D.15 and equation E.13 gives

$$\xi_\delta = \sum_{n=1}^{\infty} \frac{\xi_\nu}{n!} = \exp(\xi_\nu) - 1 \ . \tag{E.14}$$

It finally appears that λ^{-1} can be obtained analytically, showing the convenience of this example,

$$\xi_\nu = \lambda^{-1}\left[\xi_\delta\right] = \ln(\xi_\delta + 1) \ . \tag{E.15}$$

---

[1] In the vectorial field case $\vec{X}$, the moment generating function can be written as $\mathcal{M}_{\vec{X}}(\vec{t}) = \exp\left[\vec{\mu}_X^T \vec{t} + \vec{t}^T C_X^{-1} \vec{t}/2\right]$, where C_X is the basis elements covariance matrix of $\vec{X}$.

Appendix F

Tri-linear grid field interpolation

In this appendix I derive the relations allowing to populate particles in a cell following the interpolation scheme of order 1 (see section 4.1.1), analogous to the CIC assignment scheme (eq. 4.4).

The cubic volume symbolising a given voxel in tri-linear interpolation is by construction

$$\begin{aligned}\widetilde{\rho}(x,y,z) = \rho_{000}(a-x)(a-y)(a-z) + \rho_{a00}x(a-y)(a-z)+ \\ +\rho_{0a0}(a-x)y(a-z) + \rho_{00a}(a-x)(a-y)z+ \\ +\rho_{a0a}x(a-y)z + \rho_{0aa}(a-x)yz+ \\ + \rho_{aa0}xy(a-z) + \rho_{aaa}xyz \ ,\end{aligned} \tag{F.1}$$

where ρ_{ijk} represents the eight local density values defining the vertices of the cube, replacing (i,j,k) by 0 or $a = L/N_s$. The coordinates x, y and z are defined in the interval $[0,a]$. Converting it into normalised probability, the 3-D PDF is defined as

$$P_{\widetilde{\rho}}(x,y,z) = \frac{8}{a^3}\frac{\widetilde{\rho}(x,y,z)}{\sum_{i,j,k}\rho_{ijk}}. \tag{F.2}$$

Eq. F.2 shows that the three coordinates cannot be generated independently, leading to consider conditional probabilities.

More specifically, the method to sample a set of points in the cube will use the *classical inverse transform sampling method* (Devroye, 1986). It relies on the notion of cumulative distribution function (CDF) to generate random numbers from any 1-D probability distribution function $P(z)$ in the definition ensemble $[a_1,a_2]$, defined by

$$\mathrm{CDF}[P](z) = \int_{a_1}^{z} \mathrm{d}z' P(z') \ . \tag{F.3}$$

Then one has to solve analytically for z the relation $g = \mathrm{CDF}[P](z)$, calling this inversion $z = F(g)$. Finally, given F, a uniform number g in $[0,1]$ will follow $P(z)$.

The different steps for the coordinates assignation will be as follows[1]:

---

[1] Note that there is no uniqueness of the method.

1. Integrate $P_\rho(x,y,z)$ over z; z then y; z then x; y then x gives respectively the marginalised probabilities $P^{z}_{\tilde{\rho}}(x,y)$, $P^{y,z}_{\tilde{\rho}}(x)$, $P^{x,z}_{\tilde{\rho}}(y)$, $P^{x,y}_{\tilde{\rho}}(z)$
2. Generating N_{cell} pseudo-random numbers according to the inverse transform sampling method described above for $P^{y,z}_{\tilde{\rho}}(x)$ completes the coordinates x.
3. Since $P^{z}_{\tilde{\rho}}(x,y) = P^{y,z}_{\tilde{\rho}}(x)P^{z}_{\tilde{\rho}}(y|x)$, where $y|x$ means the conditional probability of y given x, a number N_{cell} of y-pseudo-random-coordinates depending on the previous ones x can be generated following $P^{z}_{\tilde{\rho}}(y|x)$
4. In the same way, as $P_{\tilde{\rho}}(x,y,z) = P^{z}_{\tilde{\rho}}(x,y)P_{\tilde{\rho}}(z|x,y)$, N_{cell} new random numbers targeting $P_{\tilde{\rho}_z}(z|x,y)$ allows to simulate the z-coordinates

Each marginalised probability calculated in point 1. shall take the following forms :

$$P^{y,z}_{\tilde{\rho}}(x) = \frac{2}{a^2}\frac{C(x)}{\sum_{i,j,k}^{0,a}\tilde{\rho}_{ijk}}, P^{x,z}_{\tilde{\rho}}(y) = \frac{2}{a^2}\frac{D(y)}{\sum_{i,j,k}^{0,a}\tilde{\rho}_{ijk}}, P^{x,y}_{\tilde{\rho}}(z) = \frac{2}{a^2}\frac{E(z)}{\sum_{i,j,k}^{0,a}\tilde{\rho}_{ijk}}, \tag{F.4}$$

where

$$\begin{aligned} C(x) &= a(\rho_{000}+\rho_{00a}+\rho_{0a0}+\rho_{0aa})+ \\ &\qquad +x(-\rho_{000}-\rho_{00a}-\rho_{0a0}-\rho_{0aa}+\rho_{a00}+\rho_{a0a}+\rho_{aa0}+\rho_{aaa}) \\ D(y) &= a(\rho_{000}+\rho_{00a}+\rho_{a00}+\rho_{a0a})+ \\ &\qquad +y(-\rho_{000}-\rho_{00a}+\rho_{0a0}+\rho_{0aa}-\rho_{a00}-\rho_{a0a}+\rho_{aa0}+\rho_{aaa}) \\ E(z) &= a(\rho_{000}+\rho_{0a0}+\rho_{a00}+\rho_{aa0})+ \\ &\qquad +z(-\rho_{000}+\rho_{00a}-\rho_{0a0}+\rho_{0aa}-\rho_{a00}+\rho_{a0a}-\rho_{aa0}+\rho_{aaa}). \end{aligned} \tag{F.5}$$

The transformations on g in order to successively obtain x, y and z are given by

$$x = aC_2^{-1}\left[-C_1+\sqrt{C_1^2+C_2C_3g}\right] , \tag{F.6}$$

$$y = -(2D_1)^{-1}\left[D_2-\sqrt{D_2^2+4gD_1D_3}\right] , \tag{F.7}$$

$$z = -(2E_1)^{-1}\left[E_2-\sqrt{E_2^2+4gE_1E_3}\right] \tag{F.8}$$

with

$$
\begin{aligned}
C_1 &= \rho_{000} + \rho_{0a0} + \rho_{00a} + \rho_{0aa} \\
C_2 &= -\rho_{000} + \rho_{a00} - \rho_{0a0} - \rho_{00a} + \rho_{a0a} - \rho_{0aa} + \rho_{aa0} + \rho_{aaa} \\
C_3 &= \rho_{000} + \rho_{a00} + \rho_{0a0} + \rho_{00a} + \rho_{a0a} + \rho_{0aa} + \rho_{aa0} + \rho_{aaa} \\
D_1 &= a(-\rho_{000} - \rho_{00a} + \rho_{0a0} + \rho_{0aa}) \\
&\qquad\qquad + x(\rho_{000} + \rho_{00a} - \rho_{0a0} - \rho_{0aa} - \rho_{a00} - \rho_{a0a} + \rho_{aa0} + \rho_{aaa}) \\
D_2 &= 2a^2(\rho_{000} + \rho_{00a}) - 2ax(\rho_{000} + \rho_{00a} - \rho_{a00} - \rho_{a0a}) \\
D_3 &= a^2[a(\rho_{000} + \rho_{00a} + \rho_{0a0} + \rho_{0aa}) \\
&\qquad\qquad + x(-\rho_{000} - \rho_{00a} - \rho_{0a0} - \rho_{0aa} + \rho_{a00} + \rho_{a0a} + \rho_{aa0} + \rho_{aaa})] \\
E_1 &= a\,[a(-\rho_{000} + \rho_{00a}) + x(\rho_{000} - \rho_{00a} - \rho_{a00} + \rho_{a0a}) + y(\rho_{000} - \rho_{00a} - \rho_{0a0} + \rho_{0aa})] + \\
&\qquad + xy(-\rho_{000} + \rho_{00a} + \rho_{0a0} - \rho_{0aa} + \rho_{a00} - \rho_{a0a} - \rho_{aa0} + \rho_{aaa}) \\
E_2 &= 2a\,[a(a\rho_{000} + x(-\rho_{000} + \rho_{a00}) + y(-\rho_{000} + \rho_{0a0})) + xy(\rho_{000} - \rho_{0a0} - \rho_{a00} + \rho_{aa0})] \\
E_3 &= a^2[a^2(\rho_{000} + \rho_{00a}) + a(y(-\rho_{000} - \rho_{00a} + \rho_{0a0} + \rho_{0aa}) \\
&\qquad + x(-\rho_{000} - \rho_{00a} + \rho_{a00} + \rho_{a0a})) \\
&\qquad + xy(\rho_{000} + \rho_{00a} - \rho_{0a0} - \rho_{0aa} - \rho_{a00} - \rho_{a0a} + \rho_{aa0} + \rho_{aaa})]
\end{aligned}
\tag{F.9}
$$

These expressions allow a verification of the method by comparison of the analytical predictions on the numerical results of a simple simulation. Taking $a = 1$ and ρ_{ijk} eight uniform random numbers in the arbitrarily chosen interval $[0, 10]$, the histograms for each of the estimated coordinates for a sample of $N_c = 10^7$ particles are represented in figure F.1. They show that at the scale of a cell, as expected, a linear relation for each coordinate is well reproduced. Moreover, a perfect matching between model and simulation without any apparent bias allows to definitely validate the objective of this method.

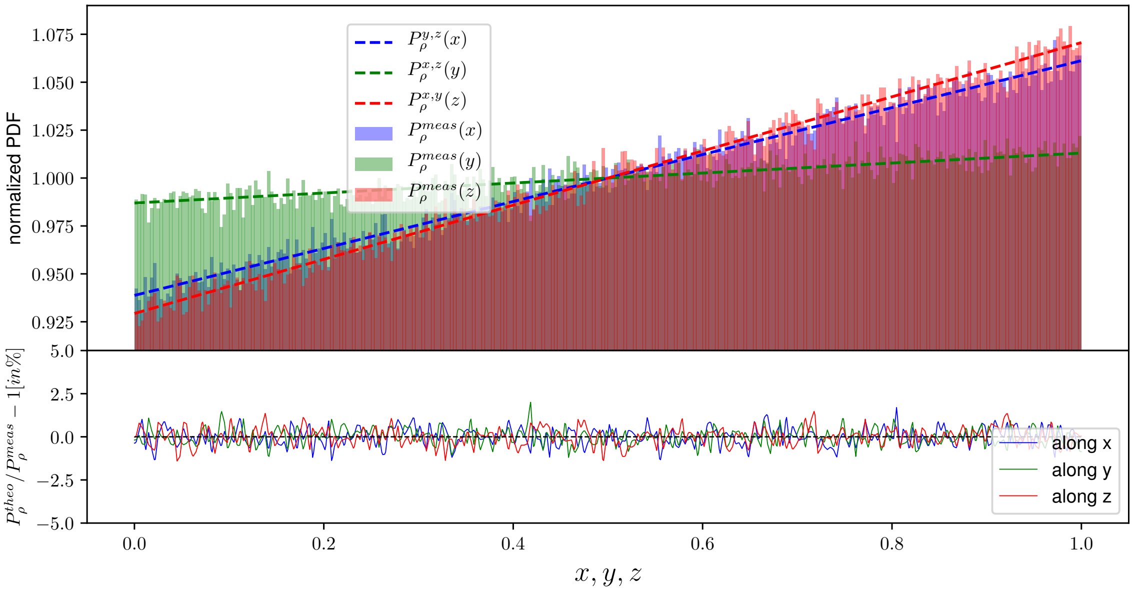


Figure F.1: *upper panel:* Standardised histograms for each coordinate estimated over a sample of $N_c = 10^7$ particles. This simulation is produced in a side box $a = 1$ whose the densities at the vertices of the cube are randomly drawn in the interval $[0, 10]$. Predictions are shown in dashed lines. *bottom panel:* Relative deviations in percent of the expected and measured histograms using the same colour-code.

Appendix G

Power spectrum shot noise

The estimated power spectrum of a point-like distribution, unlike a continuous grid-sampled field, should not be expected to take the form of eq. 4.8. Indeed, a second contribution will be added to it, called the *shot noise*. This statistical artefact corresponds to the auto-correlation of particles with themselves and this appendix presents the derivation of this contribution.

Be $\Lambda(\vec{x})$ the continuous counting density field and $N(\vec{x})$ its discrete version. Booth of them are related to the usual density filed by their mean values $\bar{\Lambda} = \bar{N} = a^3\rho_0$, where ρ_0 is the average density of the field and a the volume of each cell.
With N_s^3 being the total number of cells in the simulated box and j labelling each of them, the total number of particles is $N_t = \sum_{j=0}^{N_s^3-1} N_j$. $N(\vec{x})$ can be constructed by summation over each particle position $\vec{\epsilon}$ inside the cell labelled by the index j

$$N(\vec{x}) = a^3 \sum_{j=0}^{N_s^3-1} \sum_{i=1}^{N_j} \delta^{\mathcal{D}}(\vec{x} - (\vec{x}_j + \vec{\epsilon}_i)) \tag{G.1}$$

Starting from $\delta(\vec{x}) = N(\vec{x})/\bar{N} - 1$, we get that for $\vec{k} \neq 0$, $\delta_{\vec{k}} = N_{\vec{k}}/\bar{N}$, the definition of the grid power spectrum eq. B.5 can be rewritten as

$$P(\vec{k}) = \langle \delta_{\vec{k}} \delta^{\star}_{\vec{k}} \rangle k_F^3 = \frac{\left\langle N_{\vec{k}} N^{\star}_{\vec{k}} \right\rangle}{\bar{N}^2} k_F^3 \ , \tag{G.2}$$

for which

$$N_{\vec{k}} = \left(\frac{a}{2\pi}\right)^3 \sum_{j=0}^{N_s^3-1} \sum_{i=1}^{N_j} e^{-i\vec{k}.\vec{x}_j} e^{-i\vec{k}.\vec{\epsilon}_i} \ . \tag{G.3}$$

It follows that

$$P(\vec{k}) = v \left\langle \sum_{j=0}^{N_s^3-1} \sum_{l=0}^{N_s^3-1} e^{-i\vec{k}.(\vec{x}_j - \vec{x}_l)} \sum_{i=1}^{N_j} \sum_{n=1}^{N_l} e^{-i\vec{k}.(\vec{\epsilon}_i - \vec{\epsilon}_n)} \right\rangle \ , \tag{G.4}$$

with $v = k_f^{-3} N_t^{-2}$.
The previous relation can be restricted to $j = l$ and $i = n$, since shot noise only concerns self interacting particles. It ends up with the simple form

$$P^{\mathrm{SN}}(\vec{k}) = \frac{1}{(2\pi)^3 \rho_0} \ . \tag{G.5}$$

Finally, note that being a constant in Fourier space, this contribution vanishes in configuration space when estimating the two-point correlation function.

Appendix H

Features on velocity divergence

In this appendix, I review some velocity field properties on which I rely in the Monte Carlo procedure.

Be $\vec{v}$ a velocity field in comobile space. From it derives the velocity divergence field

$$\theta(\vec{x}) = \vec{\nabla}.\vec{u} \; , \tag{H.1}$$

where I use a slightly different convention for the scaled velocity field

$$\vec{u} \equiv -\vec{v}\frac{1+z}{100E(z)} \tag{H.2}$$

than those defined in section 4.3.

In Fourier space, eq. H.1 takes the form

$$\theta_{\vec{k}} = (2\pi)^{-3} \int_{-\infty}^{\infty} \mathrm{d}^3\vec{x} e^{-i\vec{k}.\vec{x}} \vec{\nabla}.\vec{u}(\vec{x}) \; , \tag{H.3}$$

that can be integrated by part using that for general scalar fields ϕ and vector fields $\vec{F}$, we have the identity $\vec{\nabla}.(\phi\vec{F}) = \phi\vec{\nabla}.\vec{F} + \vec{F}.\vec{\nabla}\phi$. Eq. H.3 decomposes as

$$\theta_{\vec{k}} = (2\pi)^{-3} \left[e^{-i\vec{k}.\vec{x}}\vec{u}(\vec{x})\right]_{-\infty}^{\infty} - (2\pi)^{-3} \int_{-\infty}^{\infty} \mathrm{d}^3\vec{x} \; \vec{u}(\vec{x}).\vec{\nabla}e^{-i\vec{k}.\vec{x}} \; . \tag{H.4}$$

Assuming a statistic isotropy, the first term vanishes and only remains

$$\theta_{\vec{k}} = i\vec{k}.\vec{u}_{\vec{k}} \; . \tag{H.5}$$

This expression can be used to write the power spectrum transformation between the two fields. In Cartesian coordinates, it reads

$$\begin{aligned} P_{\theta\theta}(k) = \left\langle|\theta_k|^2\right\rangle = \left\langle|i\vec{k}.\vec{u}_k|^2\right\rangle =& k_x^2 P_{u_x u_x}(k) + k_y^2 P_{u_y u_y}(k) + k_z^2 P_{u_z u_z}(k) \\ &+ 2[k_x k_y P_{u_x u_y}(k) + k_x k_z P_{u_x u_z}(k) + k_y k_z P_{u_y u_z}(k)] \; . \end{aligned} \tag{H.6}$$

Making the weak assumption that $P_{u_x u_x}(k) = P_{u_y u_y}(k) = P_{u_z u_z}(k) \equiv P_{u_i u_i}(k)$ and that $P_{u_x u_y}(k) = P_{u_x u_z}(k) = P_{u_y u_z}(k) \equiv P_{u_i u_j}(k)$ with $i \neq j$, eq. H.6 simplifies as

$$P_{\theta\theta}(k) = k^2 P_{u_i u_i}(k) + 2P_{u_i u_j}(k)[k_x k_y + k_x k_z + k_y k_z] \; . \tag{H.7}$$

In the Monte Carlo process, the idea is to deduce the velocity power spectrum from $P_{\theta\theta}$. However, inverting eq. H.5 and eq. H.7 is not an easy task. Rather, an assumption enabling these inversions is to assume that the velocity field derives from a scalar potential, *i.e.*

$$\vec{u}(\vec{x}) \equiv \vec{\nabla}\phi_v(\vec{x}) \ . \tag{H.8}$$

The Fourier transform of eq. H.8 reading

$$\vec{u}_{\vec{k}} = (2\pi)^{-3}\int_{-\infty}^{\infty} \mathrm{d}^3\vec{x} e^{-i\vec{k}.\vec{x}}\vec{\nabla}\phi_v \ , \tag{H.9}$$

it can similarly be integrated by part when noticing that for two scalar fields A and B , we have that $\vec{\nabla}(AB) = B\vec{\nabla}A + A\vec{\nabla}B$. We ends up with

$$\vec{u}_{\vec{k}} = i\vec{k}\phi_{v,\vec{k}} \ . \tag{H.10}$$

In turn, definition H.8 propagates up to the divergence field as $\theta(\vec{x}) = \vec{\nabla}.\vec{\nabla}\phi(\vec{x})$. An integration by part on its Fourier counterpart yields

$$\theta_{\vec{k}} = -k^2\phi_{v,\vec{k}} \ . \tag{H.11}$$

Finally, combining relations H.10 with H.11 gives the inversion of eq. H.5

$$\vec{u}_{\vec{k}} = -i\frac{\vec{k}}{k^2}\theta_{\vec{k}} \ , \tag{H.12}$$

obviously non bijective due to assumption H.8.

On the other hand concerning the extraction of $P_{u_iu_i}$ from eq. H.7, it can first be written that (using eq. H.12)

$$\left\langle \vec{u}_{\vec{k}}.\vec{u}^*_{\vec{k}} \right\rangle = 3P_{u_iu_i}(k) = P_{\theta\theta}(k)/k^2 \ , \tag{H.13}$$

allowing to write

$$P_{\theta\theta}(k) = 3k^2P_{u_iu_i}(k) \ . \tag{H.14}$$

In addition, some other useful relations can be derived. For instance when identifying eq. H.7 with eq. H.14, we get the relation

$$P_{u_iu_j}(k) = \frac{k^2}{k_xk_y + k_xk_z + k_yk_z}P_{u_iu_i}(k) \ , \tag{H.15}$$

resulting in the non intuitive fact that cross velocity component spectra are non zero, as it would be thought to be uncorrelated. Moreover, coming back to eq. H.10,

$$P_{u_iu_j}(k) \equiv \left(u_{k_i}u^*_{k_j} + u_{k_j}u^*_{k_i}\right)/2 = k_ik_jP_{\phi\phi}(k) \ , \tag{H.16}$$

which can be combined with eq. H.7 to give

$$P_{\theta\theta} = P_{\phi\phi}\left(k_x^4 + k_y^4 + k_z^4 + 2\left[k_x^2k_y^2 + k_x^2k_z^2 + k_y^2k_z^2\right]\right) = k^4P_{\phi\phi} \ , \tag{H.17}$$

consistent with H.11. Finally, eq. H.16 and eq. H.17 give the equivalent of eq. H.14 for the cross-velocity power spectrum, namely

$$P_{\theta\theta} = \frac{k^4}{k_ik_j}P_{u_iu_j} \ . \tag{H.18}$$

All of these derived expressions can be related to the true velocity field using definition H.2.

Appendix I

Angular power spectrum shot noise

In this appendix, I derive the shot noise prediction in the case of the angular power spectrum for various radial weighting functions.

In the case of the power spectrum in Fourier space, the shot noise computed in appendix G reads

$$P^{\mathrm{SN}}(\vec{k}) = \frac{1}{(2\pi)^3\bar{\rho}} \ , \tag{I.1}$$

where $\bar{\rho}$ is the average density of the simulated box. In this context, let relax the density parameter $\bar{\rho} \to \bar{\rho}(r(z))$ where $r(z)$ is the radial comoving position. Replacing this results in A.19 and using that

$$\int \mathrm{d}k k^2 j_\ell(kr_1) j_\ell(kr_2) = \frac{\pi}{2r^2}\delta^D(r_1 - r_2) \ , \tag{I.2}$$

or equivalently in redshift-space

$$\int \mathrm{d}k k^2 j_\ell(kr(z_1)) j_\ell(kr(z_2)) = \frac{\pi}{2}\delta^D(r(z_1) - r(z_2)) \ , \tag{I.3}$$

the shot noise contribution to the C_ℓ is (dropping the tilde notation):

$$C_\ell^{\mathrm{SN}}(r_1, r_2) = \int \mathrm{d}r \frac{W_1(r)W_2(r)}{r^2\bar{\rho}(r)} \ , \quad C_\ell^{\mathrm{SN}}(z_1, z_2) = \int \mathrm{d}z \frac{W_1(z)W_2(z)}{\bar{\rho}(z)} \tag{I.4}$$

Following these results, four cases can be discussed:

- Taking two non overlapping window functions, the shot noise contribution to any cross angular power spectrum is zero
- Let $W_1(r(z)) \neq W_2(r(z))$ defined with non zeros values in their respective intervals R_1, R_2 such that $R_1 \cup R_2 \neq 0$. The Poisson density will only contribute to the shot noise thought $\rho(R_1 \cup R_2)$
- Let $W_1(r(z)) = W_2(r(z))$, the auto angular power spectrum will take the form

$$C_\ell^{\mathrm{SN}}(r_1) = \int \mathrm{d}r \frac{W_1(r)^2}{r^2\bar{\rho}(r)} \ , \quad C_\ell^{\mathrm{SN}}(z_1) = \int \mathrm{d}z \frac{W_1(z)^2}{\bar{\rho}(z)} \ . \tag{I.5}$$

Applying the example of a Top-Hat window function in the interval $[R_1, R_2]$, and notifying that the total number of objects in this interval is given by

$N = 4\pi \int_{R_1}^{R_2} drr^2\bar{\rho}(r) = \int_{z_1}^{z_2} dz\bar{\rho}(z)$, the shot noise will contribute in the same way in booth space as

$$C_\ell^{\mathrm{SN,TH}}(r_1) = \frac{4\pi}{N} \tag{I.6}$$

- In the case of some arbitrary window functions, the shot noise terms can be obtained by analytically or numerically integrating the relations I.4. Nevertheless, defining

  $$W(r) = Ar^2 f(r), \quad W(z) = Af(z) \tag{I.7}$$

  $$w_i = \frac{f(r_i)}{\bar{\rho}(r_i)} = \frac{f(z_i)}{\bar{\rho}(z_i)}\ , \tag{I.8}$$

  where A is obtained applying A.10, a general expression of the shot noise can be obtained by discretizing relations I.4:

  $$C_\ell^{\mathrm{SN}} = 4\pi \frac{\sum_{i=1}^N w_i^2}{\left(\sum_{i=1}^N w_i^2\right)^2}\ . \tag{I.9}$$

Appendix J

Gaussian variance of the C_ℓ

In this appendix, I derive the Gaussian contribution to the covariance matrix both for the auto-C_ℓ and for the cross-C_ℓ.

The general expression of the angular power spectrum covariance matrix reads

$$\mathcal{C}_{\ell\ell'} = \left\langle \hat{C}_\ell \hat{C}_{\ell'} \right\rangle - \left\langle \hat{C}_\ell \right\rangle \left\langle \hat{C}_{\ell'} \right\rangle \tag{J.1}$$

that can be developed in the case of auto-C_ℓ (considering a single radial weighing function) thanks to its estimator

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \delta_\ell^m \delta_\ell^{m*} \text{ with } \delta_\ell^m(r) = 4\pi i^\ell \int \mathrm{d}^3\vec{k}\ \delta_{\vec{k}} Y_\ell^m(\theta_{\vec{k}}, \phi_{\vec{k}})^* j_\ell(kr)\ . \tag{J.2}$$

Using the Wick's theorem in the case where δ is a Gaussian field, the first contribution is split in three terms:

$$\begin{aligned}\left\langle \hat{C}_\ell \hat{C}_{\ell'} \right\rangle = \frac{1}{(2\ell+1)(2\ell'+1)} \sum_{m=-\ell}^{\ell} \sum_{m'=-\ell'}^{\ell'} \Bigg[&\left\langle |\delta_\ell^m|^2 \right\rangle \left\langle |\delta_{\ell'}^{m'}|^2 \right\rangle + \left\langle \delta_\ell^{m*} \delta_{\ell'}^{m'} \right\rangle \left\langle \delta_\ell^m \delta_{\ell'}^{m'*} \right\rangle + \\ &+ \left\langle \delta_\ell^m \delta_{\ell'}^{m'} \right\rangle \left\langle \delta_\ell^{m*} \delta_{\ell'}^{m'*} \right\rangle \Bigg]\end{aligned} \tag{J.3}$$

where the first term exactly corresponds to the opposite of the second term of J.1. It is therefore not necessary to calculate it as they will be suppressed. The second term can be easily tackled applying standard definitions recalled in J.2 while for the last one, although more challenging, requires the relation already obtained in A.18. At the end of the day, it yields

$$\left\langle \delta_\ell^{m*} \delta_{\ell'}^{m'} \right\rangle = \left\langle \delta_\ell^m \delta_{\ell'}^{m'*} \right\rangle = \left\langle \delta_\ell^m \delta_{\ell'}^{m'} \right\rangle = \left\langle \delta_\ell^{m*} \delta_{\ell'}^{m'*} \right\rangle = C_\ell \delta_{\ell\ell'}^K\ . \tag{J.4}$$

Putting all together, the covariance matrix of the two-point correlation function in harmonic space, in the case of an underlying Gaussian field takes the form

$$\mathcal{C}_{\ell\ell'} = \frac{2{C_\ell}^2}{\sqrt{(2\ell+1)(2\ell'+1)}} \delta_{\ell\ell'}^K \tag{J.5}$$

In the same way as in the Gaussian case of the Fourier power spectrum, the covariance matrix turns out to be also diagonal. The denominator representing, as for

M_{k_i} in equation 2.46, the number of independent angular modes, which is maximal for low ℓ.

Finally, when considering two disjoint radial weighing functions, the cross-C_ℓ estimator is changed in

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \delta_\ell^m \tilde{\delta}_\ell^{m*} \ , \tag{J.6}$$

leading to a estimated covariance matrix for the Gaussian case

$$\mathcal{C}_{\ell\ell'} = \frac{C_{\ell,1}C_{\ell,2} + XC_\ell^2}{\sqrt{(2\ell+1)(2\ell'+1)}} \delta^K_{\ell\ell'} \ . \tag{J.7}$$

Appendix K

Redshift and comoving space density relation

In this appendix, I present the relation allowing to get a constant radial density in redshift-space, from a constant radial density in comoving space.

In comoving flat space, the number of galaxy element is defined by

$$\mathrm{d}N = \bar{n}(r) r^2 \mathrm{d}r \mathrm{d}^2\Omega \tag{K.1}$$

where $\mathrm{d}^2\Omega = \sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$. This expression takes a slightly different form in redshift-space:

$$\mathrm{d}N = \bar{n}(z) \mathrm{d}z \mathrm{d}^2\Omega \ . \tag{K.2}$$

This difference makes it possible to define the infinitesimal element of volume without having to assume a metric. Identifying K.2 and K.1 allows to write

$$\bar{n}(z) = \bar{n}(r) \frac{cr^2}{H(z)} \ , \tag{K.3}$$

where the distance redshift relation

$$r(z) = \frac{c}{H_0} \int_0^z \frac{\mathrm{d}z}{\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}} = c \int_0^z \frac{\mathrm{d}z}{H(z)} \tag{K.4}$$

has been used.

Snapshots are initially sampled in comoving space where the density is constant, meaning that $\bar{n}(r) = \mathrm{cte}$. Going from comobile space to redshift-space naturally affects this property. In order to keep $\bar{n}(z) = \mathrm{cte}$ as well, lets define the quantity

$$\frac{1}{P_n(z)} = \frac{\bar{n}(z)}{\bar{n}(z_{min})} = \left(\frac{r(z)}{r_{min}}\right)^2 \sqrt{\frac{\Omega_{m,0}(1+z_{min})^3 + \Omega_{\Lambda,0}}{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}} \ . \tag{K.5}$$

For $z > z_{min}$, the quantity $P_n(z)$ is a number belonging to the interval [0,1] and is a decreasing function of r. Lets interpret it as a normalised probability. Once the mapping between comoving space and redshift-space is done, each sampled particle i is kept with a probability $P_n(z_i)$. This ensures a constant density in redshift-space. It also acts as a loss of statistics.

Bibliography

Abate A., et al., 2012, *Large Synoptic Survey Telescope: Dark Energy Science Collaboration*

Abbott B., et al., 2016, *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett., 116, 061102

Abbott B., et al., 2017, *Gravitational Waves and Gamma-rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A*, Astrophys. J. Lett., 848, L13

Abbott T., et al., 2018, *Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing*, Phys. Rev. D, 98, 043526

Abell P. A., et al., 2009, *LSST Science Book, Version 2.0*

Abellan G. F., Murgia R., Poulin V., Lavalle J., 2020, *Hints for decaying dark matter from S_8 measurements*

Adler R. J., Casey B., Jacob O. C., 1995, *Vacuum catastrophe: An elementary exposition of the cosmological constant problem*, American Journal of Physics, 63, 620

Aghamousa A., et al., 2016, *The DESI Experiment Part I: Science,Targeting, and Survey Design*

Aghanim N., et al., 2018, *Planck 2018 results. VI. Cosmological parameters*

Agrawal A., Makiya R., Chiang C.-T., Jeong D., Saito S., Komatsu E., 2017, *Generating Log-normal Mock Catalog of Galaxies in Redshift Space*, JCAP, 10, 003

Ahmad Q., et al., 2002, *Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory*, Phys. Rev. Lett., 89, 011301

Aiola S., et al., 2020, *The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters*

Akaike H., 1974, *A new look at the statistical model identification*, IEEE Transactions on Automatic Control, 19, 716

Aker M., et al., 2019, *Improved Upper Limit on the Neutrino Mass from a Direct Kinematic Method by KATRIN*, Phys. Rev. Lett., 123, 221802

Alam S., et al., 2017, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample*, Mon. Not. Roy. Astron. Soc., 470, 2617

Alam S., et al., 2020, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory*

Alpher R., Bethe H., Gamow G., 1948, *The origin of chemical elements*, Phys. Rev., 73, 803

Archidiacono M., Hannestad S., Lesgourgues J., 2020, *What will it take to measure individual neutrino mass states using cosmology?*, JCAP, 09, 021

Armendariz-Picon C., Mukhanov V. F., Steinhardt P. J., 2000, *A Dynamical solution to the problem of a small cosmological constant and late time cosmic acceleration*, Phys. Rev. Lett., 85, 4438

Ashtari Esfahani A., et al., 2017, *Determining the neutrino mass with cyclotron radiation emission spectroscopy Project 8*, J. Phys. G, 44, 054004

Asorey J., Crocce M., Gaztañaga E., Lewis A., 2012, *Recovering 3D clustering information with angular correlations*, mnras, 427, 1891

Audren B., Lesgourgues J., Benabed K., Prunet S., 2013, *Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code*, JCAP, 1302, 001

Avila S., Murray S. G., Knebe A., Power C., Robotham A. S. G., Garcia-Bellido J., 2015, *HALOGEN: A tool for fast generation of mock halo catalogues*, Mon. Not. Roy. Astron. Soc., 450, 1856

Baratta P., Bel J., Plaszczynski S., Ealet A., 2020, *High-precision Monte-Carlo modelling of galaxy distribution*, Astron. Astrophys., 633, A26

Bardeen J. M., Bond J., Kaiser N., Szalay A., 1986, *The Statistics of Peaks of Gaussian Random Fields*, Astrophys. J., 304, 15

Bassett B. A., Tsujikawa S., Wands D., 2006, *Inflation dynamics and reheating*, Rev. Mod. Phys., 78, 537

Bateman H., 1953, *Higher transcendental functions [volumes i-iii]*. Vol. 2, McGraw-Hill Book Company

Baumann D., 2011, in Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small. pp 523–686 (`arXiv:0907.5424`), doi:10.1142/9789814327183_0010

Bel J., Pezzotta A., Carbone C., Sefusatti E., Guzzo L., 2019, *Accurate fitting functions for peculiar velocity spectra in standard and massive-neutrino cosmologies*, Astron. Astrophys., 622, A109

Bellini E., Sawicki I., Zumalacárregui M., 2020, *hi_class: Background Evolution, Initial Conditions and Approximation Schemes*, JCAP, 02, 008

Bernardeau F., van de Weygaert R., 1996, *A new method for accurate estimation of velocity field statistics*, mnras, 279, 693

Bernardeau F., van de Weygaert R., Hivon E., Bouchet F. R., 1997, *The Omega dependence of the velocity divergence distribution*, mnras, 290, 566

Bernardeau F., Colombi S., Gaztanaga E., Scoccimarro R., 2002, *Large scale structure of the universe and cosmological perturbation theory*, Phys. Rept., 367, 1

Beutler F., et al., 2011, *The 6dF Galaxy Survey: Baryon Acoustic Oscillations and the Local Hubble Constant*, Mon. Not. Roy. Astron. Soc., 416, 3017

Bik A., et al., 2012, *Age Spread in W3 Main*, ApJ, 744, 87

Bird S., Viel M., Haehnelt M. G., 2012, *Massive Neutrinos and the Non-linear Matter Power Spectrum*, Mon. Not. Roy. Astron. Soc., 420, 2551

Blas D., Lesgourgues J., Tram T., 2011, *The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes*, JCAP, 07, 034

Blot L., Corasaniti P. S., Alimi J.-M., Reverdy V., Rasera Y., 2015, *Matter power spectrum covariance matrix from the DEUS-PUR ΛCDM simulations: mass resolution and non-Gaussian errors*, Mon. Not. Roy. Astron. Soc., 446, 1756

Blot L., et al., 2019, *Comparing approximate methods for mock catalogues and covariance matrices II: Power spectrum multipoles*, Mon. Not. Roy. Astron. Soc., 485, 2806

Blumenthal G. R., Faber S., Primack J. R., Rees M. J., 1984, *Formation of Galaxies and Large Scale Structure with Cold Dark Matter*, Nature, 311, 517

Boisseau B., Esposito-Farese G., Polarski D., Starobinsky A. A., 2000, *Reconstruction of a scalar tensor theory of gravity in an accelerating universe*, Phys. Rev. Lett., 85, 2236

Bond J. R., Myers S. T., 1996, *The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms*, apjs, 103, 1

Bond J., Efstathiou G., Silk J., 1980, *Massive Neutrinos and the Large Scale Structure of the Universe*, Phys. Rev. Lett., 45, 1980

Bond J., Jaffe A. H., Knox L., 1998, *Estimating the power spectrum of the cosmic microwave background*, Phys. Rev. D, 57, 2117

Bonvin C., Durrer R., 2011, *What galaxy surveys really measure*, prd, 84, 063505

Borde A., Guth A. H., Vilenkin A., 2003, *Inflationary space-times are incompletein past directions*, Phys. Rev. Lett., 90, 151301

Bosma A., 1981, *21-cm line studies of spiral galaxies. II. The distribution and kinematics of neutral hydrogen in spiral galaxies of various morphological types.*, aj, 86, 1825

Brans C., Dicke R., 1961, *Mach's principle and a relativistic theory of gravitation*, Phys. Rev., 124, 925

Buchert T., 2000, *On average properties of inhomogeneous fluids in general relativity. 1. Dust cosmologies*, Gen. Rel. Grav., 32, 105

Cahill K., 2019, *Physical Mathematics*, 2 edn. Cambridge University Press, doi:10.1017/9781108555814

Cai Y.-C., Bernstein G., 2012, *Combining weak-lensing tomography and spectroscopic redshift surveys*, mnras, 422, 1045

Caldwell R., Dave R., Steinhardt P. J., 1998, *Cosmological imprint of an energy component with general equation of state*, Phys. Rev. Lett., 80, 1582

Campagne J. E., Neveu J., Plaszczynski S., 2017, *Angpow: a software for the fast computation of accurate tomographic power spectra*, Astron. Astrophys., 602, A72

Campagne J.-E., Neveu Plaszczynski S., 2018, *AngPow: Fast computation of accurate tomographic power spectra*, Astrophysics Source Code Library (ascl:1807.012)

Carbone C., Petkova M., Dolag K., 2016, *DEMNUni: ISW, Rees-Sciama, and weak-lensing in the presence of massive neutrinos*, JCAP, 07, 034

Castorina E., Carbone C., Bel J., Sefusatti E., Dolag K., 2015, *DEMNUni: The clustering of large-scale structures in the presence of massive neutrinos*, JCAP, 07, 043

Chen H.-Y., Fishbach M., Holz D. E., 2018, *A two per cent Hubble constant measurement from standard sirens within five years*, Nature, 562, 545

Cheung C., Creminelli P., Fitzpatrick A., Kaplan J., Senatore L., 2008, *The Effective Field Theory of Inflation*, JHEP, 03, 014

Chevallier M., Polarski D., 2001, *Accelerating universes with scaling dark matter*, Int. J. Mod. Phys. D, 10, 213

Chiang C.-T., et al., 2013, *Galaxy redshift surveys with sparse sampling*, jcap, 2013, 030

Chow N., Khoury J., 2009, *Galileon Cosmology*, Phys. Rev. D, 80, 024037

Christenson J. H., Cronin J. W., Fitch V. L., Turlay R., 1964, *Evidence for the* 2π *Decay of the* K_2^0 *Meson*, Phys. Rev. Lett., 13, 138

Codis S., Pichon C., Bernardeau F., Uhlemann C., Prunet S., 2016, *Encircling the dark: constraining dark energy via cosmic density in spheres*, Mon. Not. Roy. Astron. Soc., 460, 1549

Colavincenzo M., et al., 2019, *Comparing approximate methods for mock catalogues and covariance matrices – III: bispectrum*, Mon. Not. Roy. Astron. Soc., 482, 4883

Coles P., Barrow J. D., 1987, *Non-Gaussian statistics and the microwave background radiation*, mnras, 228, 407

Coles P., Jones B., 1991, *A lognormal model for the cosmological mass distribution.*, mnras, 248, 1

Cowan C., Reines F., Harrison F., Kruse H., McGuire A., 1956, *Detection of the free neutrino: A Confirmation*, Science, 124, 103

Creminelli P., Luty M. A., Nicolis A., Senatore L., 2006, *Starting the Universe: Stable Violation of the Null Energy Condition and Non-standard Cosmologies*, JHEP, 12, 080

Creminelli P., Tambalo G., Vernizzi F., Yingcharoenrat V., 2020, *Dark-Energy Instabilities induced by Gravitational Waves*, JCAP, 05, 002

Davis M., Peebles P., 1983, *Evidence for local anisotropy of the hubble flow*, Ann. Rev. Astron. Astrophys., 21, 109

Dawson K. S., et al., 2013, *The Baryon Oscillation Spectroscopic Survey of SDSS-III*, Astron. J., 145, 10

De Felice A., Tsujikawa S., 2010, *f(R) theories*, Living Rev. Rel., 13, 3

De Sitter W., 1917.

Deffayet C., Esposito-Farese G., Vikman A., 2009, *Covariant Galileon*, Phys. Rev. D, 79, 084003

Deffayet C., Pujolas O., Sawicki I., Vikman A., 2010, *Imperfect Dark Energy from Kinetic Gravity Braiding*, JCAP, 10, 026

Devroye L., 1986, *Non-Uniform Random Variate Generation*. Springer-Verlag

Di Dio E., Montanari F., Durrer R., Lesgourgues J., 2014, *Cosmological Parameter Estimation with Large Scale Structure Observations*, JCAP, 01, 042

Di Valentino E., Melchiorri A., Silk J., 2019, *Planck evidence for a closed Universe and a possible crisis for cosmology*, Nature Astron., 4, 196

Di Valentino E., et al., 2020a, *Cosmology Intertwined II: The Hubble Constant Tension*

Di Valentino E., et al., 2020c, *Cosmology Intertwined III:* $f\sigma_8$ *and* S_8

Di Valentino E., Mukherjee A., Sen A. A., 2020b, *Dark Energy with Phantom Crossing and the* H_0 *tension*

Di Valentino E., Melchiorri A., Mena O., Vagnozzi S., 2020d, *Nonminimal dark sector physics and cosmological tensions*, Phys. Rev. D, 101, 063502

Dicke R., Peebles P., Roll P., Wilkinson D., 1965, *Cosmic Black-Body Radiation*, Astrophys. J., 142, 414

Dodelson S., 2003, *Modern Cosmology*. Academic Press, Amsterdam

Dodelson S., Schneider M. D., 2013, *The effect of covariance estimator error on cosmological parameter constraints*, Phys. Rev. D, 88, 063537

Drees M., Gerbier G., 2012, *Mini-Review of Dark Matter: 2012*

Dvali G., Gabadadze G., Porrati M., 2000, *4-D gravity on a brane in 5-D Minkowski space*, Phys. Lett. B, 485, 208

Dyson F. W., Eddington A. S., Davidson C., 1920, *A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919*, Philosophical Transactions of the Royal Society of London Series A, 220, 291

Efstathiou G., Gratton S., 2020, *The evidence for a spatially flat Universe*, Mon. Not. Roy. Astron. Soc., 496, L91

Efstathiou G., Bond J., White S. D., 1992, *COBE Background radiation anisotropies and large scale structure in the universe*, Mon. Not. Roy. Astron. Soc., 258, 1

Einstein A., 1905, *On the electrodynamics of moving bodies*, Annalen Phys., 17, 891

Einstein A., 1916, *The Foundation of the General Theory of Relativity*, Annalen Phys., 49, 769

Einstein A., 1917, *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*, Sitzungsberichte der Koniglich Preussischen Akademie der Wissenschaften (Berlin, pp 142–152

Eisenstein D. J., 1997, *An Analytic expression for the growth function in a flat universe with a cosmological constant*

Eisenstein D. J., Hu W., 1998a, *Baryonic Features in the Matter Transfer Function*, The Astrophysical Journal, 496, 605

Eisenstein D. J., Hu W., 1998b, *Baryonic features in the matter transfer function*, Astrophys. J., 496, 605

Eisenstein D. J., Hu W., Tegmark M., 1998, *Cosmic Complementarity: H 0 and Omega M from Combining Cosmic Microwave Background Experiments and Redshift Surveys*, The Astrophysical Journal Letters, 504, L57

Eisenstein D. J., Hu W., Tegmark M., 1999, *Cosmic complementarity: Joint parameter estimation from CMB experiments and redshift surveys*, Astrophys. J., 518, 2

Eisenstein D. J., et al., 2011, *SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems*, Astron. J., 142, 72

Esteban I., Gonzalez-Garcia M., Hernandez-Cabezudo A., Maltoni M., Schwetz T., 2019, *Global analysis of three-flavour neutrino oscillations: synergies and tensions in the determination of θ_{23}, δ_{CP}, and the mass ordering*, JHEP, 01, 106

Feldman H. A., Kaiser N., Peacock J. A., 1994, *Power spectrum analysis of three-dimensional redshift surveys*, Astrophys. J., 426, 23

Fogli G., Lisi E., Marrone A., Montanino D., Palazzo A., Rotunno A., 2012, *Global analysis of neutrino masses, mixings and phases: entering the era of leptonic CP violation searches*, Phys. Rev. D, 86, 013012

Forero D., Tortola M., Valle J., 2012, *Global status of neutrino oscillation parameters after Neutrino-2012*, Phys. Rev. D, 86, 073012

Fosalba P., Gaztanaga E., Castander F., Manera M., 2008, *The onion universe: all sky light-cone simulations in shells*, Mon. Not. Roy. Astron. Soc., 391, 435

Friedrich O., Seitz S., Eifler T., Gruen D., 2016, *Performance of internal Covariance Estimators for Cosmic Shear Correlation Functions*, Mon. Not. Roy. Astron. Soc., 456, 2662

Fry J. N., 1984a, *Galaxy N-point correlation functions - Theoretical amplitudes for arbitrary N*, apjl, 277, L5

Fry J. N., 1984b, *The Galaxy correlation hierarchy in perturbation theory*, Astrophys. J., 279, 499

Fry J. N., 1984c, *The Galaxy correlation hierarchy in perturbation theory*, apj, 279, 499

Fry J. N., 1985, *Dynamical measure of density in exotic cosmologies*, Phys. Lett. B, 158, 211

Fukuda Y., et al., 1998, *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett., 81, 1562

Gaztañaga E., Eriksen M., Crocce M., Castander F. J., Fosalba P., Marti P., Miquel R., Cabré A., 2012, *Cross-correlation of spectroscopic and photometric galaxy surveys: cosmology from lensing and redshift distortions*, mnras, 422, 2904

Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, apj, 622, 759

Gouyou Beauchamp S., in in prep., *paper*

Greiner M., Enßlin T. A., 2015, *Log-transforming the matter power spectrum*, aap, 574, A86

Grieb J. N., Sánchez A. G., Salazar-Albornoz S., Dalla Vecchia C., 2016, *Gaussian covariance matrices for anisotropic galaxy clustering measurements*, Monthly Notices of the Royal Astronomical Society, 457, 1577

Grieb J. N., et al., 2017, *The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the Fourier space wedges of the final sample*, Mon. Not. Roy. Astron. Soc., 467, 2085

Gubitosi G., Piazza F., Vernizzi F., 2013, *The Effective Field Theory of Dark Energy*, JCAP, 02, 032

Guth A. H., 1987, *The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems*, Adv. Ser. Astrophys. Cosmol., 3, 139

Hahn O., Angulo R. E., Abel T., 2015, *The Properties of Cosmic Velocity Fields*, Mon. Not. Roy. Astron. Soc., 454, 3920

Hamilton A., 1997, in Ringberg Workshop on Large Scale Structure. (`arXiv:astro-ph/9708102`), doi:10.1007/978-94-011-4960-0_17

Hamilton A. J., Rimes C. D., Scoccimarro R., 2006, *On measuring the covariance matrix of the nonlinear power spectrum from simulations*, Mon. Not. Roy. Astron. Soc., 371, 1188

Harrison E. R., 1970, *Fluctuations at the threshold of classical cosmology*, Phys. Rev. D, 1, 2726

Hart L., Chluba J., 2018, *New constraints on time-dependent variations of fundamental constants using Planck data*, Mon. Not. Roy. Astron. Soc., 474, 1850

Hartlap J., Simon P., Schneider P., 2007, *Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix*, aap, 464, 399

Heath D. J., 1977, *The growth of density perturbations in zero pressure Friedmann-Lemaitre universes.*, mnras, 179, 351

Henning J., et al., 2018, *Measurements of the Temperature and E-Mode Polarization of the CMB from 500 Square Degrees of SPTpol Data*, Astrophys. J., 852, 97

Heymans C., et al., 2020, *KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints*, arXiv e-prints, p. arXiv:2007.15632

Hildebrandt H., et al., 2017, *KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing*, Mon. Not. Roy. Astron. Soc., 465, 1454

Hildebrandt H., et al., 2020, *KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data*, Astron. Astrophys., 633, A69

Horndeski G. W., 1974, *Second-order scalar-tensor field equations in a four-dimensional space*, Int. J. Theor. Phys., 10, 363

Howlett C., Lewis A., Hall A., Challinor A., 2012, *CMB power spectrum parameter degeneracies in the era of precision cosmology*, Journal of Cosmology and Astroparticle Physics, 2012, 027

Hu W., Kravtsov A. V., 2003, *Sample variance considerations for cluster surveys*, Astrophys. J., 584, 702

Hu W., Eisenstein D. J., Tegmark M., 1998, *Weighing neutrinos with galaxy surveys*, Phys. Rev. Lett., 80, 5255

Hu B., Raveri M., Frusciante N., Silvestri A., 2014, *Effective Field Theory of Cosmic Acceleration: an implementation in CAMB*, Phys. Rev. D, 89, 103530

Huang C. D., et al., 2019, *Hubble Space Telescope Observations of Mira Variables in the Type Ia Supernova Host NGC 1559: An Alternative Candle to Measure the Hubble Constant*,] 10.3847/1538-4357/ab5dbd

Hubble E., 1929a, *A relation between distance and radial velocity among extra-galactic nebulae*, Proc. Nat. Acad. Sci., 15, 168

Hubble E., 1929b, *A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae*, Proceedings of the National Academy of Science, 15, 168

Hubble E., 1934, *The Distribution of Extra-Galactic Nebulae*, apj, 79, 8

Ijjas A., Steinhardt P. J., Loeb A., 2013, *Inflationary paradigm in trouble after Planck2013*, Phys. Lett. B, 723, 261

Ivanov M. M., Simonović M., Zaldarriaga M., 2020, *Cosmological Parameters from the BOSS Galaxy Power Spectrum*, JCAP, 05, 042

Izard A., Fosalba P., Crocce M., 2018, *ICE-COLA: fast simulations for weak lensing observables*, Mon. Not. Roy. Astron. Soc., 473, 3051

Jackson J., 1972, *Fingers of God: A critique of Rees' theory of primoridal gravitational radiation*, Mon. Not. Roy. Astron. Soc., 156, 1P

Jedamzik K., Pogosian L., 2020, *Relieving the Hubble tension with primordial magnetic fields*

Joudaki S., et al., 2018, *KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering*, Mon. Not. Roy. Astron. Soc., 474, 4894

Jungman G., Kamionkowski M., Griest K., 1996, *Supersymmetric dark matter*, Phys. Rept., 267, 195

Kaiser N., 1987, *Clustering in real space and in redshift space*, Mon. Not. Roy. Astron. Soc., 227, 1

Kenworthy W. D., Scolnic D., Riess A., 2019, *The Local Perspective on the Hubble Tension: Local Structure Does Not Impact Measurement of the Hubble Constant*, Astrophys. J., 875, 145

Kibble W., 1945, *An extension of a theorem of Mehler's on Hermite polynomials*, Mathematical Proceedings of the Cambridge Philosophical Society, 41, 12

Kitaura F.-S., Yepes G., Prada F., 2013, *Modelling baryon acoustic oscillations with perturbation theory and stochastic halo biasing*, Monthly Notices of the Royal Astronomical Society: Letters, 439, L21

Kitaura F.-S., Gil-Marín H., Scóccola C. G., Chuang C.-H., Müller V., Yepes G., Prada F., 2015, *Constraining the halo bispectrum in real and redshift space from perturbation theory and non-linear stochastic bias*, Monthly Notices of the Royal Astronomical Society, 450, 1836

Klypin A., Prada F., Betancort-Rijo J., Albareti F. D., 2018, *Density distribution of the cosmological matter field*, Mon. Not. Roy. Astron. Soc., 481, 4588

Kreisch C. D., Cyr-Racine F.-Y., Doré O., 2020, *Neutrino puzzle: Anomalies, interactions, and cosmological tensions*, Phys. Rev. D, 101, 123505

Kurtz S., 2005, *Hypercompact HII regions*, IAU Symp., 1, 111

Lacasa F., Kunz M., 2017, *Inadequacy of internal covariance estimation for super-sample covariance*, Astron. Astrophys., 604, A104

Lahav O., Lilje P. B., Primack J. R., Rees M. J., 1991, *Dynamical effects of the cosmological constant*, Mon. Not. Roy. Astron. Soc., 251, 128

Laureijs R., et al., 2011, *Euclid Definition Study Report*

Layzer D., 1956, *A new model for the distribution of galaxies in space*, aj, 61, 383

Leavitt H. S., Pickering E. C., 1912, *Periods of 25 Variable Stars in the Small Magellanic Cloud.*, Harvard College Observatory Circular, 173, 1

Lesgourgues J., Pastor S., 2006, *Massive neutrinos and cosmology*, Phys. Rept., 429, 307

Levi M. E., et al., 2019, *The Dark Energy Spectroscopic Instrument (DESI)*

Lewis A., Challinor A., Lasenby A., 2000, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, The Astrophysical Journal, 538, 473

Li X., Shafieloo A., 2019, *A Simple Phenomenological Emergent Dark Energy Model can Resolve the Hubble Tension*, Astrophys. J. Lett., 883, L3

Li Y., Hu W., Takada M., 2014, *Super-Sample Covariance in Simulations*, Phys. Rev. D, 89, 083519

Liddle A. R., 2004, *How many cosmological parameters?*, Mon. Not. Roy. Astron. Soc., 351, L49

Linde A. D., 1987, *A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems*, Adv. Ser. Astrophys. Cosmol., 3, 149

Linder E. V., 2003, *Exploring the expansion history of the universe*, Phys. Rev. Lett., 90, 091301

Lippich M., et al., 2019, *Comparing approximate methods for mock catalogues and covariance matrices – I. Correlation function*, Mon. Not. Roy. Astron. Soc., 482, 1786

Marinoni C., Bel J., Buzzi A., 2012, *The Scale of Cosmic Isotropy*, JCAP, 10, 036

Mehler F., 1866, *Ueber die Entwicklung einer Function von beliebig vielen Variablen nach Laplaceschen Functionen höherer Ordnung.*, Journal für die reine und angewandte Mathematik, 66, 161

Meiksin A., White M. J., 1999, *The Growth of correlations in the matter power spectrum*, Mon. Not. Roy. Astron. Soc., 308, 1179

Meszaros P., 1974, *The behaviour of point masses in an expanding cosmological substratum*, Astron. Astrophys., 37, 225

Misner C. W., Thorne K. S., Wheeler J. A., 1973, *Gravitation*

Monaco P., Theuns T., Taffoni G., 2002, *Pinocchio: pinpointing orbit-crossing collapsed hierarchical objects in a linear density field*, Mon. Not. Roy. Astron. Soc., 331, 587

Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, *An accurate tool for the fast generation of dark matter halo catalogues*, Monthly Notices of the Royal Astronomical Society, 433, 2389

Montanari F., Durrer R., 2012, *New method for the Alcock-Paczyński test*, prd, 86, 063503

Newton I., 1687, *Philosophiæ Naturalis Principia Mathematica*. England

Nicolis A., Rattazzi R., Trincherini E., 2009, *The Galileon as a local modification of gravity*, Phys. Rev. D, 79, 064036

Nojiri S., Odintsov S. D., 2005, *Modified Gauss-Bonnet theory as gravitational alternative for dark energy*, Phys. Lett. B, 631, 1

Noller J., von Braun-Bates F., Ferreira P. G., 2014, *Relativistic scalar fields and the quasistatic approximation in theories of modified gravity*, Phys. Rev. D, 89, 023521

Norberg P., Baugh C. M., Gaztanaga E., Croton D. J., 2009a, *Statistical Analysis of Galaxy Surveys - I. Robust error estimation for 2-point clustering statistics*, Mon. Not. Roy. Astron. Soc., 396, 19

Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009b, *Statistical analysis of galaxy surveys – I. Robust error estimation for two-point clustering statistics*, Monthly Notices of the Royal Astronomical Society, 396, 19

Nunes R. C., Bernui A., 2020, $\theta_{\rm BAO}$ *estimates and the* H_0 *tension*

Pauli W., 1930, *Pauli letter collection: letter to Lise Meitner*, Typed copy, `http://cds.cern.ch/record/83282`

Peacock J., Dodds S., 1994, *Reconstructing the linear power spectrum of cosmological mass fluctuations*, Mon. Not. Roy. Astron. Soc., 267, 1020

Peacock J. A., Nicholson D., 1991, *The large-scale clustering of radio galaxies*, Monthly Notices of the Royal Astronomical Society, 253, 307

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

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

Peebles P., 1982, *Large scale background temperature and mass fluctuations due to scale invariant primeval perturbations*, Astrophys. J. Lett., 263, L1

Penzias A. A., Wilson R. W., 1965, *A Measurement of Excess Antenna Temperature at 4080 Mc/s.*, apj, 142, 419

Percival W. J., et al., 2014a, *The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors*, Monthly Notices of the Royal Astronomical Society, 439, 2531

Percival W. J., et al., 2014b, *The Clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Including covariance matrix errors*, Mon. Not. Roy. Astron. Soc., 439, 2531

Perenon L., Velten H., 2019, *The effective field theory of dark energy diagnostic of linear Horndeski theories after GW170817 and GRB170817A*, Universe, 5, 138

Perenon L., Piazza F., Marinoni C., Hui L., 2015, *Phenomenology of dark energy: general features of large-scale perturbations*, JCAP, 11, 029

Perlmutter S., Schmidt B. P., 2003a, *Measuring Cosmology with Supernovae*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 195–217, doi:10.1007/3-540-45863-8_11, `https://doi.org/10.1007/3-540-45863-8_11`

Perlmutter S., Schmidt B. P., 2003b, *Measuring cosmology with supernovae*, Lect. Notes Phys., 598, 195

Perlmutter S., et al., 1999, *Measurements of Ω and Λ from 42 high redshift supernovae*, Astrophys. J., 517, 565

Piazza F., Steigerwald H., Marinoni C., 2014, *Phenomenology of dark energy: exploring the space of theories with future redshift surveys*, JCAP, 05, 043

Pontecorvo B., 1957, *Mesonium and anti-mesonium*, Sov. Phys. JETP, 6, 429

Pontecorvo B., 1958, *Inverse beta processes and non conservation of lepton charge*, Sov. Phys. JETP, 7, 172

Poulin V., Smith T. L., Karwal T., Kamionkowski M., 2019, *Early Dark Energy Can Resolve The Hubble Tension*, Phys. Rev. Lett., 122, 221301

Primack J. R., Gross M. A., 2000, *Hot dark matter in cosmology*, pp 287–308

Pueblas S., Scoccimarro R., 2009a, *Generation of Vorticity and Velocity Dispersion by Orbit Crossing*, Phys. Rev. D, 80, 043504

Pueblas S., Scoccimarro R., 2009b, *Generation of vorticity and velocity dispersion by orbit crossing*, prd, 80, 043504

Rasanen S., 2006, *Accelerated expansion from structure formation*, JCAP, 11, 003

Raveri M., Hu B., Frusciante N., Silvestri A., 2014, *Effective Field Theory of Cosmic Acceleration: constraining dark energy with CMB data*, Phys. Rev. D, 90, 043513

Reid M., Pesce D., Riess A., 2019, *An Improved Distance to NGC 4258 and its Implications for the Hubble Constant*, Astrophys. J. Lett., 886, L27

Riess A. G., et al., 1998, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, Astron. J., 116, 1009

Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, *Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond* Λ*CDM*, Astrophys. J., 876, 85

Riess A. G., Yuan W., Casertano S., Macri L. M., Scolnic D., 2020, *The Accuracy of the Hubble Constant Measurement Verified through Cepheid Amplitudes*, Astrophys. J. Lett., 896, L43

Riotto A., 2002, *Inflation and the Theory of Cosmological Perturbations*, arXiv:hep-ph/0210162v1

Riotto A., 2003, *Inflation and the theory of cosmological perturbations*, ICTP Lect. Notes Ser., 14, 317

Roberts M. S., 1969, *Integral Properties of Spiral and Irregular Galaxies*, aj, 74, 859

Rubin V. C., Ford W. Kent J., 1970, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, apj, 159, 379

Sahni V., Starobinsky A. A., 2000, *The Case for a positive cosmological Lambda term*, Int. J. Mod. Phys. D, 9, 373

Sakharov A., 1991, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Sov. Phys. Usp., 34, 392

Sawicki I., Bellini E., 2015, *Limits of quasistatic approximation in modified-gravity cosmologies*, Phys. Rev. D, 92, 084061

Schwarz G., 1978, *Estimating the Dimension of a Model*, Ann. Statist., 6, 461

Schwarzschild K., 1916, *On the gravitational field of a sphere of incompressible fluid according to Einstein's theory*, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.), 1916, 424

Scoccimarro R., 2004, *Redshift-space distortions, pairwise velocities and nonlinearities*, Phys. Rev. D, 70, 083007

Scoccimarro R., Zaldarriaga M., Hui L., 1999, *Power spectrum correlations induced by nonlinear clustering*, Astrophys. J., 527, 1

Sefusatti E., Crocce M., Scoccimarro R., Couchman H. M. P., 2016, *Accurate estimators of correlation functions in Fourier space*, mnras, 460, 3624

Sellentin E., Heavens A. F., 2015, *Parameter inference with estimated covariance matrices*, Monthly Notices of the Royal Astronomical Society: Letters, 456, L132

Simon O., 2016, PhD thesis, École Polytechnique fédérale de Lausanne

Smith R., et al., 2003, *Stable clustering, the halo model and nonlinear cosmological power spectra*, Mon. Not. Roy. Astron. Soc., 341, 1311

Smoot G. F., et al., 1992, *Structure in the COBE Differential Microwave Radiometer First-Year Maps*, apjl, 396, L1

Solà Peracaula J., Gomez-Valent A., de Cruz Pérez J., Moreno-Pulido C., 2019, *Brans-Dicke Gravity with a Cosmological Constant Smoothes Out ΛCDM Tensions*, Astrophys. J. Lett., 886, L6

Sola J., Gomez-Valent A., Perez J. d. C., Moreno-Pulido C., 2020, *Brans-Dicke cosmology with a Λ- term: a possible solution to ΛCDM tensions*

Sotiriou T. P., Faraoni V., 2010, *f(R) Theories Of Gravity*, Rev. Mod. Phys., 82, 451

Spergel D., et al., 2003, *First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters*, Astrophys. J. Suppl., 148, 175

Sprenger T., Archidiacono M., Brinckmann T., Clesse S., Lesgourgues J., 2019, *Cosmology in the era of Euclid and the Square Kilometre Array*, JCAP, 02, 047

Springel V., 2005, *The Cosmological simulation code GADGET-2*, Mon. Not. Roy. Astron. Soc., 364, 1105

Starobinsky A. A., 1987, *A New Type of Isotropic Cosmological Models Without Singularity*, Adv. Ser. Astrophys. Cosmol., 3, 130

Steigerwald H., Bel J., Marinoni C., 2014, *Probing non-standard gravity with the growth index: a background independent analysis*, JCAP, 05, 042

Suzuki N., et al., 2012, *The Hubble Space Telescope Cluster Supernova Survey: V. Improving the Dark Energy Constraints Above $z > 1$ and Building an Early-Type-Hosted Supernova Sample*, Astrophys. J., 746, 85

Takahashi R., et al., 2009, *Simulations of Baryon Acoustic Oscillations II: Covariance matrix of the matter power spectrum*, Astrophys. J., 700, 479

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, *Revising the Halofit Model for the Nonlinear Matter Power Spectrum*, Astrophys. J., 761, 152

Tanabashi M., et al., 2018, *Review of Particle Physics*, Phys. Rev. D, 98, 030001

Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, *Solving large scale structure in ten easy steps with COLA*, Journal of Cosmology and Astroparticle Physics, 2013, 036

Taylor A., Joachimi B., Kitching T., 2013, *Putting the precision in precision cosmology: How accurate should your data covariance matrix be?*, Monthly Notices of the Royal Astronomical Society, 432, 1928–1946

Tegmark M., 1997a, *How to measure CMB power spectra without losing information*, Phys. Rev. D, 55, 5895

Tegmark M., 1997b, *Measuring cosmological parameters with galaxy surveys*, Phys. Rev. Lett., 79, 3806

Troxel M., et al., 2018, *Dark Energy Survey Year 1 results: Cosmological constraints from cosmic shear*, Phys. Rev. D, 98, 043528

Uhlemann C., Codis S., Pichon C., Bernardeau F., Reimberg P., 2016, *Back in the saddle: Large-deviation statistics of the cosmic log-density field*, Mon. Not. Roy. Astron. Soc., 460, 1529

Vainshtein A., 1972, *To the problem of nonvanishing gravitation mass*, Phys. Lett. B, 39, 393

Vattis K., Koushiappas S. M., Loeb A., 2019, *Dark matter decaying in the late Universe can relieve the H0 tension*, Phys. Rev. D, 99, 121302

Viel M., Haehnelt M. G., Springel V., 2010, *The effect of neutrinos on the matter distribution as probed by the Intergalactic Medium*, JCAP, 06, 015

Villaescusa-Navarro F., et al., 2020, *The Quijote simulations*, Astrophys. J. Suppl., 250, 2

Wadekar D., Scoccimarro R., 2019, *The Galaxy Power Spectrum Multipoles Covariance in Perturbation Theory*

Wadekar D., Ivanov M. M., Scoccimarro R., 2020, *Cosmological constraints from BOSS with analytic covariance matrices*

Wall J. V., Jenkins C. R., 2003, *Practical Statistics for Astronomers*. Cambridge Observing Handbooks for Research Astronomers, Cambridge University Press, doi:10.1017/CBO9780511536618

Wang L.-M., Steinhardt P. J., 1998, *Cluster abundance constraints on quintessence models*, Astrophys. J., 508, 483

Weinberg S., 1989, *The cosmological constant problem*, Rev. Mod. Phys., 61, 1

Wilson M., Peacock J., Taylor A., de la Torre S., 2017, *Rapid modelling of the redshift-space power spectrum multipoles for a masked density field*, Mon. Not. Roy. Astron. Soc., 464, 3121

Wishart J., 1928, *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika, 20A, 32

Wong K. C., et al., 2019, *H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes*,] 10.1093/mnras/stz3094

Wu H.-Y., Huterer D., 2017, *Sample variance in the local measurements of the Hubble constant*, Mon. Not. Roy. Astron. Soc., 471, 4946

Xavier H. S., Abdalla F. B., Joachimi B., 2016a, *Improving lognormal models for cosmological fields*, mnras, 459, 3693

Xavier H. S., Abdalla F. B., Joachimi B., 2016b, *Improving lognormal models for cosmological fields*, Mon. Not. Roy. Astron. Soc., 459, 3693

Yang W., Mukherjee A., Di Valentino E., Pan S., 2018, *Interacting dark energy with time varying equation of state and the H_0 tension*, Phys. Rev. D, 98, 123527

Zel'Dovich Y. B., 1970, *Reprint of 1970A&A.....5...84Z. Gravitational instability: an approximate theory for large density perturbations.*, aap, 500, 13

Zeldovich Y., 1972, *A Hypothesis, unifying the structure and the entropy of the universe*, Mon. Not. Roy. Astron. Soc., 160, 1P

Zennaro M., Bel J., Villaescusa-Navarro F., Carbone C., Sefusatti E., Guzzo L., 2017, *Initial Conditions for Accurate N-Body Simulations of Massive Neutrino Cosmologies*, Mon. Not. Roy. Astron. Soc., 466, 3244

Zennaro M., Bel J., Dossett J., Carbone C., Guzzo L., 2018, *Cosmological constraints from galaxy clustering in the presence of massive neutrinos*, Mon. Not. Roy. Astron. Soc., 477, 491

Zennaro M., Angulo R. E., Aricò G., Contreras S., Pellejero-Ibáñez M., 2019, *How to add massive neutrinos to your ΛCDM simulation - extending cosmology rescaling algorithms*, Mon. Not. Roy. Astron. Soc., 489, 5938

Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, Journal of Open Source Software, 4, 1298

Zumalacárregui M., Bellini E., Sawicki I., Lesgourgues J., Ferreira P. G., 2017, *hi_class: Horndeski in the Cosmic Linear Anisotropy Solving System*, JCAP, 08, 019

Zwicky F., 1937, *On the Masses of Nebulae and of Clusters of Nebulae*, Astrophys. J., 86, 217

Zyla P., et al., 2020, *Review of Particle Physics*, PTEP, 2020, 083C01

de Haan T., et al., 2016, *Cosmological Constraints from Galaxy Clusters in the 2500 square-degree SPT-SZ Survey*, Astrophys. J., 832, 95

de Mattia A., et al., 2020, *The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: measurement of the BAO and growth rate of structure of the emission line galaxy sample from the anisotropic power spectrum between redshift 0.6 and 1.1*

de Rham C., Melville S., 2018, *Gravitational Rainbows: LIGO and Dark Energy at its Cutoff*, Phys. Rev. Lett., 121, 221101