

UNIVERSITE PARIS-SUD XI

Ecole doctorale « Sciences et Technologies de l'Information des Télécommunications et des Systèmes »

Discipline : *Physique*

THESE DE DOCTORAT

soutenue le 13 décembre 2012

par

ANTOINE LABATIE

Sujet :

**Etude statistique des corrélations dans la  
distribution des galaxies - application aux  
Oscillations Baryoniques Acoustiques**

Composition du jury :

M.	J. G. BARTLETT	Professeur à l'Université Paris 7	Rapporteur
M.	B. D. WANDEL	Professeur à l'Université Paris 7	Rapporteur
Mme.	N. AGHANIM	Directrice de recherche au CNRS	Examineur
M.	J.-C. HAMILTON	Enseignant-chercheur à l'Université Paris 7	Examineur
M.	O. LE FÈVRE	Directeur de recherche au CNRS	Examineur
Mme.	A. RASSAT	Post-doctorant à l'EPFL Lausanne	Examineur
M.	J.-L. STARCK	Directeur de recherche au CEA	Directeur de thèse
M.	M. LACHIÈZE-REY	Directeur de recherche au CNRS	Co-directeur de thèse



L'oeil qui scrute, qui dissèque, qui analyse,  
doit être réconcilié avec l'oeil qui vénère et qui  
contemple. — *Hubert Reeves.*



# Résumé

Les Oscillations Baryoniques Acoustiques (BAOs) correspondent au phénomène d'ondes acoustiques dans le plasma baryon-photon avant la recombinaison. Ces oscillations impriment une échelle standard, correspondant à l'horizon sonore, qui peut être observée dans les grandes structures de l'Univers. De par cette propriété d'étalon standard, les BAOs permettent d'étudier la relation distance-*redshift* dans les catalogues de galaxies et fournissent un outil très prometteur pour étudier les propriétés de l'énergie sombre.

Les BAOs peuvent être étudiés à partir des statistiques du second ordre (la fonction de corrélation ou le spectre de puissance) dans la distribution des galaxies. Dans cette thèse, on se restreint au cas de la fonction de corrélation. Les BAOs apparaissent dans la fonction de corrélation comme un faible pic à l'échelle de l'horizon sonore. Il y a deux applications principales des BAOs : la détection des BAOs et la contraintes des paramètres cosmologiques grâce à la propriété d'étalon standard. La détection du pic BAO à une échelle proche de l'échelle attendue permet de confirmer le modèle cosmologique actuel. Quant à la contrainte des paramètres cosmologiques, permettant d'étudier l'énergie sombre, il s'agit d'un objectif majeur de la cosmologie moderne.

Dans cette thèse, on s'intéresse à différents problèmes statistiques concernant l'étude de la fonction de corrélation dans la distribution des galaxies, avec un accent particulier sur l'étude des BAOs. Dans la première partie, on fait une étude à la fois théorique et pratique du biais dû à la contrainte intégrale dans les estimateurs de la fonction de corrélation. On montre que ce biais est très faible pour les relevés de galaxies actuels. Dans la seconde partie, on s'intéresse à la détection des BAOs. On montre les limitations de la méthode de détection classique, et on propose une nouvelle méthode plus rigoureuse. Notre méthode permet notamment de prendre en compte la dépendance de la matrice de la covariance de l'estimateur en fonction du modèle. Enfin dans la troisième partie, on étudie à nouveau la question la dépendance de la matrice de la covariance, cette fois pour la contrainte des paramètres cosmologiques. On estime une matrice de covariance dépendant du modèle et l'on compare nos contraintes avec les contraintes obtenues en faisant l'approximation habituelle d'une matrice de covariance constante. On montre que l'effet de cette approximation est relativement faible et diminue lorsque les relevés de galaxies grandissent.



# Abstract

Baryon Acoustic Oscillations (BAOs) correspond to the acoustic phenomenon in the baryon-photon plasma before recombination. BAOs imprint a particular scale, corresponding to the sound horizon, that can be observed in large-scale structures of the Universe. Using this standard ruler property, BAOs can be used to probe the distance-redshift relation in galaxy catalogues, thus providing a very promising tool to study dark energy properties.

BAOs can be studied from the second order statistics (the correlation function or the power spectrum) in the distribution of galaxies. In this thesis we restrict to the case of the correlation function. BAOs appear in the correlation function as a small localized bump at the scale of the sound horizon in comoving coordinates. There are two major applications of BAO study : BAO detection and cosmological parameter constraints using the standard ruler property. The detection of BAOs at the expected scale enables to confirm the current cosmological model. As for cosmological parameter constraints, enabling the study of dark energy, it is a major goal of modern cosmology.

In this thesis we tackle different statistical problems concerning the correlation function analysis in the galaxy distribution, with a focus on the study of BAOs. In the first part, we make both a theoretical and practical study of the bias due to the integral constraints in correlation function estimators. We show that this bias is very small for current galaxy surveys. In the second part we study the BAO detection. We show the limitations of the classical detection method and propose a new method, which is more rigorous. In particular our method enables to take into account the model-dependence of the covariance matrix of the estimators. In the third part, we focus again on the model-dependence of the covariance matrix, but this time for parameter constraints. We estimate a model-dependent covariance matrix and compare our constraints with constraints obtained by making the usual approximation of a constant covariance matrix. We show that the effect of this approximation is quite small for current galaxy surveys and decreases with the survey size.





# Remerciements

Je veux remercier tous ceux qui m'ont m'aidé professionnellement et personnellement durant cette thèse.

Merci à mon directeur de thèse, Jean-Luc Starck, pour avoir toujours été disponible et présent. Merci également de m'avoir laissé la liberté d'explorer des voies qui n'étaient pas forcément très bien tracées. Merci à mon autre directeur de thèse de l'APC, Marc Lachièze-Rey, pour toutes les discussions intéressantes que l'on a pu avoir.

Je suis redevable à tous mes collègues du Service d'Astrophysique pour les bons moments passés ensemble, les thésards, post docs et permanents : Florent, Jérôme, Sandrine, Paniez, Georgios, Arnaud, Jérémy, Lila, Simon, Daniel, François, Aurèle, Tatyana, Jessica, Louis, Nicolas (et tous les autres!). Un grand merci à mon ami Pierre qui était toujours là, et qui m'a aidé à terminer ma thèse.

Merci également aux collègues de l'APC, qui ont toujours été très sympathiques avec moi, notamment Mariana, Joseph, Loïc, Alexandre, Marianne, Guillaume, Sébastien.

Enfin puisque c'est l'occasion de le dire, un grand merci à ma famille et à mes amis.



# Table des figures

1.1	Représentation bidimensionnelle des différents types de géométries possibles pour l'univers. Image de la NASA. . . . .	5
1.2	Les différentes évolutions du paramètre d'échelle $a(t)$ selon la quantité de matière $\Omega_m$ et d'énergie noire $\Omega_\Lambda$ (avec $w = -1$ ) dans l'univers, avec pour le temps présent $a(t_0) = 1$ . La courbe marron en tirets correspond à un univers fermé $\Omega_m > 1$ , $\Omega_\Lambda = 0$ qui finira par s'effondrer sur lui même, la courbe noire pointillée le cas d'un univers ouvert $\Omega_m < 1$ , $\Omega_\Lambda = 0$ et la courbe bleue en tirets-points le cas d'un univers plat $\Omega_m = 1$ , $\Omega_\Lambda = 0$ (modèle d'Einstein de Sitter). La courbe continue rouge décrit l'évolution de $a(t)$ pour un modèle standard actuel $\Omega_m = 0.3$ , $\Omega_\Lambda = 0.7$ . Figure extraite de <a href="#">Sousbie (2007)</a> . . .	9
1.3	Régions de confiance à 68.3 %, 95.4 % et 99.7% pour $\Omega_\Lambda$ et $\Omega_m$ obtenues avec le CMB, les BAOs et les données d'Union SN, ainsi que leur combinaison (en supposant $w = -1$ ). On voit notamment que les observations actuelles semblent indiquer un univers plat. Ces contraintes viennent des données disponibles en 2008 et sont en constante amélioration. Figure extraite de <a href="#">Kowalski et al. (2008)</a> . . . . .	14
1.4	Carte du ciel en coordonnées galactiques du CMB mesuré par le satellite WMAP après 7 ans d'observation ( <a href="#">Jarosik et al. 2011</a> ). Cette carte dénotée <i>Internal Linear Combination Map</i> est une combinaison linéaire des différentes bandes d'observation afin de réduire la contamination galactique. Figure fournie par l'équipe scientifique de WMAP, <a href="http://lambda.gsfc.nasa.gov/product/map/current/">http://lambda.gsfc.nasa.gov/product/map/current/</a> . . . . .	14
1.5	Spectre de puissance du CMB mesuré par WMAP après 7 ans d'observation. La courbe rouge donne le modèle $\Lambda$ CDM qui maximise la vraisemblance, les barres d'erreur donnent le bruit instrumental et la région grise représente la variance cosmique. Figure extraite de <a href="#">Larson et al. (2011)</a> . . . . .	15
1.6	Diagramme de Hubble pour la compilation Union 2.1. La courbe solide représente le modèle $\Lambda$ CDM plat qui maximise la vraisemblance. Les barres d'erreur comprennent l'erreur intrinsèque et l'erreur de mesure. Figure extraite de <a href="#">Suzuki et al. (2012)</a> . . . .	16
1.7	Fonction de corrélation de l'échantillon LRG du SDSS DR3 avec barres d'erreur (noir). Les courbes en couleur montrent différents modèles, tous avec $n_s = 0.98$ pour $\Omega_m h^2 = 0.12$ (vert), $\Omega_m h^2 = 0.13$ (rouge), $\Omega_m h^2 = 0.14$ (bleu) avec $\Omega_b h^2 = 0.024$ , et pour $\Omega_m h^2 = 0.105$ avec zéro-baryon (rose). On voit que la détection du pic est statistiquement significative et que la position du pic dépend très faiblement de la valeur de $\Omega_m h^2$ (propriété d'étalon standard du pic BAO). Figure extraite de <a href="#">Eisenstein et al. (2005)</a> . . . . .	17
1.8	Dynamique d'évolution pour $\Omega_m$ et $\Omega_\Lambda$ . Les flèches indiquent la direction d'évolution des paramètres dans un univers en expansion. La droite $\Omega_m + \Omega_\Lambda = 1$ n'est pas stable et le seul point attracteur correspond à $\Omega_m = 0$ , $\Omega_\Lambda = 1$ ce qui correspond à l'univers de Sitter. Figure extraite de <a href="#">Carroll (2001)</a> . . . . .	18
1.9	Croissance de fluctuations à différentes échelles : une échelle entrant sous l'horizon pendant l'époque où la radiation domine (bleu) et une échelle entrant sous l'horizon pendant l'époque où la matière domine. Les fluctuations croissent comme le facteur de croissance $D$ , hormis celles qui entrent sous l'horizon pendant l'époque où la radiation domine (à ce moment elles cessent de croître jusqu'en fin de domination de la radiation). Figure réalisée par Andrew Jaffe. . . . .	22

- 1.10 Spectre de puissance linéaire  $P_L(k)$  pour un modèle cosmologique plat avec  $\Omega_m = 0.237$ ,  $\Omega_b = 0.041$  et  $h = 0.735$ , obtenus avec la formule de Eisenstein and Hu (1998) (tirets-pointillés verts) et avec le logiciel CMBFAST, comparés à spectre linéaire de référence obtenu avec CAMB. Figure extraite de Sánchez et al. (2008). . . . . 23
- 1.11 Fonction de transfert  $T(k)$  (en haut) et spectre de puissance linéaire  $P_L(k)$  (en bas) à *redshift*  $z = 0$  obtenu avec CAMB et les paramètres cosmologiques de Komatsu et al. (2011). La ligne en tirets verts correspond au spectre non linéaire obtenu avec l'approximation HALOFIT (Smith et al. (2003), voir section 1.3.4). On peut voir les effets de la fonction de transfert aux différentes échelles. (1) : les grandes échelles (faibles  $k$ ) sont restées en dehors de l'horizon pendant la période où la radiation domine,  $T(k) = 1$ , et le spectre de puissance est égal spectre primordial  $P_L(k) \propto k^{n_s}$  à un facteur de normalisation près. (2) : A l'échelle de l'horizon à l'égalité matière-radiation, on observe un changement de comportement,  $T(k) < 1$ , et le spectre de puissance atteint son amplitude maximum. (3) : Aux petites échelles (grands  $k$ ), les fluctuations sont fortement atténuées après leur passage sous l'horizon,  $T(k) \propto k^{-4}$ , et l'on observe aussi l'effet des BAOs comme une série d'oscillations à la fois dans  $T(k)$  et dans  $P_L(k)$ . Figure réalisée par Pablo Arnalte-Mur. . . . . 24
- 1.12 Comparaison de la fonction de corrélation obtenue sur des simulations  $N$ -corps de volume total  $120h^{-3}\text{Gpc}^3$  (plus de 4 fois le volume de Hubble), pour un modèle  $\Lambda$ CDM plat avec  $\Omega_m = 0.237$ ,  $\Omega_b = 0.041$ ,  $n_s = 0.954$ ,  $\sigma_8 = 0.77$  et  $h = 0.735$ , en espace réel à *redshift*  $z = 0$  (cercles noirs) avec : (i) la fonction de corrélation linéaire  $\xi_L(r)$  (trait rouge continu), (ii) la fonction de corrélation non linéaire  $\xi_{NL}(r)$  après la correction HALOFIT de  $\xi_L(r)$  (tirets bleus), (iii) la fonction de corrélation *dewiggled*  $\xi_{dw}(r)$  donnée par la formule (1.106) (tirets oranges), (iv) la fonction de corrélation *dewiggled*  $\xi_{dw,NL}(r)$  après la correction HALOFIT de la formule (1.107) (tirets-pointillés bleus). Les barres d'erreur donnent l'écart-type de l'estimateur sur des volumes de  $2.41h^{-1}\text{Gpc}^3$ . Figure extraite de Sánchez et al. (2008). . . . . 26
- 1.13 Evolution de la densité de masse  $r^2\rho(r)$  autour d'une surdensité adiabatique initiale. Chaque figure montre le profil de masse à un *redshift* différent, donné dans chaque coin en haut à droite, pour 4 espèces différentes : matière noire (noir), baryons (bleu), photons (rouge), neutrinos (vert). Les densités  $\rho(r)$  sont calculées par transformée de Fourier de la fonction de transfert, elle même calculée avec le programme CMBFAST. Figure extraite de Eisenstein et al. (2007b). . . . . 28
- 2.1 Les différentes figures montrent la même coupe d'épaisseur  $15h^{-1}\text{Mpc}$  en coordonnées comobiles de la simulation  $N$ -corps Millennium (Springel et al. 2005), à différents *redshift*  $z = 18.3$  (en haut à gauche),  $z = 5.7$  (en haut à droite),  $z = 1.4$  (en bas à gauche) et  $z = 0$  (en bas à droite). On peut remarquer l'évolution non linéaire qui détruit la Gaussianité du champ et permet la formation de la toile cosmique avec son réseau filaments entourant des grands vides cosmiques. Figures fournies par l'équipe scientifique de la simulation Millennium, <http://www.mpa-garching.mpg.de/millennium/>. . . . . 34
- 2.2 La statistique  $v_3(\nu)$  obtenue sur un catalogue de galaxies du relevé 2dFGRS. Les différentes courbes montrent  $v_3(\nu)$  après un lissage du champ par un noyau Gaussien de taille  $s = 1h^{-1}\text{Mpc}$  (violet),  $s = 2h^{-1}\text{Mpc}$  (bleu),  $s = 4h^{-1}\text{Mpc}$  (vert),  $s = 6h^{-1}\text{Mpc}$  (orange) et  $s = 8h^{-1}\text{Mpc}$  (rouge) comparées à l'espérance  $\mathbb{E}[v_3(\nu)]$  pour un champ Gaussien de la formule (2.17) (noir). Chaque courbe a été renormalisée pour rendre la comparaison possible. On voit que les courbes  $v_3(\nu)$  se rapprochent du cas Gaussien quand on augmente la taille du lissage, ce qui supporte l'idée que le champ de galaxies devient Gaussien à des échelles suffisamment grandes. . . . . 35
- 2.3 Fonction de corrélation bidimensionnelle  $\xi(\pi, \sigma)$  pour le relevé de galaxies 2dFGRS, en fonction de la séparation sur la ligne de visée  $\pi$  et dans la direction transverse  $\sigma$ , montrant l'effet des distorsions de *redshift*. La forme de cigare à  $\pi \approx 0$  est dû à l'effet *finger of god* qui lisse la fonction de corrélation, alors que l'effet cigare aux grands  $\pi$  (i.e. l'amplification de la corrélation) est dû à l'effet de Kaiser. Figure extraite de Peacock et al. (2001). . . . . 38

2.4	Taille d'un objet (dans la ligne de visée et dans la direction transverse) à <i>redshift</i> $z$ , d'extension $\Delta z$ en <i>redshift</i> et d'extension angulaire $\Delta\theta$ sur le ciel. La vraie taille de l'objet est donnée dans la ligne de visée par $c\frac{\Delta z}{H(z)}$ et dans la direction transverse par $D_M(z)\Delta\theta$ . Cependant si l'on utilise une cosmologie fiducielle incorrecte (comme c'est forcément le cas puisqu'on ne connaît pas exactement la vraie cosmologie) pour convertir les <i>redshift</i> en distances, les paramètres $H^{fid}(z)$ et $D_M^{fid}(z)$ sont différents et l'objet est déformé. En particulier, un objet sphérique peut devenir ellipsoïdal en coordonnées fiducielles. . . . .	43
2.5	Mesure de la dimension de Minkowski pour une ligne (gauche) et pour une surface (droite). Le nombre de cubes intersectant l'objet est $N(\delta) \propto \frac{1}{\delta}$ pour une ligne et $N(\delta) \propto \frac{1}{\delta^2}$ pour une surface. Figure extraite de Feder (1988). . . . .	45
2.6	Itérations successives pour la construction de la courbe de von Koch $F$ . Figure extraite de Falconer (1990). . . . .	46
2.7	Les dimensions de Rényi $D_q(r)$ en fonction de $q$ pour les échelles $r = 60 h^{-1}\text{Mpc}$ (gauche) et $r = 70 h^{-1}\text{Mpc}$ (droite). La figure montre les résultats sur l'échantillon LRG SDSS DR6 (points rouges), sur des catalogues aléatoires sans corrélation (ligne solide verte), sur des simulations $N$ -corps de matière noire (tirets bleus) et sur un catalogue de galaxies de la simulation Millennium (pointillés roses). Les barres d'erreur à $1\sigma$ de l'échantillon sont estimées à partir de 5 réalisations $N$ -corps. Figure extraite de Sarkar et al. (2009). . . . .	48
2.8	La dimension de corrélation $D_2(r)$ mesurée dans quatre régions de <i>redshift</i> pour le catalogue WiggleZ avec les barres d'erreur (noir) et pour un modèle $\Lambda\text{CDM}+\text{WMAP}$ avec un biais $b^2$ ajusté aux données (bleu). Interpolation polynomiale du 5 <sup>ème</sup> degré de la courbe $D_2(r)$ obtenue sur les données (rouge). L'échelle d'homogénéité $R_H$ est définie comme l'échelle où $D_2(r) = 2.97$ (i.e. homogénéité à 1%), à la fois pour l'interpolation polynomiale et pour le modèle $\Lambda\text{CDM}$ . L'intervalle de confiance pour $R_H$ est calculé à partir de réalisations $\Lambda\text{CDM}$ log-normales. Figure extraite de Scrimgeour et al. (2012). . . . .	49
3.1	Coupes de différents relevés de galaxies : le CfA2 (droite), le 2dFGRS (haut) et le SDSS (bas). La figure donne une idée de la profondeur des différents relevés. On peut voir les caractéristiques générales de la distribution de galaxies qui forme une toile cosmique. Figure extraite de Jones et al. (2004). . . . .	52
3.2	Mesures actuelles de l'échelle de dilatation $D_V(z)$ en utilisant la propriété d'étalon standard des BAOs avec les relevés 6dFGRS, SDSS et WiggleZ. Les résultats sont comparés à un modèle cosmologique $\Lambda\text{CDM}$ plat avec $\Omega_m = 0.29$ et un paramètre de Hubble $h = 0.69$ . Figure extraite de Blake et al. (2011b). . . . .	53
3.3	Couverture angulaire des catalogues Main VL et LRG VL en projection d'Aitoff. La ligne bleue représente le plan galactique qui sépare l'hémisphère galactique nord de l'hémisphère galactique sud. . . . .	55
3.4	Fonctions de corrélation estimées du catalogue Main VL (bleu) et du LRG VL (rouge) pour une cosmologie fiducielle $\Lambda\text{CDM}$ plate avec $\Omega_m = 0.27$ . Les barres d'erreur sont estimées à partir de 2,000 simulations log-normales $\Lambda\text{CDM}$ avec les paramètres $\Omega_m = 0.27$ , $h = 0.7$ , $\Omega_b h^2 = 2.205^{-2}$ , $n_s = 1$ , $\sigma_8 = 0.8$ et respectivement $b = 1.65$ pour le Main VL et $b = 2.5$ pour le LRG VL (ces simulations sont présentées plus en détails dans le chapitre 4). . . . .	56
3.5	Densité moyenne estimée $\bar{n}(z)$ du catalogue DR7-Full en fonction du <i>redshift</i> pour le choix d'une cosmologie fiducielle $\Lambda\text{CDM}$ plate avec $\Omega_m = 0.25$ . . . . .	57
3.6	Couverture angulaire du catalogue DR7-Full en projection d'Aitoff. La ligne bleue représente le plan galactique qui sépare l'hémisphère galactique nord de l'hémisphère galactique sud. . . . .	58
3.7	Fonction de corrélation estimée du catalogue DR7-Full pour une cosmologie fiducielle $\Lambda\text{CDM}$ plate avec $\Omega_m = 0.25$ . Les barres d'erreur sont obtenues à partir de 2,000 simulations log-normales $\Lambda\text{CDM}$ avec les paramètres $\Omega_m = 0.25$ , $h = 0.72$ , $\Omega_b h^2 = 2.27 \times 10^{-2}$ , $n_s = 0.966$ , $\sigma_8 = 0.81$ et $b = 2.5$ (ces simulations sont présentées plus en détails dans le chapitre 5). Le pic BAO est un peu plus large que ce qui est attendu, ce qui est expliqué par la variance d'échantillon dans Kazin et al. (2010). . . . .	60

3.8	Densité moyenne estimée $\bar{n}(z)$ du catalogue DR9 CMASS (bleu) qui s'étend entre les <i>redshifts</i> $0.43 < z < 0.7$ , pour une cosmologie fiducielle $\Lambda$ CDM plate avec $\Omega_m = 0.274$ . On montre en comparaison la densité du catalogue DR7-Full (rouge) qui s'étend entre les <i>redshifts</i> $0.16 < z < 0.47$ . Figure extraite de <a href="#">Anderson et al. (2012)</a> . . . . .	61
3.9	La couverture angulaire du catalogue DR9 CMASS en coordonnées galactiques, avec l'hémisphère nord (à gauche) et l'hémisphère sud (à droite). Les différents secteurs sont colorés en fonction de leur complétude. La zone colorée en gris montre la couverture angulaire attendue à la fin du relevé BOSS, correspondant à $10,269 \text{ deg}^2$ . Figure extraite de <a href="#">Sánchez et al. (2012)</a> . . . . .	62
3.10	A gauche : Fonction de corrélation estimée du catalogue DR9 CMASS pour une cosmologie fiducielle $\Lambda$ CDM plate avec $\Omega_m = 0.274$ . Les barres d'erreur sont obtenues à partir de $N = 600$ simulations avec la même fonction de sélection que le catalogue de données et les paramètres cosmologiques $\Omega_m = 0.274$ , $h = 0.7$ , $\Omega_b h^2 = 0.0224$ , $n_s = 0.95$ et $\sigma_8 = 0.8$ (voir <a href="#">Manera et al. (2012)</a> pour plus de détails sur ces simulations). La ligne en pointillés correspond au meilleur <i>fit</i> $\Lambda$ CDM, en combinant les données de la fonction de corrélation DR9 CMASS et du CMB avec WMAP7. A droite : Même chose en renormalisant par $(s/s_{\text{BAO}})^2$ , avec $s_{\text{BAO}} = 153.2 \text{ Mpc}$ pour mettre en évidence le pic BAO. Figure extraite de <a href="#">Sánchez et al. (2012)</a> . . . . .	63
3.11	Réalisation d'un processus de segments de Cox dans un cube de taille 50 par 50 par 50 (l'unité est ici arbitraire) avec une intensité volumique de segments $\lambda_l = 0.01$ , une longueur de segments $l = 10$ et un nombre de points par unité de longueur de segments $\lambda_l = 1.8$ . Le processus a une densité moyenne $\lambda = 0.18$ , ce qui correspond à un nombre moyen de points égal à 22,500 dans le volume considéré. . . . .	64
3.12	A gauche : coupe de l'échantillon Main VL entre les valeurs de latitude $-10^\circ < \lambda < 10^\circ$ dans les coordonnées du relevé SDSS. A droite : Même chose pour une réalisation log-normale du Main VL. On peut voir le bon accord visuel des deux distributions, bien que la structure de toile cosmique que l'on peut apercevoir sur les données ne soit pas reproduite dans les simulations. . . . .	65
3.13	Spectre de puissance du champ Gaussien sans <i>zero-padding</i> (rouge) et avec <i>zero-padding</i> (bleu), pour un spectre en entrée $P_L(k)$ à <i>redshift</i> $z = 0$ obtenu avec la formule de <a href="#">Eisenstein and Hu (1998)</a> pour les paramètres cosmologiques de <a href="#">Komatsu et al. (2011)</a> . On observe des oscillations artificielles dues aux conditions aux limites périodiques de la transformée de Fourier en l'absence de <i>zero-padding</i> . Ces oscillations n'apparaissent plus avec le <i>zero-padding</i> . . . . .	67
3.14	Spectre de puissance en entrée (noir) et spectre obtenu en inversant toutes les étapes (formule (3.13)) avec <i>zero-padding</i> (bleu), pour un spectre en entrée $P_L(k)$ à <i>redshift</i> $z = 0$ obtenu avec la formule de <a href="#">Eisenstein and Hu (1998)</a> pour les paramètres cosmologiques de <a href="#">Komatsu et al. (2011)</a> . Le spectre obtenu après inversion est extrêmement proche du spectre en entrée. . . . .	67
3.15	Distribution lissée des halos de matière noire (pondérés par leur masse) sur une coupe d'épaisseur $40 h^{-1} \text{ Mpc}$ en coordonnées comobiles pour différentes réalisations LasDamas avec des tailles de cube différentes ( $2400 h^{-1} \text{ Mpc}$ , $1000 h^{-1} \text{ Mpc}$ , $640 h^{-1} \text{ Mpc}$ et $420 h^{-1} \text{ Mpc}$ ). . . . .	70
4.1	Fonctions de corrélation estimées pour les échantillons Main VL (trait plein bleu) et LRG VL (trait plein rouge) pour une cosmologie fiducielle $\Lambda$ CDM plate avec $\Omega_m = 0.27$ , et fonctions de corrélation théoriques $b^2 \xi_{\Lambda\text{CDM}}$ avec un biais ajusté pour fitter les données. On trouve des valeurs du biais égales à $b = 1.65$ pour le Main VL et $b = 2.5$ pour le LRG VL. Les barres d'erreur sont estimées à partir de 2,000 simulations log-normales $\Lambda$ CDM avec les paramètres $\Omega_m = 0.27$ , $h = 0.7$ , $\Omega_b h^2 = 2.205^{-2}$ , $n_s = 1$ , $\sigma_8 = 0.8$ et respectivement $b = 1.65$ pour le Main VL et $b = 2.5$ pour le LRG VL. . . .	78
4.2	A gauche : Moyenne $\bar{\xi}$ et écart-type $s(\hat{\xi})$ des différents estimateurs avec $N = 2000$ réalisations de segments de Cox pour $a = 20$ . On trace la fonction de corrélation analytique (noir), les estimateurs Peebles-Hausser (violet), Davis-Peebles (bleu clair), Hamilton (vert) et Landy-Szalay (rouge). A droite : Même chose avec $N = 512$ réalisations de segments de Cox pour $a = 50$ . . . . .	80

4.3	A gauche : Moyenne $\bar{\xi}$ des différents estimateurs et incertitude sur cette moyenne $\frac{s(\hat{\xi})}{\sqrt{N}}$ avec $N = 2000$ réalisations de segments de Cox pour $a = 20$ . A droite : Même chose avec $N = 512$ réalisations de segments de Cox pour $a = 50$ . . . . .	80
4.4	A gauche : Moyenne pondérée des différents estimateurs $f(r_i)\bar{\xi}$ et incertitude $f(r_i)\frac{s(\hat{\xi})}{\sqrt{N}}$ sur cette moyenne avec $N = 2000$ réalisations de segments de Cox pour $a = 20$ . On trace en également la fonction $f(r)$ (ligne pointillée noire). A droite : Même chose avec $N = 512$ réalisations de segments de Cox pour $a = 50$ . . . . .	81
4.5	A gauche : Moyenne $\bar{\xi}$ et écart-type $s(\hat{\xi})$ des différents estimateurs avec $N = 2000$ réalisations Main VL, Peebles-Hauser (violet), Davis-Peebles (bleu clair), Hamilton (vert), Landy-Szalay (rouge) et modèle $\Lambda$ CDM à $z = 0.1$ avec $b = 1.5$ (noir). A droite : Même chose avec $N = 2000$ réalisations LRG VL, mais avec un modèle $\Lambda$ CDM à $z = 0.3$ avec $b = 2.5$ . . . . .	83
4.6	A gauche : Moyenne $\bar{\xi}$ des différents estimateurs et incertitude sur cette moyenne $\frac{s(\hat{\xi})}{\sqrt{N}}$ avec $N = 2000$ réalisations Main VL, Peebles-Hauser (violet), Davis-Peebles (bleu clair), Hamilton (vert), Landy-Szalay (rouge) et modèle $\Lambda$ CDM à $z = 0.1$ avec $b = 1.5$ (noir). A droite : Même chose avec $N = 2000$ réalisations LRG VL, mais avec un modèle $\Lambda$ CDM à $z = 0.3$ avec $b = 2.5$ . . . . .	84
5.1	Différentes ondelettes mère utilisées pour détecter les BAOs. On montre l'ondelette 'chapeau mexicain' avec les paramètres $R = 113.6 h^{-1}\text{Mpc}$ , $s = 20 h^{-1}\text{Mpc}$ (en bleu) de l'étude <a href="#">Tian et al. (2011)</a> et la BAOlet avec les paramètres $R = 116 h^{-1}\text{Mpc}$ , $s = 36 h^{-1}\text{Mpc}$ (en rouge) de l'étude <a href="#">Arnalte-Mur et al. (2012)</a> . Ces paramètres correspondent aux réponses maximum $Z_w^{max}$ dans les études respectives <a href="#">Tian et al. (2011)</a> et <a href="#">Arnalte-Mur et al. (2012)</a> . . . . .	89
5.2	Moyennes des estimateurs $\bar{\xi}_{\omega_m}$ en tirets comparées aux fonctions de corrélation attendues $\xi_{\omega_m}$ en traits pleins pour $\alpha = 1$ , $B = 2.5^2$ et pour $\omega_m = 0.08$ (violet), $0.105$ (bleu clair), $0.13$ (vert), $0.155$ (jaune), $0.18$ (rouge). . . . .	100
5.3	pdf estimée de $\chi_{\theta_0}^2$ (noir) en utilisant l'histogramme des 2,000 simulations log-normales et pdf d'une loi $\chi_{18}^2$ (rouge). Les barres d'erreur donnent l'erreur de Poisson dans l'estimation de la pdf due au nombre fini de simulations. . . . .	101
5.4	pdf estimée de $Z_w$ pour le chapeau mexicain avec les paramètres $R = 113.6 h^{-1}\text{Mpc}$ , $s = 20 h^{-1}\text{Mpc}$ en utilisant l'histogramme sur 2,000 simulations log-normales (noir) et pdf d'une loi normale centrée sur $\langle w, \xi_{\theta_0} \rangle / \sigma_w(R, s)$ et d'écart-type 1. Les barres d'erreur donnent l'erreur de Poisson sur l'histogramme due au nombre fini de simulations. . . .	102
5.5	pdf estimée de $Z_w$ pour le filtre BAOlet avec les paramètres $R = 116 h^{-1}\text{Mpc}$ , $s = 36 h^{-1}\text{Mpc}$ en utilisant l'histogramme sur 2,000 simulations log-normales (noir) et pdf d'une loi normale centrée sur $\langle w, \xi_{\theta_0} \rangle / \sigma_w(R, s)$ et d'écart-type 1. Les barres d'erreur donnent l'erreur de Poisson sur l'histogramme due au nombre fini de simulations. . . .	102
5.6	A gauche : Significativité de la détection des BAOs obtenue avec la méthode classique $\sqrt{\Delta\chi^2}$ (avec la convention $\sqrt{\Delta\chi^2} = 0$ pour $\Delta\chi^2 < 0$ ) en fonction de la significativité rigoureuse $\sigma_{\text{reel}}(\Delta\chi^2)$ , avec une hypothèse $\mathcal{H}_0$ <i>no wiggles</i> . La courbe est obtenue pour 50,000 réalisations qui prennent différentes valeurs de $\Delta\chi^2$ . On voit que pour $\Delta\chi^2 < 0$ et pour des valeurs faibles de $\Delta\chi^2$ , la méthode classique sous-estime la significativité. En revanche aux plus grandes valeurs de $\Delta\chi^2$ , la significativité devient surestimée. Dans l'ensemble, on peut quand même dire que la significativité $\sqrt{\Delta\chi^2}\sigma$ donnée par la méthode classique est une approximation très correcte de la vraie significativité. A droite : Même chose pour une hypothèse $\mathcal{H}_0$ zéro baryon. . . . .	108
5.7	A gauche : Significativité de la détection des BAOs obtenue avec la méthode classique $\sqrt{\Delta\chi^2}$ (avec la convention $\sqrt{\Delta\chi^2} = 0$ pour $\Delta\chi^2 < 0$ ) en fonction de la significativité rigoureuse $\sigma_{\text{reel}}(\Delta\chi^2)$ , avec une hypothèse $\mathcal{H}_0$ <i>no wiggles</i> . La courbe est obtenue pour 50,000 réalisations qui prennent différentes valeurs de $\Delta\chi^2$ . On voit que pour $\Delta\chi^2 < 0$ et pour des valeurs faibles de $\Delta\chi^2$ , la méthode classique sous-estime la significativité. En revanche aux plus grandes valeurs de $\Delta\chi^2$ , la significativité devient grossièrement surestimée, si bien que la méthode classique est grossièrement fautive dans le cas d'une matrice de covariance $C_\theta$ . A droite : Même chose pour une hypothèse $\mathcal{H}_0$ zéro baryon. . . . .	110

- 5.8 A gauche : Significativité de la détection des BAOs obtenue sous  $\mathcal{H}_1$  avec  $\Delta l$  en fonction de la significativité obtenue avec  $\Delta\chi^2$ , avec une hypothèse  $\mathcal{H}_0$  *no wiggles*. Pour des raisons de visibilité, on montre seulement 1,000 réalisations avec un modèle pris à chaque fois aléatoirement dans tous les modèles  $\mathcal{H}_1$ . On voit que pour la majorité des réalisations, la significativité obtenue est plus grande avec  $\Delta l$  qu'avec  $\Delta\chi^2$ . Ceci justifie notre recommandation de continuer à utiliser un rapport de vraisemblance généralisé dans le cas d'une matrice  $C_\theta$ . A droite : Même chose avec une hypothèse  $\mathcal{H}_0$  zéro baryon. 111
- 6.1 Dépendance de  $\rho_\theta$  par rapport à  $\omega_m$  dans le cas  $\alpha = 1$  et  $B = 2.5^2$ . On montre  $\rho_\theta$  pour les valeurs  $\omega_m = 0.08$  (en haut à gauche), 0.105 (en haut au milieu), 0.13 (en haut à droite), 0.155 (en bas à gauche), 0.18 (en bas au milieu). La corrélation entre les différents intervalles diminue fortement quand  $\omega_m$  augmente. On montre les  $n = 18$  intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  de  $20h^{-1}\text{Mpc}$  à  $200h^{-1}\text{Mpc}$ . . . . . 120
- 6.2 Dépendance de  $\sigma_\theta = (\sqrt{C_{\theta,ii}})$  par rapport à  $\omega_m$  dans le cas  $\alpha = 1$  et  $B = 2.5^2$ . On montre  $\sigma_\theta$  pour les valeurs  $\omega_m = 0.08$  (violet), 0.105 (bleu clair), 0.13 (vert), 0.155 (jaune), 0.18 (rouge). La variance diagonale diminue fortement quand  $\omega_m$  augmente. . . . . 120
- 6.3 Dépendance de  $\rho_\theta$  par rapport à  $\alpha$  dans le cas  $\omega_m = 0.13$  et  $B = 2.5^2$ . On montre  $\rho_\theta$  pour les valeurs  $\alpha = 0.8$  (en haut à gauche), 0.9 (en haut au milieu), 1.0 (en haut à droite), 1.1 (en bas à gauche), 1.2 (en bas au milieu). La corrélation entre les différents intervalles diminue quand  $\alpha$  augmente. On montre les  $n = 18$  intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  de  $20h^{-1}\text{Mpc}$  à  $200h^{-1}\text{Mpc}$ . . . . . 121
- 6.4 Dépendance de  $\sigma_\theta = (\sqrt{C_{\theta,ii}})$  par rapport à  $\alpha$  dans le cas  $\omega_m = 0.13$  et  $B = 2.5^2$ . On montre  $\sigma_\theta$  pour les valeurs  $\alpha = 0.8$  (violet), 0.9 (bleu clair), 1.0 (vert), 1.1 (jaune), 1.2 (rouge). La variance diagonale diminue quand  $\alpha$  augmente. . . . . 121
- 6.5 Postérieure  $p(\omega_m, D_V(0.3) | \hat{\xi})$  dans le cas d'une matrice de covariance constante  $C = C_{\theta_0}$  avec  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1.0, 2.5^2)$  (position de la croix rouge sur la figure) pour la mesure  $\hat{\xi}$  sur le SDSS DR7-Full. On montre les régions de confiance de  $1\sigma$  à  $5\sigma$  avec l'approximation que la postérieure est un Gaussienne bidimensionnelle. Les contours correspondent à des valeurs  $-2\ln(p) = -2\ln(p_{max}) + 2.29, 6.16, 11.81, 19.32, 28.74$  (voir la section 'Confidence Limits on Estimated Model Parameters' dans [Press et al. \(2007\)](#)). On obtient les contraintes unidimensionnelles  $\omega_m = 0.144 \pm 0.013$  (précision de 9.0%) et  $D_V(0.3) = 1104 \pm 83$  Mpc (précision de 7.5%). . . . . 123
- 6.6 Postérieure  $p(\omega_m, D_V(0.3) | \hat{\xi})$  dans le cas de notre matrice de covariance dépendant du modèle  $C_\theta$  pour la mesure  $\hat{\xi}$  sur le SDSS DR7-Full. On obtient les contraintes unidimensionnelles  $\omega_m = 0.140 \pm 0.011$  (précision de 7.9%) et  $D_V(0.3) = 1114 \pm 74$  Mpc (précision de 6.7%). Il y a seulement un petit déplacement de la position du maximum de la postérieure et les régions de confiance deviennent aussi un peu plus petites dans le cas de la matrice de covariance  $C_\theta$ . . . . . 123
- 6.7 pdf estimée de  $\delta\omega_m^{\max}$ ,  $\delta\sigma_{\omega_m}$ ,  $\delta\alpha^{\max}$  et  $\delta\sigma_\alpha$  en utilisant leur histogramme sur 2,000 réalisations de la formule (6.26). Les barres d'erreur donnent l'incertitude de Poisson sur l'estimation de la pdf due au nombre fini de réalisations. . . . . 125
- 6.8 Importance des différents termes intervenant dans la log-vraisemblance  $-2\ln\mathcal{L}_\theta(\hat{\xi})$  lorsque l'on fait varier un paramètre d'amplitude  $c$  dans la matrice de covariance  $\frac{1}{c}C_\theta$ , et dans le cas particulier  $n = 18$  et  $\chi_\theta^2 = 18$ . On voit que les 2 termes de la vraisemblance se compensent nettement, ce qui tend à atténuer les variations de la vraisemblance comparées aux variations de chacun des termes (i.e. comparées aux variations de la matrice de covariance). . . . . 127



# Table des matières

Résumé	V
Abstract	VII
Remerciements	IX
Table des figures	XI
Introduction	1
<b>1 Modèle cosmologique standard</b>	<b>3</b>
1.1 Modèle de Friedmann-Lemaître-Roberston-Walker (FLRW)	3
1.1.1 Le principe cosmologique	3
1.1.2 Métrique de Robertson-Walker et équations de Friedmann-Lemaître	4
1.1.3 Constituants cosmologiques et évolution de l'univers	8
1.1.4 Distances dans un univers FLRW	9
1.2 Modèle du Big Bang chaud	10
1.2.1 Matière noire et énergie sombre	11
1.2.2 Modèle $\Lambda$ CDM - Mesure des paramètres cosmologiques	12
1.3 La formation des structures	17
1.3.1 Inflation	17
1.3.2 Evolution linéaire du champ de matière	19
1.3.3 Fonction de transfert	21
1.3.4 Evolution non linéaire du champ de matière	23
1.4 Les Oscillations Baryoniques Acoustiques	27
<b>2 Analyse statistique de la corrélation des galaxies</b>	<b>31</b>
2.1 Champs aléatoires en cosmologie	31
2.1.1 Champ statistiquement homogène et isotrope	31
2.1.2 Fonction de corrélation à 2 points et spectre de puissance	32
2.1.3 Cas particulier du champ Gaussien	33
2.2 La corrélation des galaxies	35
2.2.1 Le champ de galaxies comme un processus de Cox	35
2.2.2 Biais des galaxies par rapport à la matière	36
2.2.3 Distorsions de <i>redshift</i>	37
2.2.4 Estimation de la fonction de corrélation	39
2.2.5 Modélisation de la fonction de corrélation	41
2.3 Multifractalité de la distribution des galaxies	44
2.3.1 Ensembles fractals	44
2.3.2 Ensembles multifractals	45
2.3.3 Fractalité à petite échelle et homogénéité à grande échelle de la distribution des galaxies	47

<b>3</b>	<b>Catalogues de galaxies et simulations</b>	<b>51</b>
3.1	Relevés de galaxies . . . . .	51
3.2	Les catalogues SDSS-II étudiés . . . . .	53
3.2.1	Les catalogues Main VL et LRG VL . . . . .	54
3.2.2	Le catalogue DR7-Full . . . . .	56
3.3	Les catalogues SDSS-III . . . . .	60
3.3.1	Le catalogue BOSS DR9 CMASS . . . . .	60
3.4	Simulations . . . . .	62
3.4.1	Simulations de segments de Cox . . . . .	62
3.4.2	Simulations log-normales . . . . .	63
3.4.3	Simulations log-normales pour différentes valeurs de $\alpha$ simultanément . . . . .	68
3.4.4	Simulations $N$ -corps . . . . .	69
<b>4</b>	<b>Incertitude sur l'estimation de la fonction de corrélation des galaxies</b>	<b>71</b>
4.1	Fractalité et estimation de la fonction de corrélation . . . . .	72
4.1.1	Problème conceptuel de la fonction de corrélation pour une fractale . . . . .	72
4.1.2	But du travail dans ce chapitre . . . . .	73
4.2	Estimateurs et incertitude sur la densité moyenne $\bar{n}$ . . . . .	73
4.2.1	Effet général de l'incertitude sur la densité moyenne $\bar{n}$ . . . . .	73
4.2.2	Le bias de la contrainte intégrale . . . . .	76
4.3	Simulations utilisées . . . . .	77
4.4	Effet de la contrainte intégrale sur des simulations . . . . .	79
4.4.1	Effet de la contrainte intégrale sur des processus de segments de Cox . . . . .	79
4.4.2	Prédiction de l'effet de la contrainte intégrale . . . . .	81
4.4.3	Effet de la contrainte intégrale sur des simulations Main VL et LRG VL . . . . .	82
4.5	Comparaison des différents estimateurs sur des simulations Main VL et LRG VL . . . . .	82
4.6	Conclusion . . . . .	84
<b>5</b>	<b>Détection des Oscillations Baryoniques Acoustiques</b>	<b>85</b>
5.1	Les différentes méthodes de détection des BAOs . . . . .	86
5.1.1	Principe de la détection des BAOs . . . . .	86
5.1.2	Historique de la détection des BAOs dans les grandes structures . . . . .	89
5.1.3	But du travail dans ce chapitre . . . . .	90
5.2	Méthode classique pour la détection des BAOs . . . . .	91
5.2.1	La statistique du $\chi^2$ . . . . .	91
5.2.2	La méthode du $\chi^2$ pour la détection des BAOs . . . . .	91
5.2.3	Limitations de la méthode du $\chi^2$ . . . . .	93
5.3	Nouvelle méthode pour la détection des BAOs . . . . .	94
5.4	Modélisation de $\hat{\xi}$ et simulations SDSS DR7-Full . . . . .	96
5.4.1	Modèles de fonctions de corrélation pour $\mathcal{H}_0$ et $\mathcal{H}_1$ . . . . .	96
5.4.2	Simulations log-normales SDSS DR7-Full utilisées . . . . .	98
5.4.3	Vérification de la modélisation statistique de $\hat{\xi}$ . . . . .	99
5.5	Détection des BAOs sur des simulations SDSS DR7-Full . . . . .	101
5.5.1	Détection des BAOs avec une matrice de covariance constante $C$ . . . . .	101
5.5.2	Détection des BAOs avec une matrice de covariance $C_\theta$ . . . . .	108
5.6	Conclusion . . . . .	110
<b>6</b>	<b>Effet de <math>C_\theta</math> pour contraindre les paramètres cosmologiques</b>	<b>113</b>
6.1	Utiliser les BAOs pour contraindre les paramètres cosmologiques . . . . .	114
6.1.1	Approche basée seulement sur la position du pic BAO . . . . .	114
6.1.2	Approche basée sur la forme globale de la fonction de corrélation . . . . .	115
6.1.3	But du travail dans ce chapitre . . . . .	116
6.2	Modélisation de $\hat{\xi}$ et simulations SDSS DR7-Full . . . . .	117
6.2.1	Modèles de fonctions de corrélation . . . . .	117
6.2.2	Simulations log-normales SDSS DR7-Full utilisées . . . . .	118
6.3	Dépendance de $C_\theta$ par rapport à $\omega_m$ , $\alpha$ et $B$ . . . . .	118
6.4	Effet de $C_\theta$ pour contraindre les paramètres cosmologiques . . . . .	119

---

6.4.1	Erreur de modélisation pour la mesure $\hat{\xi}$ du SDSS DR7-Full . . . . .	122
6.4.2	Quantification de l'erreur de modélisation sur des simulations du SDSS DR7-Full	124
6.4.3	Quantification de l'erreur de modélisation pour les relevés futurs . . . . .	125
6.5	Conclusion . . . . .	128
<b>Conclusion</b>		<b>129</b>
<b>Annexe</b>		<b>133</b>
<b>A Espérance et covariance d'un champ log-normal</b>		<b>133</b>
A.1	Espérance . . . . .	133
A.2	Covariance du champ log-normal . . . . .	133
<b>B Tests d'hypothèses</b>		<b>135</b>
B.1	Meilleur <i>fit</i> $\chi^2$ . . . . .	135
B.2	Différence de meilleurs <i>fits</i> $\chi^2$ pour des modèles imbriqués . . . . .	136
B.3	Optimalité du rapport de vraisemblance . . . . .	136
<b>C Combinaison linéaire optimale d'estimateurs</b>		<b>139</b>
<b>D Mes contributions</b>		<b>141</b>
<b>Bibliographie</b>		<b>143</b>

# Introduction

Cette thèse porte sur l'étude des corrélations dans la distribution des galaxies, et notamment sur l'étude des Oscillations Baryoniques Acoustiques (BAOs), qui impriment une échelle caractéristique de corrélation. Elle s'inscrit dans la problématique plus générale de la cosmologie.

La cosmologie est la science en charge de décrire l'univers de façon globale, et notamment sa formation et son évolution. Cette science a donc grandement évolué au fur et à mesure que l'homme a pris conscience de l'étendue de l'univers. Avant le 19<sup>ème</sup> siècle, l'univers connu était simplement réduit au système solaire. L'étude de la répartition des étoiles au sein de notre galaxie a ensuite été effectuée jusqu'au début du 20<sup>ème</sup> siècle. Dans les années 1920 les travaux de Hubble montrèrent la nature extragalactique des objets que l'on appelait alors les 'nébuleuses', puis montrèrent que l'univers était en réalité en expansion. Couplée aux travaux d'Einstein en 1915 sur la relativité générale, cette découverte permit de jeter les bases de la cosmologie moderne : l'univers est en expansion et suit les lois de la relativité générale.

La notion d'expansion de l'univers implique naturellement l'idée que l'univers était plus chaud et dense dans le passé, et plus précisément le modèle du *Big Bang* qui a émergé vers le milieu du 20<sup>ème</sup> siècle. La deuxième partie du 20<sup>ème</sup> siècle a permis la découverte de deux composantes nouvelles dans l'univers : la matière noire et l'énergie sombre. Il est maintenant admis que la matière ordinaire ne constitue que quelques pourcents de la masse totale dans l'univers. De plus il existe une énergie sombre encore inconnue, qui possède une pression négative, et qui cause l'accélération de l'expansion de l'univers.

Ce modèle standard permet de bien décrire les nombreuses observations cosmologiques, concernant la structure et la dynamique de l'univers. Il comporte toutefois un grand nombre de paramètres qu'il faut ajuster pour décrire les observations. En particulier la mesure des paramètres de l'énergie sombre pour décrire cette forme d'énergie mystérieuse constitue un objectif majeur de la cosmologie moderne.

Cette thèse a pour but d'apporter une contribution dans cette compréhension, à travers différentes études statistiques de la corrélation dans la distribution des galaxies. En particulier nous nous intéressons aux BAOs, qui impriment une échelle particulière dans la fonction de corrélation des galaxies. Il y a deux applications principales des BAOs. D'une part, la détection du pic BAO à une échelle proche de l'échelle attendue permet de confirmer le modèle cosmologique actuel. D'autre part, les BAOs constituent ce que l'on appelle un étalon standard permettant de mesurer la géométrie de l'univers, et donc de contraindre les paramètres cosmologiques. Leur étude fournit en particulier un moyen très prometteur d'étudier les propriétés de l'énergie sombre.

Du fait de la faiblesse du signal BAO dans la fonction de corrélation, il est important de bien tenir compte de tous les effets statistiques intervenant dans l'analyse, qui pourraient causer des effets systématiques indésirables. Cette thèse est principalement orientée vers l'étude de certains effets statistiques pouvant affecter l'analyse.

Ce manuscrit est organisé de la façon suivante :

- ★ Le premier chapitre décrit le modèle cosmologique standard ainsi que la physique du phénomène des Oscillations Baryoniques Acoustiques.
- ★ Dans le chapitre 2, on présente les différentes méthodes d'analyse statistique de la distribution des galaxies, avec un accent sur l'étude de la fonction de corrélation. On montrera comment cette étude permet d'étudier les modèles cosmologiques, e.g. pour contraindre les paramètres cosmologiques ou pour étudier l'homogénéisation de l'univers aux grandes échelles.

- ★ Dans le chapitre 3, on décrit les relevés de galaxies existants, et notamment les différents catalogues de galaxies utilisés dans cette thèse. On présente ensuite les différents types de simulations que l'on utilise pour l'étude statistique de ces différents catalogues.
- ★ On commence à présenter notre travail proprement dit dans le chapitre 4. Dans ce chapitre, on étudie l'incertitude sur les différents estimateurs  $\hat{\xi}$  de la fonction de corrélation pour les relevés de galaxies actuels. On montre comment ces estimateurs sont affectés par l'incertitude sur la densité moyenne de galaxies  $\bar{n}$ , qui cause notamment un biais connu sous le nom de 'contrainte intégrale'.
- ★ Dans le chapitre 5, on s'intéresse au problème particulier de la détection des BAOs. On montre les limitations de la méthode de détection classique, et l'on propose une nouvelle méthode plus rigoureuse. Notre méthode permet notamment de prendre en compte la dépendance de la matrice de la covariance  $C_\theta$  de l'estimateur  $\hat{\xi}$  en fonction du modèle  $\theta$ .
- ★ Enfin dans le dernier chapitre, on étudie à nouveau la question la dépendance de la matrice de la covariance  $C_\theta$ , cette fois pour la contrainte des paramètres cosmologiques. On compare les contraintes obtenues avec une matrice de covariance dépendant du modèle  $C_\theta$ , avec les contraintes obtenues en faisant l'approximation habituelle d'une matrice de covariance constante.
- ★ En annexe on présente différents résultats mathématiques dont on a eu besoin pour le travail présenté dans ce manuscrit. On liste également les publications et autres contributions qui reposent sur ce travail.

# Chapitre 1

## Modèle cosmologique standard

### Sommaire

---

<b>1.1</b>	<b>Modèle de Friedmann-Lemaître-Roberston-Walker (FLRW)</b>	<b>3</b>
1.1.1	Le principe cosmologique	3
1.1.2	Métrique de Robertson-Walker et équations de Friedmann-Lemaître	4
1.1.3	Constituants cosmologiques et évolution de l'univers	8
1.1.4	Distances dans un univers FLRW	9
<b>1.2</b>	<b>Modèle du Big Bang chaud</b>	<b>10</b>
1.2.1	Matière noire et énergie sombre	11
1.2.2	Modèle $\Lambda$ CDM - Mesure des paramètres cosmologiques	12
<b>1.3</b>	<b>La formation des structures</b>	<b>17</b>
1.3.1	Inflation	17
1.3.2	Evolution linéaire du champ de matière	19
1.3.3	Fonction de transfert	21
1.3.4	Evolution non linéaire du champ de matière	23
<b>1.4</b>	<b>Les Oscillations Baryoniques Acoustiques</b>	<b>27</b>

---

## 1.1 Modèle de Friedmann-Lemaître-Roberston-Walker (FLRW)

### 1.1.1 Le principe cosmologique

Le principe cosmologique est l'hypothèse selon laquelle l'univers est spatialement homogène, c'est à dire identique en tout point. Ceci est en contradiction avec l'existence de structures dans la distribution de matière jusqu'à de très grandes échelles. Cependant le principe peut être reformulé plus faiblement en considérant que l'univers est homogène lorsque l'on considère des échelles assez grandes (i.e. après un lissage du champ de matière à une échelle suffisamment grande).

Le principe cosmologique fut implicitement formulé par Einstein en 1917 lorsqu'il cherchait des solutions particulières aux équations de la relativité générale décrivant l'univers dans son ensemble. Comme le terme *principe* l'indique, il n'existait à l'époque aucune preuve observationnelle. Aujourd'hui de nombreuses observations sont venues corroborer l'homogénéité de l'univers à des échelles suffisamment grandes (voir [Davis \(1997\)](#) et la discussion dans la section 2.3) :

- L'isotropie des données 2D sur le ciel : catalogues de sources en rayons X, catalogues de sources radio, sursauts gamma et surtout la remarquable isotropie du CMB (avec des anisotropies ne dépassant pas  $\Delta T/T \approx 10^{-5}$ )
- L'homogénéité des grands relevés de galaxies récents dont le *2 degree Field Galaxy Redshift Survey* et le *Sloan Digital Sky Survey* à des échelles au delà de  $70h^{-1}\text{Mpc}$

Notons que l'isotropie de l'univers n'implique pas nécessairement l'homogénéité, car on peut imaginer un univers avec une symétrie sphérique qui ne soit pas homogène. Cependant cela impliquerait que l'on occupe une place privilégiée dans l'univers. Donc si l'on combine l'hypothèse d'isotropie avec le principe de Copernic, selon laquelle nous n'occupons pas une position spéciale dans l'univers, on retrouve l'hypothèse d'homogénéité de l'univers.

### 1.1.2 Métrique de Robertson-Walker et équations de Friedmann-Lemaître

La relativité générale élaborée par Albert Einstein entre 1907 et 1915 est la pierre de voûte de la cosmologie moderne. Aucun des nombreux tests expérimentaux effectués à ce jour n'a pu la mettre en défaut. Elle énonce que la gravitation ne doit pas être considérée comme une force mais comme la déformation de l'espace-temps, et que cette déformation dépend elle-même du contenu énergétique sous forme de masse ou d'énergie cinétique.

Les déformations de l'espace-temps sont exprimées à l'aide d'une métrique qui permet d'exprimer la distance entre deux événements. Dans un espace Euclidien à trois dimensions par exemple, deux points séparés par un vecteur  $\mathbf{dx} = (dx, dy, dz) = (dx^1, dx^2, dx^3)$  sont séparés par une distance  $ds$  :

$$ds^2 = dx^2 + dy^2 + dz^2 = \sum_{i,j} g_{i,j} dx^i dx^j \quad (1.1)$$

avec  $g_{ij} = \delta_{ij}$  et  $\delta_{ij}$  le symbole de Kronecker. Dans le contexte de la relativité, les distances que l'on considère incluent également le paramètre temps, et les événements sont repérés par trois coordonnées d'espace et une coordonnée de temps. Les distances s'expriment alors :

$$ds^2 = \sum_{\mu,\nu} g_{\mu\nu} dx^\mu dx^\nu \quad (1.2)$$

où l'on a adopté la convention habituelle d'utiliser des indices grecs au lieu d'indices latins pour les quadri-vecteurs (vecteurs dans l'espace-temps). Dans le cadre de la relativité restreinte, l'espace-temps est un espace de Minkowski où la distance propre entre deux événements est donnée par :

$$ds^2 = c^2 dt^2 - d\mathbf{x}^2 \quad (1.3)$$

avec  $c = 299,792,458 \text{ m/s}$  la vitesse de la lumière. Dans la suite on utilisera un système d'unités avec  $c = 1$  pour simplifier les notations. La métrique dans un espace de Minkowski est donc égale à :

$$g_{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (1.4)$$

Regardons d'abord comment la mécanique newtonienne se généralise dans ce cas. La première loi de Newton énonce qu'une particule libre (i.e. qui n'est soumise à aucune force) suit l'équation du mouvement suivante :

$$\frac{d^2 x^i}{dt^2} = 0 \quad (1.5)$$

Ceci est en fait équivalent à dire que les particules libres suivent des lignes droites dans l'espace-temps 4D. Il se trouve que c'est aussi équivalent à décrire la trajectoire des objets aux repos comme des géodésiques<sup>1</sup> dans l'espace de Minkowski<sup>2</sup>.

En relativité générale, les particules au repos suivent également des géodésiques dans l'espace-temps. La différence avec la relativité restreinte est que la métrique ( $g_{\mu\nu}$ ) peut être bien plus compliquée que dans l'expression (1.3). Les déformations de l'espace-temps, qui interviennent au travers de la métrique, peuvent donc influencer la trajectoire des objets de façon complexe.

Heureusement le principe cosmologique se traduit par un grand nombre de symétries de la métrique, ce qui simplifie grandement son expression. En effet, l'expression générale de la métrique dans

1. Une géodésique est une trajectoire qui maximise localement la distance propre  $\Delta s$  intégrée sur la trajectoire entre deux événements.

2. En effet les géodésiques sont des droites dans un espace de Minkowski.

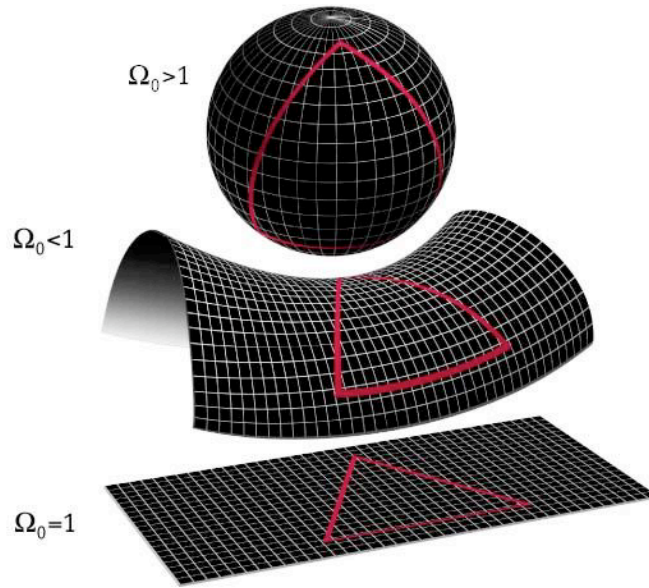


FIGURE 1.1 – Représentation bidimensionnelle des différents types de géométries possibles pour l'univers. Image de la NASA.

un espace-temps homogène et isotrope fut découverte par Robertson et Walker de façon indépendante vers 1935. En coordonnées sphériques  $(r, \theta, \phi)$  :

$$ds^2 = dt^2 - a(t)^2 \left( \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right) \quad (1.6)$$

où  $a(t)$  est le facteur d'échelle qui décrit l'expansion de l'univers au temps  $t$  (ou plus simplement le facteur d'expansion de l'univers), et  $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$  l'angle solide. La courbure de l'espace  $k$  peut être pris dans  $\{-1, 0, 1\}$ , ce qui correspond respectivement à un espace hyperbolique, euclidien ou sphérique (voir figure 1.1). Il faut quand même noter que cette forme de métrique n'est valable que lorsque le principe cosmologique est parfaitement vérifié (i.e. que l'univers est parfaitement homogène).

Comme nous l'avons vu dans la section 1.1.1, l'univers n'est en fait homogène qu'à des échelles suffisamment grandes. L'équation (1.6) ne décrit donc pas parfaitement la métrique aux petites échelles, mais donne une métrique moyenne. Ceci est responsable du 'problème de moyennage' en relativité générale, qui apparaît lorsque l'on mesure des quantités par moyennage (e.g. la fonction de corrélation, le spectre de puissance ou des paramètres comme la constante de Hubble) dans un échantillon. Lorsque l'on fait cela en utilisant la métrique de Robertson-Walker, on suppose que l'échantillon est homogène, ce qui n'est pas exactement le cas. Comme les équations d'Einstein (1.10) sont non linéaires, les résultats attendus pour les quantités moyennées ne sont pas forcément les mêmes pour la métrique moyennée de l'équation (1.6) que pour la vraie métrique (phénomène de *backreaction*, voir [Räsänen \(2011\)](#)). Cette forme de métrique est toutefois à la base de la cosmologie moderne car elle peut être combinée avec les équations d'Einstein (1.10) pour décrire l'évolution globale de l'Univers comme nous allons le voir.

Effectuons le changement de coordonnées  $r \rightarrow \chi$ , où  $\chi$  permet de déterminer les distances comobiles<sup>3</sup> :

$$\begin{cases} r = \sin(\chi) & \text{si } k = 1 \\ r = \chi & \text{si } k = 0 \\ r = \sinh(\chi) & \text{si } k = -1 \end{cases} \quad (1.7)$$

On obtient alors :

$$ds^2 = dt^2 - a(t)^2 (d\chi^2 + S_k^2(\chi) d\Omega^2) \quad (1.8)$$

3. La distance comobile entre deux objets fait abstraction de l'expansion de l'univers, c'est-à-dire utilise une unité de longueur qui suit l'expansion de l'univers.



où l'on a :

$$S_k(\chi) = \begin{cases} \sin(\chi) & \text{si } k = 1 \\ \chi & \text{si } k = 0 \\ \sinh(\chi) & \text{si } k = -1 \end{cases} \quad (1.9)$$

Jusqu'ici nous avons présenté le concept de métrique exprimant les déformations de l'espace-temps. Mais nous n'avons pas bien précisé comment sont générées ces déformations. En mécanique newtonienne, la force de gravitation est générée par la distribution de masse. De façon analogue, en relativité générale, les déformations de l'espace-temps sont générées par la distribution d'énergie. Ceci s'exprime à travers l'équation d'Einstein :

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi GT_{\mu\nu} \quad (1.10)$$

avec le tenseur  $G_{\mu\nu}$  dépendant de la métrique et de ses dérivées, et le tenseur  $T_{\mu\nu}$  correspondant au tenseur énergie-impulsion. Dans le cas d'un gaz parfait homogène et isotrope, on obtient :

$$T_{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} \quad (1.11)$$

En se servant de cette forme pour  $T_{\mu,\nu}$  ainsi que de la forme de la métrique (1.6), l'équation d'Einstein (1.10) se simplifie grandement. On obtient les équations suivantes, découvertes par Friedman et Lemaître de façon indépendante, respectivement en 1922-1924 et en 1927 :

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \quad (1.12)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \quad (1.13)$$

A l'heure actuelle, ces équations constituent toujours la base de la cosmologie moderne. On voit dans l'équation (1.12) qu'il existe nécessairement une valeur critique de la densité telle que l'univers est plat (i.e.  $k = 0$ ) :

$$\rho_c = \frac{3H^2}{8\pi G} \quad (1.14)$$

Lorsque Einstein formula la théorie de la relativité, il était alors admis que l'univers était statique (voir section 1.2). Or on voit immédiatement avec l'équation (1.13) que l'univers ne peut pas être statique à moins d'avoir  $\rho = -3P$ , c'est à dire à moins d'avoir une densité ou une pression négative. Ceci conduit Einstein à modifier l'équation (1.10) en introduisant la constante cosmologique  $\Lambda$  :

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu} \quad (1.15)$$

On peut voir ceci comme rajouter un terme au tenseur-énergie impulsion  $\tilde{T}_{\mu\nu} = T_{\mu\nu} + \Lambda g_{\mu\nu}$ . Ceci correspond à une nouvelle composante dont l'équation d'état reliant pression et énergie est  $\mathcal{P}_\Lambda = w\rho_\Lambda$  avec  $w = -1$ . La constante cosmologique a donc l'étonnante propriété d'exercer une pression négative. La pression et densité globales sont alors données par :

$$\tilde{\mathcal{P}} = P + P_\Lambda = P - \frac{\Lambda}{8\pi G} \quad (1.16)$$

$$\tilde{\rho} = \rho + \rho_\Lambda = \rho + \frac{\Lambda}{8\pi G} \quad (1.17)$$

Les équations (1.12) et (1.13) deviennent :

$$H^2 = \frac{8\pi G}{3}\tilde{\rho} - \frac{k}{a^2} \quad (1.18)$$

$$= \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2} \quad (1.19)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\tilde{\rho} + 3\tilde{\mathcal{P}}) \quad (1.20)$$

$$= -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3} \quad (1.21)$$

Avec une constante cosmologique, il devient possible de construire un univers statique (i.e. avec  $\dot{a} = 0$  et  $\ddot{a} = 0$ ). Cependant un tel univers statique est en réalité instable parce que les inhomogénéités locales conduisent inévitablement, soit à une expansion, soit à une contraction de l'univers. De plus lorsque Hubble découvrit en 1929 que l'univers est en expansion, Einstein décrivit l'introduction de  $\Lambda$  comme sa plus grande erreur. Plus précisément, Hubble constata que les galaxies environnantes ont une vitesse de récession proportionnelle à leur distance, i.e. suivent la loi de Hubble (Hubble 1929) :

$$v = H_0 d \quad (1.22)$$

Les vitesses de récession des galaxies qui expriment une dilatation de l'espace lui-même engendrent un accroissement de la longueur d'onde des photons, i.e. un décalage vers le rouge (*redshift*). Le *redshift* d'un objet observé à travers des photons émis au temps  $t$  est donc lié au facteur d'expansion de l'univers  $a$  :

$$1 + z = \frac{a(t_0)}{a(t)} = \frac{1}{a(t)} \quad (1.23)$$

où  $t_0$  représente le temps actuel et l'on prend la convention habituelle que le facteur d'expansion actuel  $a(t_0)$  est égal à 1.

La constante cosmologique n'était plus nécessaire et fut abandonnée jusqu'à la fin des années 90, où cette composante fut à nouveau nécessaire pour expliquer l'expansion accélérée mesurée avec des supernovae (Riess et al. 1998; Perlmutter et al. 1999). Plus précisément cette accélération ne peut être expliquée qu'avec l'addition d'une composante telle que  $w < -1/3$  (i.e. avec une pression négative), connue sous le nom d'énergie sombre, où  $w$  est défini par l'équation d'état  $P = P(\rho)$  de la façon suivante :

$$P = w\rho \quad (1.24)$$

Le cas de la constante cosmologique (aussi appelée énergie du vide) avec  $w = -1$  n'est qu'un cas particulier d'énergie sombre, bien qu'il semble indiqué par les plus récentes observations provenant du CMB, de l'analyse des grands relevés de galaxies et des supernovae.

On peut combiner les équations (1.19) et (1.21) afin d'obtenir l'équation d'évolution de la densité d'énergie  $\rho$  de l'univers :

$$\dot{\rho} + 3(\rho + P) \frac{\dot{a}}{a} = 0 \quad (1.25)$$

En combinant avec l'équation d'état (1.24), on obtient alors l'évolution de la densité d'énergie de l'univers en fonction du facteur de croissance :

$$\rho(t) \propto a(t)^{-3(1+w)} \quad (1.26)$$

En réalité l'équation d'état est différente pour chaque composante de l'univers, de sorte que pour chaque espèce l'évolution de la densité est donnée par :

$$\rho_i(t) \propto a(t)^{-3(1+w_i)} \quad (1.27)$$

avec  $w_\Lambda = -1$  pour une constante cosmologique (cas particulier d'énergie sombre),  $w_r = 1/3$  pour la radiation et  $w_m = 0$  pour la matière non relativiste. Pour la matière relativiste comme les neutrinos,  $w_i$  n'est pas constant et varie en fonction des époques mas cette composante peut être négligée pour les époques récentes.

On peut alors écrire les différentes densités en fonction du *redshift*  $z$  :

$$\rho_m = \rho_{m,0}(1+z)^3 \quad (1.28)$$

$$\rho_r = \rho_{r,0}(1+z)^4 \quad (1.29)$$

$$\rho_\Lambda = \rho_{\Lambda,0} \quad (1.30)$$

Ceci explique notamment pourquoi la densité de rayonnement devient négligeable au cours du temps par rapport à la densité de matière. Cela explique aussi pourquoi l'influence de l'énergie sombre ne se fait ressentir qu'aux époques récentes.

On peut aussi définir artificiellement une densité de courbure :

$$\rho_k = -\frac{3k}{8\pi G a^2} \quad (1.31)$$

$$\rho_k = \rho_{k,0}(1+z)^2 \quad (1.32)$$

de sorte que l'on puisse réécrire l'équation (1.19) :

$$H^2 = \frac{8\pi G}{3}(\rho_m + \rho_r + \rho_\Lambda + \rho_k) \quad (1.33)$$

Il est important de noter que toutes les densités  $\rho_i$  sont décroissantes au cours du temps, hormis pour l'énergie sombre  $\rho_\Lambda = \text{const}$ , ce qui explique que cette forme d'énergie finit forcément par dominer l'évolution.

### 1.1.3 Constituants cosmologiques et évolution de l'univers

Comme nous l'avons vu dans la section 1.1.2, l'évolution de l'univers est dictée par la densité  $\rho$  et la pression  $P$  des constituants cosmiques, ainsi que de la courbure  $k$  de l'univers. Nous avons vu dans l'équation (1.14) que la densité critique d'énergie pour obtenir un univers plat avec  $k = 0$  est donnée par :

$$\rho_c(t) = \frac{3H(t)^2}{8\pi G} \quad (1.34)$$

On peut alors comprendre la contribution de chaque constituant en terme de rapport avec la densité critique :

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)} = \frac{8\pi G\rho}{3H(t)^2} \quad (1.35)$$

Les constituants principaux de l'univers sont la matière (matière noire et matière baryonique), la radiation et l'énergie sombre. On adopte les notations respectives :

$$\Omega_m = \frac{\rho_{m,0}}{\rho_{c,0}}, \quad \Omega_r = \frac{\rho_{r,0}}{\rho_{c,0}}, \quad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{c,0}} = \frac{\Lambda}{3H_0^2} \quad (1.36)$$

De même on peut définir :

$$\Omega_k = \frac{\rho_{k,0}}{\rho_{c,0}} = -\frac{k}{H_0^2} \quad (1.37)$$

si bien que l'on obtient avec l'équation (1.33) :

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1 \quad (1.38)$$

Notons que l'on a omis les indices 0, i.e. on a écrit  $\Omega = \Omega_0$  pour simplifier les notations. Si l'on veut décrire les densités des différents constituants à un *redshift*  $z$  quelconque on écrira :

$$\Omega_i(z) = \frac{\rho_i(z)}{\rho_c(z)} \quad (1.39)$$

On peut réutiliser l'équation (1.33) pour obtenir :

$$H(z)^2 = \frac{H_0^2}{\rho_{c,0}}(\rho_m(z) + \rho_r(z) + \rho_\Lambda(z) + \rho_k(z)) \quad (1.40)$$

Et finalement en utilisant les équation d'évolution des densités (1.29), (1.30), (1.30) et (1.32) :

$$H(z) = H_0 E(z) \quad (1.41)$$

$$E(z) = \sqrt{\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda} \quad (1.42)$$

L'accélération de l'expansion est décrite par le paramètre de décélération  $q = -\frac{a\ddot{a}}{\dot{a}^2}$ . Le signe de cette quantité est choisie de sorte qu'elle soit positive quand l'expansion décélère. A partir de l'équation (1.21) on obtient le facteur de décélération actuel :

$$q_0 = \Omega_r + \frac{\Omega_m}{2} - \Omega_\Lambda \quad (1.43)$$

La figure 1.2 montre l'évolution du facteur d'échelle  $a(t)$  en fonction du temps pour différent choix de densités  $\Omega_m, \Omega_\Lambda$ . Un univers fermé avec  $\Omega_m > 1, \Omega_\Lambda = 0$  finira par s'effondrer sur lui-même. Un univers avec  $\Omega_m < 1, \Omega_\Lambda = 0$  est ouvert et un univers  $\Omega_m = 1, \Omega_\Lambda = 0$  (modèle d'Einstein de Sitter) est plat. Ces deux modèles auront une expansion qui décélère mais qui est infinie. En revanche, pour un modèle standard avec  $\Omega_m = 0.3, \Omega_\Lambda = 0.7$ , l'expansion s'accélère après un temps suffisant, de par l'effet de la constante cosmologique (en particulier  $q_0 < 0$ ).

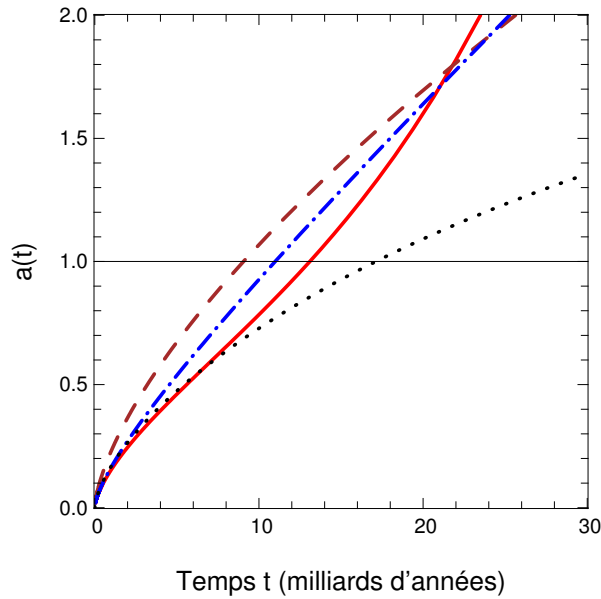


FIGURE 1.2 – Les différentes évolutions du paramètre d'échelle  $a(t)$  selon la quantité de matière  $\Omega_m$  et d'énergie noire  $\Omega_\Lambda$  (avec  $w = -1$ ) dans l'univers, avec pour le temps présent  $a(t_0) = 1$ . La courbe marron en tirets correspond à un univers fermé  $\Omega_m > 1$ ,  $\Omega_\Lambda = 0$  qui finira par s'effondrer sur lui même, la courbe noire pointillée le cas d'un univers ouvert  $\Omega_m < 1$ ,  $\Omega_\Lambda = 0$  et la courbe bleue en tirets-points le cas d'un univers plat  $\Omega_m = 1$ ,  $\Omega_\Lambda = 0$  (modèle d'Einstein de Sitter). La courbe continue rouge décrit l'évolution de  $a(t)$  pour un modèle standard actuel  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ . Figure extraite de [Sousbie \(2007\)](#).

### 1.1.4 Distances dans un univers FLRW

#### Redshift

Dans un univers FLRW il y a de nombreuses façons de définir la distance entre 2 points. Une quantité très importante est le décalage vers le rouge d'un objet, ou *redshift*, qui décrit comment les spectres émis par les objets sont décalés du fait de l'expansion de l'univers entre le moment de l'émission  $t_e$  et celui de l'observation  $t_o$ . Comme le reste de l'univers, la longueur d'onde  $\lambda$  d'un photon subit l'expansion, i.e.  $\lambda \propto a(t)$ .

Le *redshift*  $z$  d'un objet est alors défini ainsi :

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)} = 1 + z \quad (1.44)$$

Le *redshift* est quasiment toujours défini par rapport à nous à l'instant  $t_o = t_0$ . Avec  $a(t_0) = 1$  on obtient alors :

$$z = \frac{1}{a} - 1 \quad (1.45)$$

#### Distance comobile

Une petite distance comobile  $\delta D_C$  entre 2 objets est la distance qui reste constante lorsque les objets suivent le flot de Hubble, i.e. l'expansion de l'univers. En d'autres termes, c'est la distance propre entre les 2 objets telle que mesurée par une règle au moment de l'émission, et multipliée par  $1/a(t_e)$  (de sorte que cette distance ne dépend effectivement pas du moment de la mesure). La distance comobile d'un objet dans la ligne de visée est donc donnée par l'intégration de  $\delta D_C$  dans la direction radiale, entre l'instant d'émission et d'observation :

$$D_C = \int_{t_e}^{t_0} \frac{c}{a(t)} dt \quad (1.46)$$

Etant donné que  $dz = -\frac{\dot{a}}{a^2}dt = -\frac{1}{a}H(z)dt$  on peut réécrire la distance comme une intégrale sur le *redshift* :

$$D_C = \int_0^z \frac{c}{H(z')} dz' \quad (1.47)$$

$$= D_H \int_0^z \frac{1}{E(z')} dz' \quad (1.48)$$

où l'on a défini à nouveau  $H(z) = H_0 E(z)$  et  $D_H = \frac{c}{H_0}$  la distance de Hubble. On rappelle l'expression de  $E(z)$  en fonction des différents constituants cosmiques :

$$E(z) = \sqrt{\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda} \quad (1.49)$$

La distance comobile entre 2 objets au même *redshift*  $z$ , mais séparé sur le ciel par un angle  $\delta\theta$ , est par définition égale à  $D_M \delta\theta$ , où  $D_M$  est la distance comobile transverse à *redshift*  $z$ .  $D_M$  est relié à la distance comobile radiale  $D_C$  par la relation :

$$D_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_k}} \sinh \left[ \sqrt{\Omega_k} D_C / D_H \right] & \text{pour } \Omega_k > 0 \\ D_C & \text{pour } \Omega_k = 0 \\ D_H \frac{1}{\sqrt{|\Omega_k|}} \sin \left[ \sqrt{|\Omega_k|} D_C / D_H \right] & \text{pour } \Omega_k < 0 \end{cases} \quad (1.50)$$

### Distance angulaire

La distance angulaire  $D_A$  entre deux objets est définie comme le ratio de la distance propre entre les deux objets et de la distance angulaire. Elle est reliée à la distance comobile transverse par la relation :

$$D_A = \frac{D_M}{1+z} \quad (1.51)$$

### Distance lumineuse

La distance lumineuse est définie comme un rapport entre la luminosité absolue d'un objet  $L$  et sa luminosité apparente  $l$  :

$$D_L = \sqrt{\frac{L}{4\pi l}} \quad (1.52)$$

Dans un espace euclidien, une onde sphérique se dilue avec le carré de la distance parcourue, et donc la distance lumineuse correspond à la définition usuelle de la distance. Dans le cas général d'un espace en expansion on peut montrer :

$$D_L = (1+z)D_M = (1+z)^2 D_A \quad (1.53)$$

## 1.2 Modèle du Big Bang chaud

La découverte de l'expansion de l'univers fut le premier argument en faveur du modèle de Big Bang chaud. Selon ce modèle, l'univers initial était constitué d'un plasma extrêmement chaud et dense, composé de photons, fermions et quarks. L'univers s'est ensuite dilaté; sa densité et sa température ont diminué ce qui permit la nucléosynthèse primordiale, i.e. la formation des premiers noyaux légers (hydrogène, hélium, deutérium, lithium).

Environ 380,000 ans après le Big Bang la température est devenue suffisamment basse ( $\approx 3,000K$ ) pour que les premiers atomes se forment dans un processus que l'on nomme recombinaison (bien que les atomes ne se soient pas combinés auparavant). A cette époque, les baryons se découplent des photons et la lumière commence à se propager librement. Cette radiation, qui va simplement se refroidir au cours de l'évolution, est connue sous le nom de fond diffus cosmologique (CMB pour *Cosmic Microwave Background*). A partir de cette époque, les photons et la matière évoluent indépendamment. Les baryons et la matière noire froide évoluent de façon jointe sous l'action de la gravitation, et forment

les structures que l'on observe actuellement, les galaxies et les étoiles.

Le modèle du Big Bang chaud s'appuie principalement sur 3 principales observations :

- la loi de Hubble pour la vitesse de récession apparente des galaxies donnée par l'équation (1.22). Cette loi indique que l'univers est en expansion, et que la récession des galaxies est causée par l'expansion de l'espace lui-même. Aujourd'hui de nombreuses observations (notamment avec les supernovae Ia) ont confirmé et généralisé cette loi.
- la détection du fond diffus cosmologique par Penzias et Wilson en 1965. L'explication naturelle pour la présence de cette radiation est que l'univers est passé par une phase beaucoup plus chaude et dense où baryons et photons étaient couplés. Le couplage a duré jusqu'à ce que la température devienne suffisamment basse, qu'il y ait la recombinaison et que cette radiation soit émise. Par ailleurs, la température de la radiation égale à  $T = 2.73K$ , son spectre de corps noir quasi-parfait et son isotropie à un facteur  $10^{-5}$  sont également compatibles avec cette interprétation.
- Les mesures d'abondance des éléments légers dans l'univers sont en excellent accord avec les prédictions d'une nucléosynthèse primordiale. Celle-ci prévoit par exemple que l'abondance massique de l'hélium 4 est située entre 23% et 30%, en parfait accord avec les observations.

### 1.2.1 Matière noire et énergie sombre

La matière noire est un type de matière qui n'interagit pas électromagnétiquement (et qui ne peut donc pas être vue avec des télescopes). Son existence et ses propriétés sont en fait déduites de ses effets gravitationnels sur la matière visible. On pense que la matière noire constitue 84% de la masse de l'univers. Le premier à postuler l'existence d'une telle matière fut Jan Oort en 1932 qui étudia les orbites des étoiles dans la galaxie et conclut que la voie Lactée devait contenir plus de matière que ce qui était visible (Oort 1932). Peu après, Zwicky appliqua le théorème de viriel sur le superamas de Coma et conclut que la masse totale devait être 400 fois plus importante que la masse de la matière visible (Zwicky 1937).

Depuis, de nombreuses observations ont corroboré l'existence de matière noire dans l'univers : les vitesses rotationnelles dans les galaxies, les effets de lentilles gravitationnelles (e.g. par l'amas du Boulet), la distribution de température du gaz chaud dans les amas de galaxies, l'analyse du CMB, des supernovae de type Ia ainsi que des BAOs (voir section 1.2.2). Il existe 3 types différents de matière noire : la matière noire froide (CDM pour *Cold Dark Matter*), la matière noire tiède (WDM pour *Warm Dark Matter*) et la matière noire chaude (HDM pour *Hot Dark Matter*). Les différences viennent de la masse des particules (les particules de matière noire froide étant massives) et par conséquent de leur vitesse.

Le meilleur candidat pour la matière chaude est le neutrino dont la masse n'est pas encore bien contrainte. Concernant la matière noire froide les meilleurs candidats sont les WIMP (pour *Weak Interacting Massive Particle*), qui sont prédits par de nombreuses théories supersymétriques, et les MACHO (pour *Massive Astronomical Compact Halo Object*).

La vitesse de déplacement de ces particules intervient dans l'ordre de formation des grandes structures de l'univers. Un univers dominé par la matière noire chaude aurait un scénario de formation des structures 'du haut vers le bas' : formation des superamas de galaxies qui ensuite se fragmentent en amas, puis en galaxies, etc. Au contraire, un univers dominé par la matière noire froide aurait un scénario 'du bas vers le haut' : formation des galaxies (à partir de nuages de gaz) qui se regroupent ensuite en amas puis en superamas. Aujourd'hui la théorie de la matière noire froide est privilégiée car les galaxies sont à l'équilibre dynamique, à la différence de certains amas - ce qui laisse supposer qu'elles se sont formées antérieurement.

Un constituant cosmologique encore plus mystérieux est l'énergie sombre. Comme discuté dans la section 1.1.2, l'idée d'une constante cosmologique qui permettrait un univers statique fut envisagé par Einstein peu après sa formulation de la théorie de la relativité générale. Mais l'idée fut vite abandonnée lorsque Hubble mesura que l'univers était en expansion. L'énergie sombre est une forme d'énergie emplissant tout l'univers et dotée d'une pression négative et qui se comporte comme une

force gravitationnelle répulsive. Elle est l'explication la plus simple pour expliquer l'accélération de l'univers mesurée pour la première fois dans [Riess et al. \(1998\)](#) et confirmée dans [Perlmutter et al. \(1999\)](#). Depuis de nombreuses observations ont renforcé la vision d'un univers en expansion accéléré comme l'étude du CMB et des BAOs (voir section 1.2.2).

La nature de l'énergie sombre est encore spéculative, et les preuves de son existence ne sont qu'indirectes à travers la relation *redshift*-distance. Une des propriétés de l'énergie sombre est qu'elle possède une pression négative (plus précisément le paramètre  $w$  de son équation d'état  $P = w\rho$  vérifie  $w < -1/3$ ) afin d'expliquer l'accélération de l'expansion de l'univers ( $\ddot{a} > 0$ , voir équation (1.13)).

La forme la plus simple d'énergie sombre est une constante cosmologique  $\Lambda$ , qui correspond à l'énergie du vide prédite par la plupart des théories de physique des particules. On peut voir simplement avoir des arguments de thermodynamique classique pourquoi une constante cosmologique possède une pression négative. Un changement de volume  $dV$  requiert un travail (un changement d'énergie)  $-P_\Lambda dV$ , où  $P_\Lambda$  est la pression. Mais la quantité d'énergie du vide va en fait augmenter lorsque le volume augmente de  $dV > 0$  puisque l'énergie est égale à  $\rho_\Lambda V$ . La pression  $P$  doit donc être négative, et plus précisément  $P_\Lambda = -\rho_\Lambda$ . Un problème majeur est que la théorie quantique des champs prévoit une valeur infiniment trop grande par rapport à la valeur déduite des observations cosmologiques (plus de 100 ordres de grandeurs trop grande, [Carroll \(2001\)](#)).

Une autre possibilité d'énergie sombre est donnée par la quintessence. La quintessence est un champ scalaire qui a une équation d'état  $P_Q = w_Q \rho_Q$ , avec  $w_Q$  évoluant lentement au cours du temps (elle peut connaître des variations plus rapides d'une valeur palier à une autre). La quintessence est une quantité dynamique, dont la densité et l'équation d'état varient au cours du temps, et qui peut présenter des fluctuations dans l'espace. Aucune preuve de cette théorie n'a été trouvée pour l'instant, mais elle n'a pas non plus été éliminée.

D'autres explications ont été avancées pour expliquer l'accélération de l'expansion de l'univers (et donc préciser la nature de l'énergie sombre), notamment à travers la théorie des cordes ou la cosmologie branaire. Par exemple la théorie des cordes rend possible le fait que notre univers observable ne soit qu'un parmi  $10^{500}$  univers à l'intérieur d'un grand multivers ([Hogan 2007](#)). L'énergie du vide aurait des valeurs différentes dans les différents univers et cette énergie devrait être très grande dans la plupart des univers. Mais elle doit être faible dans le nôtre pour la simple raison que c'est une condition pour l'existence d'observateurs tels que nous (c'est le principe anthropique faible). Cependant cette explication de l'énergie noire en tant que combinaison d'aléatoire du principe anthropique peut sembler décevante ([Hogan 2007](#)).

Une autre explication de l'accélération de l'expansion, qui est aussi marginale, serait que cette accélération ne soit qu'une illusion créée par notre mouvement relatif par rapport au reste de l'univers ([Tsagas 2011](#)). Finalement une autre explication est basée sur le phénomène de *backreaction* ([Räsänen 2011](#)) décrit dans la section 1.1.2. Ce phénomène est dû aux différences dans les quantités observées lorsque l'on prend en compte les fluctuations de la métrique dans les équations d'Einstein au lieu de considérer la métrique moyenne de Robertson-Walker. On peut montrer que ce phénomène peut causer une accélération de l'expansion, sans avoir besoin d'une forme additionnelle d'énergie sombre ([Räsänen 2011](#)). Cet effet semble toutefois trop faible par rapport à l'accélération observée.

### 1.2.2 Modèle $\Lambda$ CDM - Mesure des paramètres cosmologiques

Le modèle  $\Lambda$ CDM (pour  $\Lambda$  - *Cold Dark Matter*) a émergé vers la fin des années 1990 comme le modèle de concordance, après une période où diverses observations de l'univers apparaissaient incompatibles, et où il n'y avait pas de consensus concernant la densité d'énergie dans l'univers. C'est le modèle le plus simple qui est en accord avec les différentes observations cosmologiques : CMB, distribution des grandes structures de l'univers (galaxies et amas de galaxies), abondance des éléments légers (hydrogène, hélium, deutérium, lithium), expansion accélérée.

Au sens propre, le modèle  $\Lambda$ CDM est un modèle contenant une constante cosmologique, de la matière noire froide et de la matière baryonique. Ce modèle implique l'existence d'une singularité initiale, le *Big Bang*. Un modèle  $\Lambda$ CDM plat est de courbure nulle  $\Omega_k = 0$  et un modèle  $\Lambda$ CDM non plat autorise une courbure  $\Omega_k \neq 0$ . Il est aussi accepté qu'il y eut une période connue comme l'inflation cosmique environ  $10^{-36}$  secondes après le *Big Bang*, qui fut une période d'expansion exponentielle de l'espace d'un facteur  $10^{26}$  ou plus. Le modèle  $\Lambda$ CDM dépend de 6 paramètres :  $\Omega_\Lambda$ ,  $\Omega_m$ ,  $\Omega_b$ ,  $H_0$ ,  $n_s$  et  $\sigma_8$ .

Une extension habituelle est de considérer des formes plus générales d'énergie sombre dont la densité est alors dénotée  $\Omega_{DE}$  (au lieu de  $\Omega_\Lambda$  dans le cas d'une constante cosmologique). Le modèle  $w$ CDM correspond à une énergie noire avec  $w$  constant dans l'équation d'état (la constante cosmologique correspond au cas particulier  $w = -1$ ). Un autre modèle très souvent utilisé suppose une dépendance linéaire de  $w(a)$  en fonction du facteur d'échelle  $a$  (Linder 2003) :

$$w(a) = w_0 + w_a(1 - a) \quad (1.54)$$

$$= w_0 + w_a \left( \frac{z}{1+z} \right) \quad (1.55)$$

Cette simple paramétrisation permet en effet de déceler un comportement non constant de  $w$ , comme c'est le cas par exemple dans les modèles de quintessence.

La contrainte des différents paramètres du modèle  $\Lambda$ CDM ainsi que des ses extensions, notamment  $\Omega_m$ ,  $\Omega_{DE}$  et  $w_{DE}$ , a été l'objet principal de la recherche en cosmologie ces dernières années. Ces paramètres peuvent être contraints à travers différentes sondes cosmologiques, comme le CMB ou bien des sondes exploitant la relation *redshift*-distance comme les supernovae Ia en tant que chandelles standards ou les Oscillations Baryoniques Acoustiques (BAOs pour *Baryon Acoustic Oscillations*) en tant qu'étalon standard dans la corrélation des grandes structures.

Les études les plus récentes combinant ces différentes sondes (Komatsu et al. 2011; Blake et al. 2011b; Sánchez et al. 2012; Amanullah et al. 2010) supportent le modèle  $\Lambda$ CDM plat,  $\Omega_k \approx 0$ , avec une grande partie du contenu sous la forme d'une constante cosmologique  $\Omega_\Lambda \approx 0.72$ , et le reste sous forme de matière  $\Omega_m \approx 0.28$ . Une grande partie de cette matière est de la matière noire et seulement une fraction  $\Omega_b \approx 0.045$  est de la matière baryonique. La constante de Hubble à l'heure actuelle est aussi bien déterminée  $H_0 \approx 69 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , et l'équation d'état de l'énergie sombre est compatible avec une constante cosmologique  $w = -1$ .

La figure 1.3, extraite de Kowalski et al. (2008), montre comment ces différentes observations peuvent être combinées pour contraindre les paramètres cosmologiques (ici  $\Omega_m$  et  $\Omega_\Lambda$ ). On voit que les différentes sondes cosmologiques ont des directions de dégénérescence différentes, si bien qu'on peut les combiner pour exploiter pleinement leur potentiel. Nous donnons ci-dessous un bref aperçu du principe de ces différentes observations afin de comprendre les mécanismes permettant de contraindre les paramètres cosmologiques.

## CMB

Le fond diffus cosmologique a été découvert fortuitement en 1964 par les radio-astronomes Penzias et Wilson en utilisant une antenne qui devait servir initialement à la mesure du rayonnement de la Voie Lactée dans le domaine radio. En voulant étalonner le bruit de fond, ils découvrirent un bruit d'origine inconnue correspondant à une température de 2.7K et ne présentant pas de variations saisonnières, désormais connu comme étant le CMB.

L'existence du CMB est parfaitement expliquée par le modèle du Big Bang chaud. Dans ce modèle, l'univers primordial était en équilibre thermique, avec une distribution d'énergie suivant une loi de Planck (ou loi de corps noir). Tant que la température était plus élevée que la température d'ionisation de l'hydrogène, les photons étaient en interaction avec les baryons, i.e. que l'univers se comportait comme un plasma où les photons ne pouvaient pas se propager librement. Lorsque la température est devenue assez faible pour qu'il y ait découplage, les photons ont pu se propager librement tout en gardant leur distribution de corps noir. Ce découplage eut lieu environ 380,000 ans après le Big Bang, à *redshift*  $z \approx 1100$  et à la température  $T \approx 3,000\text{K}$ . Du fait de l'expansion de l'univers, leur longueur d'onde s'est allongée (du domaine des rayons gamma au domaine micro-ondes), i.e. que leur température a diminué.

Ce découplage ayant eu lieu partout dans l'univers au même moment, le CMB observé aujourd'hui vient de la surface de dernière diffusion, qui est la sphère dont nous sommes le centre et dont la taille correspond à la distance parcourue par la lumière de cette époque à aujourd'hui. Le CMB apparaît comme un corps noir quasi parfait avec une température  $T = 2.73\text{K}$  (voir figure 1.4) et seulement de très faibles anisotropies sur le ciel de l'ordre de  $10^{-5}$  (mesurées pour la première fois avec le satellite COBE dans les années 1990).



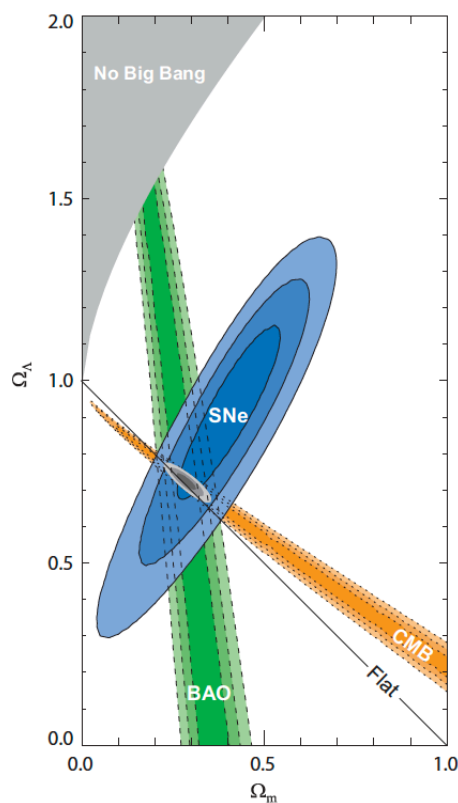


FIGURE 1.3 – Régions de confiance à 68.3 %, 95.4 % et 99.7% pour  $\Omega_\Lambda$  et  $\Omega_m$  obtenues avec le CMB, les BAOs et les données d’Union SN, ainsi que leur combinaison (en supposant  $w = -1$ ). On voit notamment que les observations actuelles semblent indiquer un univers plat. Ces contraintes viennent des données disponibles en 2008 et sont en constante amélioration. Figure extraite de [Kowalski et al. \(2008\)](#).

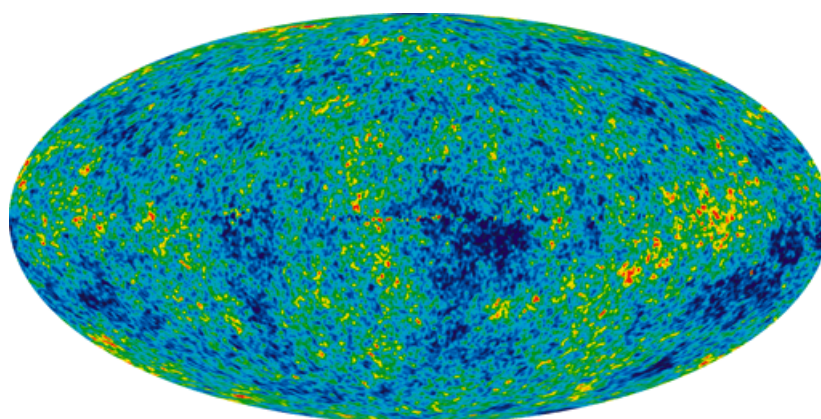


FIGURE 1.4 – Carte du ciel en coordonnées galactiques du CMB mesuré par le satellite WMAP après 7 ans d’observation ([Jarosik et al. 2011](#)). Cette carte dénotée *Internal Linear Combination Map* est une combinaison linéaire des différentes bandes d’observation afin de réduire la contamination galactique. Figure fournie par l’équipe scientifique de WMAP, <http://lambda.gsfc.nasa.gov/product/map/current/>.

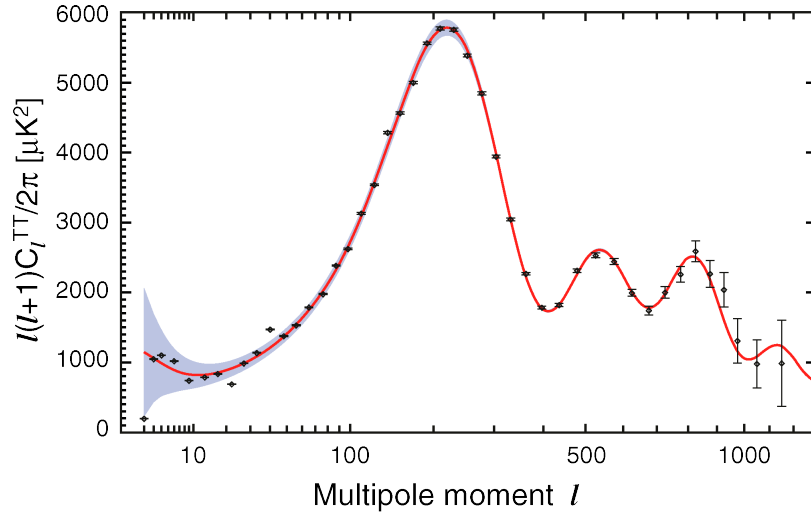


FIGURE 1.5 – Spectre de puissance du CMB mesuré par WMAP après 7 ans d’observation. La courbe rouge donne le modèle  $\Lambda$ CDM qui maximise la vraisemblance, les barres d’erreur donnent le bruit instrumental et la région grise représente la variance cosmique. Figure extraite de Larson et al. (2011).

Les données plus récentes venant du satellite WMAP (Jarosik et al. 2011) ont permis d’obtenir une très grande précision de mesure. Une précision encore plus grande est attendu avec les données de Planck lancé en mai 2009, encore non publiées à la date de l’écriture de ce manuscrit. Nous n’expliquerons pas ici la théorie du CMB mais on peut montrer que si l’on décompose les anisotropies en harmoniques sphériques :

$$\delta T(\theta, \phi) = \sum_{l,m} a_{lm} Y_{lm}(\theta, \phi) \quad (1.56)$$

alors les valeurs des coefficients  $C_l = \langle |a_{lm}|^2 \rangle_m$  (appelé spectre de puissance angulaire) présentent une série de pics dont l’amplitude et la position peuvent être reliées aux paramètres cosmologiques (voir figure 1.5). Les résultats les plus récents (Komatsu et al. 2011) provenant des données WMAP après 7 ans d’observation donnent pour un modèle  $\Lambda$ CDM plat les valeurs suivantes :

$$H_0 = 70.4 \pm 2.5 \text{ km.s}^{-1} \text{ Mpc}^{-1} \quad (1.57)$$

$$\Omega_m h^2 = 0.1345 \pm 0.0056 \quad (1.58)$$

$$\Omega_b = 0.0455 \pm 0.0028 \quad (1.59)$$

$$\Omega_\Lambda = 0.727 \pm 0.03 \quad (1.60)$$

$$n_s = 0.967 \pm 0.014 \quad (1.61)$$

$$(1.62)$$

où l’on a utilisé la notation  $H_0 = 100h \text{ km.s}^{-1} \text{ Mpc}^{-1}$ , et  $n_s$  est l’indice spectral défini section 1.3.1.

### Supernovae Ia

On pense actuellement que les supernovae de Type Ia fournissent des ‘chandelles standard’, dans le sens que l’énergie lumineuse qu’elles émettent est proportionnelle à leur temps d’extinction qui est facile à mesurer. Ainsi on peut connaître la magnitude absolue  $M$ , et grâce à l’observation de la magnitude apparente  $m$  et du *redshift* on peut en déduire la distance lumineuse avec la formule suivante :

$$m - M = 5 \log_{10} \frac{D_L}{1 \text{ pc}} - 5 \quad (1.63)$$

En faisant cette mesure pour un grand nombre de supernovae à des *redshift*  $z$  différents, on obtient une courbe  $D_L = f(z)$ , ce qui permet de contraindre les paramètres cosmologiques à travers les for-

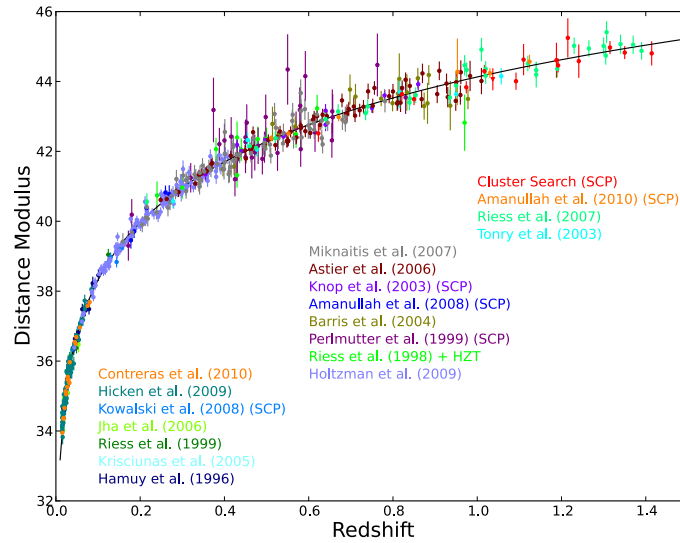


FIGURE 1.6 – Diagramme de Hubble pour la compilation Union 2.1. La courbe solide représente le modèle  $\Lambda$ CDM plat qui maximise la vraisemblance. Les barres d’erreur comprennent l’erreur intrinsèque et l’erreur de mesure. Figure extraite de Suzuki et al. (2012).

mules de la section 1.1.4. Cette méthode fut notamment utilisée dans Riess et al. (1998) et Perlmutter et al. (1999), qui montrèrent pour la première fois une accélération de l’expansion de l’univers.

La plus récente compilation Union 2.1 de supernovae regroupe 580 supernovae allant jusqu’à *redshift*  $z = 1.415$ . Elle est accessible à l’adresse <http://supernova.lbl.gov/Union>. La figure 1.6 montre le diagramme de Hubble construit à partir de cette compilation de supernovae (Suzuki et al. 2012).

## BAOs

Les Oscillations Baryoniques Acoustiques, ou BAOs, correspondent aux ondes acoustiques qui étaient présentes dans le plasma baryon-photon avant le découplage et l’émission du CMB. Alors que les surdensités attireraient gravitationnellement la matière, la pression des photons entraînait une résistance à ces compressions, générant des ondes acoustiques dans le plasma primordial.

Lorsque le découplage eut lieu, les photons ont pu se diffuser librement, et les ondes sonores se sont figées. La distance que les ondes ont pu parcourir avant le découplage, l’horizon sonique  $r_s$ , imprime une échelle particulière dans la distribution photon-matière à cette époque. Cette échelle est notamment responsable des différents pics que l’on peut voir dans le spectre angulaire du CMB, qui permet de bien contraindre l’horizon sonique,  $r_s = 153.3 \pm 2.0$  Mpc (Komatsu et al. 2009). On donnera plus d’informations sur la physique des Oscillations Baryoniques Acoustiques dans la section 1.4.

Après le découplage, l’échelle  $r_s$  est resté imprimée dans la distribution des baryons, et donc de la matière qui n’a ensuite évolué que gravitationnellement. Cette échelle peut être vu comme un pic localisé dans la fonction de corrélation de la matière (voir section 2.1.2 pour une définition précise de la fonction de corrélation). Ce pic a été détecté pour la première fois de façon convaincante dans l’échantillon de *Luminous Red Galaxies* (LRG) du *Sloan Digital Sky Survey Data Release 3* (SDSS DR3) à un niveau de  $3.4\sigma$  (Eisenstein et al. 2005).

De même que les supernovae Ia fournissent des chandelles standard pour étudier la relation *redshift*-distance lumineuse, le pic BAO fournit un étalon standard (*standard ruler*) pour étudier la relation *redshift*-distance comobile. En effet, la localisation du pic BAO dans la fonction de corrélation fournit une échelle caractéristique dans la distribution des grandes structures.

Pour comprendre comment utiliser cet étalon standard, supposons qu’on observe un objet de taille  $\delta D$  situé à *redshift*  $z$ , à la fois dans la ligne de visée et dans la direction transverse. D’après les formules de la section 1.1.4 cet objet a une extension en *redshift*  $\Delta z$  avec  $\Delta z \frac{c}{H(z)} = \delta D$ , et une extension angulaire  $\Delta\theta$  avec  $\Delta\theta D_M(z) = \delta D$ .

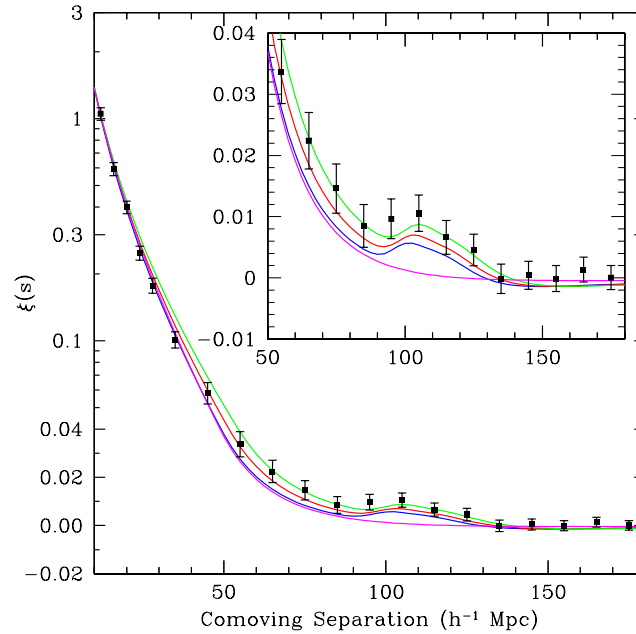


FIGURE 1.7 – Fonction de corrélation de l'échantillon LRG du SDSS DR3 avec barres d'erreur (noir). Les courbes en couleur montrent différents modèles, tous avec  $n_s = 0.98$  pour  $\Omega_m h^2 = 0.12$  (vert),  $\Omega_m h^2 = 0.13$  (rouge),  $\Omega_m h^2 = 0.14$  (bleu) avec  $\Omega_b h^2 = 0.024$ , et pour  $\Omega_m h^2 = 0.105$  avec zéro-baryon (rose). On voit que la détection du pic est statistiquement significative et que la position du pic dépend très faiblement de la valeur de  $\Omega_m h^2$  (propriété d'étalon standard du pic BAO). Figure extraite de Eisenstein et al. (2005).

Les BAOs fournissent une échelle caractéristique dans la distribution; ils permettent donc en principe de contraindre à la fois  $H(z)$  (observation dans la ligne de visée) et  $D_M(z)$  (observation transverse). Toutefois la faiblesse du pic BAO rend difficile de contraindre indépendamment ces 2 paramètres, et jusqu'à présent le pic BAO a été principalement étudié dans le monopole de la fonction de corrélation. Il en résulte que la position du pic BAO a seulement permis de contraindre l'échelle de dilatation  $D_V(z)$ , qui est une combinaison de  $H(z)$  et  $D_M(z)$  (voir Eisenstein et al. (2005); Padmanabhan and White (2008)) :

$$D_V(z) = \left[ D_M(z)^2 \frac{cz}{H(z)} \right]^{1/3} \quad (1.64)$$

On montrera plus concrètement dans la section 6.1 comment ce paramètre est contraint à travers l'estimation de la fonction de corrélation ou du spectre de puissance.

## 1.3 La formation des structures

### 1.3.1 Inflation

L'inflation est un modèle cosmologique, fortement lié au modèle du Big Bang, qui est apparu à la fin des années 70. La période d'inflation consiste en une phase d'expansion exponentielle de l'univers primordial par un facteur d'au moins  $10^{26}$ , et pouvant être égal à  $10^{1000000}$  dans certains modèles. Ce modèle offre une solution à deux problèmes :

- le problème de l'horizon, à savoir que l'univers observable est homogène et isotrope, ce qui semble incompatible avec le fait que différentes régions de l'univers observable n'aient jamais été en contact causal.

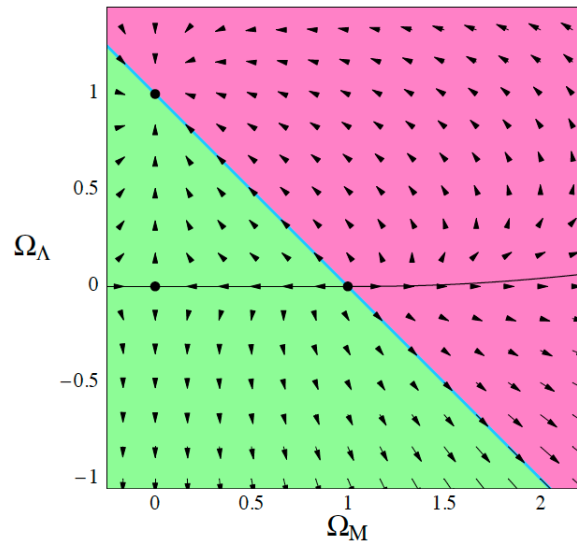


FIGURE 1.8 – Dynamique d'évolution pour  $\Omega_m$  et  $\Omega_\Lambda$ . Les flèches indiquent la direction d'évolution des paramètres dans un univers en expansion. La droite  $\Omega_m + \Omega_\Lambda = 1$  n'est pas stable et le seul point attracteur correspond à  $\Omega_m = 0$ ,  $\Omega_\Lambda = 1$  ce qui correspond à l'univers de Sitter. Figure extraite de [Carroll \(2001\)](#).

- le problème de la platitude, à savoir que si l'univers actuel est proche d'être plat ( $\Omega_k \approx 0$ ), cela veut dire qu'il était extraordinairement proche d'être plat dans le passé. En effet en combinant les équations (1.32) et (1.35) on obtient :

$$\Omega_k(z) = -k \frac{(1+z)^2}{H^2(z)} \quad (1.65)$$

D'autre part l'équation (1.42) montre que  $H^2(z) \propto (1+z)^4$  lorsque la radiation domine et  $H^2(z) \propto (1+z)^3$  lorsque la matière domine. Si bien que l'on a dans ces deux régimes  $\Omega_k(z) \propto 1/(1+z)^2$  et  $\Omega_k(z) \propto 1/(1+z)$ . Si l'univers est presque plat à *redshift*  $z = 0$ , alors il était extraordinairement proche d'être plat à grand *redshift*. Ceci est illustré sur la figure 1.8, où l'on voit que la droite  $\Omega_m + \Omega_\Lambda = 1$  sur laquelle notre univers semble se situer n'est pas stable.

Ces deux problèmes ont une origine commune : dans un univers en décélération un observateur découvre toujours des régions plus vastes au cours du temps (i.e. l'univers observable grandit), mais certains objets qui apparaissent en même temps dans l'univers observable n'ont pas eu le moment d'échanger de l'information. L'inflation résout ce problème en créant une très grande accélération de l'expansion aux temps primordiaux, ce qui consiste schématiquement à agrandir si fortement une petite région de l'univers, qu'elle devienne de la taille de l'univers observable.

L'inflation peut en outre être testée observationnellement, notamment avec le CMB. En particulier, une mesure précise des modes B de la polarisation du CMB permettrait de mettre en évidence la radiation gravitationnelle produite par l'inflation. Une des principale prédiction de l'inflation consiste en des fluctuations primordiales de densité gaussiennes très petites ( $\delta \ll 1$ ), avec un spectre de puissance généralement de la forme suivante (on définira précisément le spectre de puissance dans la section 2.1.2) :

$$P(k) \propto k^{n_s} \quad (1.66)$$

avec  $n_s$  l'indice spectral légèrement inférieur 1. Ces fluctuations sont supposées adiabatiques, i.e. qu'elles affectent tous les constituants cosmologiques de la même façon. Elles ont ensuite été amplifiées par effet gravitationnel pour finalement donner naissance aux structures que l'on connaît aujourd'hui.

Le cas  $n_s = 1$  est appelé spectre d'Harrison Zel'dovich et a la particularité d'être invariant d'échelle. Cette propriété signifie en fait que l'amplitude des fluctuations par unité de  $\ln k$  à différentes échelles

$k$  est la même lorsque celles-ci entrent dans l'horizon, de taille  $\sim 1/(aH)$ . L'amplitude des fluctuations par unité de  $\ln k$  est donnée par :

$$\Delta^2(k, t) = \frac{k^3 P(k, t)}{2\pi^2} \propto k^{n_s+3} \quad (1.67)$$

La condition d'invariance d'échelle est donc donnée par :

$$\Delta^2(k = aH, t) = \text{const} \quad (1.68)$$

On peut montrer que cette condition implique un spectre primordial de la forme  $P(k) \propto k$ .

### 1.3.2 Evolution linéaire du champ de matière

Si l'on note  $\rho(\mathbf{x}, t)$  la densité massique à la position  $\mathbf{x}$  et au temps  $t$ , et  $\delta(\mathbf{x}, t)$  la fluctuation de densité :

$$\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}} \quad (1.69)$$

avec  $\bar{\rho}$  la densité massique moyenne. Il faut noter que l'existence d'une densité moyenne  $\bar{\rho}$  n'est pas à priori évidente, mais c'est notamment une conséquence du principe cosmologique. Cette question sera discutée plus en détails dans la section 2.3.3.

La notation de coordonnée  $\mathbf{x}$  réfère aux coordonnées comobiles, qui sont reliées aux coordonnées eulériennes (physiques)  $\mathbf{r}$  :

$$\mathbf{r} = a(t)\mathbf{x} \quad (1.70)$$

Ainsi si une particule suit le flot de Hubble, sa coordonnée comobile  $\mathbf{x}$  reste inchangée au cours de l'évolution. On peut différencier l'équation précédente pour obtenir :

$$\frac{\partial \mathbf{r}}{\partial t} = \dot{a}(t)\mathbf{x} + a(t)\mathbf{u} = H(t)\mathbf{r} + a(t)\mathbf{u} \quad (1.71)$$

où l'on a introduit la vitesse particulière  $\mathbf{u}$  comme l'excès de vitesse par rapport au flot de Hubble  $H(t)\mathbf{r}$ .

Nous allons brièvement décrire l'évolution du champ de densité matière dans le cas linéaire, qui est valide pour  $\delta \ll 1$ , i.e. en dehors des petites échelles aux temps récents. Pour une revue complète dans le contexte de la théorie des perturbations, nous renvoyons le lecteur à [Bernardeau et al. \(2002\)](#). Pour la matière non relativiste, la pression peut être négligée, et les équation d'évolution sont données en fonction de  $\delta$ , du champ de vitesse particulière  $\mathbf{v} = a(t)\mathbf{u}$  (en coordonnées eulériennes) et du potentiel gravitationnel perturbé  $\Phi'$  :

$$\Phi' = \Phi - \bar{\Phi} \quad (1.72)$$

$$\Delta\Phi' = 4\pi G(\rho - \bar{\rho}) \quad (1.73)$$

On obtient alors (voir e.g. [Martínez and Saar \(2002\)](#)) les équations d'évolution :

– L'équation de continuité qui assure la conservation de la masse :

$$\dot{\delta} + \frac{1}{a}\nabla \cdot [(1 + \delta)\mathbf{v}] = 0 \quad (1.74)$$

– L'équation d'Euler pour le mouvement d'un fluide :

$$\dot{\mathbf{v}} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}[\mathbf{v} \cdot \nabla \mathbf{v} + \nabla \cdot \Phi] = 0 \quad (1.75)$$

– L'équation de Poisson reliant la distribution de matière au potentiel gravitationnel  $\Phi$  :

$$\Delta\Phi = 4\pi G\bar{\rho}a^2\delta \quad (1.76)$$

Il faut noter que ces équations (issues de la mécanique Newtonienne) ne sont pas valables pour des échelles proches du rayon de courbure de l'univers ou dans les champs extrêmes comme les trous noirs. En revanche, elles sont parfaitement applicables pour décrire l'évolution de la densité et des vitesses dans l'univers observable.

Dans le cas où  $\delta \ll 1$  et où les vitesses particulières sont faibles, i.e.  $|\mathbf{u}| \ll 1$ , on obtient l'évolution suivante des fluctuations :

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}\delta = 0 \quad (1.77)$$

Bien que la gravité soit une force non locale, elle n'intervient dans l'approximation linéaire qu'à travers le Laplacien du potentiel qui est local. La solution de cette équation différentielle du second ordre inclut deux modes :

$$\delta(\mathbf{x}, t) = A(\mathbf{x})D^+(t) + B(\mathbf{x})D^-(t) \quad (1.78)$$

où  $A(\mathbf{x})$ ,  $B(\mathbf{x})$  sont des fonctions quelconques dépendant du champ de densité initial et les solutions partielles  $D^+(t)$  et  $D^-(t)$  sont appelés les facteurs de croissance linéaires.  $D^+(t)$  correspond à une fonction croissante alors que  $D^-(t)$  correspond à une fonction décroissante, qui tend à devenir négligeable.

Si l'on considère le *redshift* comme variable au lieu du temps, les modes  $D^+(z)$  et  $D^-(z)$  peuvent s'exprimer analytiquement (voir par exemple [Dodelson \(2003\)](#)). Utilisons la notation habituelle pour la constante de Hubble mesurée par un observateur hypothétique à *redshift*  $z$  :

$$H(z) = H_0 E(z) \quad (1.79)$$

avec  $H_0$  la constante de Hubble à *redshift*  $z = 0$ . Les facteurs de croissance se réécrivent alors :

$$D^+(z) = E(z) \int_z^\infty \frac{(1+z')}{E^3(z')} dz' \quad (1.80)$$

$$D^-(z) = E(z) \quad (1.81)$$

Pour les modèles avec  $\Omega_\Lambda = 0$ , les facteurs de croissance peuvent être obtenus analytiquement. Pour  $\Omega_\Lambda > 0$ , ils doivent être calculés de façon numérique. On peut quand même obtenir une bonne approximation analytique dans ce cas (voir [Bernardeau et al. \(2002\)](#)) :

– ( $\Omega_m = 1, \Omega_\Lambda = 0$ )

$$D^+(z) = (1+z)^{-1} = a \quad (1.82)$$

$$D^-(z) = (1+z)^{3/2} = a^{-3/2} \quad (1.83)$$

– ( $\Omega_m < 1, \Omega_\Lambda = 0$ )

$$x = \frac{1}{\Omega_m(z)} - 1 \quad (1.84)$$

$$D^+(z) = 1 + \frac{3}{x} + \sqrt{\frac{1+x}{x^3}} \ln(\sqrt{1+x} + \sqrt{x}) \quad (1.85)$$

$$D^-(z) = \sqrt{\frac{1+x}{x^3}} \quad (1.86)$$

– ( $\Omega_m, \Omega_\Lambda$ )

$$D^+(z) \approx \frac{5}{2(1+z)} \frac{\Omega_m(z)}{\Omega_m^{4/7}(z) - \Omega_\Lambda(z) + (1 + \Omega_m(z)/2)(1 + \Omega_\Lambda(z)/70)} \quad (1.87)$$

$$D^-(z) = (1+z)H(z) \quad (1.88)$$

### Facteur de croissance sans dimension

En négligeant le facteur de  $D^-$  par rapport à  $D^+$ , on peut réécrire l'évolution du contraste  $\delta$  :

$$\dot{\delta} = \dot{a} \frac{d\delta}{da} = \dot{a} \frac{\delta}{a} \frac{d \ln \delta}{d \ln a} = H(a) \delta \frac{d \ln D^+}{d \ln a} \quad (1.89)$$

Le facteur de croissance sans dimension  $f(a)$  est défini comme :

$$f(a) = \frac{d \ln D^+}{d \ln a} \quad (1.90)$$

L'équation de continuité (1.74) avec  $\delta \ll 1$  donne alors :

$$\delta = -\frac{1}{H(a)f(a)} \nabla \cdot \mathbf{u} \quad (1.91)$$

en se servant du fait que  $\nabla \cdot \mathbf{u} = \frac{1}{a} \nabla \cdot \mathbf{v}$ .

Avec cette relation on peut donner une interprétation des modes croissants  $D^+(t)$  et décroissants  $D^-(t)$ . Les modes croissants correspondent au cas où le champ de densité et le champ de vitesse sont en phase, i.e. que la matière tombe dans les puits de potentiel. Au contraire, les modes décroissants correspondent au cas où le champ de vitesse tend à faire s'échapper la matière des puits de potentiel, ce qui explique l'atténuation de ces modes.

La fonction  $f(a)$  peut être bien approximée à quelques pour cents pour  $z = 0$  et  $\Omega_\Lambda = 0$  :

$$f(\Omega_m, z = 0) \approx \Omega_m^{0.6} \quad (1.92)$$

Lahav et al. (1991) a montré que dans le cas général, la fonction  $f$  ne dépend quasiment pas de  $\Omega_\Lambda$  :

$$f(\Omega_m, \Omega_\Lambda, z = 0) \approx \Omega_m^{0.6} + \frac{\Omega_\Lambda}{70} \left(1 + \frac{\Omega_m}{2}\right) \quad (1.93)$$

Cette étude a aussi montré que l'approximation (1.92) peut être utilisée à toutes les époques :

$$f(\Omega_m, z) \approx \Omega_m(z)^{0.6} \quad (1.94)$$

Notons que la relation (1.91) peut être utilisée pour déduire les vitesses particulières à partir du champ de densité.

### 1.3.3 Fonction de transfert

Lors de l'évolution linéaire, les différents modes de Fourier  $\delta_k$  évoluent indépendamment. On peut alors écrire le spectre de puissance linéaire :

$$P_L(k) = T_i^2(k) P(k, t_i) \quad (1.95)$$

où  $t_i$  est un temps primordial (e.g. juste après la période d'inflation), i.e. que le spectre primordial vérifie  $P(k, t_i) = Ak^{n_s}$ .  $T_i^2(k)$  est appelé la fonction de transfert. Elle inclut l'évolution linéaire du champ de matière décrit dans la section 1.3.2 mais également certains effets en dessous de l'horizon dans la période où la radiation domine. Ce sont ces effets qui donnent une dépendance en  $k$  à  $T_i^2(k)$ . L'effet principal est que, durant la période où la radiation domine, les modes plus petits que l'horizon ne peuvent croître à cause de la pression des photons.

On a vu dans la section 1.3.2 que le facteur de croissance vérifie  $D^+ \propto a$  quand la matière domine. On peut aussi montrer que le facteur de croissance vérifie  $D^+ \propto a^2$  quand la radiation domine. Pour récapituler, les fluctuations croissent proportionnellement à  $a^2$  durant la période où la radiation domine, tant que ces fluctuations sont au delà de l'horizon. Si une fluctuation arrive sous l'horizon durant cette période alors elle cesse de croître tant que la radiation domine. Finalement toutes les fluctuations croissent proportionnellement à  $a$  durant la période où la matière domine. Ceci est montré sur la figure 1.9.

La transition entre la période où la radiation domine et la période où la matière domine arrive au *redshift*  $1 + z_{eq} = \Omega_m/\Omega_r$ , et l'échelle de transition  $k_{eq}$  est donnée par  $k_{eq} = a_{eq}H_{eq}$  (l'horizon au moment de la transition étant de taille  $1/(a_{eq}H_{eq})$ ). On voit sur la figure 1.9 que les fluctuations telles que  $k < k_{eq}$  suivent simplement l'évolution du facteur de croissance et donc :

$$T(k) = C \text{ si } k < k_{eq} \quad (1.96)$$



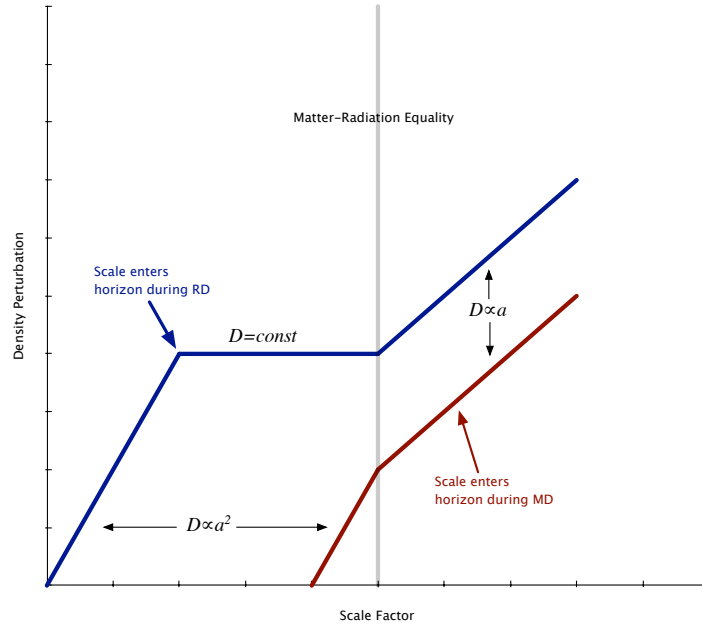


FIGURE 1.9 – Croissance de fluctuations à différentes échelles : une échelle entrant sous l’horizon pendant l’époque où la radiation domine (bleu) et une échelle entrant sous l’horizon pendant l’époque où la matière domine. Les fluctuations croissent comme le facteur de croissance  $D$ , hormis celles qui entrent sous l’horizon pendant l’époque où la radiation domine (à ce moment elles cessent de croître jusqu’en fin de domination de la radiation). Figure réalisée par Andrew Jaffe.

En revanche, pour les échelles plus petites on peut écrire :

$$\delta(t) = \frac{D(t_{en})}{D(t_i)} \frac{D(t)}{D(t_{eq})} \delta(t_i) \quad (1.97)$$

$$= \frac{a_{en}^2}{a_i^2} \frac{a_0}{a_{eq}} \delta(t_i) \quad (1.98)$$

où  $t_{en}$  et  $a_{en}$  désignent respectivement le temps et le facteur d’échelle au moment où le mode entre sous l’horizon. Pour un tel mode  $k$ , on a  $k = (aH)_{en} \propto 1/a_{en}$  car durant la période où la matière domine on a  $a \propto t^{1/2}$ , i.e.  $H \propto 1/a^2$ . On obtient donc :

$$\delta(t) \propto k^{-2} \delta(t_i) \quad (1.99)$$

On obtient donc finalement :

$$T(k) = const \times \begin{cases} 1 & \text{si } k < k_{eq} \\ k^{-4} & \text{si } k > k_{eq} \end{cases} \quad (1.100)$$

Un autre effet plus faible intervenant dans la fonction de transfert est dû aux oscillations baryoniques acoustiques (BAOs). Elles apparaissent comme une série d’oscillations dans la fonction de transfert  $T(k)$ , dues à la propagation d’ondes acoustiques dans le plasma baryon-photon avant la recombinaison. Elles impriment une échelle caractéristique dans le spectre de puissance, correspondant à l’horizon sonique au moment de la recombinaison, et qui peut être utilisé pour étudier la géométrie de l’univers. On donnera plus de détails sur la physique des Oscillations Baryoniques Acoustiques dans la section 1.4.

Pour calculer la fonction de transfert de façon précise on doit résoudre numériquement les équations de Boltzmann pour les différentes espèces. Différents codes existent pour résoudre ces équations, les plus utilisés étant CAMB (Lewis et al. 2000) et CMBFAST (Seljak and Zaldarriaga 1996). Il existe aussi différentes approximations analytiques pour  $T(k)$ , qui sont assez précises dans beaucoup d’applications (e.g. Bond and Efstathiou (1984); Eisenstein and Hu (1998, 1999)). Ces approximations ne doivent

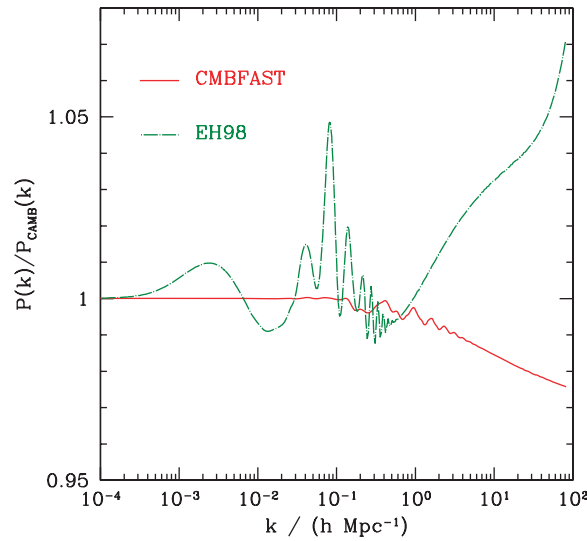


FIGURE 1.10 – Spectre de puissance linéaire  $P_L(k)$  pour un modèle cosmologique plat avec  $\Omega_m = 0.237$ ,  $\Omega_b = 0.041$  et  $h = 0.735$ , obtenus avec la formule de Eisenstein and Hu (1998) (tirets-pointillés verts) et avec le logiciel CMBFAST, comparés à spectre linéaire de référence obtenu avec CAMB. Figure extraite de Sánchez et al. (2008).

cependant pas être utilisées pour des utilisations demandant une grande précision, e.g. pour obtenir des contraintes cosmologiques. En effet Sánchez et al. (2008) montre que l'approximation analytique de Eisenstein and Hu (1999) peut donner une imprécision de 5% sur le spectre linéaire  $P_L(k)$  (voir figure 1.10).

On montre sur la figure 1.11 la fonction de transfert  $T(k)$  et le spectre  $P_L(k)$  de l'équation (1.95), obtenus avec CAMB pour des valeurs standards des paramètres cosmologiques. La figure montre les différents effets décrits précédemment :  $T(k) \propto 1$  aux faibles  $k$ ,  $T(k) \propto k^{-4}$  aux grands  $k$  et les BAOs créant de faibles oscillations dans la fonction de transfert.

### 1.3.4 Evolution non linéaire du champ de matière

#### Approximation de Zel'dovich

Différentes méthodes existent pour calculer l'évolution non linéaire du champ de densité. Une méthode très utilisée est l'approximation de Zel'dovich (Zel'dovich 1970) qui est notamment utilisé dans beaucoup de simulations  $N$ -corps pour le déplacement des particules à l'instant initial de la simulation. Si l'on suppose la matière non collisionnelle, la position Eulérienne comobile d'une particule à l'instant  $t$  est donnée en fonction du champ initial Lagrangien  $\mathbf{q}$  :

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + D(t)\nabla_{\mathbf{q}}\Psi(\mathbf{q}) \quad (1.101)$$

où  $D(t)$  est le facteur de croissance linéaire et  $\Psi(\mathbf{q})$  est le potentiel des vitesses initiales du champ.

La conservation de la masse implique  $\rho(\mathbf{x}, t) d^3\mathbf{x} = \bar{\rho} d^3\mathbf{q}$  avec  $\bar{\rho}$  la densité moyenne. On obtient alors la densité comobile :

$$\rho = \bar{\rho} \left| \frac{\partial \mathbf{x}}{\partial \mathbf{q}} \right|^{-1} = \bar{\rho} |\delta_{ik} + D(t)\Psi_{,ik}|^{-1} \quad (1.102)$$

où  $\delta_{ik}$  est le symbole de Kronecker et  $\Psi_{,ik}$  est la matrice jacobienne de  $\Psi$ . Si l'on note  $-\alpha_1$ ,  $-\alpha_2$  et  $-\alpha_3$  les valeurs propres de  $\Psi_{,ik}$  on obtient :

$$\rho = \frac{\bar{\rho}}{(1 - D(t)\alpha_1)(1 - D(t)\alpha_2)(1 - D(t)\alpha_3)} \quad (1.103)$$

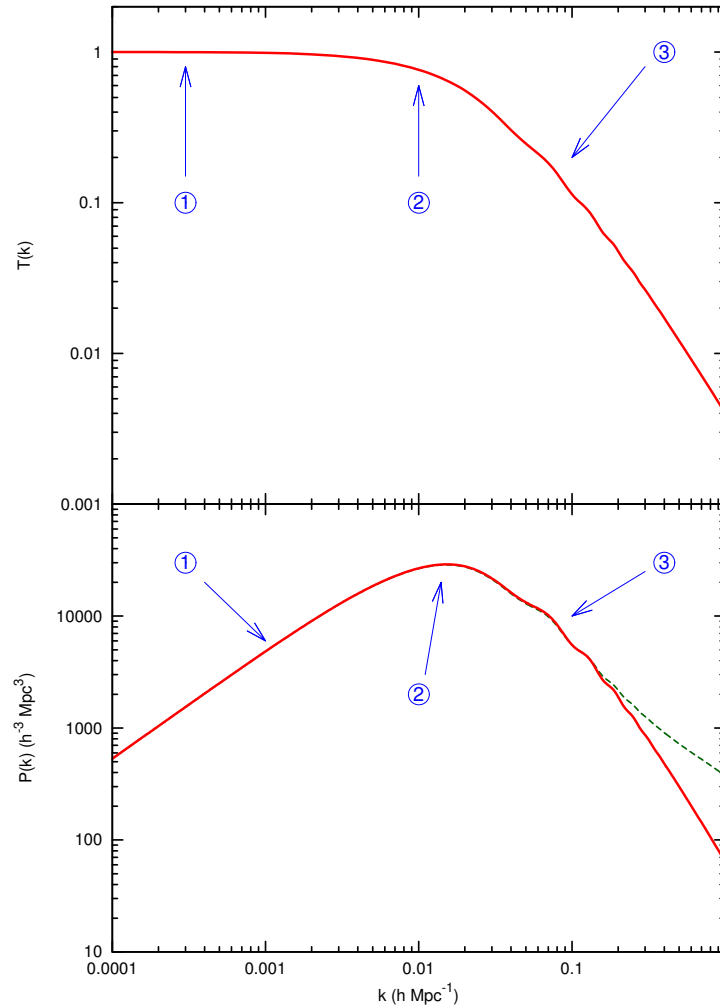


FIGURE 1.11 – Fonction de transfert  $T(k)$  (en haut) et spectre de puissance linéaire  $P_L(k)$  (en bas) à *redshift*  $z = 0$  obtenu avec CAMB et les paramètres cosmologiques de Komatsu et al. (2011). La ligne en tirets verts correspond au spectre non linéaire obtenu avec l'approximation HALOFIT (Smith et al. (2003), voir section 1.3.4). On peut voir les effets de la fonction de transfert aux différentes échelles. (1) : les grandes échelles (faibles  $k$ ) sont restées en dehors de l'horizon pendant la période où la radiation domine,  $T(k) = 1$ , et le spectre de puissance est égal spectre primordial  $P_L(k) \propto k^{n_s}$  à un facteur de normalisation près. (2) : A l'échelle de l'horizon à l'égalité matière-radiation, on observe un changement de comportement,  $T(k) < 1$ , et le spectre de puissance atteint son amplitude maximum. (3) : Aux petites échelles (grands  $k$ ), les fluctuations sont fortement atténuées après leur passage sous l'horizon,  $T(k) \propto k^{-4}$ , et l'on observe aussi l'effet des BAOs comme une série d'oscillations à la fois dans  $T(k)$  et dans  $P_L(k)$ . Figure réalisée par Pablo Arnalte-Mur.

D'où l'expression de la fluctuation :

$$1 + \delta = \frac{1}{(1 - D(t)\alpha_1)(1 - D(t)\alpha_2)(1 - D(t)\alpha_3)} \quad (1.104)$$

Si l'on a ordonné les valeurs propres telles que  $\alpha_1 > \alpha_2 > \alpha_3$ , on obtient alors un effondrement de la matière suivant la direction correspondant à  $\alpha_1$  et la création de feuillet.

### Correction invariante d'échelle + dégradation non linéaire des BAOs

Une méthode plus récente pour modéliser les non-linéarités est basée sur le formalisme du *halo model* (Seljak 2000; Peacock and Smith 2000; Cooray and Sheth 2002). Dans ce modèle, la distribution de matière est décomposée en populations de halos virialisés, i.e. en équilibre gravitationnel. Cette approche a été utilisée dans Smith et al. (2003) avec un grand nombre de simulations  $N$ -corps pour calculer les corrections non linéaires à appliquer à un spectre linéaire  $P_L(k)$ . Les résultats de cette étude constituent l'approximation HALOFIT (correction donnée en tirets verts sur la figure 1.11). Toutefois cette étude n'a été faite que pour des spectres  $P_L(k)$  invariants d'échelle, et l'approximation HALOFIT ne décrit pas bien l'évolution non linéaire de la signature BAO dans le spectre de puissance.

Eisenstein et al. (2007b) a montré que la dégradation non linéaire de la signature BAO était dominée par les mouvements relatifs de paires de particules. Cette étude a plus précisément montré que cela pouvait être très bien approximé par un lissage Gaussien de la fonction de corrélation (i.e. la multiplication du spectre de puissance par une fonction Gaussienne centrée) au niveau du pic BAO, à la fois en espace réel et en espace *redshift* (voir section 2.2.3).

Il en résulte l'ansatz simple estimer le spectre de puissance après la dégradation non linéaire des BAOs (*dewiggled spectrum*) :

$$P_{dw}(k) = [P_L(k) - P_{nw,L}(k)]G(k) + P_{nw,L}(k) \quad (1.105)$$

$$= P_L(k)G(k) + P_{nw,L}(k)[1 - G(k)] \quad (1.106)$$

où  $G(k) = \exp\left[-\frac{(ka)^2}{2}\right] = \exp\left[-\frac{k^2}{2k_*^2}\right]$  est un noyau Gaussien avec une taille de lissage  $a$ , et  $P_{nw,L}(k)$  un spectre *no wiggles* lisse, avec la même forme que  $P_L(k)$ , mais sans la signature BAO (voir Eisenstein and Hu (1998)). Cette ansatz permet de conserver la puissance aux petites échelles et de n'appliquer la dégradation non linéaire que sur la signature BAO.

Finalement on peut rajouter les corrections non linéaires indépendantes du pic BAO, qui sont importantes aux petites échelles. Pour ce faire on applique la même correction à  $P_{dw}(k)$  que celle donnée par HALOFIT pour le spectre lisse  $P_{nw}(k)$  :

$$P_{dw,NL}(k) = \left(\frac{P_{nw,NL}(k)}{P_{nw,L}(k)}\right) P_{dw}(k) \quad (1.107)$$

Ce type d'approche a été souvent utilisé dans la littérature pour obtenir un modèle non linéaire du champ de matière en espace *redshift* (e.g. dans Eisenstein et al. (2005); Tegmark et al. (2006); Reid et al. (2010); Blake et al. (2011a,b)). Sánchez et al. (2008) a montré que cette procédure donne un excellent accord avec les résultats de larges simulations  $N$ -corps (voir figure 1.12). Cependant le modèle n'est pas parfait et il entraîne un très léger biais dans les contraintes de paramètres cosmologiques, pouvant affecter l'analyse pour les relevés futurs très grands comme EUCLID (Robberto et al. 2007). En revanche, Sánchez et al. (2008) a montré qu'un modèle simple basé sur la théorie des perturbations renormalisée n'entraîne pas ce faible biais.

### Théorie des perturbations renormalisée

De nouveaux développements en théories des perturbations, comme la Théorie des Perturbations Renormalisée (*Renormalized Perturbation Theory*, RPT, Crocce and Scoccimarro (2006)), ont apporté des progrès substantiels dans la compréhension théorique des effets non linéaires, qui peuvent maintenant être modélisés de façon précise (Crocce and Scoccimarro 2006; Matsubara 2008a,b; Taruya et al. 2009). En se basant sur RPT, Crocce and Scoccimarro (2008) propose un ansatz simple pour décrire la forme globale de la corrélation :

$$\xi_{NL}(r) = \xi_L(r) \otimes \tilde{G}(r) + A_{mc} \xi'_L \xi_L^{(1)} \quad (1.108)$$

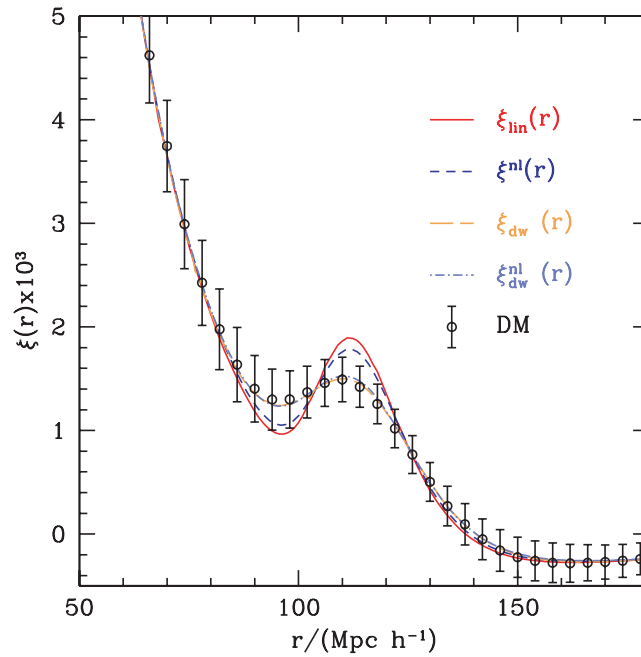


FIGURE 1.12 – Comparaison de la fonction de corrélation obtenue sur des simulations  $N$ -corps de volume total  $120h^{-3}\text{Gpc}^3$  (plus de 4 fois le volume de Hubble), pour un modèle  $\Lambda\text{CDM}$  plat avec  $\Omega_m = 0.237$ ,  $\Omega_b = 0.041$ ,  $n_s = 0.954$ ,  $\sigma_8 = 0.77$  et  $h = 0.735$ , en espace réel à *redshift*  $z = 0$  (cercles noirs) avec : (i) la fonction de corrélation linéaire  $\xi_L(r)$  (trait rouge continu), (ii) la fonction de corrélation non linéaire  $\xi_{NL}(r)$  après la correction HALOFIT de  $\xi_L(r)$  (tirets bleus), (iii) la fonction de corrélation *dewiggled*  $\xi_{dw}(r)$  donnée par la formule (1.106) (tirets oranges), (iv) la fonction de corrélation *dewiggled*  $\xi_{dw,NL}(r)$  après la correction HALOFIT de la formule (1.107) (tirets-pointillés bleus). Les barres d'erreur donnent l'écart-type de l'estimateur sur des volumes de  $2.41h^{-1}\text{Gpc}^3$ . Figure extraite de [Sánchez et al. \(2008\)](#).

où l'on retrouve la dégradation non linéaire par une convolution (symbole  $\otimes$ ) avec une Gaussienne  $\tilde{G}(r) = a \exp\left[-\frac{r^2}{2a^2}\right]$ . Le second terme  $\xi_{mc} = A_{mc}\xi'_L\xi_L^{(1)}$  est un terme de couplage de modes (*mode coupling*) avec  $\xi'_L$  la dérivée de  $\xi_L$ , et  $\xi_L^{(1)}$  définie comme :

$$\xi_L^{(1)} = \hat{\mathbf{r}} \cdot \nabla^{-1}\xi_L(r) = 4\pi \int P_L(k)j_1(kr)kdk \quad (1.109)$$

où  $j_1$  est la fonction de Bessel sphérique d'ordre 1. Le facteur d'amplitude du couplage de modes  $A_{mc}$  est en général laissé libre dans le modèle, tout comme l'échelle de lissage  $a = 1/k_*$ . [Sánchez et al. \(2008\)](#) a montré que ce modèle est en excellent accord avec les résultats de simulations  $N$ -corps et qu'il fournit des contraintes non biaisés des paramètres cosmologiques. Ce modèle a été utilisé à de nombreuses reprises pour modéliser la fonction de corrélation (e.g. dans [Sánchez et al. \(2009\)](#); [Beutler et al. \(2011\)](#); [Blake et al. \(2011b\)](#)).

## 1.4 Les Oscillations Baryoniques Acoustiques

Nous allons décrire ici le phénomène des Oscillations Baryoniques Acoustiques dues à la propagation d'ondes acoustiques dans le plasma baryon-photon avant la recombinaison. Ces ondes impriment une échelle caractéristique dans la distribution de matière (et donc dans la distribution des galaxies). Elles peuvent donc être utilisées comme un étalon standard pour étudier la géométrie de l'univers (voir section 6.1). La détection de ces ondes à l'échelle et l'amplitude attendues dans les grandes structures de l'univers fournit aussi une très forte confirmation du modèle standard de la cosmologie (voir chapitre 5).

Pour expliquer la physique des BAOs, on suit la présentation de [Eisenstein et al. \(2007b\)](#) et l'on raisonne dans l'espace de configuration<sup>4</sup> (et non en espace de Fourier). Bien que la description soit totalement équivalente, les processus physiques sont plus intuitifs.

Avant la recombinaison, l'énergie des photons était assez élevée pour éviter la formation d'atomes d'hydrogène. Cela entraînait un couplage entre baryons et photons par diffusion Compton ainsi que des interactions électromagnétiques entre protons et électrons, formant un plasma. Dans ce fluide deux phénomènes agissent de façon opposée : l'interaction gravitationnelle tend à compresser le plasma au niveau des surdensités, alors que la pression des photons tend à diluer toute surdensité. La combinaison de ces deux phénomènes entraîne la formation d'ondes acoustiques. Ce phénomène acoustique est stoppé au moment de la recombinaison, i.e. du découplage entre baryons et photons. Durant toute cette période avant la recombinaison, la matière noire a évolué de façon indépendante, interagissant seulement gravitationnellement avec les autres espèces.

Afin de comprendre quelle est l'empreinte de ces oscillations dans la distribution de matière, on peut étudier l'évolution de la densité des différentes espèces autour d'une surdensité initiale ponctuelle. On représente différentes étapes de cette évolution (i.e. différents *redshift*) sur la figure 1.13 extraite de [Eisenstein et al. \(2007b\)](#). A chaque *redshift*, la figure montre le profil de densité pour chaque espèce en fonction de la distance comobile  $r$  par rapport à la position de la surdensité initiale. Le profil de densité est multiplié par  $r^2$  pour éviter l'effet de dilution volumique (i.e. que la figure montre en fait la distribution de masse  $r^2\rho(r)$  autour de la surdensité initiale). On suppose que la fluctuation primordiale est adiabatique, i.e. qu'elle est initialement la même pour toutes les espèces.

Nous allons décrire brièvement les différents phénomènes montrés sur la figure 1.13. La surdensité initiale implique une surdensité de la radiation, i.e. une pression plus forte que la moyenne dans cette région. Pour compenser cette surpression, une onde acoustique sphérique est générée vers l'extérieur. Cette onde acoustique entraîne également les baryons, qui sont couplés aux photons. Cependant les baryons subissent l'effet gravitationnel de la matière noire au centre de la perturbation, causant un étirement de l'onde acoustique. Quant à la matière noire, elle subit l'influence gravitationnelle des baryons qui s'échappent de la perturbation, ce qui résulte aussi en un léger déplacement et un étirement de son profil initial. On peut voir ces différents phénomènes sur les figures 1.13(a) et 1.13(b).

4. On utilise ici la dénomination espace de configuration et non espace réel, pour ne pas confondre avec la dénomination espace réel que l'on utilise pour différencier de l'espace *redshift* (voir section 2.2.3).

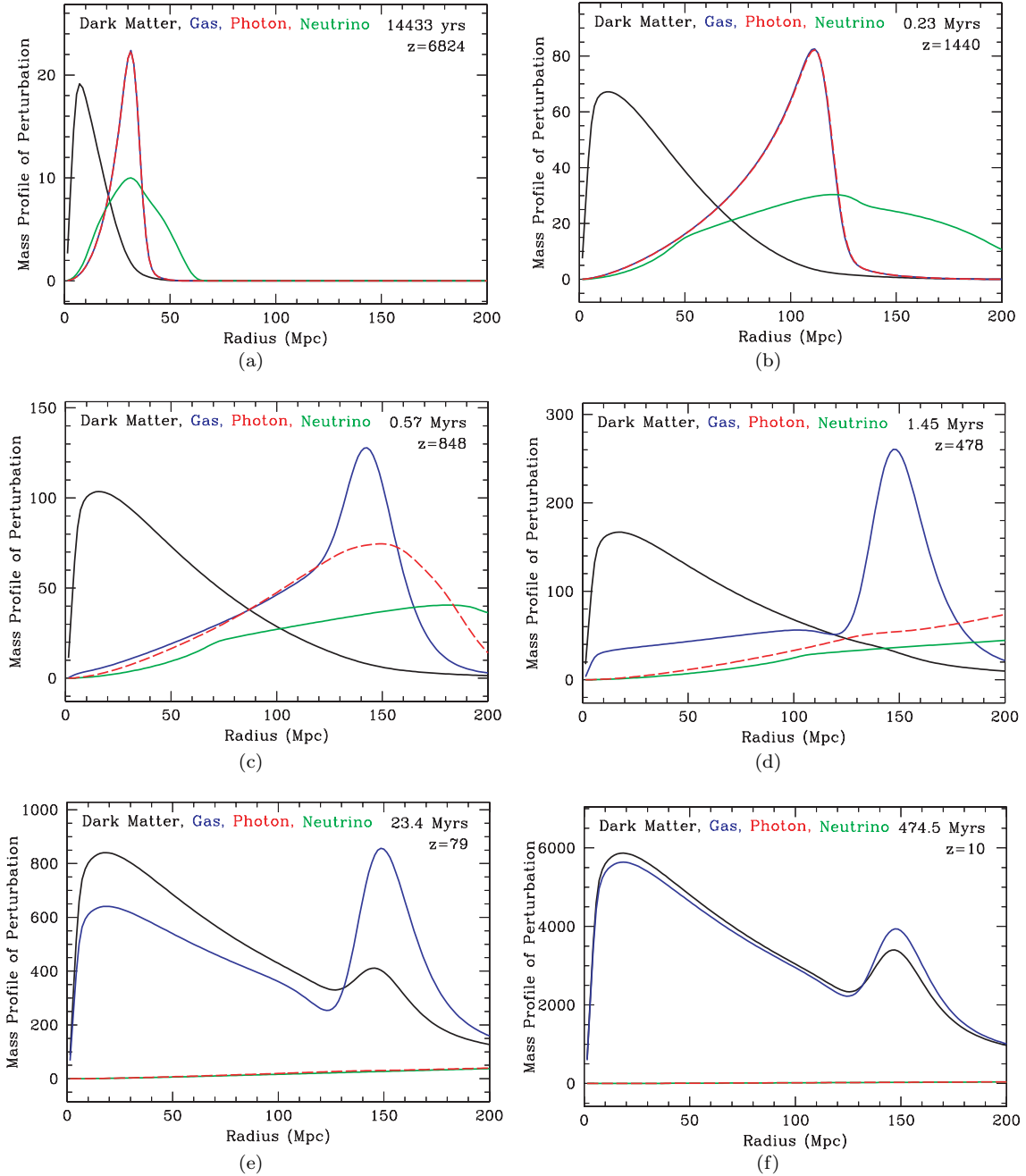


FIGURE 1.13 – Evolution de la densité de masse  $r^2\rho(r)$  autour d’une surdensité adiabatique initiale. Chaque figure montre le profil de masse à un *redshift* différent, donné dans chaque coin en haut à droite, pour 4 espèces différentes : matière noire (noir), baryons (bleu), photons (rouge), neutrinos (vert). Les densités  $\rho(r)$  sont calculées par transformée de Fourier de la fonction de transfert, elle même calculée avec le programme CMBFAST. Figure extraite de [Eisenstein et al. \(2007b\)](#).

L'onde sonore se propage à une vitesse  $c_s$  donnée par (Eisenstein et al. 2007b) :

$$c_s = \frac{c}{\sqrt{3(1+R)}} \quad (1.110)$$

avec  $R = \frac{3\rho_b}{4\rho_\gamma}$ . Cette onde sonore continue à se propager jusqu'à la recombinaison, à *redshift*  $z_{rec} \approx 1,100$  (correspondant à  $t_{rec} \approx 380,000$  ans), où la température est devenue suffisamment faible pour permettre la formation d'atomes d'hydrogène. À partir de la recombinaison, les baryons sont découplés des photons et ne sont donc plus affectés par la pression de radiation. Ainsi, comme on peut le voir sur les figures 1.13(c) et 1.13(d), la propagation sonore des baryons est stoppée. Ils se retrouvent concentrés dans une enveloppe sphérique autour de la surdensité initiale. Quant aux photons, ils se propagent désormais librement dans toutes les directions, ce qui a tendance à lisser les inhomogénéités de la distribution. Les photons qui ont pu se propager librement à partir de cette époque peuvent être observés aujourd'hui puisqu'ils constituent le CMB. En regardant le CMB dans différentes directions, on peut observer les inhomogénéités de la distribution de photons au moment de la recombinaison. Les ondes acoustiques peuvent donc être clairement vues comme une série de pics dans le spectre de puissance du CMB.

Après la recombinaison, la matière évolue simplement gravitationnellement à partir de sa distribution à la recombinaison comme on peut le voir sur les figures 1.13(e) et 1.13(f). Cette distribution est composée de deux surdensités : une surdensité de matière noire proche de la surdensité initiale et une surdensité sphérique de baryons à une distance  $r_s \approx 150$  Mpc correspondant à l'horizon sonore. L'évolution gravitationnelle entraîne une accréation de la matière au niveau de ces deux surdensités, ainsi qu'une homogénéisation entre les distributions de matière noire et de matière baryonique. À faible *redshift*, il ne reste qu'une forte surdensité proche de la surdensité initiale et une faible surdensité sphérique à la distance  $r_s$ , la distance parcourue par l'onde sonore avant la recombinaison :

$$r_s = \int_0^{t_{rec}} c_s(t) \frac{dt}{a(t)} \quad (1.111)$$

La taille de l'horizon sonore dépend de façon compliquée des paramètres cosmologiques, mais une approximation utile est donnée dans Eisenstein and Hu (1998) :

$$r_s \approx \frac{44.5 \ln\left(\frac{9.83}{\Omega_m h^2}\right)}{\sqrt{1 + 10(\Omega_b h^2)^{3/4}}} \text{Mpc} \quad (1.112)$$

Avec des valeurs raisonnables pour les paramètres cosmologiques,  $\Omega_m = 0.27$ ,  $\Omega_b = 0.045$  et  $h = 0.7$ , on obtient  $r_s \approx 152.9$  Mpc. Avec l'unité  $h^{-1}$  habituellement utilisée dans l'étude des grandes structures, on obtient  $r_s \approx 107.0h^{-1}$  Mpc.

On peut voir sur la figure 1.13(f) que la surdensité sphérique à distance  $r_s$  est bien plus faible que la surdensité centrale. Ceci est dû à la prépondérance de la matière noire par rapport à la matière baryonique. La différence d'amplitude entre les deux surdensités est encore plus importante quand on regarde le profil de densité  $\rho(r)$  au lieu du profil massique total  $r^2\rho(r)$ . Le ratio des deux amplitudes dépend principalement du ratio entre la densité baryonique et la densité de matière noire, et de l'ordre de  $\approx 0.01$  pour le profil  $\rho(r)$ .

On peut considérer l'univers primordial comme une superposition de telles densités ponctuelles adiabatiques (avec des amplitudes variables). Dans l'approximation linéaire, la distribution actuelle correspond donc à la superposition des solutions montrées sur la figure 1.13 à différentes positions et différentes amplitudes.

Le pic BAO peut être clairement identifié dans la fonction de corrélation de la matière  $\xi(r)$  comme un faible pic à une échelle  $r \approx r_s$ . Cela vient de la corrélation entre la surdensité centrale et la surdensité sphérique à distance  $\approx r_s$ . Le pic localisé dans la fonction de corrélation se traduit par une série de pics dans le spectre de puissance  $P(k)$ . L'amplitude de la corrélation au niveau du pic BAO est bien plus faible dans le cas du champ de matière que dans le cas de CMB, car seulement une faible partie de la matière totale est impliquée dans le phénomène acoustique, alors que la totalité des photons est impliquée dans le cas du CMB.



La largeur du pic BAO est due à différents phénomènes, notamment l'étirement de l'onde acoustique pendant sa propagation, le *Silk damping* au moment de la recombinaison et la corrélation des perturbations initiales. L'évolution non linéaire cause également un lissage du pic BAO, principalement à cause des mouvements relatifs de paires de particules séparées d'une distance  $\approx r_s$  (voir section 1.3.4).

# Chapitre 2

## Analyse statistique de la corrélation des galaxies

### Sommaire

---

<b>2.1 Champs aléatoires en cosmologie</b>	<b>31</b>
2.1.1 Champ statistiquement homogène et isotrope	31
2.1.2 Fonction de corrélation à 2 points et spectre de puissance	32
2.1.3 Cas particulier du champ Gaussien	33
<b>2.2 La corrélation des galaxies</b>	<b>35</b>
2.2.1 Le champ de galaxies comme un processus de Cox	35
2.2.2 Biais des galaxies par rapport à la matière	36
2.2.3 Distorsions de <i>redshift</i>	37
2.2.4 Estimation de la fonction de corrélation	39
2.2.5 Modélisation de la fonction de corrélation	41
<b>2.3 Multifractalité de la distribution des galaxies</b>	<b>44</b>
2.3.1 Ensembles fractals	44
2.3.2 Ensembles multifractals	45
2.3.3 Fractalité à petite échelle et homogénéité à grande échelle de la distribution des galaxies	47

---

## 2.1 Champs aléatoires en cosmologie

### 2.1.1 Champ statistiquement homogène et isotrope

Un champ aléatoire est défini comme statistiquement homogène<sup>1</sup> (ou plus simplement stationnaire) lorsque ses densités de probabilité jointes  $p(\delta_1, \delta_2, \dots)$  (ou de façon équivalente ses moments) sont invariantes translation des coordonnées de l'espace  $\mathbf{x}_1, \mathbf{x}_2, \dots$  (où l'on a défini  $\delta_i = \delta(\mathbf{x}_i)$ ). Ainsi les différentes probabilités ne dépendent que des positions relatives des points les uns par rapport aux autres.

Un champ est dit statistiquement isotrope lorsque les quantités  $p(\delta_1, \delta_2, \dots)$  sont invariantes par rotation de l'espace. On supposera que les champs cosmologiques sont stationnaires et isotropes, comme exprimé par le principe cosmologie (voir section 1.1.1). Toutefois ceci n'est vrai qu'en espace réel et n'est plus strictement vrai en espace *redshift*, qui est utilisé pour l'analyse des relevés de galaxies.

---

1. à ne pas confondre avec un champ homogène qui a la même valeur partout dans l'espace

### 2.1.2 Fonction de corrélation à 2 points et spectre de puissance

En cosmologie la fonction de corrélation à 2 points est définie comme la covariance du champ de fluctuations  $\delta$ . On rappelle que le champ de fluctuations  $\delta \in [-1, \infty]$  est défini à partir du champ de densité  $\rho$  par la relation :

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \quad (2.1)$$

Dans le cas d'un champ stationnaire, la fonction de corrélation est définie pour tout vecteur de séparation  $\mathbf{r}$  :

$$\xi(\mathbf{r}) = \mathbb{E}[\delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r})] \quad (2.2)$$

Ceci correspond effectivement à la covariance du champ  $\delta$  entre 2 positions séparées par un vecteur  $\mathbf{r}$ . En effet on a  $\mathbb{E}[\delta(\mathbf{x})] = 0$ , i.e. que la covariance entre  $\delta(\mathbf{x})$  et  $\delta(\mathbf{x} + \mathbf{r})$  s'écrit :

$$\text{Cov}[\delta(\mathbf{x}), \delta(\mathbf{x} + \mathbf{r})] = \mathbb{E}[\delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r})] \quad (2.3)$$

Dans le cas où le champ est statistiquement isotrope, la fonction de corrélation ne dépend que de la norme de la séparation  $r = \|\mathbf{r}\|$ . Cette hypothèse n'est pas exactement vérifiée lorsque l'on considère la distribution des galaxies en espace *redshift*. Cependant si l'on se place dans l'approximation 'plan-parallèle' où la séparation entre deux points est faible par rapport à la distance entre l'observateur et ces points, le champ peut être approximé comme stationnaire avec simplement une anisotropie entre la direction de ligne de visée et la direction transverse. La fonction de corrélation bidimensionnelle  $\xi(\pi, \sigma)$  est alors bien définie, où  $\pi$  représente la séparation dans la ligne de visée et  $\sigma$  la séparation transverse (voir section 2.2.3)

De plus, nous allons voir dans la section 2.2.3 que le monopole de la fonction de corrélation bidimensionnelle (ou du spectre de puissance bidimensionnel) en espace *redshift* est liée à la fonction de corrélation en espace réel par un simple effet multiplicatif. On peut donc simplement étudier les prédictions des modèles cosmologiques en espace réel, et les comparer au monopole de la fonction de corrélation en espace *redshift*.

Continuons donc la présentation dans le cas de champs stationnaires et statistiquement isotropes. On peut de façon équivalente considérer le champ de fluctuations dans l'espace de Fourier  $\tilde{\delta}$  :

$$\tilde{\delta}(\mathbf{k}) = \int \delta(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}) d^3\mathbf{x} \quad (2.4)$$

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int \tilde{\delta}(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}) d^3\mathbf{k} \quad (2.5)$$

Les quantités  $\tilde{\delta}(\mathbf{k})$  sont les coefficients de Fourier de  $\delta(\mathbf{x})$ , et sont des nombres complexes. Comme  $\delta(\mathbf{x})$  est réel, la formule (2.5) implique que :

$$\tilde{\delta}(\mathbf{k}) = \tilde{\delta}^*(-\mathbf{k}) \quad (2.6)$$

Tout comme pour  $\delta(\mathbf{x})$ , le champ est déterminé entièrement par les propriétés statistiques de  $\tilde{\delta}(\mathbf{k})$ . Calculons la covariance du champ dans l'espace de Fourier :

$$\mathbb{E}[\tilde{\delta}(\mathbf{k})\tilde{\delta}(\mathbf{k}')] = \iint \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle \exp[-i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{x} - i\mathbf{k}' \cdot \mathbf{r}] d^3\mathbf{x} d^3\mathbf{r} \quad (2.7)$$

$$= \iint \xi(r) \exp[-i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{x} - i\mathbf{k}' \cdot \mathbf{r}] d^3\mathbf{x} d^3\mathbf{r} \quad (2.8)$$

$$= \delta_D(\mathbf{k} + \mathbf{k}') \int \xi(r) \exp(i\mathbf{k} \cdot \mathbf{r}) d^3\mathbf{r} \quad (2.9)$$

$$= (2\pi)^3 \delta_D(\mathbf{k} + \mathbf{k}') P(k) \quad (2.10)$$

où  $\delta_D$  représente la distribution de Dirac et  $P(k)$  est par définition le spectre de puissance du champ. Notons que la stationnarité du champ entraîne que des modes de Fourier distincts ont une covariance égale à 0 (i.e. sont non corrélés). Le spectre de puissance est donné en trois dimensions par la transformée de Fourier de la fonction de corrélation tri-dimensionnelle  $\xi(\mathbf{r}) = \xi(r)$  :

$$P(\mathbf{k}) = \int \xi(r) \exp(i\mathbf{k} \cdot \mathbf{r}) d^3\mathbf{r} \quad (2.11)$$

$$\xi(\mathbf{r}) = \frac{1}{(2\pi)^3} \int P(k) \exp(-i\mathbf{k} \cdot \mathbf{r}) d^3\mathbf{k} \quad (2.12)$$

Les fonctions monodimensionnelles  $\xi(r)$  et  $P(k)$  sont donc reliés par transformée de Hankel :

$$\xi(r) = \frac{1}{(2\pi)^3} \int 4\pi P(k) \frac{\sin(kr)}{kr} k^2 dk \quad (2.13)$$

### 2.1.3 Cas particulier du champ Gaussien

#### Définition et propriétés

Un champ Gaussien  $\delta$  est tel que pour n'importe quel  $N$ -uplet de points  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$ , le vecteur de dimension  $N$ ,  $\boldsymbol{\delta} = (\delta(\mathbf{x}_1), \dots, \delta(\mathbf{x}_N))$ , suit une loi multivariée Gaussienne. De plus pour un champ stationnaire et isotrope, la densité de probabilité de  $\boldsymbol{\delta}$  est simplement donnée en fonction de la moyenne  $\bar{\delta}$  et de la covariance du champ  $\mathbf{C}$  :

$$p(\boldsymbol{\delta}) \propto \frac{1}{\sqrt{|\mathbf{C}|}} \exp \left[ -\frac{1}{2} (\boldsymbol{\delta} - \bar{\delta}) \mathbf{C}^{-1} (\boldsymbol{\delta} - \bar{\delta}) \right] \quad (2.14)$$

avec  $C$  la matrice de covariance du vecteur  $\boldsymbol{\delta}$  liée à la covariance du champ  $\mathbf{C}$  de la façon suivante :

$$C_{ij} = \mathbf{C}(|\mathbf{x}_i - \mathbf{x}_j|) \quad (2.15)$$

Pour un champ de moyenne nulle  $\bar{\delta} = 0$ , comme c'est le cas pour le champ de fluctuations, le champ est entièrement déterminé par sa fonction de covariance  $\mathbf{C}$ , égale à la fonction de corrélation définie dans l'équation (2.2) :

$$\xi(r) = \mathbf{C}(r) \quad (2.16)$$

Comme la transformée de Fourier est linéaire, les modes  $\tilde{\delta}(\mathbf{k})$  sont aussi Gaussiens. Comme les différents modes de Fourier sont non corrélés d'après l'équation (2.10) ils sont également indépendants (ceci est une spécificité du cas Gaussien).

#### Champ de fluctuations $\delta$ en cosmologie et Gaussianité

Comme expliqué dans la section 1.3.1, l'inflation prédit des fluctuations primordiales du champ de matière  $\delta_m(t_i)$  quasi-Gaussiennes (conséquence du théorème central-limite), et cette Gaussianité est préservée par l'évolution linéaire. Le CMB révèle les fluctuations dans un régime linéaire, et c'est pourquoi ses fluctuations sont Gaussiennes. En revanche l'évolution non linéaire détruit la Gaussianité et permet l'apparition de structures planaires ou filamentaires qui ne sont pas attendues pour un champ Gaussien. En effet les grandes structures de l'univers à l'époque actuelle semblent constituées de grands vides séparés par des feuilletés et des filaments, constituant un réseau plus connu sous le nom de toile cosmique (*Cosmic Web*, [Bond et al. \(1996\)](#); [Springel et al. \(2005\)](#), voir figure 2.1).

L'approximation de l'évolution linéaire est d'autant plus valide que l'on va vers les grandes échelles ou les hauts *redshift* (en effet l'approximation linéaire est basée sur la condition  $\delta \ll 1$ ). Dans la distribution des galaxies actuelles, on peut donc regarder en dessous de quelle échelle cette approximation n'est plus valable, ou de façon équivalente en dessous de quelle échelle le champ n'est plus Gaussien.

Une méthodologie très utile pour tester la Gaussianité est basée sur les fonctions de Minkowski, qui quantifient la géométrie et la topologie des ensembles d'excursion,  $F_u = \{\mathbf{x}, \delta(\mathbf{x}) \geq u\}$ , du champ  $\delta$ . Un grand intérêt des fonctions de Minkowski est qu'elles sont connues analytiquement pour des champs Gaussiens ([Tomita 1990](#)). Par exemple, la troisième fonction de Minkowski  $v_3(\nu)$  (liée à la caractéristique d'Euler  $\chi$  ou au genus  $G$ ), a une espérance donnée par :

$$\mathbb{E}[v_3(\nu)] \propto (\nu^2 - 1) \exp \left( \frac{-\nu^2}{2} \right) \quad (2.17)$$

avec  $\nu = \frac{u-\mu}{\sigma}$ , où  $u$  est le seuil du champ correspond à l'ensemble d'excursion mesuré,  $\mu$  est la moyenne du champ et  $\sigma$  correspond à l'écart-type en un point donné.

On montre sur la figure 2.2 la fonction  $v_3(\nu)$  obtenue sur le catalogue de galaxies du 2 *degree Field Galaxy Redshift Survey* (2dFGRS) décrit dans [Martínez et al. \(2005\)](#) et lissé à différentes échelles  $s = 1, 2, 4, 6$  et  $8 h^{-1} \text{Mpc}$ . On voit qu'aux échelles  $s = 4, 6, 8 h^{-1} \text{Mpc}$ , la fonction  $v_3(\nu)$  est proche de celle d'un champ Gaussien alors qu'aux échelles  $s = 1, 2 h^{-1} \text{Mpc}$  il y a des larges différences.

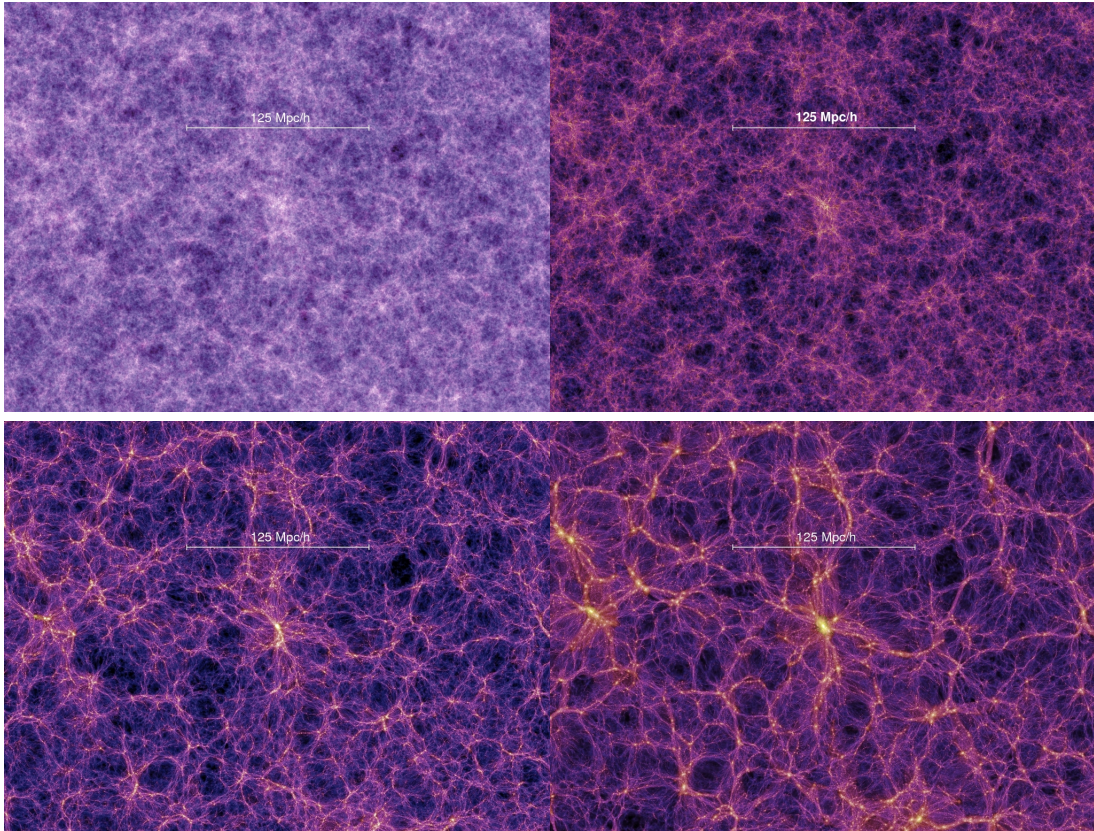


FIGURE 2.1 – Les différentes figures montrent la même coupe d'épaisseur  $15h^{-1}\text{Mpc}$  en coordonnées comobiles de la simulation  $N$ -corps Millennium (Springel et al. 2005), à différents *redshift*  $z = 18.3$  (en haut à gauche),  $z = 5.7$  (en haut à droite),  $z = 1.4$  (en bas à gauche) et  $z = 0$  (en bas à droite). On peut remarquer l'évolution non linéaire qui détruit la Gaussianité du champ et permet la formation de la toile cosmique avec son réseau filaments entourant des grands vides cosmiques. Figures fournies par l'équipe scientifique de la simulation Millennium, <http://www.mpa-garching.mpg.de/millennium/>.

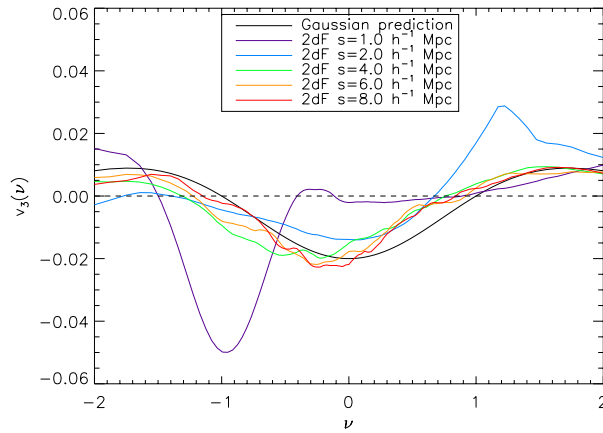


FIGURE 2.2 – La statistique  $v_3(\nu)$  obtenue sur un catalogue de galaxies du relevé 2dFGRS. Les différentes courbes montrent  $v_3(\nu)$  après un lissage du champ par un noyau Gaussien de taille  $s = 1 h^{-1}\text{Mpc}$  (violet),  $s = 2 h^{-1}\text{Mpc}$  (bleu),  $s = 4 h^{-1}\text{Mpc}$  (vert),  $s = 6 h^{-1}\text{Mpc}$  (orange) et  $s = 8 h^{-1}\text{Mpc}$  (rouge) comparées à l'espérance  $\mathbb{E}[v_3(\nu)]$  pour un champ Gaussien de la formule (2.17) (noir). Chaque courbe a été renormalisée pour rendre la comparaison possible. On voit que les courbes  $v_3(\nu)$  se rapprochent du cas Gaussien quand on augmente la taille du lissage, ce qui supporte l'idée que le champ de galaxies devient Gaussien à des échelles suffisamment grandes.

## 2.2 La corrélation des galaxies

### 2.2.1 Le champ de galaxies comme un processus de Cox

Une caractéristique importante de la distribution des galaxies est que celle-ci est discrète et non continue, i.e. qu'elle est décrite par un processus ponctuel. Pour une revue de statistiques spatiales nous renvoyons le lecteur à [Stoyan et al. \(1995\)](#) et pour un aperçu des applications en cosmologie à [Martínez and Saar \(2002\)](#). En pratique un catalogue de galaxies peut être représenté par une suite aléatoire de  $N$  points  $\Psi = \{\mathbf{x}_i\}_{i=1}^N$ , où  $N$  est aussi aléatoire et les  $\mathbf{x}_i$  correspondent aux positions observées des galaxies à l'intérieur d'une région de l'espace  $W \subset \mathbb{R}^3$ . De façon équivalente, la distribution des galaxies peut être vue comme une somme de distributions de Dirac  $\delta_D$  centrées sur les galaxies observées :

$$n_g(\mathbf{x}) = \sum_{i=1}^N \delta_D(\mathbf{x} - \mathbf{x}_i) \quad (2.18)$$

Un modèle réaliste pour la distribution des galaxies  $n_g$  est celui d'un processus de Cox (connu aussi sous le nom de double processus stochastique de Poisson). C'est un processus de Poisson avec une intensité donnée par un champ continu  $\rho_g(\mathbf{x})$  qui est lui-même un champ aléatoire. Connaissant  $\rho_g(\mathbf{x})$ , le nombre de galaxies dans un volume  $dV$  autour de  $\mathbf{x}$  est une réalisation de Poisson avec intensité  $\rho_g(\mathbf{x})dV$ . Ce processus est donc doublement stochastique ; d'un côté avec la stochasticité du champ continu  $\rho_g(\mathbf{x})$ , et d'un autre côté à travers l'échantillonnage de Poisson de ce champ.

On peut vérifier que la fonction de corrélation de ce processus ponctuel est la même que celle du champ continu sous-jacent  $\rho_g$ , plus une fonction de Dirac en 0  $\frac{1}{\bar{n}}\delta_D$  due au caractère ponctuel du champ (i.e. au bruit de Poisson).

Dans le cas d'un processus ponctuel, la fonction de corrélation de  $n_g$  (vue comme un processus continu avec l'équation (2.18)) définie dans la section 2.1.2 possède une interprétation statistique simple. Si l'on considère deux volumes infinitésimaux  $dV_1$  et  $dV_2$  centrés en  $\mathbf{x}_1$  et  $\mathbf{x}_2$  tels que  $\mathbf{x}_1 = \mathbf{x}_2 + \mathbf{r}$ , la probabilité  $dP_{12}$  qu'à la fois  $dV_1$  et  $dV_2$  contiennent un point du processus est égale à :

$$dP_{12} = \bar{n}^2[1 + \xi(\mathbf{r})]dV_1dV_2 \quad (2.19)$$

où  $\bar{n}$  est la densité moyenne du champ, i.e.  $\bar{n} = \mathbb{E}[n_g]$ . Une autre façon d'exprimer ceci est de regarder la probabilité conditionnelle  $dP_{2|1}$  de trouver un point dans le volume  $dV_2$  sachant qu'un

point est présent dans le volume  $dV_1$  :

$$dP_{2|1} = \bar{n}[1 + \xi(\mathbf{r})]dV_2 \quad (2.20)$$

Lorsque le champ est statistiquement isotrope, la fonction de corrélation ne dépend que de la norme de la séparation  $r = \|\mathbf{r}\|$ . On dira d'un champ aléatoire pour lequel  $\xi = 0$  que c'est un champ non corrélé. La fonction de corrélation peut donc être interprétée comme l'excès relatif ( $\xi > 0$ ) ou le défaut relatif ( $\xi < 0$ ) du nombre de paires à une distance  $\mathbf{r}$  par rapport à un processus non corrélé.

On voit avec la formule (2.20) qu'un champ pour lequel deux volumes distincts de l'espace sont indépendants, i.e.  $dP_{2|1} = \bar{n}dV_2$ , est non corrélé. Un cas particulier de champ non corrélé est celui d'un processus de Poisson d'intensité  $\lambda$  sans corrélation. Pour un tel processus le nombre de points dans une région  $A$  de volume  $V(A)$  suit une loi de Poisson d'intensité  $\lambda V(A)$ , et les nombres de points à l'intérieur de deux régions distinctes  $A$  et  $B$  sont indépendants. On se servira notamment de ce type de processus pour estimer la fonction de corrélation (voir section 2.2.4).

## 2.2.2 Biais des galaxies par rapport à la matière

Le champ de galaxies continu  $\rho_g$  défini dans la section 2.2.1 est évidemment lié au champ de matière continu  $\rho_m$ . En réalité la relation entre ces 2 champs est complexe et dépend des processus de formation et d'évolution des galaxies. Un modèle simple et très souvent utilisé consiste à supposer une relation linéaire entre ces deux champs à travers le paramètre de biais  $b$  (biais 'masse-luminosité') défini de la façon suivante :

$$\delta_g(\mathbf{x}) = b \delta_m(\mathbf{x}) \quad (2.21)$$

avec  $\delta_m$  le champ de fluctuations de la matière défini par l'équation (2.1) et  $\delta_g$  le champ de fluctuations défini de la même façon par rapport à  $\rho_g$ . Il faut noter que le terme biais ici a une signification différente du biais d'un estimateur statistique. Le biais masse-luminosité quantifie comment les fluctuations sont amplifiées dans la distribution des galaxies par rapport au champ de matière, alors que le biais d'un estimateur quantifie la différence entre l'espérance de l'estimateur et la quantité estimée.

Dans le cas 'non biaisé' où  $b = 1$ , on dit que les galaxies suivent le champ de matière. On peut aussi noter que le cas  $b > 1$  pose théoriquement un problème pour ce simple modèle de biais, puisque l'on doit à la fois avoir  $\delta_m \in [-1, \infty]$  et  $\delta_g \in [-1, \infty]$ . Ce modèle simple donne facilement la relation entre la fonction de corrélation des galaxies  $\xi_g$  et celle de la matière  $\xi_m$  :

$$\xi_g(r) = b^2 \xi_m(r) \quad (2.22)$$

Un autre modèle simple de 'biais mass-luminosité' qui permet d'avoir une intuition des processus physiques sous-jacents, consiste à supposer que les galaxies sont formées dans les pics du champ de matière, i.e. dans les maxima locaux. Dans ce cas, le biais du champ obtenu en sélectionnant ces pics par rapport au champ de matière peut être calculé analytiquement, au moins dans le cas d'un champ de matière Gaussien (Kaiser 1984; Bardeen et al. 1986). Il existe des modèles plus généraux où le biais peut être stochastique et dépendre de la position et de l'échelle (e.g. Fry and Gaztanaga (1993); Mo et al. (1997); Matsubara (1999)).

Le paradigme actuel pour la formation des galaxies est basé sur le *halo model* selon lequel les galaxies se forment à l'intérieur de halos de matière noire virialisés. Selon le *halo model*, le biais des halos de matière noire est déterministe, où l'on définit le biais :

$$b_h^2 = \frac{\xi_h}{\xi_m} = \frac{P_h}{P_m} \quad (2.23)$$

où  $\xi_h$  et  $\xi_m$  ( $P_h$  et  $P_m$ ) dénotent la fonction de corrélation (le spectre de puissance) des halos et de la matière respectivement. Ce biais peut être calculé comme la moyenne du biais  $b(m)$  en tant que fonction de la masse  $m$  du halo, pondéré suivant la sélection des halos (voir Sánchez et al. (2008)) :

$$b_h = \frac{\int_0^\infty \psi(m)b(m)n(m)dm}{\int_0^\infty \psi(m)n(m)dm} \quad (2.24)$$

où  $n(m)$  est la fonction de masse des halos de matière noire, donnant la densité spatiale de halos dans l'intervalle de masse  $[m, m + dm]$  (Sheth et al. 2001),  $b(m)$  le biais des halos de masse  $m$  (Sheth

et al. 2001) et  $\psi(m)$  la fonction de sélection des halos de masse  $m$  dans la construction de l'échantillon. Sánchez et al. (2008) trouve que cette approximation d'un biais constant (i.e. indépendant de l'échelle) est très précise pour différentes sélections de halos aux échelles  $r > 60h^{-1}\text{Mpc}$ . Si l'on veut modéliser la fonction de corrélation aux plus petites échelles il faut toutefois tenir compte de la dépendance d'échelle, et ceci est généralement fait grâce à des simulations  $N$ -corps. Eisenstein et al. (2005) trouve que la correction par rapport à un biais constant est de seulement 10% pour  $r > 10h^{-1}\text{Mpc}$  à  $z = 0.3$ . L'effet diminue rapidement à plus haut *redshift* et Blake et al. (2011a) ne trouve qu'une correction de 1% pour  $r > 10h^{-1}\text{Mpc}$  à  $z = 0.6$ .

Comme on étudie en général la corrélation d'une certaine population de galaxies et non des halos de matière noire, une étude précise requiert un raffinement du *halo model* basé sur une *Halo Occupation Distribution* (HOD). Dans ce modèle les galaxies se forment à l'intérieur des halos de matière noire, avec une distribution pour un halo de masse  $m$  donnée par  $p(N_g|m)$  (Scoccimarro et al. 2001; Berlind and Weinberg 2002).

Il faut noter qu'un tel modèle HOD ne donne plus tout à fait le champ de galaxies  $n_g$  suivant un processus Cox puisque les distributions  $p(N_g|m)$  peuvent être quelconques (et en particulier non Poissoniennes).

Etant donné que le 'biais masse-luminosité' est lié à la formation des galaxies et leur évolution, il est logique que ce biais dépende de la population de galaxies considérée. En d'autres termes, la corrélation d'une population de galaxies dépend des propriétés intrinsèques telles que la masse, et évolue avec la temps. Ce phénomène connu sous le nom de ségrégation des galaxies, peut être observé lorsque l'on étudie la dépendance de la corrélation en fonction des propriétés des galaxies. En général, les galaxies rouges et lumineuses sont plus fortement corrélées que les galaxies bleues, peu lumineuses et spirales. A l'intérieur des modèles décrits précédemment cela veut dire concrètement que les paramètres  $b$  ou  $p(N_g|m)$  ne sont pas absolus mais dépendent de la population de galaxies.

Dans nos simulations nous utiliserons un modèle très simple, avec un biais constant aux grandes échelles  $b$  et une correction dépendante d'échelle étalonnée à partir de simulations  $N$ -corps.

### 2.2.3 Distorsions de *redshift*

Afin d'estimer la fonction de corrélation  $\xi(r)$  (ou le spectre de puissance  $P(k)$ ) on a besoin de connaître la position en 3D de chaque objet de l'échantillon. En cosmologie, un problème vient de la façon différente dont est estimée la position d'un objet : sa position angulaire sur le ciel est facilement estimée, mais sa distance dans la ligne de visée est obtenue en général grâce à son *redshift*. Comme on l'a vu dans la section 1.1.4 la distance comobile d'un objet en fonction de son *redshift* dépend du modèle cosmologique, si bien que pour construire un échantillon il faut supposer une cosmologie fiducielle. L'effet d'une cosmologie fiducielle erronée peut être modélisé, et c'est cet effet qui permet notamment de contraindre les paramètres cosmologiques.

Cependant il y a un effet supplémentaire qui vient des distorsions de *redshift* (*redshift distortions*) dues aux vitesses particulières des objets. En effet, avec le bon modèle cosmologique, on peut calculer la distance comobile d'un objet à partir de son *redshift* lorsque celui-ci suit le flot de Hubble. Mais cette estimation est faussée car le *redshift* de l'objet est aussi affecté par sa vitesse particulière. Si l'on prend le cas particulier des relevés de galaxies spectroscopiques, le *redshift* est mesuré avec une grande précision (typiquement,  $\sigma(z) \sim 10^{-4}$ ). Cependant les vitesses particulières créent des déformations dans le champ de densité (on dit qu'on observe le champ en espace *redshift* et non en espace réel). Ces déplacements particuliers peuvent être séparés en deux types :

- Un déplacement dû à des vitesses aléatoires à l'intérieur des amas de galaxies, qui crée un étirement de l'amas en espace *redshift* (effet *finger of god*). Cet effet a pour conséquence de lisser la fonction de corrélation approximativement à l'échelle du déplacement aléatoire.
- L'effet Kaiser (Kaiser 1987) qui est plus subtil. Cet effet décrit les vitesses particulières d'objets qui s'effondrent gravitationnellement vers un centre de masse. Il diffère de l'effet *finger of god* car les vitesses des objets sont cohérentes et non aléatoires. Du fait de la cohérence des vitesses avec le champ de densité, cet effet amplifie la corrélation aux grandes échelles.

Ces deux effets impliquent que la fonction de corrélation en espace *redshift* n'est plus isotrope (alors qu'elle est isotrope dans l'espace réel selon le principe cosmologique). Il est donc utile de considérer



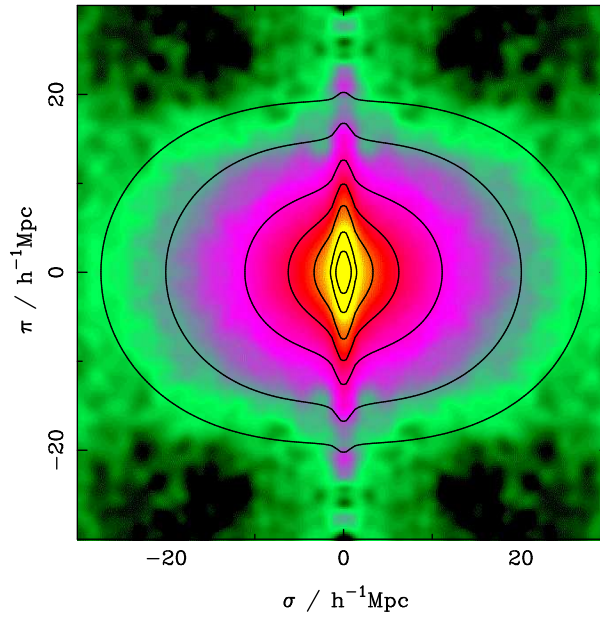


FIGURE 2.3 – Fonction de corrélation bidimensionnelle  $\xi(\pi, \sigma)$  pour le relevé de galaxies 2dFGRS, en fonction de la séparation sur la ligne de visée  $\pi$  et dans la direction transverse  $\sigma$ , montrant l'effet des distorsions de *redshift*. La forme de cigare à  $\pi \approx 0$  est dû à l'effet *finger of god* qui lisse la fonction de corrélation, alors que l'effet cigare aux grands  $\pi$  (i.e. l'amplification de la corrélation) est dû à l'effet de Kaiser. Figure extraite de Peacock et al. (2001).

la fonction en bidimensionnelle  $\xi(\pi, \sigma)$ , qui dépend de la séparation de la paire  $\mathbf{r}$  projetée sur la ligne de visée  $\pi$  et projetée sur la direction transverse  $\sigma$ . On peut voir sur la figure 2.3, pour le relevé de galaxies 2dFGRS, les deux effets mentionnés précédemment dans la direction de  $\pi$  à  $\sigma \approx 0$  : la forme de cigare à  $\pi \approx 0$  est dû à l'effet *finger of god* qui lisse la fonction de corrélation, alors que l'effet cigare aux grands  $\pi$  (i.e. l'amplification de la corrélation) est dû à l'effet de Kaiser.

L'effet Kaiser (Kaiser 1987) peut être quantifié facilement dans le régime linéaire en supposant que l'on est dans l'approximation 'plan-parallèle' (*plane-parallel*), i.e. que l'observateur est suffisamment loin pour que les vitesses des objets dans la ligne de visée puissent effectivement être considérées comme parallèles. On obtient que le spectre de puissance en espace *redshift*  $P_s(\mathbf{k})$  est relié au spectre de puissance en espace réel  $P(k)$  par la formule :

$$P_s(\mathbf{k}) = (1 + \beta \mu_k^2)^2 P(k) \quad (2.25)$$

avec  $\mu_k \in [-1, 1]$  égal à la projection du vecteur unitaire  $\hat{\mathbf{k}}$  sur la ligne de visée, et  $\beta$  proportionnel au facteur de croissance sans dimension introduit dans la section 1.3.2 :

$$\beta = \frac{f(\Omega_m)}{b} = \frac{1}{b} \frac{d \ln D}{d \ln a} \quad (2.26)$$

avec  $b$  le biais (supposé constant) de la population d'objets. En particulier pour le champ de matière lui-même, on a  $b = 1$ . De plus, comme on l'a vu dans la section 1.3.2, on peut en général approximer  $f(\Omega_m, z) \approx \Omega_m(z)^{0.6}$ .

Il est très courant dans l'étude des catalogues de galaxies de ne s'intéresser qu'au monopole de la fonction de corrélation  $\xi_0(s)$  (ou du spectre de puissance) en espace *redshift*, de par sa simplicité et de par le fait qu'il soit moins bruité que la fonction de corrélation bidimensionnelle  $\xi(\pi, \sigma)$  :

$$\xi_0(s) = \frac{1}{4\pi} \int_{\Omega} \xi(\mathbf{s}) d^2 o_s \quad (2.27)$$

$$P_0(k) = \frac{1}{4\pi} \int_{\Omega} P_s(\mathbf{k}) d^2 o_k \quad (2.28)$$

Les monopoles correspondent aux moyennes sphériques pour une norme donnée  $s = \|\mathbf{s}\|$  ou  $k = \|\mathbf{k}\|$ . Dans la littérature, le monopole de la fonction de corrélation est souvent dénoté fonction de corrélation par abus de langage (de même l'indice 0 du monopole est souvent oublié mais en général le contexte est assez clair pour faire la différence). A partir de la relation (2.25) on peut montrer que le monopole de la fonction de corrélation et du spectre de puissance en espace *redshift* sont reliés à leur contrepartie en espace réel par les relations :

$$\xi_0(s) = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right) \xi(r) \quad (2.29)$$

$$P_0(k) = \left(1 + \frac{2\beta}{3} + \frac{\beta^2}{5}\right) P(k) \quad (2.30)$$

L'effet Kaiser consiste donc en un simple effet multiplicatif sur la fonction de corrélation et le spectre de puissance, i.e. que c'est un biais indépendant de l'échelle.

### 2.2.4 Estimation de la fonction de corrélation

Comme on l'a vu dans la section 2.2.1, la fonction de corrélation  $\xi(r)$  d'un processus ponctuel homogène et isotrope peut être interprétée comme l'excès de probabilité de trouver une paire de galaxies dans deux volumes  $dV_1$  et  $dV_2$  à une séparation  $\mathbf{r}$ , par rapport à une distribution sans corrélation :

$$dP_{12} = \bar{n}^2 [1 + \xi(\mathbf{r})] dV_1 dV_2 \quad (2.31)$$

La fonction de corrélation étant une quantité géométrique, elle requiert de connaître la position des galaxies en 3D. Comme on l'a déjà expliqué dans la section 2.2.3, on ne connaît pas exactement la position des galaxies à l'intérieur des relevés. On ne peut obtenir un volume 3D qu'après avoir supposé une cosmologie fiducielle pour convertir les *redshift* en distances avec la relation suivante, déjà donnée dans la section 1.1.4 :

$$D_C^{fid}(z) = \frac{c}{H_0^{fid}} \int_0^z \frac{1}{E^{fid}(z')} dz' \quad (2.32)$$

avec l'expression de  $E^{fid}(z)$  pour un modèle  $\omega$ CDM :

$$E^{fid}(z) = \sqrt{\Omega_r^{fid}(1+z)^4 + \Omega_m^{fid}(1+z)^3 + \Omega_k^{fid}(1+z)^2 + \Omega_{DE}^{fid}(1+z)^{3(1+w^{fid})}} \quad (2.33)$$

On voit que la paramètre  $H_0^{fid}$  intervient seulement comme une constante multiplicative dans la distance des objets. Il est quasiment systématique dans l'étude de la corrélation des galaxies d'utiliser la notation  $H_0^{fid} = 100 h^{fid} \text{ km s}^{-1} \text{ Mpc}^{-1}$  et de considérer les distances à l'intérieur du relevé dans l'unité  $h^{-1} \text{ Mpc}$  (où l'on omettra la notation  $^{fid}$  par simplicité). En utilisant cette unité on obtient que les distances des objets sont indépendantes du choix de  $H_0^{fid}$ , si bien que l'estimateur de la fonction de corrélation est également indépendant du choix de  $H_0^{fid}$ .

Même si l'on suppose que la cosmologie fiducielle est correcte, les vitesses particulières des objets introduisent des distortions de *redshift* (voir section 2.2.3). On ne calcule donc pas la fonction de corrélation en espace réel, mais en espace *redshift*. Toutefois lorsque la cosmologie fiducielle est incorrecte, la relation *redshift*-distance est également incorrecte, si bien que les galaxies ne sont pas estimées à la bonne distance de l'observateur.

Ces deux phénomènes rendent la distribution de galaxies non statistiquement isotrope dans l'espace *redshift* en coordonnées fiducielles. La corrélation est différente dans la ligne de visée et dans la direction transverse, si bien que la fonction de corrélation  $\xi(\pi, \theta)$  est bi-dimensionnelle. Dans cette thèse, on s'intéressera au monopole de la fonction de corrélation  $\xi_0(s)$ , donné par :

$$\xi_0(s) = \frac{1}{4\pi} \int_{\Omega} \xi(\mathbf{s}) d^2 o_{\mathbf{s}} \quad (2.34)$$

Nous allons maintenant expliquer comment estimer le monopole de la fonction de corrélation. Supposons que l'on connaisse le champ de galaxies  $n_g$  sur un relevé de volume géométrique  $\mathcal{V} \subset \mathbb{R}^3$ .

Faisons l'intégrale double de l'équation (2.31) pour toutes les paires de volume du relevé à distance  $\|\mathbf{r}\| \in I_r = [r - dr/2, r + dr/2]$  :

$$\int_{\mathcal{V}} \int_{\mathcal{V}} \mathbb{1}_{\|\mathbf{x}_1 - \mathbf{x}_2\| \in I_r} dP_{12} = \bar{n}_g^2 \int_{\mathcal{V}} \int_{\mathcal{V}} \mathbb{1}_{\|\mathbf{x}_1 - \mathbf{x}_2\| \in I_r} [1 + \xi(\mathbf{r})] dV_1 dV_2 \quad (2.35)$$

On voit facilement que le côté gauche de l'équation (2.35) correspond à l'espérance du nombre de paires à une distance  $r \in I_r$  dans le relevé, nombre que l'on notera  $DD(r)$ . Notons  $V_{2,r}$  le double volume de points  $\mathbf{x}_1$  et  $\mathbf{x}_2$  du relevé situés à une distance  $\|\mathbf{x}_1 - \mathbf{x}_2\| \in I_r$  :

$$V_{2,r} = \int_{\mathcal{V}} \int_{\mathcal{V}} \mathbb{1}_{\|\mathbf{x}_1 - \mathbf{x}_2\| \in I_r} dV_1 dV_2 \quad (2.36)$$

En général le relevé est assez régulier pour que la simple approximation suivante soit très bien vérifiée :

$$\int_{\mathcal{V}} \int_{\mathcal{V}} \mathbb{1}_{\|\mathbf{x}_1 - \mathbf{x}_2\| \in I_r} \xi(\mathbf{r}) dV_1 dV_2 \approx V_{2,r} \xi_0(r) \quad (2.37)$$

i.e. que toutes les directions sur la sphère du vecteur de séparation  $\mathbf{x}_1 - \mathbf{x}_2$  sont représentées de façon égale sur l'ensemble de paires de points  $\mathbf{x}_1$  et  $\mathbf{x}_2$  du relevé situés à distance  $\|\mathbf{x}_1 - \mathbf{x}_2\| \in I_r$ . On peut donc réécrire l'équation (2.35) :

$$\mathbb{E}[DD(r)] = \bar{n}_g^2 V_{2,r} [1 + \xi_0(r)] \quad (2.38)$$

$$\mathbb{E}[DD(r)] = \mathbb{E}[N_{DD}] \frac{V_{2,r}}{V^2} [1 + \xi_0(r)] \quad (2.39)$$

où  $N_{DD}$  est le nombre total de paires de galaxies dans le relevé, d'espérance  $\mathbb{E}[N_{DD}] = \bar{n}_g^2 V^2$ , avec  $V = |\mathcal{V}|$  le volume du relevé.

Voyons comment on peut utiliser l'équation (2.39) pour estimer la fonction de corrélation. La méthode habituelle est de considérer des catalogues de Poisson aléatoires, i.e. de densité  $\bar{n}_r$  uniforme, pour lesquels tous les volumes distincts sont indépendants. Comme on s'y attend, ce processus a une corrélation nulle (voir équation (2.31)) et l'on peut donc mesurer la corrélation des galaxies comme l'excès de probabilité de trouver une paire à une séparation  $r$  comparé à un tel processus. Réécrivons l'équation (2.39) pour un tel processus :

$$\mathbb{E}[RR(r)] = \mathbb{E}[N_{RR}] \frac{V_{2,r}}{V^2} \quad (2.40)$$

En combinant les équations (2.39) et (2.40) on obtient l'estimateur de Peebles-Hauser du monopole de la fonction de corrélation  $\xi_0(r)$  (Peebles and Hauser 1974) :

$$\hat{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 1 \quad (2.41)$$

Il existe des estimateurs plus sophistiqués du monopole de la fonction de corrélation (Davis and Peebles 1983; Hamilton 1993; Landy and Szalay 1993) :

$$\hat{\xi}_{DP}(r) = \frac{N_{DR}}{N_{DD}} \frac{DD(r)}{DR(r)} - 1 \quad (2.42)$$

$$\hat{\xi}_{HAM}(r) = \frac{N_{DR}^2}{N_{DD} N_{RR}} \frac{DD(r) RR(r)}{DR^2(r)} - 1 \quad (2.43)$$

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 2 \frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)} \quad (2.44)$$

Dans le chapitre 4, on comparera ces estimateurs et on expliquera pourquoi ces différentes extensions améliorent la qualité de l'estimateur.

Jusqu'à présent on a supposé que l'on connaissait la distribution de galaxies dans un relevé de volume  $\mathcal{V}$ . En pratique il peut y avoir des effets de sélection compliqués à l'intérieur de relevé. En effet

celui-ci peut avoir une couverture angulaire complexe, avec éventuellement une sélection angulaire qui varie sur le ciel. En général il y a également des effets de sélection en fonction de la distance dus à la façon de sélectionner les galaxies. Ces différents effets rendent le relevé non statistiquement homogène, i.e. avec une densité  $\Phi(\mathbf{x})\bar{n}_g$  qui varie selon la position. Ceci peut être évité en changeant la façon de sélectionner les galaxies et en se restreignant à des zones du ciel avec une sélection constante. On obtient alors un relevé statistiquement homogène (échantillon *volume-limited*), mais cela induit une perte d'information par rapport au relevé initial.

Pour estimer la fonction de corrélation dans le cas non statistiquement homogène, on peut simplement introduire la même sélection dans le catalogue aléatoire que dans le relevé de galaxies. Lorsque l'échantillon est non statistiquement homogène, il n'est plus optimal d'accorder le même poids à toutes les galaxies dans les estimateurs de la fonction de corrélation. En effet on voit bien que les paires de galaxies sont plus redondantes dans les zones de forte densité, et qu'il faut donc diminuer le poids de ces galaxies par rapport aux zones de faible densité. Une fonction de poids approximativement optimale, qui dépend de la fonction de sélection  $\Phi_i$  à la position de la galaxie  $i$  et de la distance  $r$  où l'on estime la fonction de corrélation, est donnée dans [Hamilton \(1993\)](#) :

$$w_i = \frac{1}{1 + \bar{n}_g \Phi_i J(r)} \quad (2.45)$$

avec  $J(r)$  l'intégrale de la vraie fonction de corrélation :

$$J(r) = 4\pi \int_0^r \xi(r') r'^2 dr' \quad (2.46)$$

Il y a toutefois une contrainte pour ne pas introduire de biais dans l'estimateur, à savoir que la densité moyenne pondérée dans l'échantillon de données et dans l'échantillon aléatoire doivent être proportionnelles (i.e. qu'il doit y avoir un facteur multiplicatif constant entre les deux). Lorsque l'on introduit des poids comme ceux de la formule (2.45), les quantités de comptage de paires  $DD$ ,  $RR$  et  $DR$  sont modifiées de la façon suivante dans les estimateurs (2.41), (2.42), (2.43) et (2.44) : au lieu d'ajouter +1 pour chaque paire, on ajoute  $w_i w_j$ , avec  $w_i$  et  $w_j$  les poids respectifs des galaxies de la paire.

Avec le modèle de processus de Cox de la section 2.2.1, on voit qu'il existe deux types de bruit dans l'estimation de la fonction de corrélation. On rappelle que dans ce modèle, la distribution de galaxies discrète  $n_g$  est vue comme une réalisation de Poisson sur un processus aléatoire continu  $\rho_g$ . L'effet du relevé est d'introduire une fonction de sélection  $\Phi$  ( $\Phi = \mathbb{1}_V$  pour un échantillon *volume-limited*) de sorte que l'on peut considérer le champ de galaxies observé comme une réalisation de Poisson sur le processus continu  $\Phi\rho_g$ . Parce que  $\Phi$  est non nulle seulement dans un volume fini, même si l'on connaissait le champ continu sous-jacent  $\Phi\rho_g$ , il y aurait une incertitude sur la fonction de corrélation du champ  $\rho_g$ . Ce premier type de bruit est appelé bruit d'échantillon (*sample variance*). Le second type de bruit est dû à l'échantillonnage de Poisson du champ continu, i.e. que le champ discret  $n_g$  ne donne accès au champ continu  $\Phi\rho_g$  qu'à une incertitude près. Ce bruit de Poisson est plus fort dans les zones de faible densité, et on le nomme également bruit de grenaille (*shot noise*).

## 2.2.5 Modélisation de la fonction de corrélation

### Variations de $\xi(r)$ en coordonnées comobiles

Pour modéliser la fonction de corrélation estimée, la première étape est de modéliser la fonction de corrélation linéaire. Pour tout modèle  $w$ CDM, la fonction de corrélation linéaire ne dépend que de la densité de matière  $\omega_m = \Omega_m h^2$ , de la densité de baryon  $\omega_b = \Omega_b h^2$ , de l'indice spectral  $n_s$  et de  $\sigma_8$  (en principe il y a aussi une dépendance par rapport à la température du CMB  $\theta_{CMB}$ , mais celle-ci est connue avec une très grande précision,  $\theta_{CMB} = 2.7260 \pm 0.0013$ , voir [Fixsen \(2009\)](#)).

La figure 1.7 montre la corrélation linéaire pour différentes valeurs de  $\omega_m$  (avec  $\omega_b$ ,  $n_s$  et  $\sigma_8$  fixés). On voit notamment comment l'amplitude du pic BAO augmente lorsque la proportion de baryons augmente par rapport à la matière totale (i.e.  $\omega_m$  diminue à  $\omega_b$  fixé).

L'étape suivante de la modélisation consiste à appliquer les corrections non linéaires (voir section 1.3.4). Ces corrections dépendent légèrement de l'évolution de l'univers pour un *redshift* donné, i.e. des paramètres  $H_0$ ,  $\Omega_\Lambda$ ,  $\Omega_k$  et  $w$  pour un modèle  $w$ CDM. Il y a finalement l'effet du biais des galaxies par

rapport à la matière dont il faut tenir compte. Comme on l'a vu dans la section 2.2.2, ce biais est très proche d'un facteur multiplicatif constant dans la fonction de corrélation à des échelles suffisamment grandes. Cependant si l'on veut modéliser la corrélation à des échelles faibles et intermédiaires, on peut se servir de simulations  $N$ -corps pour étalonner ce biais.

### Effets dus à la variation de la géométrie

A ce stade, on obtient une modélisation de la fonction de corrélation en espace réel et en coordonnées comobiles. Mais comme on l'a vu précédemment, on estime la fonction de corrélation en espace *redshift* et dans des coordonnées fiducielles. On peut d'abord modéliser l'effet du passage de l'espace réel vers l'espace *redshift*, qui se traduit par deux effets (voir section 2.2.3) :

- Un lissage de la fonction de corrélation suivant la ligne de visée, dû à l'effet *finger of gods* (qui se traduit aussi par un lissage du monopole de la fonction de corrélation)
- Une amplification du spectre de puissance suivant la ligne de visée, due à l'effet Kaiser (qui se traduit par un facteur constant dans le monopole de la fonction de corrélation)

Finalement il faut tenir compte de l'effet de la cosmologie fiducielle, qui crée des distortions par rapport à la cosmologie réelle. La cosmologie fiducielle intervient à travers la relation distance-*redshift* (relations (2.32) et (2.33)) pour convertir les *redshift* des galaxies en distances. Les distortions sont d'autant plus grandes que la relation fiducielle est 'éloignée' de la vraie relation, qui dépend de  $H_0$ ,  $\Omega_\Lambda$ ,  $\Omega_k$  et  $w$  pour un modèle  $w$ CDM.

On peut voir les distortions dues à la cosmologie fiducielle sur la figure 2.4. La figure montre la taille d'un objet situé à *redshift*  $z$ , d'extension  $\Delta z$  en *redshift* et d'extension angulaire  $\Delta\theta$  sur le ciel. Dans la vraie cosmologie (que l'on ne connaît pas), l'objet a une extension  $c\frac{\Delta z}{H(z)}$  dans la ligne de visée et  $D_M(z)\Delta\theta$  dans la direction transverse. En revanche, lorsque l'on choisit une cosmologie fiducielle, l'objet a une extension  $c\frac{\Delta z}{H^{fid}(z)}$  dans la ligne de visée et  $D_M^{fid}(z)\Delta\theta$  dans la direction transverse. On voit en particulier qu'un objet sphérique peut devenir un ellipsoïde parce que la cosmologie fiducielle est incorrecte.

On peut toutefois quantifier cet effet de distortion dans la fonction de corrélation. L'effet est particulièrement clair dans la fonction de corrélation bi-dimensionnelle  $\xi(\pi, \sigma)$  puisqu'une taille réelle  $c\frac{\Delta z}{H(z)}$  dans la ligne de visée apparaît comme une taille  $c\frac{\Delta z}{H^{fid}(z)}$ , et une taille réelle  $D_M(z)\Delta\theta$  dans la direction transverse apparaît comme une taille  $D_M^{fid}(z)\Delta\theta$ . Autrement dit on observe un relevé transformé par rapport au relevé réel, avec une dilatation  $\frac{H(z)}{H^{fid}(z)}$  dans la ligne de visée et  $\frac{D_M^{fid}(z)}{D_M(z)}$  dans la direction transverse. Ceci se traduit donc immédiatement comme une dilatation de la fonction de corrélation bi-dimensionnelle. Si l'on considère l'espérance de l'estimateur de la fonction de corrélation calculée en coordonnées fiducielles, on obtient :

$$\mathbb{E} \left[ \hat{\xi}(\pi, \sigma) \right] = \xi \left( \frac{H^{fid}(z)}{H(z)} \pi, \frac{D_M(z)}{D_M^{fid}(z)} \sigma \right) \quad (2.47)$$

où  $\xi(\pi, \sigma)$  est la vraie fonction de corrélation en espace *redshift* 2D. On voit donc que l'étude de la géométrie de l'univers se fait à travers les paramètres  $H(z)$  et  $D_M(z)$  qui influencent la relation *redshift*-distance et qui transforment la fonction de corrélation observée. A cause de l'incertitude inhérente aux observations (variance cosmique et bruit de Poisson pour la distribution des galaxies, voir section 2.2.4), on ne peut pas mesurer ces paramètres exactement, mais seulement les contraindre statistiquement.

Il y a une subtilité à laquelle il faut faire attention à ce stade. En effet un relevé a toujours une certaine extension en *redshift*, si bien que la vision d'une dilatation unique dans tout le relevé suivant la ligne de visée et une dilatation unique suivant la direction transverse n'est pas rigoureusement vraie. Pour des raisons de simplicité on suppose quand même que cela est vrai en général en prenant  $H(\langle z \rangle)$  et  $D_M(\langle z \rangle)$  au *redshift* moyen du relevé  $\langle z \rangle$ .

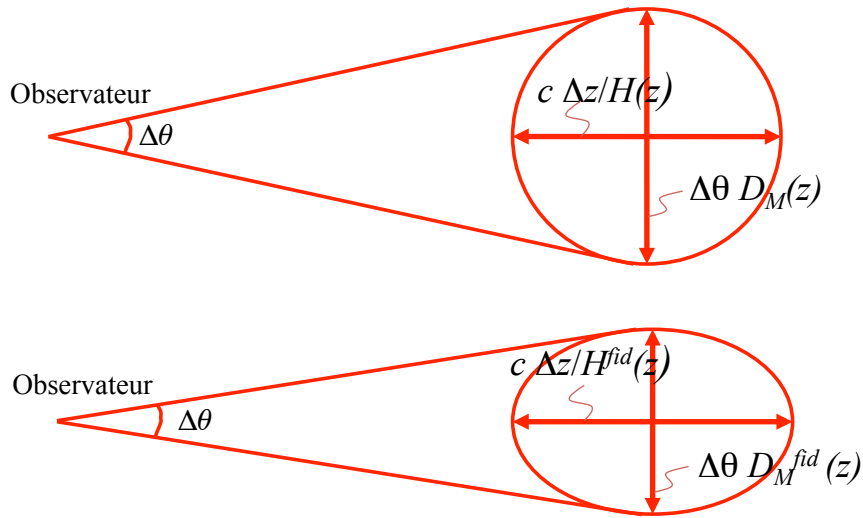


FIGURE 2.4 – Taille d'un objet (dans la ligne de visée et dans la direction transverse) à *redshift*  $z$ , d'extension  $\Delta z$  en *redshift* et d'extension angulaire  $\Delta\theta$  sur le ciel. La vraie taille de l'objet est donnée dans la ligne de visée par  $c \frac{\Delta z}{H(z)}$  et dans la direction transverse par  $D_M(z)\Delta\theta$ . Cependant si l'on utilise une cosmologie fiducielle incorrecte (comme c'est forcément le cas puisqu'on ne connaît pas exactement la vraie cosmologie) pour convertir les *redshift* en distances, les paramètres  $H^{fid}(z)$  et  $D_M^{fid}(z)$  sont différents et l'objet est déformé. En particulier, un objet sphérique peut devenir ellipsoïdal en coordonnées fiducielles.

On va même au-delà de cette approximation dans les études habituelles qui s'intéressent au monopole de la fonction de corrélation  $\xi_0(s)$  (voir section 2.2.4) :

$$\xi_0(s) = \frac{1}{4\pi} \int_{\Omega} \xi(\mathbf{s}) d^2 o_{\mathbf{s}} \quad (2.48)$$

où l'on utilise à nouveau la notation  $s$  pour la séparation en espace *redshift* par souci de clarté. On voit dans ce cas que les dilatations dans la ligne de visée et dans la direction transverse vont se 'mélanger' dans le monopole de la fonction de corrélation, ce qui va changer la forme globale de la fonction de corrélation. Toutefois Eisenstein et al. (2005) et Padmanabhan and White (2008) ont montré qu'on peut à nouveau approximer ce phénomène par une simple dilatation par rapport à la fonction de corrélation réelle avec un facteur  $\alpha = D_V(\langle z \rangle) / D_V^{fid}(\langle z \rangle)$  avec  $D_V$  l'échelle de dilatation (*dilation scale*) donnée par :

$$D_V(z) = \left[ (1+z)^2 D_M(z)^2 \frac{cz}{H(z)} \right]^{1/3} \propto D_M(z)^{2/3} \frac{1}{H(z)^{1/3}} \quad (2.49)$$

On voit que cette échelle de dilatation est donnée par une moyenne géométrique des facteurs de dilatation dans la ligne de visée et dans la direction transverse (respectivement  $D_M(z)$  et  $\frac{1}{H(z)}$ ). La dilatation transverse a un poids plus grand (puissance 2/3) comparée à la dilatation dans la ligne de visée (puissance 1/3). En effet la direction transverse a aussi plus de poids que la direction de la ligne de visée dans le monopole de la fonction de corrélation. L'espérance du monopole de la fonction de corrélation en coordonnées fiducielles (souvent noté  $\xi(r)$  par abus de notation et appelé fonction de corrélation par abus de langage) est reliée à la vraie fonction de corrélation par la relation :

$$\mathbb{E} \left[ \hat{\xi}_0(s) \right] = \xi_0(\alpha s) \quad (2.50)$$

où  $\xi_0(s)$  est cette fois le vrai monopole de la fonction de corrélation en espace *redshift*. C'est le paramètre  $\alpha$  qui mesure maintenant la géométrie de l'univers (i.e. la relation *redshift*-distance). La validité

de l'approximation (2.50) a été étudiée dans [Padmanabhan and White \(2008\)](#). Cette étude montre que dans l'approximation 'plan-parallèle', la formule (2.50) est très bien vérifiée pour le monopole aux échelles suffisamment élevées, et notamment à l'échelle du pic BAO.

## 2.3 Multifractalité de la distribution des galaxies

### 2.3.1 Ensembles fractals

#### Le concept de fractale

L'idée d'un univers fractal est très ancien, et déjà au 18<sup>ème</sup> siècle Emmanuel Kant et Johann Lambert ont imaginé un univers comme une infinité de mondes hiérarchisés. Plus tard ce type d'univers a été mis en avant pour expliquer le paradoxe d'Olbers, à savoir que le ciel apparaît noir durant la nuit, en contradiction avec l'hypothèse d'un nombre infini d'étoiles.

La première description mathématique des fractales a été donnée par Mandelbrot ([Mandelbrot 1982](#)), qui a donné une nouvelle interprétation de ces objets avec le concept d'invariance par changement d'échelle. La dénomination des fractales est un néologisme de Mandelbrot créée à partir du mot latin *fractus* signifiant 'brisé' ou 'irrégulier'. Une définition exacte des fractales est délicate car elle pourrait exclure des cas intéressants. Voici une tentative de définition d'une fractale  $F$  extraite de [Falconer \(1990\)](#), que nous avons traduit de l'anglais :

1.  $F$  possède une structure fine, i.e. des détails à des échelles arbitrairement petites.
2.  $F$  est trop irrégulière pour être décrite par la géométrie habituelle, à la fois localement et globalement.
3.  $F$  possède souvent une forme d'invariance par changement d'échelle, peut être de façon approximative ou statistique.
4. La 'dimension fractale' de  $F$  est souvent plus grande que sa dimension topologique.
5. Dans la plupart des cas intéressants,  $F$  peut être définie de façon très simple, peut être récursivement."

Les fractales peuvent être soit déterministes, i.e. construites avec un procédé itératif ou une récursion, soit aléatoires, i.e. générées par un processus stochastique. En général, seules les fractales construites avec un processus itératif possèdent de façon rigoureuse une propriété d'invariance par changement d'échelle. Les fractales aléatoires sont les plus couramment rencontrées dans la pratique, et peuvent décrire de nombreux objets irréguliers dans la nature, e.g. les nuages, les montagnes, les côtes ou les fluides turbulents.

#### Les dimensions fractales

Les dimensions fractales sont des extensions de la notion de dimension habituelle pour décrire des objets simples comme des lignes, surfaces ou volumes. Une première dimension fractale est la dimension de Minkowski calculée par 'comptage de boîte' (*box-counting dimension*). Elle est calculée en remplissant l'espace avec des petits cubes de taille  $\delta \rightarrow 0$  et en comptant le nombre de cubes qui intersectent l'objet (figure 2.5). Pour une ligne de longueur  $l$ , le nombre de cubes intersectant l'objet va être approximativement  $N(\delta) \approx \frac{l}{\delta}$ , et pour une surface d'aire  $A$  ce nombre va être approximativement  $N(\delta) \approx \frac{A}{\delta^2}$ .

La dimension de Minkowski est donc définie pour un ensemble  $F$  de la façon suivante :

$$\dim_{\text{box}}(F) = \lim_{\delta \rightarrow 0} \frac{\ln N(\delta)}{\ln 1/\delta} \quad (2.51)$$

Si cette limite n'existe pas, on parle de dimension de Minkowski supérieure ( $\dim_{\text{upperbox}}$ ) et de dimension de Minkowski inférieure ( $\dim_{\text{lowerbox}}$ ), comme les limites supérieure et inférieure.

Une autre dimension importante est la dimension de Hausdorff. Elle est définie à partir de recouvrements de  $F$  par un ensemble dénombrable de parties  $\mathcal{A} = \bigcup_i \mathcal{A}_i$  qui ont toutes un diamètre inférieur à  $r$ . Pour un réel  $d$  on calcule la limite suivante :

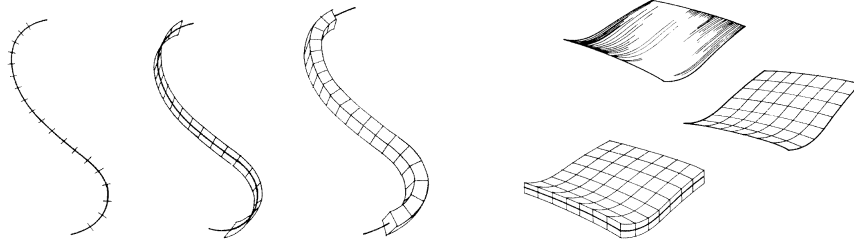


FIGURE 2.5 – Mesure de la dimension de Minkowski pour une ligne (gauche) et pour une surface (droite). Le nombre de cubes intersectant l’objet est  $N(\delta) \propto \frac{1}{\delta}$  pour une ligne et  $N(\delta) \propto \frac{1}{\delta^2}$  pour une surface. Figure extraite de Feder (1988).

$$H^d(F) = \liminf_{r \rightarrow 0} \sum_{\mathcal{A}} \text{diam}(\mathcal{A}_i)^d \quad (2.52)$$

La dimension de Hausdorff  $d_H$  est alors définie comme :

$$H^d(F) = \begin{cases} 0 & \text{pour } d > d_H(F) \\ \infty & \text{pour } d < d_H(F) \end{cases} \quad (2.53)$$

On peut montrer que la dimension de Hausdorff est toujours plus petite que la dimension de Minkowski :

$$d_H(F) \leq \text{dim}_{\text{lowerbox}} \leq \text{dim}_{\text{upperbox}} \quad (2.54)$$

Une autre dimension fractale, qui est bien adaptée pour décrire les motifs de points aléatoires, est la dimension de corrélation. En particulier, cette dimension est très utile pour mesurer la fractalité de la distribution des galaxies. En notant  $N(< r)$  le nombre moyen de voisins à une distance  $r$  d’un point donné, l’ensemble de points est dit fractal avec une dimension de corrélation  $D_2$  lorsque :

$$N(< r) \propto r^{D_2} \quad (2.55)$$

Un tel ensemble est localement invariant d’échelle, i.e. que le nombre de points varie en loi de puissance autour de chaque point. Par exemple, si  $D_2 = 1$  les points sont localement situés sur des structures filamenteuses, et si  $D_2 = 2$  les structures sont localement planaires.

Pour de nombreuses fractales, toutes ces dimensions sont égales. Pour les objets géométriques habituels (lignes, surfaces, volumes), elles sont entières. Cependant pour des ensembles purement fractals, ces dimensions sont non entières. La figure 2.6 montre un tel exemple avec la construction de la courbe de von Koch  $F$ . A chaque itération, chaque segment de la courbe est subdivisé en 4 nouveaux sous-segments. Pour cet ensemble, les dimensions fractales sont toutes égales à  $\frac{\ln 3}{\ln 2}$ . En ce qui concerne la longueur de la courbe  $E_k$  à l’itération  $k$ , elle est égale à  $\left(\frac{4}{3}\right)^k$ , i.e. que la courbe limite a une longueur infinie. D’un autre côté, la courbe occupe une surface égale à 0 dans le plan, i.e. que les notions habituelles de longueur de de surface ne donnent pas une description utile de l’objet.

### 2.3.2 Ensembles multifractals

Une mesure multifractale  $\mu$  sur un ensemble  $A$  est une généralisation d’une mesure fractale. Elle est utilisée lorsqu’un unique exposant (la dimension fractale) n’est pas suffisant pour décrire la mesure. Une multifractale est localement invariante d’échelle en tout point. Un ensemble  $F$  de points  $(\mathbf{x}_1, \dots, \mathbf{x}_N)$  est multifractal, si la mesure  $\mu_F$  naturellement associée à  $F$  est multifractale :

$$\mu_F(A) = |\{\mathbf{x}_i \in A\}| \quad (2.56)$$



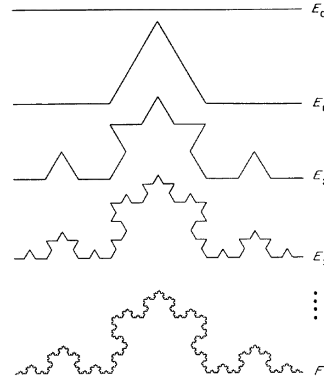


FIGURE 2.6 – Itérations successives pour la construction de la courbe de von Koch  $F$ . Figure extraite de Falconer (1990).

### Exposant de Hölder

L'exposant de Hölder en un point  $x_0$  est défini de la façon suivante :

$$\alpha(x_0) = \lim_{\delta \rightarrow 0} \frac{\ln \mu(B_{x_0}(\delta))}{\ln \delta} \quad (2.57)$$

où  $B_{x_0}(\delta)$  est une boule centrée en  $x_0$  et de rayon  $\delta$ . On a :

$$\mu(B_{x_0}(\delta)) \propto \delta^{\alpha(x_0)} \quad (2.58)$$

Plus la valeur de  $\alpha(x_0)$  est grande, plus la mesure est régulière autour de  $x_0$ .

### Spectre de singularité

Comme mentionné précédemment, une mesure multifractale n'est pas caractérisée par un unique exposant. Pour chaque valeur d'exposant  $\alpha$ , une quantité intéressante est la dimension fractale  $f(\alpha)$  de l'ensemble de points où la mesure a un exposant de Hölder égal à  $\alpha$  :

$$f(\alpha) = d_F(\{x_0 \in A \mid \alpha(x_0) = \alpha\}) \quad (2.59)$$

La fonction  $f(\alpha)$  est appelée 'spectre de singularité' de la mesure  $\mu$ . Le spectre de singularité décrit la fraction de points avec un exposant  $\alpha$  donné. Donc si on recouvre  $A$  avec des boîtes de taille  $\delta$ , le nombre de boîtes contenant un point avec l'exposant  $\alpha$  aura un comportement  $N_\alpha(\delta) \propto \delta^{-f(\alpha)}$  quand  $\delta \rightarrow 0$ . La fonction  $f(\alpha)$  est souvent une fonction unimodale avec  $\max_\alpha f(\alpha) = D$ , où  $D$  est la dimension fractale du support de la mesure. Dans le cas particulier d'une mesure monofractale, la fonction  $f(\alpha)$  se réduit à un seul point :  $\alpha = f(\alpha) = D$ , où  $D$  est la dimension fractale de la mesure.

Pour calculer le spectre de singularité, la méthode de comptage de boîte consiste à calculer la fonction de partition :

$$Z(q, \delta) = \sum_{i=0}^{N(\delta)} \mu_i^q(\delta) \quad (2.60)$$

où  $N(\delta)$  est le nombre de boîtes de taille  $\delta$  qui intersectent l'ensemble  $A$ , et  $\mu_i(\delta) = \mu(B_i(\delta))$ . Si la mesure  $\mu$  est multifractale,  $Z$  suit une loi de puissance quand  $\delta \rightarrow 0$ , avec :

$$Z(q, \delta) \propto \delta^{\tau(q)} \quad (2.61)$$

où la fonction  $\tau(q)$  est la transformée de Legendre du spectre de singularité :

$$\begin{cases} f(\alpha) &= q\alpha - \tau(q) \\ \alpha &= \frac{d\tau(q)}{dq} \end{cases} \quad (2.62)$$

Dans le cas d'une monofractale, on a vu que  $\alpha = f(\alpha) = D$ , où  $D$  est la dimension fractale de  $\mu$ . La transformée de Legendre est alors égale à  $\tau(q) = D(q-1)$ , i.e. que  $\tau(q)$  est simplement une droite de pente égale à la dimension fractale.

### Dimensions de Rényi

La dimension fractale généralisée, aussi appelée dimension de Rényi d'ordre  $q$  est donnée par :

$$D_q = \frac{\tau(q)}{q-1} \quad (2.63)$$

$D_0$  correspond à la dimension de Hausdorff,  $D_1$  est appelée la dimension d'information, et  $D_2$  correspond à la dimension de corrélation.

En pratique, à cause des limites en résolution et du bruit de mesure, les différents exposants ne peuvent être calculés que pour une plage limitée d'échelles. Les dimensions de Rényi dépendent alors des échelles considérées. Pour la distribution des galaxies, on peut montrer que la fonction de partition peut être définie de façon équivalente :

$$Z(q, r) = \frac{1}{N} \sum_{i=0}^N n_i(r)^{q-1} \quad (2.64)$$

où  $n_i(r)$  est le nombre de voisins de la galaxies  $i$  à l'intérieur d'une sphère de rayon  $r$ . Les dimensions de Rényi avec une dépendance d'échelle  $D_q(r)$  sont alors définies comme :

$$D_q(r) = \frac{1}{q-1} \tau(q, r) = \frac{1}{q-1} \frac{d \ln Z(q, r)}{d \ln r} \quad (2.65)$$

Les dimensions de Rényi peuvent notamment être utilisées pour tester l'homogénéité de la distribution des galaxies à différentes échelles.

### 2.3.3 Fractalité à petite échelle et homogénéité à grande échelle de la distribution des galaxies

Alors qu'il est clair que le champ de matière est inhomogène aux petites échelles, l'homogénéité aux grandes échelles a été longtemps débattue. Cela était dû principalement au manque de données pouvant fournir une réponse définitive (voir [Davis \(1997\)](#) pour une discussion détaillée de cette question en 1996). C'est une question importante parce que le principe cosmologique, qui est à la base de la cosmologie moderne, suppose l'homogénéité de l'univers (au moins aux grandes échelles).

En raison de la fractalité de l'univers aux petites échelles, les premiers relevés de galaxies révélaient des structures toujours plus grandes et ne montraient pas de tendance à l'homogénéisation. Comme on le montre dans l'équation (4.6), la fonction de corrélation n'est pas une quantité bien définie pour une distribution fractale car elle suppose que l'on peut définir une densité moyenne  $\bar{n}$ , ce qui n'est pas le cas pour un champ fractal. Si l'on essaye quand même d'estimer la 'fonction de corrélation' d'un champ fractal, l'espérance de l'estimateur a une sévère dépendance avec le volume considéré. En particulier la longueur de corrélation  $r_0$ , définie telle que  $\xi(r_0) = 1$ , dépend linéairement de la taille du volume (voir équation (4.7)). Cette dépendance de  $r_0$  avec la taille de l'échantillon fut observée dans les premiers catalogues de galaxies, supportant l'idée d'un univers fractal. Un autre argument en faveur de la fractalité était donné par les galaxies plus lumineuses et les amas, observés comme étant plus fortement corrélés, et formant donc des plus grandes structures.

Bien que tous ces arguments soient valides, ils sont aussi expliqués dans un modèle cosmologique standard, avec une homogénéité à grande échelle. Les variations de  $r_0$  pour les relevés de petite taille sont expliqués par la fractalité à petite échelle, et les différences de corrélation entre différentes populations d'objets sont expliquées par le mécanisme de biais (voir section 2.2.2). En revanche il existe de solides arguments en faveur d'une homogénéisation à grand échelle qui contredisent un caractère purement fractal de l'univers. Ce phénomène est souvent appelé 'fin des grandeurs' (*End of Greatness*), i.e. qu'à des échelles suffisamment grandes on ne peut plus observer de structures cohérentes.

Présentons brièvement les nombreuses observations en faveur de cette vision standard de l'homogénéisation à grande échelle. Le premier type d'observations est fourni par les observations bidimensionnelles sur le ciel : comptage d'événements en rayons X, de sources radio, de sursauts gamma, CMB.

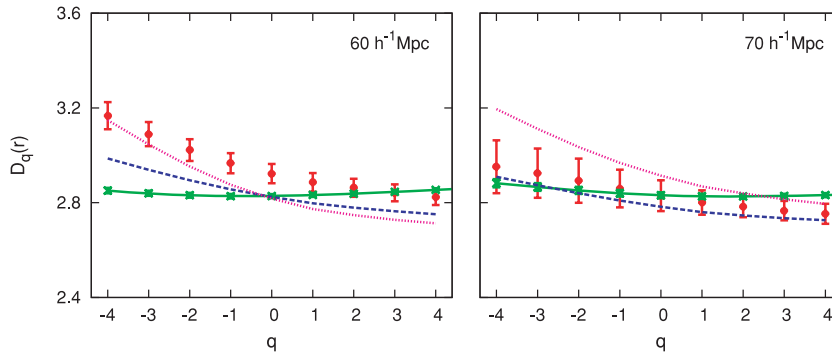


FIGURE 2.7 – Les dimensions de Rényi  $D_q(r)$  en fonction de  $q$  pour les échelles  $r = 60 h^{-1}\text{Mpc}$  (gauche) et  $r = 70 h^{-1}\text{Mpc}$  (droite). La figure montre les résultats sur l'échantillon LRG SDSS DR6 (points rouges), sur des catalogues aléatoires sans corrélation (ligne solide verte), sur des simulations  $N$ -corps de matière noire (tirets bleus) et sur un catalogue de galaxies de la simulation Millennium (pointillés roses). Les barres d'erreur à  $1\sigma$  de l'échantillon sont estimées à partir de 5 réalisations  $N$ -corps. Figure extraite de [Sarkar et al. \(2009\)](#).

Ces observations présentent toutes un degré d'isotropie remarquable. Si l'on suppose que notre position dans l'univers n'est pas spéciale (principe de Copernic), alors l'isotropie implique l'homogénéité.

Le second type d'observations est fourni par les observations tri-dimensionnelles. Un premier type de données 3D vient de la raie d'absorption de la forêt Lyman- $\alpha$  dans le spectre de quasars lointains, due à la présence de nuages d'hydrogène neutre sur la ligne de visée. On observe que ces nuages sont omniprésents et presque uniformément distribués dans l'univers.

Un autre type de données 3D provient des relevés de galaxies les plus récents comme le 2dFGRS, le SDSS ou WiggleZ. [Hogg et al. \(2005\)](#) étudie le nombre moyen  $\langle N(r) \rangle$  de *Luminous Red Galaxies* (LRG) du SDSS, sur une plage de *redshift*  $0.2 < z < 0.35$ , dans des sphères de rayon  $r$  centrées sur les LRG elles-mêmes. L'étude montre que ce nombre  $\langle N(r) \rangle \propto r^3$  pour les échelles  $r > 70 h^{-1}\text{Mpc}$ , ce qui n'est pas compatible avec un modèle purement fractale où l'on aurait  $\langle N(r) \rangle \propto r^D$ , avec  $D < 3$  la dimension fractale.

Certains travaux ont appliqué une autre approche basée sur le formalisme multifractal. La méthode consiste à étudier les dimensions de Rényi  $D_q(r)$  pour différentes échelles  $r$  (voir équation (2.65)). Si l'on trouve que ces exposants sont presque constants et égaux à la dimension euclidienne, cela veut dire que la distribution des galaxies est homogène à l'échelle considérée.

Dans [Yadav et al. \(2005\)](#), les exposants  $D_q(r)$  sont calculés sur des coupes 2D du SDSS DR1, ainsi que pour des simulations  $\Lambda\text{CDM}$  avec différents choix de biais mass-luminosité. Entre les échelles  $60 - 70 h^{-1}\text{Mpc}$  et  $150 h^{-1}\text{Mpc}$ , les exposants  $D_q(r)$  varient entre 1.7 et 2.2, i.e. sont proches de la dimension de l'espace 2D. [Sarkar et al. \(2009\)](#) fait la même analyse sur un échantillon LRG du relevé SDSS DR6, cette fois en utilisant le volume 3D pour réduire l'incertitude. Les résultats à deux échelles différentes  $r = 60 h^{-1}\text{Mpc}$  et  $r = 70 h^{-1}\text{Mpc}$  sont montrés sur la figure 2.7. A l'échelle  $r = 70 h^{-1}\text{Mpc}$ , les dimensions de Rényi sont presque constantes et proches de la valeur  $D = 3$ , ce qui est compatible avec une distribution aléatoire sans corrélation.

[Scrimgeour et al. \(2012\)](#) combine les deux approches (comptage de voisins dans des sphères  $N(r)$  et mesure de la dimension de corrélation  $D_2(r)$ ) avec le relevé plus récent WiggleZ ( $0.1 < z < 0.9$ ). Cette étude montre que les différentes mesures sont en accord excellent avec un modèle  $\Lambda\text{CDM} + \text{WMAP}$ , et que l'échelle d'homogénéité du champ de matière se situe entre  $72 h^{-1}\text{Mpc}$  et  $85 h^{-1}\text{Mpc}$  selon le *redshift* (voir figure 2.8). L'étude montre aussi que ces résultats sont incompatibles avec des distributions purement fractales de dimension  $D = 2.7$  à  $D = 2.95$ .

Il faut noter que tous ces résultats ne sont pas totalement indépendants de l'hypothèse d'homogénéité, car les catalogues 3D de galaxies sont construits à partir des catalogues en *redshift*, en supposant une métrique de Robertson-Walker. Cette complication est présente pour n'importe quelle mesure d'homogénéité car les observations sont nécessairement situées dans des cônes de lumière,

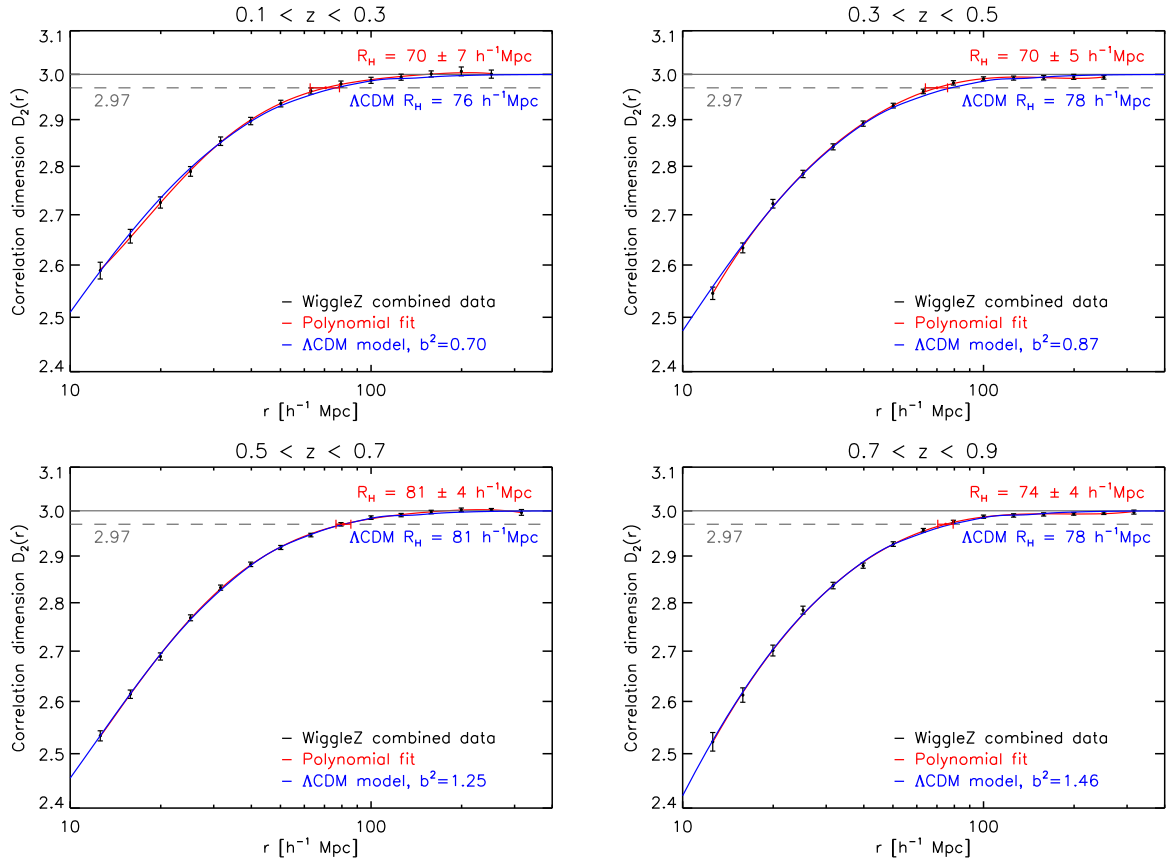


FIGURE 2.8 – La dimension de corrélation  $D_2(r)$  mesurée dans quatre régions de *redshift* pour le catalogue WiggleZ avec les barres d'erreur (noir) et pour un modèle  $\Lambda$ CDM+WMAP avec un biais  $b^2$  ajusté aux données (bleu). Interpolation polynomiale du 5<sup>ème</sup> degré de la courbe  $D_2(r)$  obtenue sur les données (rouge). L'échelle d'homogénéité  $R_H$  est définie comme l'échelle où  $D_2(r) = 2.97$  (i.e. homogénéité à 1%), à la fois pour l'interpolation polynomiale et pour le modèle  $\Lambda$ CDM. L'intervalle de confiance pour  $R_H$  est calculé à partir de réalisations  $\Lambda$ CDM log-normales. Figure extraite de Scrimgeour et al. (2012).

et non dans des coupes spatiales de l'univers. [Maartens \(2011\)](#) montre qu'il est donc impossible de faire une mesure d'homogénéité, sans faire des hypothèses comme une métrique de Robertson-Walker ou le principe cosmologique. Cependant, il serait très improbable que les observations soient en si bon accord avec un modèle  $\Lambda$ CDM si la distribution de matière était en réalité inhomogène et que l'hypothèse d'une métrique de Robertson-Walker était fausse.

Pour récapituler, la vision largement acceptée aujourd'hui est que l'univers a un comportement fractale aux petites échelles, tout en étant homogène aux grandes échelles.

# Chapitre 3

## Catalogues de galaxies et simulations

### Sommaire

---

<b>3.1 Relevés de galaxies</b> . . . . .	<b>51</b>
<b>3.2 Les catalogues SDSS-II étudiés</b> . . . . .	<b>53</b>
3.2.1 Les catalogues Main VL et LRG VL . . . . .	54
3.2.2 Le catalogue DR7-Full . . . . .	56
<b>3.3 Les catalogues SDSS-III</b> . . . . .	<b>60</b>
3.3.1 Le catalogue BOSS DR9 CMASS . . . . .	60
<b>3.4 Simulations</b> . . . . .	<b>62</b>
3.4.1 Simulations de segments de Cox . . . . .	62
3.4.2 Simulations log-normales . . . . .	63
3.4.3 Simulations log-normales pour différentes valeurs de $\alpha$ simultanément . . . . .	68
3.4.4 Simulations $N$ -corps . . . . .	69

---

### 3.1 Relevés de galaxies

Les relevés de galaxies constituent un des principaux outils de la cosmologie observationnelle. Le principe de base de ces relevés est d’obtenir des informations observationnelles sur les galaxies occupant une certaine région du ciel. Le critère de sélection des galaxies est habituellement donné par la magnitude apparente dans une bande donnée, mais il peut y avoir des critères de sélection plus complexes. Le but des relevés de galaxies est de fournir un échantillon d’une population de galaxies afin d’étudier ses propriétés statistiques. Ils fournissent des informations à la fois sur les processus de formation et d’évolution des galaxies, et sur les propriétés de la distribution spatiale des galaxies.

Les premiers relevés de galaxies tels que les relevés du *Lick Observatory* ou du *Palomar observatory* fournissaient simplement la position angulaire des galaxies. C’est au cours des années 1980 que sont apparus les premiers relevés de galaxies spectroscopiques. L’utilisation de la spectroscopie permet d’obtenir une mesure précise du *redshift* des galaxies, qui peuvent ensuite être utilisés pour calculer les distances radiales pour un choix donné de cosmologie fiducielle. Le premier relevé spectroscopique a été produit au *Center for Astrophysics* (CfA, [Huchra et al. \(1983\)](#)) et de nombreux autres ont suivi, comme le *Southern Sky Redshift Survey* (SSRS, [da Costa et al. \(1991\)](#)), le *Las Campanas Redshift Survey* (LCRS, [Sheth et al. \(1996\)](#)) ou le *Point Source Catalogue Survey* (PSCz, [Saunders et al. \(2000\)](#)).

Des relevés de galaxies plus récents ont continué à cartographier la distribution des galaxies dans des volumes toujours plus grands. Le *2 degree Field Galaxy Redshift Survey* (2dFGRS) a permis d’obtenir le *redshift* de  $\approx 250,000$  galaxies sur  $\approx 1,500$  degrés carrés jusqu’à *redshift*  $z = 0.3$ . Le *Sloan Digital Sky Survey* (SDSS) a permis d’obtenir le *redshift* de presque 1 million de galaxies sur une

surface de  $\approx 8,000$  degrés carrés jusqu'à une *redshift*  $z = 0.5$ . Ces deux relevés ont permis la première détection des Oscillations Baryoniques Acoustiques dans la distribution des galaxies, respectivement dans Eisenstein et al. (2005) et Cole et al. (2005) (voir section 5.1.2). Sur la figure 3.1 on montre deux coupes extraites respectivement du 2dFGRS et du SDSS, comparées à une coupe du relevé CfA2. On peut voir la grande évolution des relevés de galaxies sur une échelle de temps d'environ vingt ans.

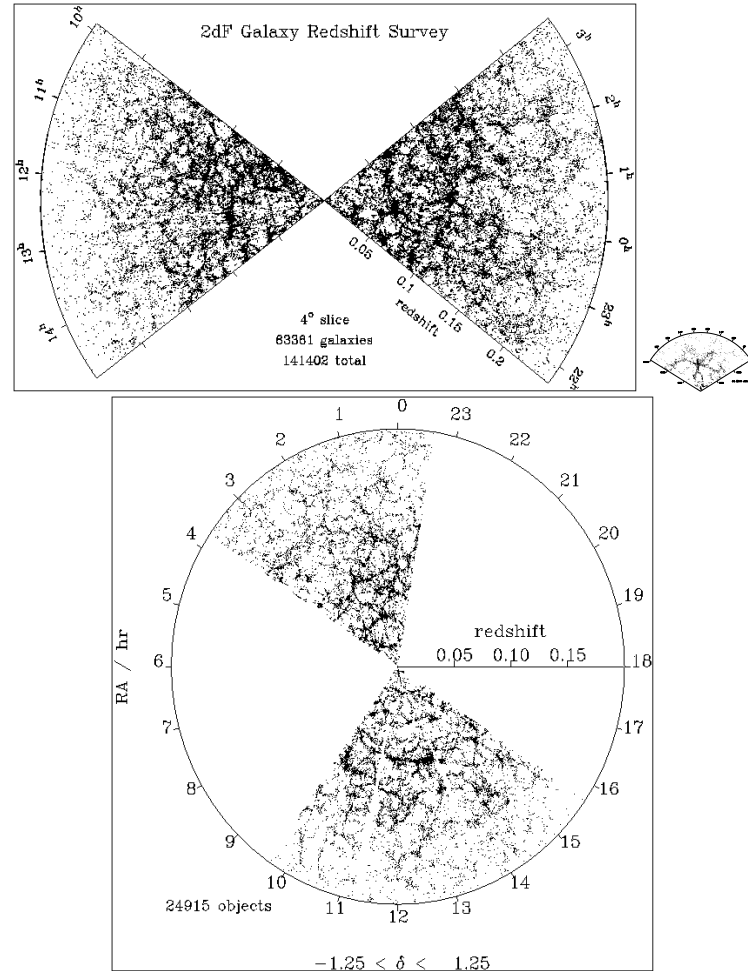


FIGURE 3.1 – Coupes de différents relevés de galaxies : le CfA2 (droite), le 2dFGRS (haut) et le SDSS (bas). La figure donne une idée de la profondeur des différents relevés. On peut voir les caractéristiques générales de la distribution de galaxies qui forme une toile cosmique. Figure extraite de Jones et al. (2004).

Parmi les relevés de galaxies récents, on peut citer deux autres relevés de galaxies qui se distinguent par des *redshifts* différents. Le 6dFGRS a permis d'obtenir le spectre d'environ 100,000 galaxies sur une grande surface de 17,000 degrés carrés, mais à faible *redshift*  $z < 0.2$ . D'un autre côté le relevé WiggleZ a obtenu le spectre de 200,000 galaxies sur une surface d'environ 1,000 degrés carrés mais à plus haut *redshift*  $z < 1$ . Il est intéressant de pouvoir étudier des relevés de galaxies à différents *redshifts*, notamment pour l'étude des oscillations baryoniques acoustiques. En effet ceux-ci permettent de contraindre l'échelle de dilatation  $D_V(z)$  (voir figure 3.2 et section 2.2.5) à différents *redshifts*, ce qui est très important pour pouvoir étudier l'énergie noire (et notamment une possible évolution de son équation d'état).

Enfin dans les relevés de galaxies en cours, on peut citer le relevé de galaxies BOSS qui s'intègre au projet SDSS-III, faisant suite au SDSS-II. Il devrait cartographier 1,5 million de galaxies sur une

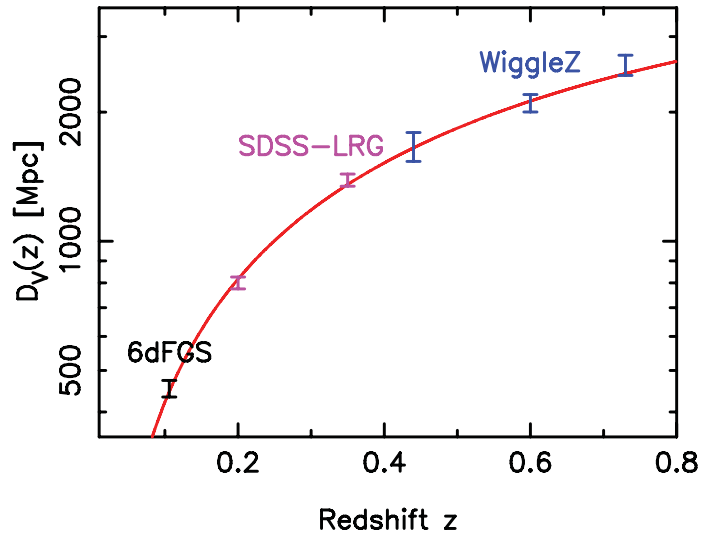


FIGURE 3.2 – Mesures actuelles de l'échelle de dilatation  $D_V(z)$  en utilisant la propriété d'étalon standard des BAOs avec les relevés 6dFGRS, SDSS et WiggleZ. Les résultats sont comparés à un modèle cosmologique  $\Lambda$ CDM plat avec  $\Omega_m = 0.29$  et un paramètre de Hubble  $h = 0.69$ . Figure extraite de [Blake et al. \(2011b\)](#).

surface de 10,000 degrés carrés jusqu'à *redshift*  $z = 0.7$ . La première publication de données DR9, a eu lieu en août 2012, et a fourni le spectre de 540,000 nouvelles galaxies à *redshift* médian  $z = 0.52$  par rapport au SDSS-II. Les observations vont continuer jusqu'en 2014 et devraient fournir des contraintes de 1.5% à 1.7% sur la mesure de  $D_A(z = 0.6)$  et  $H(z = 0.6)$  respectivement.

## 3.2 Les catalogues SDSS-II étudiés

Les observations du relevé SDSS-I ont été faites entre 2000 à 2005, et se sont prolongées avec les observations du SDSS-II jusqu'en 2008. Elles utilisent un télescope de 2.5 mètres de diamètre au *Apache Point Observatory* (APO) au Nouveau-México (Etats-Unis), équipé d'un caméra et de deux spectrographes, spécialement conçus pour cette application. Le projet SDSS consiste en de nombreux relevés complémentaires, mais on va simplement détailler le relevé connu sous le nom de *Legacy*. Les données finales de ce relevé ont été publiées dans la *Data Release 7* (DR7, [Abazajian et al. \(2009\)](#)).

Le relevé global comprend des observations photométriques dans 5 bandes optiques ( $u$ ,  $g$ ,  $r$ ,  $i$  et  $z$ ), ainsi que des observations spectroscopiques qui utilisent les observations photométriques pour la sélection des galaxies cibles. La surface total du relevé est de 8000 degrés carrés, dont la majeure partie correspond à une large zone contiguë dans l'hémisphère galactique nord, et le reste correspond à trois bandes dans l'hémisphère sud. Le relevé est composé de deux catalogues de galaxies, le catalogue 'Main' et le catalogue de *Luminous Red Galaxies*.

Le catalogue de galaxies 'Main' est un échantillon *magnitude-limited* (i.e. avec une sélection des galaxies basée sur une coupe en magnitude apparente), avec un critère de sélection sur la bande de magnitude  $r$ ,  $r < 17.77$  ([Strauss et al. 2002](#)). Cet échantillon s'étend jusqu'à *redshift*  $z \approx 0.3$  avec un *redshift* médian  $z_m \approx 0.1$  et fournit un spectre fiable pour  $\approx 560,000$  galaxies.

Un second catalogue est constitué de *Luminous Red Galaxies* (LRG), qui sont sélectionnés avec un ensemble de coupes en magnitude et en couleur ([Eisenstein et al. 2001](#)). Les galaxies sélectionnées avec cet algorithme sont situées principalement aux centres de halos de matière noire massifs, sous l'hypothèse d'un modèle HOD ([Zheng et al. 2009](#)), et sont fortement biaisés par rapport au champ de matière (voir section 2.2.2). L'échantillon est approximativement *volume-limited* (i.e. de densité constante) jusqu'à *redshift*  $z \approx 0.36$  et s'étend de façon *magnitude-limited* jusqu'à *redshift*  $z \approx 0.47$ . L'échantillon contient au total  $\approx 100,000$  galaxies (en incluant les galaxies à faible *redshift* qui sont aussi présentes dans le 'Main'), et il est particulièrement bien adapté à l'étude de la fonction de



corrélation grâce au large volume couvert (plus de  $1 h^{-1} \text{Gpc}^3$ ).

Donnons maintenant plus précisément les caractéristiques des différents catalogues que l'on va utiliser dans cette thèse.

### 3.2.1 Les catalogues Main VL et LRG VL

On détaille dans cette section les caractéristiques des échantillons SDSS *volume-limited* que l'on va utiliser dans le chapitre 4.

#### Construction des catalogues Main VL et LRG VL

Pour construire un catalogue *volume-limited* à partir de l'échantillon 'Main' global, on s'est servi du catalogue accessible sur la page du programme Mangle<sup>1</sup> (Hamilton and Tegmark 2004). Ce catalogue est basé sur le *New York University Value-Added Galaxy Catalog* (Blanton et al. 2005). Il fournit les magnitudes absolues  $M_r$  dans la bande  $r$  pour toutes les galaxies, sur lesquelles ont déjà été appliquées les  $K$ -corrections<sup>2</sup> et les corrections d'évolution<sup>3</sup> à *redshift*  $z = 0.1$  suivant Blanton et al. (2003). Les distances comobiles et les magnitudes absolues étant calculées dans la cosmologie fiducielle du programme Mangle, on les a converties dans la cosmologie fiducielle que l'on a choisi pour l'étude de ces catalogues *volume-limited*, à savoir une cosmologie  $\Lambda$ CDM plate avec  $\Omega_m = 0.27$  (i.e.  $\Omega_\Lambda = 0.73$ ).

On a aussi construit un échantillon *volume-limited* de galaxies LRG, directement à partir des données du SDSS DR7 accessibles en ligne<sup>4</sup>. On a calculé les magnitudes absolues  $M_g$  dans la bande  $g$ ,  $K$ -corrigées en suivant la méthode décrite dans Eisenstein et al. (2001).

Dans chaque cas, on obtient des échantillons *volume-limited* en procédant à une coupe en magnitudes absolue, puis en procédant à des coupes en *redshift* de sorte que la densité des échantillons reste approximativement constante.

Enfin on restreint les échantillons à une région du ciel qui est presque complète en dehors de quelques zones qui sont masquées par des étoiles brillantes. Pour cela on coupe les échantillons dans le système de coordonnées du relevé  $(\eta, \lambda)$  avec les limites  $-31.25^\circ < \eta < 28.75^\circ$  et  $-54.8^\circ < \lambda < 51.8^\circ$ , correspondant à une grande part de la partie contiguë de l'hémisphère nord. A cause de ce choix, on perd de l'information pour estimer la fonction de corrélation (i.e. que les barres d'erreur deviennent un peu plus grandes), mais des simulations de ces échantillons deviennent bien plus simples à générer. On montre la couverture angulaire de ces catalogues sur la figure 3.3. Celle-ci est entièrement incluse dans l'hémisphère galactique nord (la ligne bleue représente le plan galactique).

L'échantillon *volume-limited* Main VL que l'on a construit avec cette procédure est similaire à l'échantillon correspondant à  $M_r < -21$  dans Zehavi et al. (2005), et l'échantillon LRG VL est similaire à celui utilisé dans Martínez et al. (2009). On donne dans la table 3.1 les limites en magnitude et en *redshift* utilisées pour construire les deux échantillons. On donne aussi pour chaque échantillon le nombre total de galaxies  $N_g$ , le volume  $V$  et la densité moyenne  $\bar{n}$  (ces deux dernières quantités étant calculées pour notre cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.27$ ).

#### Fonction de corrélation du Main VL et du LRG VL

On montre ici les fonction de corrélation estimées sur les échantillons Main VL et LRG VL. On utilise l'estimateur de Landy-Szalay (voir section 2.2.4) et des catalogues de Poisson aléatoires de 100,000 points qui suivent précisément la couverture angulaire des catalogues, i.e. qui prennent en compte les quelques zones masquées par des étoiles brillantes. Pour générer ces catalogues aléatoires, on a généré des catalogues suivant la couverture angulaire globale du SDSS DR7 avec le programme Mangle, auxquels on a appliqué les mêmes coupes angulaires que nos catalogues  $-31.25^\circ < \eta < 28.75^\circ$  et  $-54.8^\circ < \lambda < 51.8^\circ$ .

1. <http://space.mit.edu/~molly/mangle/>

2. La  $K$ -correction 'corrige' le fait que les différentes sources sont observées à des *redshifts* différents. Plus précisément, elle convertit les spectres du référentiel d'observation vers le référentiel d'émission (voir Hogg et al. (2002)).

3. Les corrections d'évolution permettent de prendre en compte l'évolution temporelle des galaxies (et donc de leur spectre), afin de convertir les spectres de l'époque correspondant au *redshift* observé à une époque commune pour toutes les galaxies.

4. <http://cas.sdss.org/astrodr7/en/>

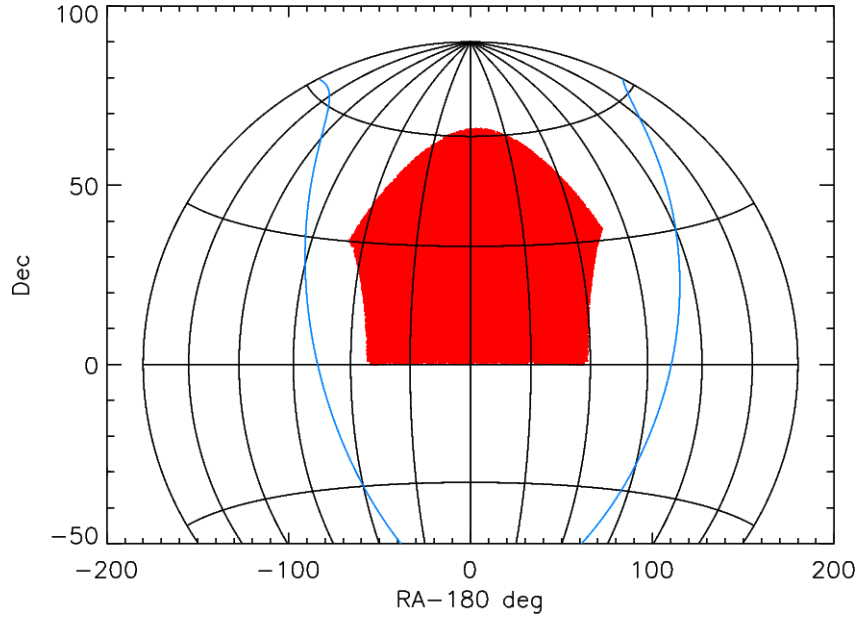


FIGURE 3.3 – Couverture angulaire des catalogues Main VL et LRG VL en projection d’Aitoff. La ligne bleue représente le plan galactique qui sépare l’hémisphère galactique nord de l’hémisphère galactique sud.

TABLE 3.1 –

	Limites de magnitude	Limites de <i>redshift</i>	Surface (deg <sup>2</sup> )	$N_g$	$V$ ( $h^{-1}$ Gpc) <sup>3</sup>	$\bar{n}$ ( $h^{-1}$ Mpc) <sup>-3</sup>
Main VL	$M_r < -21$	$0.059 < z < 0.168$	4,809	67,189	$61.2 \times 10^{-3}$	$1.098 \times 10^{-3}$
LRG VL	$-23.544 < M_g < -21.644$	$0.14 < z < 0.42$	4,809	34,347	0.79	$4.345 \times 10^{-5}$

NOTES.—Caractéristiques des échantillons SDSS *volume-limited*, Main VL et LRG VL. Le volume ainsi que la densité moyenne sont données pour le choix d’une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.27$ .

Sur la figure 3.4 on montre les fonctions de corrélation estimées du Main VL et du LRG VL. Dans chaque cas, on obtient les barres d'erreur à partir de 2,000 simulations log-normales  $\Lambda$ CDM avec les paramètres  $\Omega_m = 0.27$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 2.205^{-2}$ ,  $n_s = 1$ ,  $\sigma_8 = 0.8$  et respectivement  $b = 1.65$  pour le Main VL et  $b = 2.5$  pour le LRG VL (ces simulations sont présentées plus en détails dans le chapitre 4).

On peut faire plusieurs remarques par rapport à la figure 3.4. D'une part, la fonction de corrélation du Main VL a une amplitude plus faible que celle du LRG VL, comme on peut le voir clairement aux petites échelles. Ceci confirme notamment que le biais masse-luminosité augmente avec la luminosité des galaxies (voir section 2.2.2). D'autre part, l'estimation de la fonction de corrélation est plus bruitée dans le cas du Main VL que dans le cas du LRG VL (i.e. les barres d'erreur sont plus grandes). Or comme on vient de le dire, la corrélation est plus importante dans le cas du LRG VL, ce qui devrait causer une plus grande incertitude sur l'estimation de la fonction de corrélation pour un même volume. Cela veut donc dire que le plus grand volume de l'échantillon LRG VL compense cet effet.

On peut aussi voir que les fonctions de corrélation estimées tendent vers 0 aux grandes échelles, ce qui est compatible avec une homogénéisation à grande échelle.

Quant au pic BAO, il est difficilement visible pour le Main VL, notamment à cause de la taille des barres d'erreur. Le pic BAO est plus visible pour le LRG VL, mais il semble trop large par rapport à ce qui est attendu. Ceci est également observé dans [Martínez et al. \(2009\)](#). On verra dans la section 3.2.2 que cet effet se retrouve sur l'échantillon SDSS DR7-Full, mais qu'il ne semble pas être dû à des effets systématiques dans l'estimation.

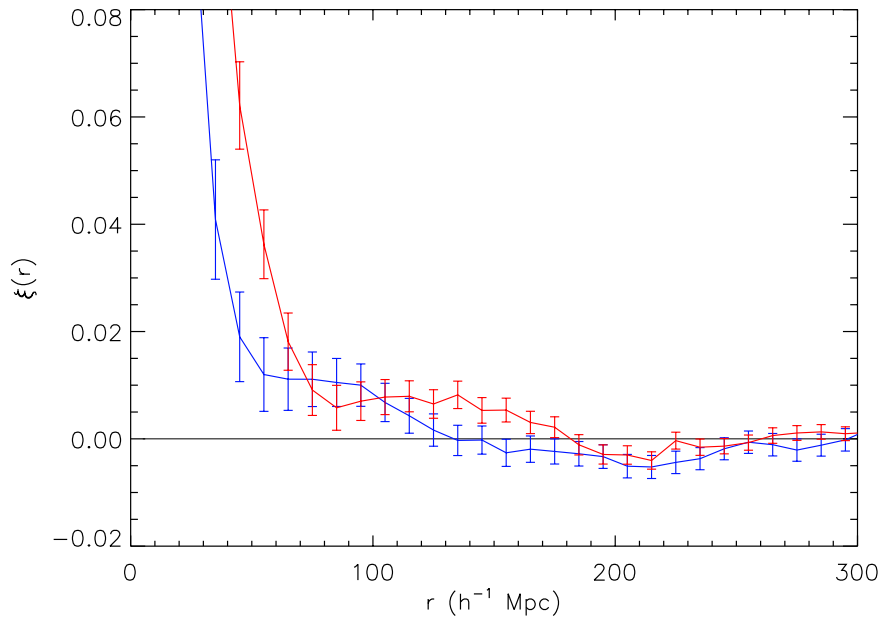


FIGURE 3.4 – Fonctions de corrélation estimées du catalogue Main VL (bleu) et du LRG VL (rouge) pour une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.27$ . Les barres d'erreur sont estimées à partir de 2,000 simulations log-normales  $\Lambda$ CDM avec les paramètres  $\Omega_m = 0.27$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 2.205^{-2}$ ,  $n_s = 1$ ,  $\sigma_8 = 0.8$  et respectivement  $b = 1.65$  pour le Main VL et  $b = 2.5$  pour le LRG VL (ces simulations sont présentées plus en détails dans le chapitre 4).

### 3.2.2 Le catalogue DR7-Full

Dans cette section on détaille les caractéristiques de l'échantillon SDSS DR7-Full, que l'on va utiliser dans les chapitres 5 et 6.

### Construction du catalogue DR7-Full

Le catalogue SDSS DR7-Full que l'on utilise provient de l'étude [Kazin et al. \(2010\)](#) et est disponible en ligne<sup>5</sup>. Il est constitué d'une grande partie des galaxies LRG du relevé SDSS DR7. En plus de la sélection initiale des LRG provenant de [Eisenstein et al. \(2001\)](#), il y a simplement une contrainte sur la complétude des secteurs<sup>6</sup> inclus (les secteurs doivent avoir une complétude de plus de 60%) ainsi qu'un critère sur la magnitude absolue  $M_g$  dans la bande  $g$ . Le catalogue obtenu est quasiment *volume-limited* jusqu'à *redshift*  $z \approx 0.36$  et s'étend de façon *magnitude-limited* jusqu'à  $z \approx 0.47$ . Pour cet échantillon, on utilise la même cosmologie fiducielle que [Kazin et al. \(2010\)](#), à savoir une cosmologie  $\Lambda$ CDM plate avec  $\Omega_m = 0.25$  (i.e.  $\Omega_\Lambda = 0.75$ ).

Sur la figure 3.5 on montre la densité moyenne estimée  $\bar{n}(z)$  du catalogue en fonction du *redshift* pour notre choix de cosmologie fiducielle. L'échantillon est principalement contiguë, avec seulement 9.8% en dehors de la partie principale dans l'hémisphère galactique nord. Le nombre de LRG est égal à 96,763 dans l'hémisphère nord et 9,068 dans l'hémisphère sud. On montre la couverture angulaire du catalogue sur la figure 3.6, où l'on voit la partie contiguë de l'hémisphère nord ainsi que les quelques bandes dans l'hémisphère sud (la ligne bleue représente le plan galactique).

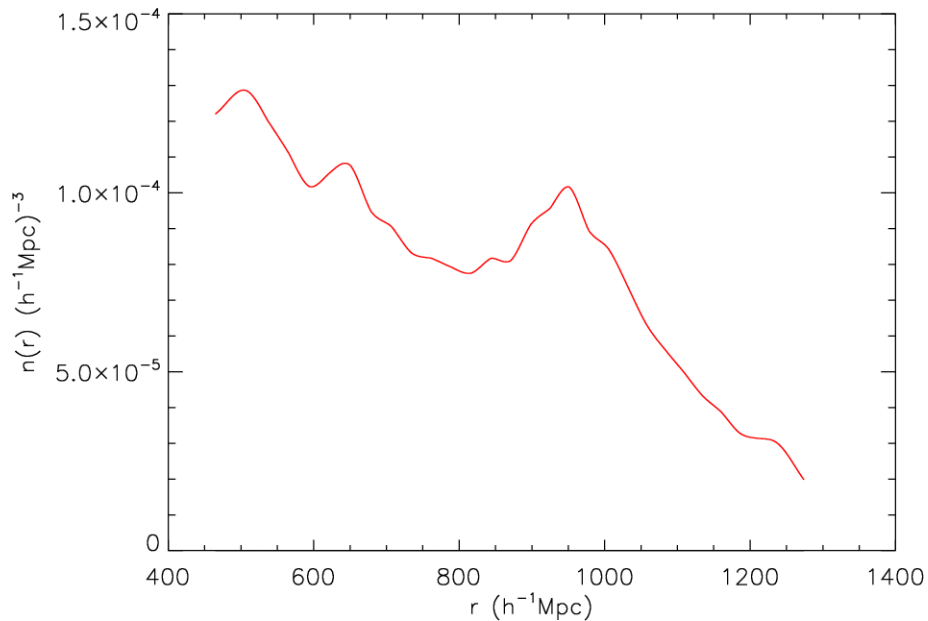


FIGURE 3.5 – Densité moyenne estimée  $\bar{n}(z)$  du catalogue DR7-Full en fonction du *redshift* pour le choix d'une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.25$ .

On donne les caractéristiques de l'échantillon dans la table 3.2.

### Fonction de corrélation du DR7-Full

Pour estimer la fonction de corrélation du catalogue SDSS DR7-Full on utilise l'estimateur de Landy-Szalay (voir section 2.2.4) et le même catalogue aléatoire de Poisson que dans [Kazin et al. \(2010\)](#), qui est aussi disponible en ligne<sup>7</sup>. Ce catalogue comprend  $\approx 1.66$  million de points, i.e.  $\approx 16$  fois le nombre de galaxies dans le catalogue de données.

Dans le cas du catalogue DR7-Full où la densité moyenne  $\bar{n}(z)$  dépend du *redshift*, on a vu dans la section 2.2.4 que l'on pouvait choisir une fonction de poids optimale  $w_i$  qui dépend du *redshift*  $z_i$  de

5. <http://cosmo.nyu.edu/~eak306/SDSS-LRG.html>

6. Un secteur est défini comme une entité élémentaire sur la couverture angulaire du relevé, i.e. comme l'intersection de plusieurs fenêtres d'observation élémentaires du télescope.

7. <http://cosmo.nyu.edu/~eak306/SDSS-LRG.html>

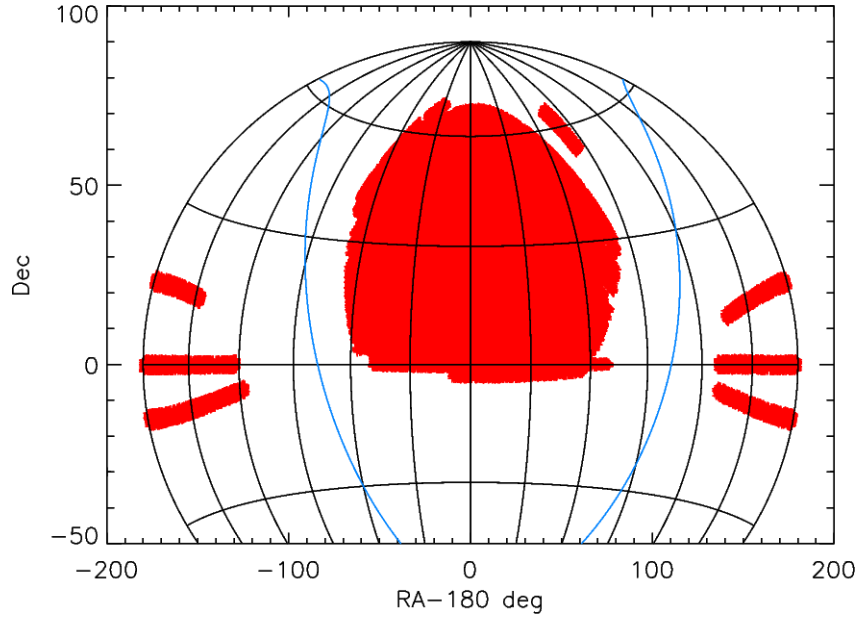


FIGURE 3.6 – Couverture angulaire du catalogue DR7-Full en projection d’Aitoff. La ligne bleue représente le plan galactique qui sépare l’hémisphère galactique nord de l’hémisphère galactique sud.

TABLE 3.2 –

	Limites de magnitude	Limites de <i>redshift</i>	Surface (deg <sup>2</sup> )	$N_g$	$V$ ( $h^{-1}$ Gpc) <sup>3</sup>	$\bar{n}$ ( $h^{-1}$ Mpc) <sup>-3</sup>
DR7-Full	$-23.2 < M_g < -21.2$	$0.16 < z < 0.47$	6,918	105,831	1.58	$6.70 \times 10^{-5}$

NOTES.—Caractéristiques de l’échantillon SDSS DR7-Full de [Kazin et al. \(2010\)](#). Le volume ainsi que la densité moyenne sont données pour le choix d’une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.25$ .

la galaxie  $i$ . Ici on a fait le choix de ne pas appliquer ces poids optimaux pour les raisons suivantes : l'effet de ces poids reste faible dans l'estimation de la fonction de corrélation et l'introduction de ces poids rendrait bien plus complexe l'estimation de la fonction de corrélation dans les simulations log-normales que l'on va utiliser dans les chapitres 5 et 6.

On doit toutefois tenir compte du problème des collisions de fibres<sup>8</sup> (*fiber collisions*), qui intervient dans le relevé SDSS. Pour cela on utilise les mêmes poids que dans Kazin et al. (2010), i.e. qu'on augmente le poids des galaxies ayant un voisinage avec des galaxies non spectrographiées à cause des collisions de fibres. Concernant l'incomplétude angulaire du catalogue ainsi que les variations de  $\bar{n}(z)$ , elles sont prises en compte dans le catalogue aléatoire, si bien que les densités moyennes du catalogue de données et du catalogue aléatoires sont proportionnelles (ce qui est requis pour une estimation correcte de la fonction de corrélation, voir section 2.2.4).

Sur la figure 3.7 on montre la fonction de corrélation estimée du catalogue DR7-Full. Les barres d'erreur sont obtenues à partir de 2,000 simulations log-normales avec les paramètres  $\Omega_m = 0.25$ ,  $h = 0.72$ ,  $\Omega_b h^2 = 2.27 \times 10^{-2}$ ,  $n_s = 0.966$ ,  $\sigma_8 = 0.81$  et  $b = 2.5$  (ces simulations sont présentées plus en détails dans le chapitre 5). Comme pour le catalogue LRG VL, on remarque que le pic BAO est plus large que ce qui est attendu. Toutefois Kazin et al. (2010) conclut que cet effet n'est pas dû à des systématiques, mais seulement à la variance d'échantillon. Notons que les relevés ultérieurs, WiggleZ (Blake et al. 2011a), 6dF (Beutler et al. 2011) ainsi que BOSS (Sánchez et al. 2012) n'ont pas confirmé cette caractéristique.

Notons aussi que la technique de reconstruction des BAOs (*BAO reconstruction*, Eisenstein et al. (2007a)) utilisée dans Padmanabhan et al. (2012) aboutit à un rétrécissement du pic BAO sur ce même catalogue. Cependant, sans appliquer cette technique et sans introduire des paramètres de nuisance pour la forme globale de la fonction de corrélation sur lesquels on peut marginaliser (e.g. des fonctions spline dans Percival et al. (2010) ou des inverses de polynôme dans Xu et al. (2012)), la largeur du pic BAO induit un faible niveau de détection, ainsi qu'un déplacement vers des valeurs  $\alpha < 1$  (voir chapitre 6).

---

8. Ce problème vient du fait que le spectrographe ne peut acquérir en même temps le spectre de 2 objets qui sont séparés d'un angle inférieur à  $55^\circ$ , dû à la largeur des fibres du spectrographe.

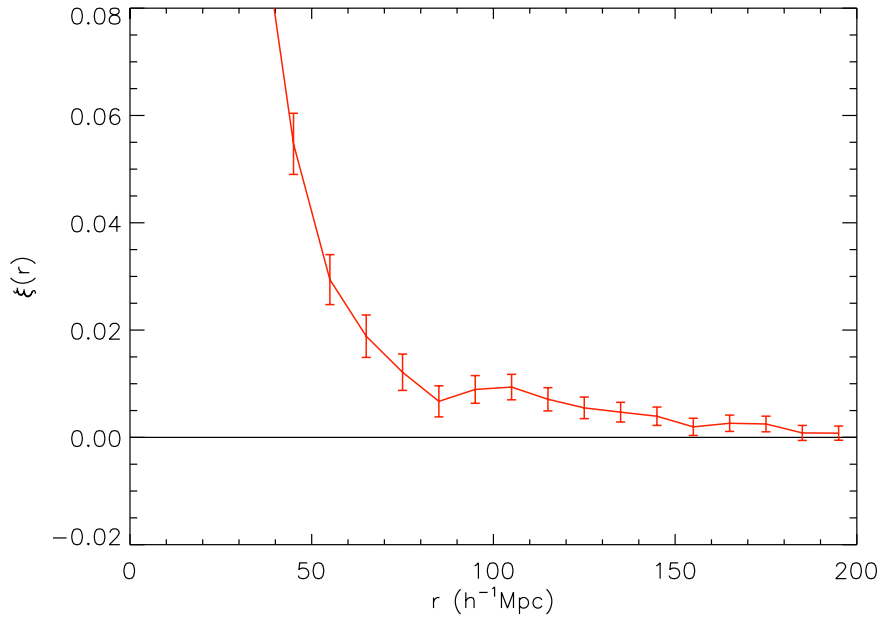


FIGURE 3.7 – Fonction de corrélation estimée du catalogue DR7-Full pour une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.25$ . Les barres d'erreur sont obtenues à partir de 2,000 simulations log-normales  $\Lambda$ CDM avec les paramètres  $\Omega_m = 0.25$ ,  $h = 0.72$ ,  $\Omega_b h^2 = 2.27 \times 10^{-2}$ ,  $n_s = 0.966$ ,  $\sigma_8 = 0.81$  et  $b = 2.5$  (ces simulations sont présentées plus en détails dans le chapitre 5). Le pic BAO est un peu plus large que ce qui est attendu, ce qui est expliqué par la variance d'échantillon dans Kazin et al. (2010).

### 3.3 Les catalogues SDSS-III

#### 3.3.1 Le catalogue BOSS DR9 CMASS

Le projet SDSS-II est suivi du projet SDSS-III, incluant notamment le relevé de galaxies BOSS. Celui-ci étend les relevés du SDSS-II à des plus hauts *redshifts*, et se base sur des améliorations de la caméra ainsi que du spectrographe. On décrit ici les résultats obtenus avec l'échantillon DR9 CMASS dans les études Anderson et al. (2012) et Sánchez et al. (2012) auxquelles on a participé.

#### Construction du catalogue BOSS DR9 CMASS

L'échantillon CMASS est un des deux échantillons du relevé BOSS, qui ont chacun des critères de sélection spécifiques pour les galaxies cibles (Eisenstein et al. 2011). Il est dominé par des 'galaxies précoces' (*early type galaxies*), bien qu'il contienne une fraction significative de galaxies spirales massives ( $\approx 26\%$ , Masters et al. (2011)). La plupart des galaxies de cet échantillon sont situées au centre de halos de matière noire avec  $\approx 10\%$  qui sont des galaxies satellites (White et al. 2011).

Anderson et al. (2012) présente une description détaillée de la construction d'un catalogue basé sur cet échantillon, ainsi que du calcul de la complétude dans chaque secteur angulaire, pour l'étude de la fonction de corrélation. Le catalogue inclut seulement les secteurs avec une complétude plus grande que 75%. De plus il est restreint à des limites en *redshift*  $0.43 < z < 0.7$ , ce qui donne un catalogue final composé de  $\approx 262,000$  galaxies avec  $\approx 206,000$  galaxies dans l'hémisphère galactique nord et  $\approx 56,000$  galaxies dans l'hémisphère galactique sud.

Sur la figure 3.8 on montre la densité moyenne estimée  $\bar{n}(z)$  du catalogue en fonction du *redshift* pour une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.274$ , avec une comparaison avec l'échantillon SDSS DR7-Full. On peut voir que le relevé BOSS sonde une plage de *redshifts* plus élevés, qui est quasiment distincte du SDSS-II. On montre la couverture angulaire du catalogue en coordonnées

galactiques sur la figure 3.9, avec l'hémisphère nord (gauche) et l'hémisphère sud (droite), ainsi que les couleurs associées à la complétude des différents secteurs.

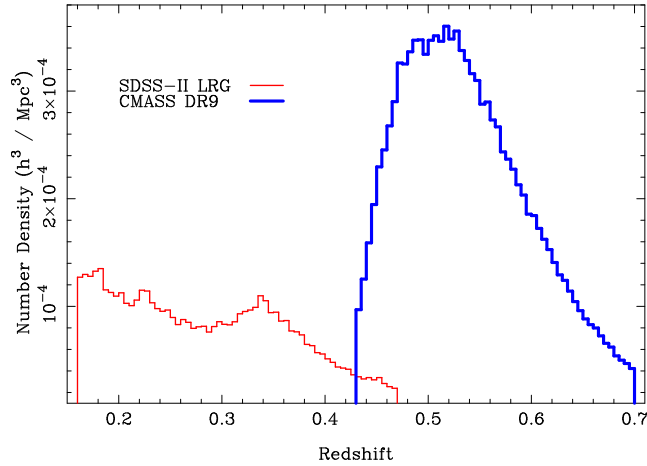


FIGURE 3.8 – Densité moyenne estimée  $\bar{n}(z)$  du catalogue DR9 CMASS (bleu) qui s'étend entre les *redshifts*  $0.43 < z < 0.7$ , pour une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.274$ . On montre en comparaison la densité du catalogue DR7-Full (rouge) qui s'étend entre les *redshifts*  $0.16 < z < 0.47$ . Figure extraite de [Anderson et al. \(2012\)](#).

### Fonction de corrélation du BOSS DR9 CMASS

La fonction de corrélation est calculée avec l'estimateur de Landy-Szalay en utilisant 70 fois plus de points dans le catalogue aléatoire que dans le catalogue de données, et des poids presque optimaux (voir section 2.2.4) :

$$w_i = \frac{1}{1 + \bar{n}(z_i)P_0} \quad (3.1)$$

avec une valeur  $P_0$  indépendante de l'échelle  $P_0 = 2 \times 10^4 h^{-3} \text{Mpc}^3$ . L'effet des collisions de fibres est pris en compte de façon similaire à la procédure décrite pour l'échantillon DR7-Full. Le volume effectif de l'échantillon peut être estimé de la façon suivante :

$$V_{\text{eff}} = \int_{\mathcal{V}} \left( \frac{\bar{n}(\mathbf{r})P_0}{1 + \bar{n}(\mathbf{r})P_0} \right)^2 d^3\mathbf{r} \quad (3.2)$$

Le concept de volume effectif a été introduit à partir de l'étude [Feldman et al. \(1994\)](#). De la même manière que [Hamilton \(1993\)](#) a calculé les poids optimaux à affecter aux galaxies pour estimer la fonction de corrélation, [Feldman et al. \(1994\)](#) a calculé les poids optimaux à affecter aux galaxies pour estimer le spectre de puissance sous l'approximation d'un champ de galaxies Gaussien. Il est intéressant de noter que ces deux études aboutissent à la même formule de poids optimaux. De plus [Feldman et al. \(1994\)](#) montre que la variance de l'estimateur du spectre de puissance varie comme  $1/V_{\text{eff}}$ , d'où l'appellation de volume effectif.

On peut noter que le volume effectif est toujours plus petit que le vrai volume, et pour des valeurs de  $\bar{n}(\mathbf{r})P_0$  assez grandes, le volume effectif coïncide avec le vrai volume. Le volume effectif donne une bonne idée du contenu statistique du catalogue. On obtient pour l'échantillon CMASS DR9 un volume effectif égal à  $V_{\text{eff}} = 2.2 \text{Gpc}^3$ .

On obtient la fonction de corrélation estimée montrée sur la figure 3.10. Les barres d'erreur sont obtenues à partir de  $N = 600$  simulations générées avec une méthode similaire à PTHalos ([Scoccimarro and Sheth 2002](#)). Ces simulations ont la même fonction de sélection que le catalogue de données et les



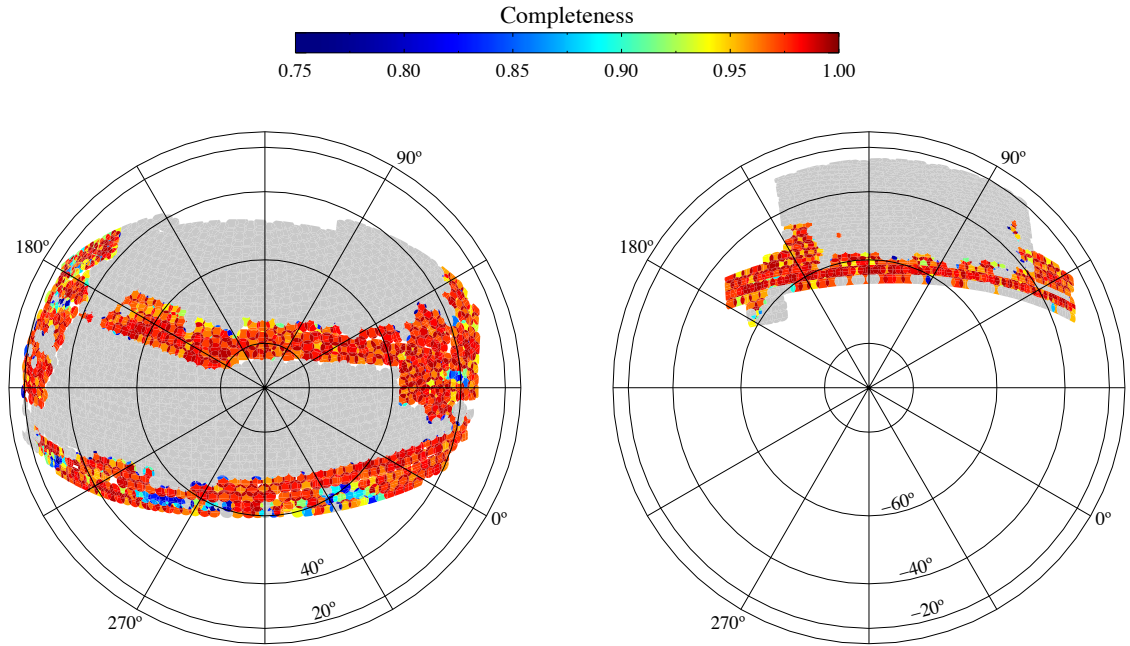


FIGURE 3.9 – La couverture angulaire du catalogue DR9 CMASS en coordonnées galactiques, avec l’hémisphère nord (à gauche) et l’hémisphère sud (à droite). Les différents secteurs sont colorés en fonction de leur complétude. La zone colorée en gris montre la couverture angulaire attendue à la fin du relevé BOSS, correspondant à  $10,269 \text{ deg}^2$ . Figure extraite de [Sánchez et al. \(2012\)](#).

paramètres cosmologiques  $\Omega_m = 0.274$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 0.0224$ ,  $n_s = 0.95$  et  $\sigma_8 = 0.8$  (voir [Manera et al. \(2012\)](#) pour plus de détails sur ces simulations).

Le pic BAO est très clairement visible sur la figure 3.10, où l’on n’observe pas la même tendance d’élargissement du pic que pour le SDSS-II. L’étude [Anderson et al. \(2012\)](#) obtient une détection des BAOs avec une significativité de plus de  $5\sigma$  en utilisant la méthode classique du  $\chi^2$ . On verra toutefois dans le chapitre 5 que la significativité donnée par cette méthode n’est pas rigoureuse.

## 3.4 Simulations

### 3.4.1 Simulations de segments de Cox

Un processus de segments de Cox est un processus ponctuel qui a l’avantage d’avoir une forme analytique pour la fonction de corrélation, et qui est également facile à générer. C’est un cas particulier de processus de Cox (voir section 2.2.1), i.e. un processus stochastique en deux étapes. La première étape est la génération d’un champ aléatoire continu et la deuxième étape est un échantillonnage de Poisson de ce champ continu.

Dans le cas particulier du processus de segments de Cox, le champ continu est constitué de segments de taille  $l$  qui sont générés aléatoirement dans le volume. Plus précisément, dans chaque élément de volume  $A$  le nombre de centres de segments suit une loi de Poisson d’intensité  $\lambda_s |A|$ , et chaque segment possède une orientation aléatoire. La longueur de segments par unité de volume est donc en moyenne égale à  $L_V = \lambda_s l$ . La seconde étape du processus consiste en un échantillonnage de Poisson de ce champ continu. Les points sont donc générés sur les segments du champ continu, avec une intensité de Poisson égale à  $\lambda_l$  par unité de longueur. La densité moyenne du processus est égale à :

$$\lambda = \lambda_l L_V = \lambda_l \lambda_s l \quad (3.3)$$

La fonction de corrélation de ce processus ponctuel est donnée analytiquement par la formule

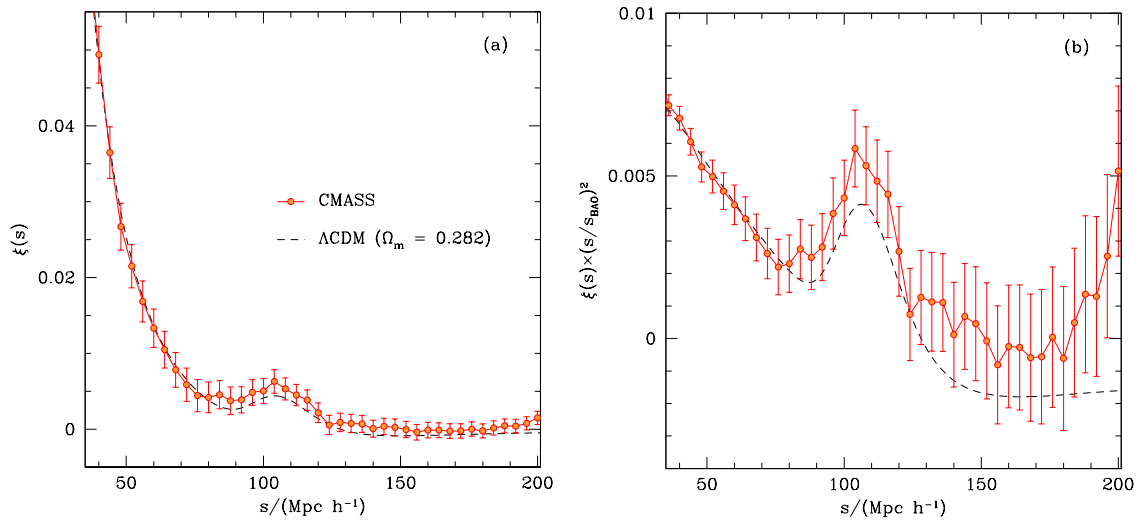


FIGURE 3.10 – A gauche : Fonction de corrélation estimée du catalogue DR9 CMASS pour une cosmologie fiducielle  $\Lambda$ CDM plate avec  $\Omega_m = 0.274$ . Les barres d’erreur sont obtenues à partir de  $N = 600$  simulations avec la même fonction de sélection que le catalogue de données et les paramètres cosmologiques  $\Omega_m = 0.274$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 0.0224$ ,  $n_s = 0.95$  et  $\sigma_8 = 0.8$  (voir [Manera et al. \(2012\)](#) pour plus de détails sur ces simulations). La ligne en pointillés correspond au meilleur *fit*  $\Lambda$ CDM, en combinant les données de la fonction de corrélation DR9 CMASS et du CMB avec WMAP7. A droite : Même chose en renormalisant par  $(s/s_{\text{BAO}})^2$ , avec  $s_{\text{BAO}} = 153.2$  Mpc pour mettre en évidence le pic BAO. Figure extraite de [Sánchez et al. \(2012\)](#).

suivante (voir [Stoyan et al. \(1995\)](#)) :

$$\xi(r) = \begin{cases} \frac{1}{2\pi r^2 L_V} - \frac{1}{2\pi r l L_V} & \text{pour } r \leq l \\ 0 & \text{pour } r \geq l \end{cases} \quad (3.4)$$

On voit que la corrélation s’annule pour  $r \geq l$ , ce qui est logique car deux points à une distance plus grande que  $l$  appartiennent nécessairement à deux segments différents et sont donc indépendants.

Il est facile de générer des réalisations de segments de Cox dans des cubes. Il suffit de créer un nombre de centres de segments suivant une loi de Poisson d’intensité  $\lambda_s V$ , avec  $V$  le volume total. Chacun de ces centres est alors placé aléatoirement dans le volume. Pour chaque centre, on considère alors un segment de taille  $l$  avec une orientation aléatoire. Finalement pour chacun de ces segments, on crée un nombre de points suivant une loi de Poisson d’intensité  $\lambda_l l$  que l’on répartit aléatoirement sur le segment.

Il y a simplement une difficulté venant du fait qu’un point dans le cube peut appartenir à un segment dont le centre est à l’extérieur du cube. On résout ce problème facilement de la façon suivante : au lieu de générer les centres dans un cube de côté  $a$ , on génère les centres dans un cube de côté  $a + 2l$ . Une fois tous les points créés dans le volume avec la procédure détaillée ci-dessus, on ne conserve que la partie intérieure du cube.

On donne un exemple de réalisation sur la figure 3.11 pour un cube de taille 50 par 50 par 50 (l’unité est ici arbitraire) avec une intensité volumique de segments  $\lambda_l = 0.01$ , une longueur de segments  $l = 10$  et un nombre de points par unité de longueur de segments  $\lambda_l = 1.8$ . Le processus a une densité moyenne  $\lambda = 0.18$ , ce qui correspond à un nombre moyen de points égal à 22,500 dans le volume considéré.

### 3.4.2 Simulations log-normales

#### Le modèle log-normal

[Coles and Jones \(1991\)](#) propose de modéliser le champ de galaxies  $\rho_g$  comme un champ log-normal.

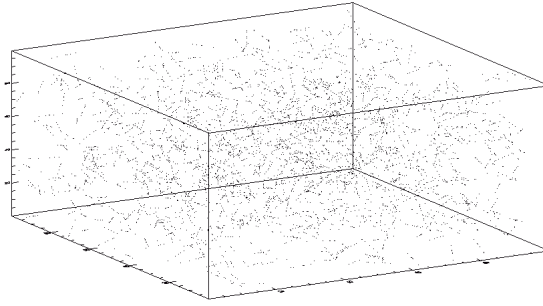


FIGURE 3.11 – Réalisation d’un processus de segments de Cox dans un cube de taille 50 par 50 par 50 (l’unité est ici arbitraire) avec une intensité volumique de segments  $\lambda_l = 0.01$ , une longueur de segments  $l = 10$  et un nombre de points par unité de longueur de segments  $\lambda_l = 1.8$ . Le processus a une densité moyenne  $\lambda = 0.18$ , ce qui correspond à un nombre moyen de points égal à 22,500 dans le volume considéré.

Un champ log-normal  $\rho_{LN}$  avec une moyenne  $\bar{\rho}_{LN} = 1$  est obtenu à partir d’un champ Gaussien  $\rho_G$  de la façon suivante (voir annexe A) :

$$\rho_{LN} = e^{\rho_G - \frac{\sigma_G^2}{2}}. \quad (3.5)$$

Ce modèle a été utilisé avec succès pour la reconstruction du champ de matière continu  $\rho_m$  dans [Kitaura et al. \(2010\)](#), où il entre comme un modèle *a priori* pour le champ de matière. Le modèle log-normal est relativement simple et il possède d’autres propriétés intéressantes (voir [Coles and Jones \(1991\)](#)) :

- Il décrit bien la distribution du nombre de galaxies présentes sur un élément du ciel comme observé par [Hubble \(1934\)](#), ou sur un élément de volume comme observé dans [Kitaura et al. \(2010\)](#) après un lissage du champ de  $5 h^{-1} \text{Mpc}$  ou  $10 h^{-1} \text{Mpc}$
- Le modèle assure la positivité du champ ( $\rho_{LN} > 0$ ) à la différence du modèle Gaussien
- De nombreuses quantités peuvent être calculées aussi facilement que pour un champ Gaussien (e.g. les statistiques des pics ou le genus)
- Le champ est arbitrairement proche d’un champ Gaussien aux temps primordiaux où  $\sigma \approx 0$
- Ce modèle est la solution des équations d’évolution de  $\rho$  en supposant que le champ des vitesses particulières initial est Gaussien

De plus [Kayo et al. \(2001\)](#) a montré que non seulement le modèle log-normal reproduit bien la distribution en un point donné  $P(\delta_1)$  par rapport à des simulations  $N$ -corps, mais il reproduit aussi très bien les distributions conditionnelles  $P(\delta_1 | \delta_2)$  pour  $\delta_2 > -0.5$ .

On montre sur la figure 3.12 un exemple de réalisation log-normale pour l’échantillon Main VL décrit dans la section 3.2.1. Sur la même figure, on montre aussi les données du catalogue Main VL lui-même. On peut voir le bon accord visuel de ces deux distributions, même si le catalogue log-normal ne semble pas reproduire la structure de toile cosmique qui peut s’apercevoir sur les données elles-mêmes.

Ainsi, les simulations log-normales permettent une très bonne approximation du champ de galaxies, et sont rapides à générer comme on va le voir dans la section suivante. De ce fait, il est très courant d’utiliser le modèle log-normal lorsqu’un grand nombre de simulations est requis, e.g. pour estimer la matrice de covariance de l’estimateur du spectre de puissance  $\hat{P}$  ou de la fonction de corrélation  $\hat{\xi}$  ([Percival et al. 2007, 2010](#); [Blake et al. 2011a,b](#); [Beutler et al. 2011](#); [Labatie et al. 2012c,a,b](#))

Dans cette thèse, on va étudier de nombreux effets statistiques dans l’analyse de la fonction de corrélation estimée  $\hat{\xi}$ . On va donc principalement se servir de simulations log-normales qui sont à la fois réalistes et rapides à générer.

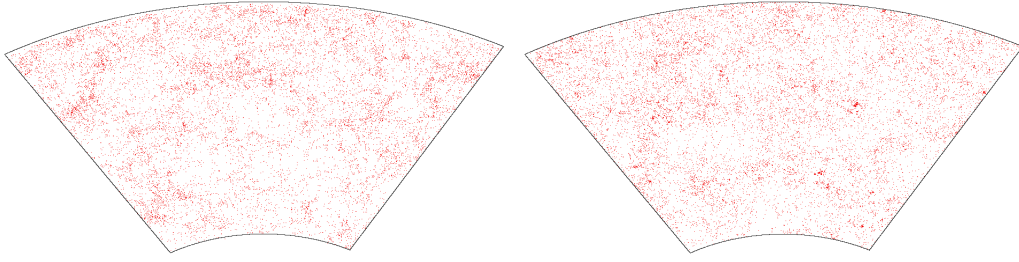


FIGURE 3.12 – A gauche : coupe de l'échantillon Main VL entre les valeurs de latitude  $-10^\circ < \lambda < 10^\circ$  dans les coordonnées du relevé SDSS. A droite : Même chose pour une réalisation log-normale du Main VL. On peut voir le bon accord visuel des deux distributions, bien que la structure de toile cosmique que l'on peut apercevoir sur les données ne soit pas reproduite dans les simulations.

### Procédure pour générer des simulations log-normales

Avant de commencer cette section, nous avons besoin de clarifier certaines dénominations. On emploiera dans cette section le terme de 'spectre de puissance' d'un champ  $\rho$  pour désigner la variance des modes de Fourier de ce champ, et le terme de fonction de covariance pour désigner la covariance entre 2 points séparés d'un vecteur  $\mathbf{r}$  :

$$P(k) = \text{Var} [\tilde{\rho}(\mathbf{k})] \quad (3.6)$$

$$= \mathbb{E} [\tilde{\rho}(\mathbf{k})^2] \quad (3.7)$$

$$\mathbf{C}(\mathbf{r}) = \text{Cov} [\rho(\mathbf{x}), \rho(\mathbf{x} + \mathbf{r})] \quad (3.8)$$

$$= \mathbb{E} [(\rho(\mathbf{x}) - \bar{\rho})(\rho(\mathbf{x} + \mathbf{r}) - \bar{\rho})] \quad (3.9)$$

En particulier le spectre de puissance et la fonction de corrélation de la distribution des galaxies (ou du champ de matière), tels qu'on les entend en cosmologie, correspondent à la définition précédente du 'spectre de puissance' et de la 'matrice de covariance' pour le cas particulier du champ de fluctuations  $\delta = \frac{\rho - \bar{\rho}}{\bar{\rho}}$  (voir section 2.1.2). Ceci est également vrai dans le cas particulier où la moyenne du champ  $\bar{\rho} = 1$ , puisque dans ce cas  $\delta = \rho - 1$  et que les définitions précédentes ne dépendent pas de la moyenne du champ.

Pour simuler des catalogues log-normaux, la procédure habituelle est de générer le champ Gaussien  $\rho_G$ , et d'en déduire le champ log-normal  $\rho_{LN}$  avec l'équation (3.5). En effet ce sont les champs Gaussiens que l'on sait facilement générer, avec des modes de Fourier Gaussiens, indépendants, et de variance  $P_G(k)$  pour un mode de Fourier  $\mathbf{k}$  (avec  $k = \|\mathbf{k}\|$ ).

On part donc du spectre de puissance  $P_{LN}$  voulu<sup>9</sup>, et l'on utilise les relations liant le champ log-normal  $\rho_{LN}$  au champ Gaussien  $\rho_G$  pour déterminer le spectre de puissance de  $\rho_G$ . La relation entre le champ log-normal et Gaussien est simple en terme de fonction de covariance (voir annexe A) :

$$\mathbf{C}_G(r) = \ln[1 + \mathbf{C}_{LN}(r)] \quad (3.10)$$

Il faut donc commencer par transformer le spectre de puissance du champ log-normal pour calculer sa fonction covariance<sup>10</sup>. Le fonction de covariance s'obtient à partir du spectre de puissance par  $\text{FFT}^{-1}$  en 3 dimensions, i.e. par transformée de Hankel du spectre 1D car le spectre 3D est isotrope :

$$\mathbf{C}_{LN} = \mathcal{H}^{-1}(P_{LN}) \quad (3.11)$$

La difficulté de la transformée de Hankel est que le spectre de puissance est habituellement donné avec un espacement exponentiel en  $k$ , i.e. que les  $\ln k_i$  sont espacés linéairement. En effet c'est dans cet espace que le spectre de puissance est lisse. Nous allons montrer plus tard comment résoudre ce problème.

9. Comme on considère pour l'instant un champ log-normal de moyenne  $\bar{\rho}_{LN} = 1$ , le 'spectre de puissance' de  $\rho_{LN}$  correspond bien au spectre de puissance comme on l'entend en cosmologie.

10. De même, la fonction de covariance  $\mathbf{C}_{LN}$  de  $\rho_{LN}$  est égale à la fonction de corrélation  $\xi_{LN}$  tel qu'on l'entend habituellement en cosmologie.

Une fois la fonction  $\mathbf{C}_{LN}$  calculée, celle du champ Gaussien  $\mathbf{C}_G$  est obtenue par la relation (3.10). Finalement on récupère le spectre de puissance du champ Gaussien  $P_G$  par transformée de Hankel directe de  $\mathbf{C}_G$  :

$$P_G = \mathcal{H}(\mathbf{C}_G) \quad (3.12)$$

Les erreurs de calcul et la discrétisation peuvent conduire à des valeurs négatives pour le spectre de puissance. Pour éviter cela on seuille simplement les valeurs du spectre  $P_G$  à 0. Pour vérifier que le programme de transformation  $P_{LN} \rightarrow P_G$  est précis, on peut inverser toutes les étapes à partir du spectre de puissance Gaussien en sortie :

$$P_G \rightarrow \mathbf{C}_G \rightarrow \mathbf{C}_{LN} \rightarrow P_{LN} \quad (3.13)$$

et l'on compare le spectre  $P_{LN}$  ainsi récupéré avec le spectre donné en entrée du programme.

### Transformée de Hankel en espace logarithmique et précision du programme

Pour faire la transformée de Hankel avec un espacement logarithmique, on utilise le code FFTlog d'Hamilton<sup>11</sup> qui permet d'effectuer la transformée de Hankel avec ce type d'espacement.

Le programme utilise le fait qu'en considérant les fonctions ne dépendant plus de  $k$  mais de  $\ln k$ , la transformée de Fourier est encore une convolution entre la nouvelle fonction  $g(\ln k) = f(k)$  et une autre fonction liée à la fonction de Bessel  $J_\mu$ , où  $\mu$  est l'ordre de la fonction de Bessel qui dépend de la dimension de l'espace ( $\mu = 0.5$  en 3 dimensions). La convolution est calculée en passant par une FFT classique sur ces fonctions.

Le programme est toujours inversible, c'est à dire qu'en appliquant 2 fois la transformation à une fonction donnée on retrouve la même fonction. Il est aussi exact (i.e. il donne les mêmes valeurs qu'une transformée continue) lorsque la fonction est périodique avec une période plus petite que l'intervalle de points discrets et qu'elle est suffisamment bien échantillonnée.

Ici cependant le spectre n'est pas une fonction périodique, et pour se rapprocher d'une transformée de Fourier continue il faut limiter les effets de conditions aux limites périodiques. Pour cela on ajoute dans le programme une option de *zero-padding* par défaut, consistant à rajouter de chaque côté autant de zéros qu'il y a de points entre  $k_{\min}$  et  $k_{\max}$ . On obtient de bons résultats avec ce choix, mais rajouter plus de points peut créer des problèmes de stabilité.

Donnons un exemple avec un spectre en entrée  $P_L(k)$  à *redshift*  $z = 0$  obtenu avec la formule de Eisenstein and Hu (1998) pour les paramètres cosmologiques de Komatsu et al. (2011). La transformation avec *zero-padding* semble plus stable, i.e. qu'elle n'entraîne pas les d'oscillations artificielles dans le spectre de puissance du champ Gaussien, qui sont présentes si l'on ne fait pas de *zero-padding* (figure 3.13). Dans le cas où l'on effectue un *zero-padding*, l'inversion de toutes les étapes donnée dans l'équation (3.13) permet de retrouver un spectre très proche du spectre en entrée (figure 3.14).

### Génération du champ log-normal

Une fois que l'on dispose du spectre Gaussien  $P_G$  on peut générer le champ Gaussien sur une grille cubique.

Les modes sont discrets et limités dans la partie haute fréquence par le nombre de points par unité de volume. Chaque mode de Fourier  $(k_x, k_y, k_z)$  possède une variance donnée par la valeur du spectre  $P_G(k)$  en  $k = \|(k_x, k_y, k_z)\|$  et multipliée par  $\frac{(\Delta k)^3}{(2\pi)^3} = \frac{1}{L^3}$  où  $L$  est la longueur du cube (chaque mode correspond à un volume  $(\Delta k)^3$  dans l'espace des  $k$ ).

On utilise la librairie *fftw*<sup>12</sup> pour construire la densité dans l'espace de Fourier avec des variables Gaussiennes indépendantes et avec des variances données par  $P_G$ . Une petite difficulté est qu'en réalité tous les modes ne sont pas indépendants pour un champ réel. Les coefficients de Fourier sont complexes et possèdent donc deux fois plus de variables que le champ réel. Ainsi ces variables ne peuvent pas toutes être indépendantes. Un mode donné est soit contraint à être réel, soit il existe un autre mode dont il est toujours le conjugué (il s'agit des paires  $(\mathbf{k}, -\mathbf{k})$ ). En dehors de ces relations, tous les modes sont indépendants.

11. <http://casa.colorado.edu/~ajsh/FFTLog/>

12. <http://www.fftw.org/>

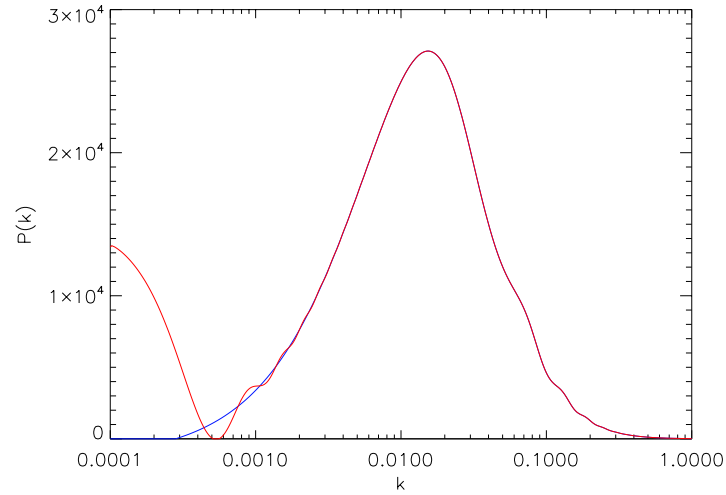


FIGURE 3.13 – Spectre de puissance du champ Gaussien sans *zero-padding* (rouge) et avec *zero-padding* (bleu), pour un spectre en entrée  $P_L(k)$  à *redshift*  $z = 0$  obtenu avec la formule de [Eisenstein and Hu \(1998\)](#) pour les paramètres cosmologiques de [Komatsu et al. \(2011\)](#). On observe des oscillations artificielles dues aux conditions aux limites périodiques de la transformée de Fourier en l'absence de *zero-padding*. Ces oscillations n'apparaissent plus avec le *zero-padding*.

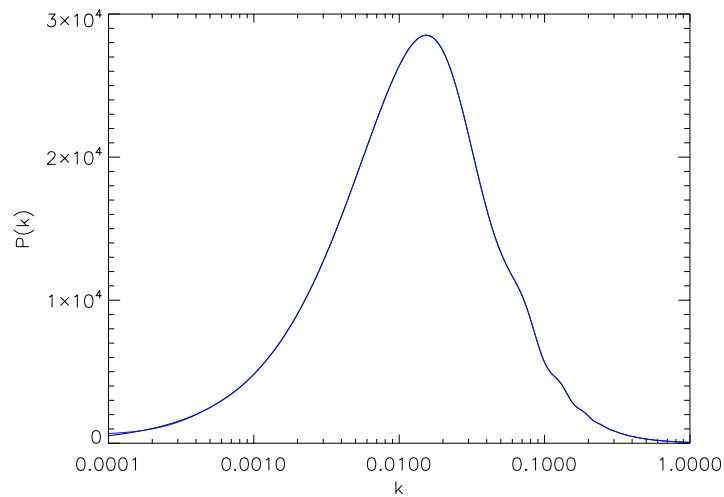


FIGURE 3.14 – Spectre de puissance en entrée (noir) et spectre obtenu en inversant toutes les étapes (formule (3.13)) avec *zero-padding* (bleu), pour un spectre en entrée  $P_L(k)$  à *redshift*  $z = 0$  obtenu avec la formule de [Eisenstein and Hu \(1998\)](#) pour les paramètres cosmologiques de [Komatsu et al. \(2011\)](#). Le spectre obtenu après inversion est extrêmement proche du spectre en entrée.

Il faut prendre ceci en compte pour incrémenter une variable de variance globale  $\sigma_G^2$  du champ Gaussien  $\rho_G$ . En effet on a besoin de cette variance à la fin de notre procédure, afin renormaliser le champ pour que celui-ci soit bien d'espérance égale à 1 (voir formule (3.5)).

Une fois que l'on a généré le champ Gaussien  $\rho_G$ , on utilise la formule (3.5) pour générer le champ log-normal en chaque point. On obtient donc un champ log-normal avec la fonction de corrélation  $\xi_{LN}$  voulue. Il ne reste plus qu'à multiplier le champ log-normal par la densité moyenne  $\bar{\rho}_{LN}$  voulue et finalement de faire un échantillonnage de Poisson sur ce champ continu.

### 3.4.3 Simulations log-normales pour différentes valeurs de $\alpha$ simultanément

On montre ici une méthode pour générer de façon très efficace des simulations log-normales en faisant varier le paramètre  $\alpha$ . On a défini ce paramètre dans la section 2.2.5 comme un facteur de dilatation du catalogue de galaxies entre les coordonnées fiducielles et les coordonnées comobiles. Plus précisément, le facteur de dilatation est égal à  $\alpha = D_V(\langle z \rangle) / D_V^{fid}(\langle z \rangle)$  avec  $\langle z \rangle$  le *redshift* moyen du catalogue et  $D_V$  l'échelle de dilatation (*dilation scale*) donnée par :

$$D_V(z) = \left[ (1+z)^2 D_M(z)^2 \frac{cz}{H(z)} \right]^{1/3} \propto D_M(z)^{2/3} \frac{1}{H(z)^{1/3}} \quad (3.14)$$

Le paramètre  $\alpha$  crée donc simplement une dilatation du catalogue de galaxies et de la fonction de corrélation apparente<sup>13</sup>. Cet effet est indépendant de la vraie fonction de corrélation en coordonnées comobiles. Et comme on va le voir, il est possible de prendre en compte ce paramètre en utilisant un seul ensemble de simulations.

Supposons que l'on veuille simuler un catalogue de galaxies qui s'étend entre les *redshifts*  $z_{min}$  et  $z_{max}$ . Notons  $r_{min}$  et  $r_{max}$  les distances minimum et maximum correspondant à cette extension en *redshift* dans la cosmologie fiducielle utilisée pour estimer la fonction de corrélation du catalogue, et  $\bar{n}(r)$  la densité moyenne estimée du catalogue dans cette cosmologie fiducielle.

L'effet de  $\alpha$  est de créer une dilatation du catalogue entre les coordonnées fiducielles et les coordonnées comobiles. Comme on connaît les limites en distance  $r_{min}$ ,  $r_{max}$  et la densité  $\bar{n}(r)$  du catalogue en coordonnées fiducielles, l'effet de  $\alpha$  est simplement de changer ces limites de distance ainsi que la densité de la façon suivante :

- Si le catalogue s'étend entre une distance minimum  $r_{min}$  et une distance maximum  $r_{max}$  en coordonnées fiducielles, il s'étend entre  $\alpha r_{min}$  et  $\alpha r_{max}$  en coordonnées comobiles.
- Si le catalogue possède une densité moyenne  $\bar{n}(r)$  en coordonnées fiducielles, il possède une densité moyenne en coordonnées comobiles égale à

$$\bar{n}_\alpha(r) = \frac{1}{\alpha^3} \bar{n}\left(\frac{r}{\alpha}\right) \quad (3.15)$$

Grâce à cette constatation, on va pouvoir générer simultanément des simulations pour différentes valeurs de  $\alpha$  en coordonnées comobiles. En d'autres termes, on va pouvoir générer simultanément des simulations pour différentes valeurs de  $D_V(\langle z \rangle) = \alpha D_V^{fid}(\langle z \rangle)$ , avec tous les autres paramètres de la simulation constants par ailleurs

Pour cela on génère simplement des simulations log-normales (avec la méthode présentée dans la section 3.4.2) entre les distances  $\alpha_{min} r_{min}$  et  $\alpha_{max} r_{max}$  et avec une densité  $\bar{n}_{max}(r) = \max_\alpha \bar{n}_\alpha(r)$ . Définissons maintenant une fonction de sélection pour chaque valeur de  $\alpha$  :

$$\Phi_\alpha(r) = \frac{\bar{n}_\alpha(r)}{\bar{n}_{max}(r)} \quad (3.16)$$

On applique cette fonction de sélection pour chaque valeur de  $\alpha$  de la façon suivante : pour chaque galaxie à une distance  $r$  en coordonnées comobiles dans la simulation originale, on génère une variable aléatoire uniforme  $u \in [0, 1]$ . Alors la galaxie appartient à la simulation pour la valeur  $\alpha$  si l'on a  $u < \Phi_\alpha(r)$ . Ainsi si l'on considère une grille  $\alpha \in \{\alpha_1, \dots, \alpha_n\}$ , pour chaque galaxie  $\mathbf{x}_i$  on obtient une

13. Plus précisément cette dilatation consiste à transformer les coordonnées  $\mathbf{x}_i$  des galaxies en coordonnées fiducielles en des coordonnées  $\alpha \mathbf{x}_i$  en coordonnées comobiles. Ainsi l'espérance de la fonction de corrélation estimée en coordonnées fiducielles est donnée par  $\mathbb{E}[\xi(r)] = \xi(\alpha r)$ , avec  $\xi$  la vraie fonction de corrélation en coordonnées comobiles.

suite d'intervalles  $\{\alpha_{i_1}, \dots, \alpha_{i_2}\}$  pour lesquelles la galaxie appartient à la simulation. Pour des raisons d'optimisation du temps de calcul que l'on va détailler ci-dessous, on considère une nouvelle galaxie pour chaque intervalle distinct  $\{\alpha_{i_1}, \dots, \alpha_{i_2}\}$  placée au même endroit que la galaxie originale (de sorte que chaque nouvelle galaxie corresponde à un unique intervalle  $\{\alpha_{i_1}, \dots, \alpha_{i_2}\}$ ).

Avec cette procédure, on obtient simultanément des simulations en coordonnées comobiles pour différentes valeurs de  $\alpha$ . Toutefois, il faut théoriquement transformer ces simulations pour les exprimer en coordonnées fiducielles avant d'estimer la fonction de corrélation, comme c'est le cas du catalogue de données. Mais parce que la fonction de corrélation est estimée par comptage de paires, on peut en fait l'estimer en coordonnées comobiles (i.e. avant la dilatation), et ensuite appliquer la dilatation sur les intervalles de distances. Les quantités  $DD^{com}(r)$ ,  $RR^{com}(r)$  et  $DR^{com}(r)$  correspondant à un intervalle  $[r - dr/2, r + dr/2]$  en coordonnées comobiles deviennent  $DD(r/\alpha)$ ,  $RR(r/\alpha)$  et  $DR(r/\alpha)$  et correspondent à un intervalle  $[r/\alpha - dr/(2\alpha), r/\alpha + dr/(2\alpha)]$  en coordonnées fiducielles.

On va maintenant montrer comment la fonction de corrélation peut être estimée très efficacement pour toutes les valeurs de  $\alpha$  simultanément. Pour cela, voyons comment calculer les termes de comptage de paires  $DD_\alpha$ , avec la même procédure qui peut être appliquée pour  $DR_\alpha$  and  $RR_\alpha$ .

Pour chaque valeur de  $r$  on considère le tableau  $(DD_{\alpha_i, raw}^{com}(r))_{i=1, \dots, n}$ , correspondant à une grille  $\alpha \in \{\alpha_1, \dots, \alpha_n\}$ . Ce tableau va compter le nombre de paires qu'il faut ajouter à  $DD_{\alpha_i}^{com}(r)$  pour obtenir  $DD_{\alpha_{i+1}}^{com}(r)$ . Pour chaque paire de points  $(\mathbf{x}_k, \mathbf{x}_l)$ , appartenant respectivement à la simulation pour  $\alpha \in \{\alpha_{k_1}, \dots, \alpha_{k_2}\}$  et  $\alpha \in \{\alpha_{l_1}, \dots, \alpha_{l_2}\}$ , la paire appartient à la simulation pour  $\alpha \in \{\max(\alpha_{k_1}, \alpha_{l_1}), \dots, \min(\alpha_{k_2}, \alpha_{l_2})\} = \{\alpha_{\max(k_1, l_1)}, \dots, \alpha_{\min(k_2, l_2)}\}$ . Ainsi on va ajouter +1 à  $DD_{\alpha, raw}^{com}(r)$  pour  $\alpha = \alpha_{\max(k_1, l_1)}$  et ajouter -1 pour  $\alpha = \alpha_{\min(k_2, l_2)+1}$ <sup>14</sup>.

Finalement on obtient le terme de comptage de paires  $DD_\alpha^{com}(r)$  de la façon suivante :

$$DD_{\alpha_i}^{com}(r) = \sum_{j=0}^i DD_{\alpha_j, raw}^{com}(r) \quad (3.17)$$

Il ne reste plus qu'à appliquer la dilatation sur les intervalles de distance :

$$DD_\alpha(r) = DD_\alpha^{com}(\alpha r) \quad (3.18)$$

Cette procédure complète permet d'obtenir les quantités  $DD_\alpha(r)$ ,  $DR_\alpha(r)$  and  $RR_\alpha(r)$  pour chaque valeur  $r$  et de  $\alpha$  avec un temps de calcul multiplié seulement par  $\approx 4$  pour les simulations du catalogue DR7-Full utilisées dans les chapitres 5 et 6, au lieu d'être multiplié par la taille de la grille en  $\alpha$ .

### 3.4.4 Simulations $N$ -corps

Comme on l'a vu dans la section 1.3, la formation des grandes structures de l'univers se fait à travers différentes étapes successives. La dernière étape d'évolution non linéaire est notamment responsable de la création de filaments et des halos de matière noire. Cette évolution n'est pas solvable analytiquement, et l'on doit avoir recours à des simulations numériques pour la simuler de façon réaliste. On appelle ces simulations numériques des simulations  $N$ -corps, puisqu'elles décrivent l'évolution de  $N$  particules sous l'influence de leurs forces de gravitation mutuelles dans un univers en expansion. Bien que ces simulations contiennent en général des millions, voire des milliards de particules, ces particules correspondent à de très grandes masses de l'ordre de  $10^9 - 10^{11} M_\odot$ . Du fait de la divergence de la force gravitationnelle lorsque la distance entre deux particules tend vers 0, les simulations introduisent une distance de lissage (*softening parameter*) qui atténue la force gravitationnelle à courte portée.

On distingue deux grands types de simulations :

- Les simulations de matière noire pure, qui est non-collisionnelle. En effet les modèles cosmologiques actuelles indiquent qu'environ 90% de la matière totale de l'univers est sous forme de matière noire, et que les galaxies apparaissent dans les halos de matière noire virialisés.

14. Notons que cela requiert le choix d'une fonction de poids constante, comme ce sera le cas dans nos simulations du catalogue DR7-Full dans les chapitres 5 et 6. En effet, si la fonction de poids dépendait de la densité  $n(\|\mathbf{x}_1\|/\alpha)$  en coordonnées fiducielles, alors cette galaxie aurait des poids différents en fonction de  $\alpha$ .



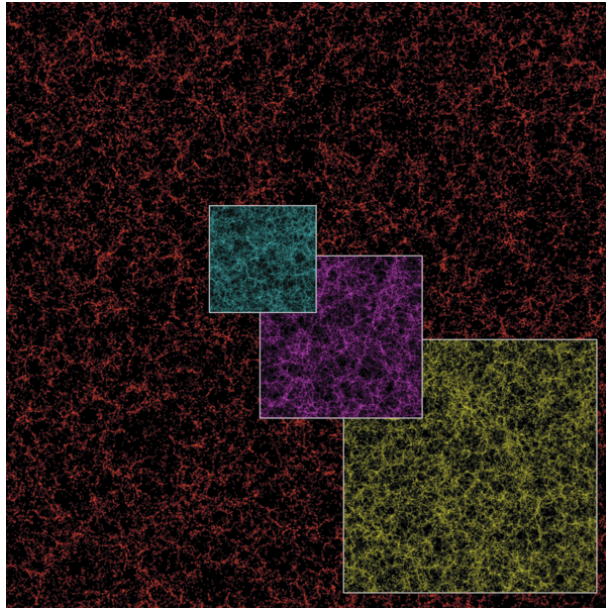


FIGURE 3.15 – Distribution lissée des halos de matière noire (pondérés par leur masse) sur une coupe d'épaisseur  $40 h^{-1}\text{Mpc}$  en coordonnées comobiles pour différentes réalisations LasDamas avec des tailles de cube différentes ( $2400 h^{-1}\text{Mpc}$ ,  $1000 h^{-1}\text{Mpc}$ ,  $640 h^{-1}\text{Mpc}$  et  $420 h^{-1}\text{Mpc}$ ).

- Les simulations hydrodynamiques qui introduisent notamment les effets de la matière baryonique liés aux processus spécifiques du gaz et des étoiles.

L'inclusion des phénomènes hydrodynamiques accroît considérablement la complexité des simulations. Aussi pour de grands volumes cosmologiques, les simulations  $N$ -corps n'incluent que la matière noire non-collisionnelle. A partir de ces simulations de matière noire pure, les populations de galaxies sont introduites dans les halos virialisés de matière noire<sup>15</sup>. Pour ce faire, on peut utiliser une extension du *halo model* basé sur un choix de *Halo Occupation Distribution* (HOD, voir section 2.2.2). Dans ce modèle le nombre de galaxies peuplant un halo de masse  $m$  est simplement donné par une distribution  $p(N_g|m)$ , qui dépend de la population de galaxies considérée.

Dans cette thèse on se servira uniquement des simulations Large Suite of Dark Matter Simulations (LasDamas, McBride et al., in prep.), qui ont pour but de modéliser les échantillons de galaxies du SDSS DR7 pour des gammes de luminosité variées. Ces simulations sont des simulations de matière noire pure avec  $1280^3$  particules de masse  $45.73 \times 10^{10} M_\odot$  et une distance de lissage égale à  $53 h^{-1}\text{kpc}$ . Les galaxies sont observées dans un cône de lumière et en espace *redshift*, i.e. avec inclusion de distortions de *redshift*.

Les galaxies sont placées artificiellement à l'intérieur des halos de matière noire en utilisant un modèle HOD avec des paramètres ajustés sur les échantillons SDSS réels. Sur la figure 3.15 on montre la distribution lissée des halos de matière noire (pondérés par leur masse) sur une coupe d'épaisseur  $40 h^{-1}\text{Mpc}$  en coordonnées comobiles pour différentes réalisations LasDamas avec des tailles de cube différentes ( $2400 h^{-1}\text{Mpc}$ ,  $1000 h^{-1}\text{Mpc}$ ,  $640 h^{-1}\text{Mpc}$  et  $420 h^{-1}\text{Mpc}$ ).

15. Les halos de matière noire sont en général identifiés avec un algorithme *Friends of Friends*. Ce type d'algorithme crée des groupes de points qui peuvent fusionner à chaque itération quand ils sont séparés d'une distance suffisamment faible.

# Chapitre 4

## Incertitude sur l'estimation de la fonction de corrélation des galaxies

### Sommaire

---

<b>4.1</b>	<b>Fractalité et estimation de la fonction de corrélation</b>	<b>72</b>
4.1.1	Problème conceptuel de la fonction de corrélation pour une fractale	72
4.1.2	But du travail dans ce chapitre	73
<b>4.2</b>	<b>Estimateurs et incertitude sur la densité moyenne <math>\bar{n}</math></b>	<b>73</b>
4.2.1	Effet général de l'incertitude sur la densité moyenne $\bar{n}$	73
4.2.2	Le biais de la contrainte intégrale	76
<b>4.3</b>	<b>Simulations utilisées</b>	<b>77</b>
<b>4.4</b>	<b>Effet de la contrainte intégrale sur des simulations</b>	<b>79</b>
4.4.1	Effet de la contrainte intégrale sur des processus de segments de Cox	79
4.4.2	Prédiction de l'effet de la contrainte intégrale	81
4.4.3	Effet de la contrainte intégrale sur des simulations Main VL et LRG VL	82
<b>4.5</b>	<b>Comparaison des différents estimateurs sur des simulations Main VL et LRG VL</b>	<b>82</b>
<b>4.6</b>	<b>Conclusion</b>	<b>84</b>

---

Dans ce chapitre on étudie l'incertitude sur les différents estimateurs de la fonction de corrélation dans les relevés de galaxies actuels. On va notamment voir dans la section 4.1.1 que la fonction de corrélation n'est pas un concept bien défini pour les distributions fractales, puisqu'elle requiert l'existence d'une densité moyenne  $\bar{n}$ , qui n'a pas de sens pour une fractale. D'une manière plus générale, même si la distribution est homogène aux grandes échelles, les estimateurs de la fonction de corrélation sont affectés par l'incertitude sur la densité moyenne  $\bar{n}$  (et notamment biaisés).

On peut alors se demander si l'analyse classique de la fonction de corrélation est valide, et notamment dans quelle mesure les estimateurs sont biaisés. Ceci est important puisque l'analyse de la fonction de corrélation (ou de façon équivalente, du spectre de puissance) est utilisée pour tester le modèle cosmologique standard ou pour contraindre les paramètres cosmologiques (voir par exemple Eisenstein et al. (2005); Sánchez et al. (2009); Beutler et al. (2011); Blake et al. (2011a,b); Cole et al. (2005); Tegmark et al. (2006); Padmanabhan et al. (2007); Reid et al. (2010); Ho et al. (2012); Percival et al. (2007, 2010); Kazin et al. (2010); Mehta et al. (2012)). Un exemple particulier de cette analyse est l'étude des Oscillations Baryoniques Acoustiques. Etant donné la faiblesse du signal BAO, cette étude requiert une analyse statistique très fine de la fonction de corrélation.

Dans la section 4.2 on va montrer analytiquement comment les estimateurs sont affectés par l'incertitude sur la densité moyenne  $\bar{n}$ . Un effet particulier de cette incertitude est le biais de la contrainte intégrale qui impose aux estimateurs l'égalité suivante :

$$\frac{1}{V^2} \int_{V^2} \hat{\xi}(r) d^3r = 0 \quad (4.1)$$

avec  $\mathcal{V}$  le domaine du catalogue, et  $V$  son volume.

Dans la suite du chapitre, on étudiera ces différents effets à partir de simulations. On présentera nos simulations dans la section 4.3. Enfin dans les section 4.4 et 4.5, on quantifiera l'effet de l'incertitude sur  $\bar{n}$ , et notamment l'effet de la contrainte intégrale, pour les relevés de galaxies actuels.

## 4.1 Fractalité et estimation de la fonction de corrélation

### 4.1.1 Problème conceptuel de la fonction de corrélation pour une fractale

Il y a une hypothèse implicite derrière la notion de fonction de corrélation, à savoir que la distribution a une densité moyenne  $\bar{n}$  bien définie. En effet la fonction de corrélation mesure l'excès de probabilité de trouver un point à une distance  $r$  d'un autre point donné, comparé à une distribution non corrélée (voir section 2.2.1). La probabilité  $dP_r$  de trouver une galaxie dans un volume  $dV$  séparé d'une distance  $r$  d'une galaxie donnée est :

$$dP_r = \bar{n}[1 + \xi(r)]dV \quad (4.2)$$

On voit donc qu'il y a un problème conceptuel avec la fonction de corrélation pour un champ fractal. En effet, pour un champ fractal la densité moyenne n'est pas un concept bien défini, et donc la fonction de corrélation n'est également pas bien définie. Comme nous allons le voir, si l'on essaye quand même d'estimer la 'fonction de corrélation' d'un champ fractal, le résultat va dépendre de la taille du volume.

Montrons cela dans le cas particulier où le volume est une sphère de rayon  $R_s$  centrée sur un point de la distribution. En utilisant la dimension de corrélation du champ fractal, selon l'équation (2.55) le nombre moyen de points dans la sphère est égal à :

$$N(< R_s) = AR_s^D \quad (4.3)$$

Ainsi la densité moyenne de points estimée (i.e. la densité moyenne empirique) est égale à :

$$\bar{n} = \frac{3A}{4\pi} R_s^{D-3} \quad (4.4)$$

Quant à la densité moyenne de points conditionnelle (i.e. la densité moyenne de points autour d'un point de la distribution), elle est donnée par  $\frac{1}{4\pi r^2} \frac{dN(<r)}{dr} = \frac{AD}{4\pi} r^{D-3}$ . En identifiant ceci avec le terme  $dP_r/dV$  de l'équation (4.2), on obtient :

$$\bar{n}[1 + \xi(r)] = \frac{AD}{4\pi} r^{D-3} \quad (4.5)$$

En combinant les expressions (4.4) et (4.5), on obtient que la 'fonction de corrélation' estimée est égale à :

$$\xi(r) = \frac{D}{3} \left( \frac{r}{R_s} \right)^{D-3} - 1 \quad (4.6)$$

On voit que cette fonction dépend de la taille de la sphère  $R_s$ . Ce comportement des estimateurs de la fonction de corrélation est observé sur les relevés de galaxies mais seulement lorsque le volume du relevé est petit, où l'on obtient alors  $\xi(r) \propto r^{-\gamma}$  ( $\gamma \approx 1.7$ ).

La longueur de corrélation  $r_0$  est définie comme la séparation où l'on obtient  $\xi(r_0) = 1$ , i.e. comme la distance de transition entre forte corrélation et faible corrélation. Lorsque l'on estime  $\xi$  sur un champ fractal, la longueur de corrélation dépend de la taille du volume observé (voir aussi [Pietronero \(1987\)](#)). En utilisant la formule (4.6) on obtient :

$$r_0 = \left( \frac{6}{D} \right)^{1/D-3} R_s \quad (4.7)$$

La longueur de corrélation est donc proportionnelle à la taille de la sphère  $R_s$ . Ce comportement a été observé dans les relevés de galaxies initiaux, mais les relevés de galaxies plus récents du relevé *Infrared Astronomical Satellite* (IRAS) jusqu'au 2dFGRS et au SDSS ont contredit cette tendance. La longueur de corrélation est devenue une quantité stable pour une population de galaxies donnée.

Malgré ces observations, les partisans d'un univers fractal soulignent quand même un usage trompeur de la fonction de corrélation qui suppose l'homogénéité à grand échelle (pour une revue des arguments en faveur d'un univers fractal voir [Sylos Labini et al. \(1998\)](#)). A la place de la fonction de corrélation, ils proposent d'utiliser la statistique de densité conditionnelle  $\Gamma(r)$  :

$$\Gamma(r) = \frac{\langle n(\mathbf{x})n(\mathbf{x} + \mathbf{r}) \rangle}{\hat{n}} \quad (4.8)$$

avec  $\hat{n}$  la densité moyenne empirique dans l'échantillon. Ici  $\hat{n}$  est simplement un facteur multiplicatif dans  $\Gamma$  (ce qui n'est pas le cas pour la fonction de corrélation), ce qui ne change pas le comportement en loi de puissance pour une distribution fractale. [Sylos Labini et al. \(1998\)](#) introduit aussi la densité conditionnelle moyenne :

$$\Gamma^*(r) = \frac{3}{r^3} \int_0^r \Gamma(s)s^2 ds \quad (4.9)$$

Cette statistique est moins bruitée et suit également une loi de puissance pour une distribution fractale,  $\Gamma^*(r) \propto r^{D-3}$ . De nombreux catalogues en *redshift* ont confirmé ce comportement aux faibles échelles, mais ils ont aussi montré une cassure dans le comportement invariant d'échelle à partir des échelles 10 à 15  $h^{-1}$  Mpc.

### 4.1.2 But du travail dans ce chapitre

Comme expliqué dans la section 2.3.3, la question de l'homogénéité de l'univers aux grandes échelles à longterm a été débattue. C'est une question importante car le principe cosmologique qui est à la base de la cosmologie moderne suppose l'homogénéité, au moins aux grandes échelles.

On a vu que les observations récentes ont confirmé cette vision standard de l'homogénéisation à grande échelle. Il existe toutefois quelques groupes de recherche qui soutiennent encore le vision d'un univers fractal, et qui mettent en avant la présence d'un biais dans l'analyse de la fonction de corrélation. Comme expliqué dans la section 4.1.1, le concept de densité moyenne n'a pas de sens pour une distribution fractale, si bien que la fonction de corrélation n'est pas bien définie. D'une manière générale, l'incertitude sur  $\bar{n}$  cause un biais dans les estimateurs même lorsque la distribution est homogène aux grandes échelles comme on va le voir dans la section 4.2.1. Cet effet engendre notamment le biais connu sous le nom de contrainte intégrale, que l'on va expliquer dans la section 4.2.2.

Ainsi dans [Sylos Labini et al. \(2009\)](#) les auteurs soutiennent que les fluctuations de densité<sup>1</sup> dans les échantillons du SDSS DR7 ne permettent pas une analyse fiable de la corrélation aux échelles  $r > 50h^{-1}$ Mpc, remettant en cause la détection des BAOs ainsi que les affirmations de compatibilité avec le modèle  $\Lambda$ CDM.

Dans ce chapitre on va étudier le biais dans les estimateurs de la fonction de corrélation dû à l'incertitude sur la densité moyenne  $\bar{n}$ , et notamment l'effet de la contrainte intégrale. Le but est de quantifier ce biais dans le cadre du modèle cosmologique standard  $\Lambda$ CDM et pour les relevés de galaxies actuels. Pour cela on utilisera des simulations log-normales du relevé SDSS DR7.

Notre objectif est de vérifier qu'il n'y a pas d'incohérence dans l'analyse habituelle de la fonction de corrélation dans le cadre du modèle  $\Lambda$ CDM, et que les résultats obtenus sont non biaisés.

## 4.2 Estimateurs et incertitude sur la densité moyenne $\bar{n}$

### 4.2.1 Effet général de l'incertitude sur la densité moyenne $\bar{n}$

Il existe différents estimateurs de la fonction de corrélation. La majorité des estimateurs se servent de catalogues aléatoires avec une géométrie identique à l'échantillon pour mesurer les excès de paires en fonction de la séparation  $r$ . On définit  $DD(r)$ ,  $RR(r)$  et  $DR(r)$  comme le nombre de paires séparés

1. Ici le terme de fluctuations de densité réfère aux fluctuations de  $\hat{n}$ , i.e. les fluctuations statistiques de la moyenne empirique de la densité dues à la variance d'échantillon. Cette fluctuation correspond à l'incertitude sur la vraie densité moyenne  $\bar{n}$ . En effet, on ne connaît pas *a priori* cette densité moyenne, et on peut seulement l'estimer à partir de la moyenne empirique  $\hat{n}$  sur l'échantillon.

par une distance dans  $[r \pm dr/2]$  pour les couples de catalogues *Data-Data*, *Random-Random*, *Data-Random*. On définit également  $N_{DD}$ ,  $N_{RR}$  et  $N_{DR}$  comme le nombre total de paires totales entre ces couples de catalogues. Avec la convention de compter les paires une fois seulement, on a :

$$N_{DD} = \frac{N_D(N_D - 1)}{2} \quad (4.10)$$

$$N_{RR} = \frac{N_R(N_R - 1)}{2} \quad (4.11)$$

$$N_{DR} = N_R N_D \quad (4.12)$$

Dans cette étude nous allons utiliser 4 estimateurs différents. Ce sont les estimateurs de Peebles-Hauser (Peebles and Hauser 1974), Davis-Peebles (Davis and Peebles 1983), Hamilton (Hamilton 1993) et Landy-Szalay (Landy and Szalay 1993), qui ont les expressions suivantes :

$$\hat{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 1 \quad (4.13)$$

$$\hat{\xi}_{DP}(r) = \frac{N_{DR}}{N_{DD}} \frac{DD(r)}{DR(r)} - 1 \quad (4.14)$$

$$\hat{\xi}_{HAM}(r) = \frac{N_{DR}^2}{N_{DD} N_{RR}} \frac{DD(r) RR(r)}{[DR(r)]^2} - 1 \quad (4.15)$$

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 2 \frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)} \quad (4.16)$$

On montre les calculs effectués dans l'étude Hamilton (1993) dans un cas simple où les catalogues sont *volume-limited* (i.e. avec une densité moyenne<sup>2</sup> constante dans le volume). Dans ce cas, la stratégie optimale pour le calcul de la fonction de corrélation est de pondérer toutes les galaxies de façon égale. La densité du catalogue  $n(\mathbf{x})$  est donnée par une somme de fonctions de Dirac centrées sur les galaxies du catalogue. On note  $\bar{n}$  la densité moyenne et  $\delta(\mathbf{x})$  le contraste de densité dans la catalogue :

$$\bar{n} = \mathbb{E}[n(\mathbf{x})] \quad (4.17)$$

$$\delta(\mathbf{x}) = \frac{n(\mathbf{x}) - \bar{n}}{\bar{n}} \quad (4.18)$$

En notant  $W$  la fonction indicatrice du catalogue et  $\langle \cdot \rangle$  l'intégrale sur le volume du catalogue. Par exemple  $\langle W(\mathbf{x}) n(\mathbf{x}) \rangle$  correspond à l'intégrale sur le volume de la densité du catalogue, et est donc égale au nombre de points dans le catalogue. On introduit les quantités suivantes d'espérance nulle :

$$\hat{\delta} = \frac{\langle W(\mathbf{x}) \delta(\mathbf{x}) \rangle}{\langle W(\mathbf{x}) \rangle} = \frac{1}{V} \langle W(\mathbf{x}) \delta(\mathbf{x}) \rangle \quad (4.19)$$

$$\Psi(r) = \frac{\langle \delta(\mathbf{x}) W(\mathbf{x}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \quad (4.20)$$

$$\hat{\xi}(r) = \frac{\langle \delta(\mathbf{x}) W(\mathbf{x}) \delta(\mathbf{y}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \quad (4.21)$$

où  $\langle \cdot \rangle_r$  représente une double intégrale sur le volume, restreinte à  $\mathbf{x}$  et  $\mathbf{y}$  séparés par une distance dans  $[r \pm dr/2]$ .  $\hat{\xi}$  est un estimateur non biaisé de la vraie fonction de corrélation  $\xi$ . Mais on ne peut pas le calculer en général car on ne connaît pas la densité moyenne  $\bar{n}$ , et donc on ne connaît pas  $\delta(\mathbf{x})$ .

On peut exprimer les différents estimateurs en fonction de  $\hat{\xi}$ ,  $\hat{\delta}$  et  $\Psi$  comme cela est fait dans Hamilton (1993). Cela requiert l'approximation de catalogues aléatoires avec un grand nombre de

2. On rappelle que l'on utilise le terme de densité moyenne pour désigner l'espérance de la densité  $\bar{n} = \mathbb{E}[n(\mathbf{x})]$ . Il ne faudra toutefois pas confondre cette 'moyenne' avec la moyenne empirique de la densité dans l'échantillon, qui présente des variations statistiques autour de l'espérance de la densité.

points, de sorte que les quantités  $\frac{1}{N_{RR}}RR(r)$  et  $\frac{1}{N_{DR}}DR(r)$  peuvent se réécrire :

$$\frac{1}{N_{RR}}RR(r) = \frac{\# \text{ paires de points aléatoires à distance } r' \in [r \pm dr/2]}{\# \text{ paires de points aléatoires}} \quad (4.22)$$

$$\frac{1}{N_{RR}}RR(r) \approx \frac{1}{V^2} \langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r \quad (4.23)$$

$$\frac{1}{N_{DR}}DR(r) = \frac{1}{N_D} \sum_{\mathbf{d}_j} \frac{\# \text{ points aléatoires } \mathbf{r}_i \text{ t.q. } |\mathbf{r}_i - \mathbf{d}_j| \in [r \pm dr/2]}{\# \text{ points aléatoires}} \quad (4.24)$$

$$\frac{1}{N_{DR}}DR(r) \approx \frac{1}{V} \frac{1}{N_D} \langle n(\mathbf{x}) W(\mathbf{x}) W(\mathbf{y}) \rangle_r \quad (4.25)$$

De la même manière le terme  $\frac{1}{N_{DD}}DD(r)$  se réécrit :

$$\frac{1}{N_{DD}}DD(r) = \frac{1}{N_D^2} \langle n(\mathbf{x}) W(\mathbf{x}) n(\mathbf{y}) W(\mathbf{y}) \rangle_r \quad (4.26)$$

On peut aussi réécrire le terme  $\hat{\delta}$  :

$$\hat{\delta} = \frac{1}{V} \langle W(\mathbf{x}) \delta(\mathbf{x}) \rangle = \frac{1}{\bar{n}V} \langle W(\mathbf{x}) n(\mathbf{x}) - \bar{n}(\mathbf{x}) \rangle \quad (4.27)$$

$$= \frac{1}{\bar{n}V} N_D - 1 \quad (4.28)$$

On obtient désormais :

$$\frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} = \frac{1}{\bar{n}^2 [1 + \hat{\delta}]^2} \frac{\langle n(\mathbf{x}) W(\mathbf{x}) n(\mathbf{y}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \quad (4.29)$$

$$\frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)} = \frac{1}{\bar{n} [1 + \hat{\delta}]} \frac{\langle n(\mathbf{x}) W(\mathbf{x}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \quad (4.30)$$

Finalement ces quantités se réécrivent :

$$\frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} = \frac{1}{[1 + \hat{\delta}]^2} \frac{\langle [\delta(\mathbf{x}) + 1] W(\mathbf{x}) [\delta(\mathbf{y}) + 1] W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \quad (4.31)$$

$$= \frac{\hat{\xi}(r) + 2\Psi(r) + 1}{[1 + \hat{\delta}]^2} \quad (4.32)$$

$$\frac{N_{RR}}{N_{DR}} \frac{DR(r)}{RR(r)} = \frac{1}{[1 + \hat{\delta}]} \frac{\langle [\delta(\mathbf{x}) + 1] W(\mathbf{x}) W(\mathbf{y}) \rangle_r}{\langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r} \quad (4.33)$$

$$= \frac{\Psi(r) + 1}{[1 + \hat{\delta}]} \quad (4.34)$$

$$\hat{\xi}_{PH}(r) = \frac{\hat{\xi}(r) + 2\Psi(r) - 2\hat{\delta} - \hat{\delta}^2}{[1 + \hat{\delta}]^2} \quad (4.35)$$

$$\hat{\xi}_{DP}(r) = \frac{\hat{\xi}(r) + \Psi(r) - \hat{\delta} - \Psi(r) \hat{\delta}}{[1 + \hat{\delta}] [1 + \Psi(r)]} \quad (4.36)$$

$$\hat{\xi}_H(r) = \frac{\hat{\xi}(r) - \Psi(r)^2}{[1 + \Psi(r)]^2} \quad (4.37)$$

$$\hat{\xi}_{LS}(r) = \frac{\hat{\xi}(r) - 2\hat{\delta}\Psi(r) + \hat{\delta}^2}{[1 + \hat{\delta}]^2} \quad (4.38)$$

Ces formules expliquent la supériorité des estimateurs d'Hamilton et de Landy-Szalay, avec  $\Psi$  et  $\hat{\delta}$  qui apparaissent au second ordre au numérateur. Les termes au dénominateur ne sont pas importants

car ils génèrent une faible erreur relative, alors que les termes au numérateur peuvent générer une forte erreur relative lorsque leur valeur devient non négligeable par rapport à  $\hat{\xi}$ . Pour les estimateurs d'Hamilton et de Landy-Szalay, l'erreur est dominée par l'erreur sur  $\hat{\xi}$  et peu affectée par les termes  $\Psi$  et  $\hat{\delta}$ , qui sont liées à l'incertitude sur  $\bar{n}$ .

Avec ces formules on voit que les estimateurs sont biaisés dans le cas général. En effet  $\hat{\delta}$  et  $\Psi(r)$  ont une espérance égale à 0, et  $\hat{\xi}(r)$  a une espérance égale à  $\xi(r)$ . Toutefois les termes sont combinées en multiplication et divisions. Donc l'espérance du membre de gauche ne s'obtient pas en remplaçant chacun des termes du membre de droite par son espérance dans les équations (4.35), (4.36), (4.37), (4.38).

#### 4.2.2 Le bias de la contrainte intégrale

Le catalogue aléatoire est utilisé pour mesurer un excès de paires du catalogue de galaxies par rapport à un catalogue aléatoire<sup>3</sup>. De façon équivalente, le catalogue aléatoire peut être vu comme un outil pour calculer des volumes. En notant à nouveau  $\mathcal{V}$  le domaine du catalogue et  $V$  son volume, si l'on prend la limite  $N_R \rightarrow \infty$ , on obtient comme dans la section 4.2.1 :

$$\frac{RR(r)}{N_{RR}} = \frac{\# \text{ paires à distance } r' \in [r \pm dr/2]}{\# \text{ paires}} \quad (4.39)$$

$$f(r) \stackrel{def}{=} \lim_{N_R \rightarrow \infty} \frac{RR(r)}{N_{RR}} = \frac{1}{V^2} \langle W(\mathbf{x}) W(\mathbf{y}) \rangle_r \quad (4.40)$$

$$= \frac{1}{V^2} \int_{\mathcal{V}} d^3 \mathbf{x} \int_{\mathcal{V}} d^3 \mathbf{y} \mathbb{1}_{|\mathbf{y}-\mathbf{x}| \in [r \pm dr/2]} \quad (4.41)$$

Pour simplifier le texte, on va écrire  $I$  et  $\hat{I}_{PH}$ ,  $\hat{I}_{DP}$ ,  $\hat{I}_H$ ,  $\hat{I}_{LS}$  ( $\hat{I}$  pour un estimateur en général) pour désigner l'intégration de la vraie fonction de corrélation et des différents estimateurs contre la fonction  $f(r)$  :

$$I \stackrel{def}{=} \int_0^{r_{\max}} f(r) \xi(r) \quad (4.42)$$

$$\hat{I}_{PH} \stackrel{def}{=} \int_0^{r_{\max}} f(r) \hat{\xi}_{PH}(r) \quad (4.43)$$

avec  $r_{\max}$  la distance maximum entre 2 points du domaine du catalogue.

On va montrer qu'il existe une contrainte sur l'estimateur de Peebles-Hauser  $\hat{\xi}_{PH}(r)$ , qui impose l'égalité suivante quelque soit la vraie fonction  $\xi(r)$  qui est estimée :

$$\hat{I}_{PH} = 0 \quad (4.44)$$

Pour un domaine de catalogue assez 'régulier' et pour une séparation  $r$  petite, la seconde intégrale dans l'équation (4.41) est égale pour quasiment toutes les points  $\mathbf{x}$  au volume de l'enveloppe sphérique  $\mathcal{V}_r$  avec  $|\mathbf{y} - \mathbf{x}| \in [r \pm dr/2]$ . Donc pour  $r$  petit, on obtient  $f(r) \approx \frac{V_r}{V} = \frac{4\pi r^2 dr}{V}$ , et si c'était le cas pour tout  $r$ , la contrainte (4.44) deviendrait :

$$\int_{\mathbb{R}^3} \hat{\xi}(\mathbf{r}) d^3 \mathbf{r} = 0. \quad (4.45)$$

Cependant quand  $r$  devient non négligeable par rapport à la taille de l'échantillon, on a  $f(r) \neq \frac{V_r}{V}$ . Donc la contrainte (4.44) est différente de (4.45) et dépend de la géométrie du domaine et de son volume.

3. Il faut noter que les catalogues de galaxies sont aussi aléatoires par nature mais ils sont corrélés spatialement. Dans ce contexte, le terme 'catalogue aléatoire' est donc utilisé de façon abusive pour désigner des catalogues aléatoires de Poisson sans corrélation.

Montrons la relation (4.44) pour l'estimateur de Peebles-Hauser :

$$\hat{\xi}_{PH}(r) = \frac{N_{RR}}{N_{DD}} \frac{DD(r)}{RR(r)} - 1 \approx \frac{1}{f(r)} \frac{1}{N_{DD}} DD(r) - 1 \quad (4.46)$$

En pratique, l'intégrale consiste à faire la somme sur tous les intervalles centrées en  $r_i$  (*bins*) de la corrélation estimée jusqu'en  $r_{max}$  :

$$\hat{I}_{PH} = \sum_i f(r_i) \hat{\xi}_{PH}(r_i) \approx \sum_i f(r_i) \left[ \frac{1}{f(r_i)} \frac{1}{N_{DD}} DD(r_i) - 1 \right] \quad (4.47)$$

$$= \frac{1}{N_{DD}} \sum_i DD(r_i) - \sum_i f(r_i) = 1 - 1 = 0 \quad (4.48)$$

$$(4.49)$$

Il est possible de montrer que la même contrainte  $\hat{I} = 0$  est approximativement vérifiée par les autres estimateurs. Pour cela, on a besoin de simplifier  $DR(r)$  dans la limite  $N_R \rightarrow \infty$  :

$$\frac{1}{N_D N_R} DR(r) = \frac{1}{N_D} \sum_{\mathbf{d}_j} \frac{\# \text{ points aléatoires } \mathbf{r}_i \text{ t.q. } |\mathbf{r}_i - \mathbf{d}_j| \in [r \pm dr/2]}{\# \text{ points aléatoires}} \quad (4.50)$$

$$g(r) = \lim_{N_R \rightarrow \infty} \frac{1}{N_D N_R} DR(r) = \frac{1}{N_D} \sum_{\mathbf{d}_j} \frac{1}{V} \int_{\mathcal{V}} d^3 \mathbf{y} \ 1_{|\mathbf{y} - \mathbf{d}_j| \in [r \pm dr/2]} \quad (4.51)$$

Cette fonction  $g(r)$  dépend de la position des galaxies dans le catalogue. On peut faire une autre approximation, à savoir que la taille de la corrélation est petite comparée au domaine du catalogue et qu'il y a un nombre assez grand de galaxies dans le catalogue. Alors les galaxies peuvent être considérées comme distribuées uniformément dans le domaine du catalogue, et on peut remplacer la moyenne par rapport aux points du catalogue par la moyenne sur le volume :

$$g(r) \approx \frac{1}{V} \int_{\mathcal{V}} d^3 \mathbf{x} \left( \frac{1}{V} \int_{\mathcal{V}} d^3 \mathbf{y} \ 1_{|\mathbf{y} - \mathbf{x}| \in [r \pm dr/2]} \right) = f(r) \quad (4.52)$$

Avec cette approximation, tous les estimateurs sont équivalents et vérifient la contrainte intégrale. Mais la dernière approximation n'est pas aussi bonne que pour l'estimateur de Peebles-Hauser, et la contrainte n'est pas aussi bien vérifiée comme on va le voir dans la section 4.4.

On voit encore que les estimateurs sont biaisés dans le cas général. La vraie fonction de corrélation ne vérifie pas nécessairement la contrainte intégrale, alors que les estimateurs la vérifient (approximativement) et donc leurs espérances la vérifient également.

### 4.3 Simulations utilisées

Dans la section 4.4.1 on va utiliser des réalisations de processus de segments de Cox avec la procédure détaillée dans la section 3.4.1. Ces processus sont très utiles car ils sont très simples à générer et on connaît leur fonction de corrélation analytique. De plus leur fonction de corrélation est strictement positive, ce qui va être utile pour tester la contrainte intégrale qui force des valeurs négatives dans les estimateurs de la fonction de corrélation (voir section 4.4).

Dans les sections 4.4.3 et 4.5, nous utiliserons des simulations log-normales des catalogues *volume-limited* Main VL et LRG VL du SDSS-II, détaillés dans la section 3.2.1. Pour générer des réalisations log-normales on utilise la procédure décrite dans la section 3.4.2.

On donne en entrée des simulations un spectre de puissance  $\Lambda$ CDM plat avec un spectre linéaire  $P_L$  avec les paramètres  $\Omega_m = 0.27$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 2.205^{-2}$ ,  $n_s = 1$ ,  $\sigma_8 = 0.8$ . Ces spectres sont calculés respectivement à *redshift*  $z = 0.1$  pour le Main VL et  $z = 0.3$  pour le LRG VL. On utilise un modèle simple pour les corrections non linéaires de l'évolution du champ de matière. Plus précisément on utilise la procédure décrite dans la section 1.3.4, avec une correction invariante d'échelle donnée par HALOFIT (Smith et al. 2003) et une dégradation non linéaire des BAOs. A partir du spectre



de puissance linéaire on modélise la dégradation non linéaire du pic BAO par un lissage Gaussien, appliqué seulement sur la signature BAO.

Pour isoler la signature BAO dans le spectre de puissance, on se sert du spectre de puissance *no wiggles* de Eisenstein and Hu (1998), qui est un spectre lisse avec la même forme que le spectre de puissance linéaire mais sans la signature BAO. On applique donc le lissage Gaussien à la différence entre le spectre linéaire et le spectre *no wiggles*,  $P_L(k) - P_{nw,L}(k)$ , et l'on restaure la forme générale du spectre en rajoutant  $P_{nw,L}(k)$  après le lissage :

$$P_{dw,L}(k) = P_{nw,L}(k) + e^{-a^2 k^2/2} [P_L(k) - P_{nw,L}(k)] \quad (4.53)$$

Pour prendre en compte les corrections non linéaires invariantes d'échelle, on applique à  $P_{dw,L}$  les corrections non linéaires de HALOFIT correspondant au spectre invariant d'échelle  $P_{nw,L}(k)$  :

$$P_{dw,NL}(k) = \frac{P_{nw,NL}(k)}{P_{nw,L}(k)} P_{dw,L}(k) \quad (4.54)$$

Pour calculer ces spectres, on utilise la librairie IDL *iCosmo* (Refregier et al. 2011), qui donne les spectres de puissances  $P_L$  et  $P_{nw,L}$  avec l'approximation de Eisenstein and Hu (1998). On utilise la valeur de  $a = 7 h^{-1}$  Mpc à la fois pour les simulations Main VL et LRG VL. Finalement on modélise le biais des galaxies par rapport à la matière par un biais constant. Les simulations ont donc une fonction de corrélation théorique égale à :

$$\xi_{\Lambda\text{CDM}}(r) = b^2 \xi_{dw,NL}(r) \quad (4.55)$$

avec  $\xi_{dw,NL}$  la transformée de Hankel de  $P_{dw,NL}$  (voir section 2.1.2 pour les détails de la relation entre le spectre de puissance et la fonction de corrélation). Les valeurs du biais ont été ajustées pour trouver un bon accord entre les estimateurs de la fonction de corrélation  $\hat{\xi}$  du Main VL et du LRG VL montrés sur la figure 3.4 et les fonctions de corrélation  $\hat{\xi}_{\Lambda\text{CDM}}$  (voir figure 4.1).

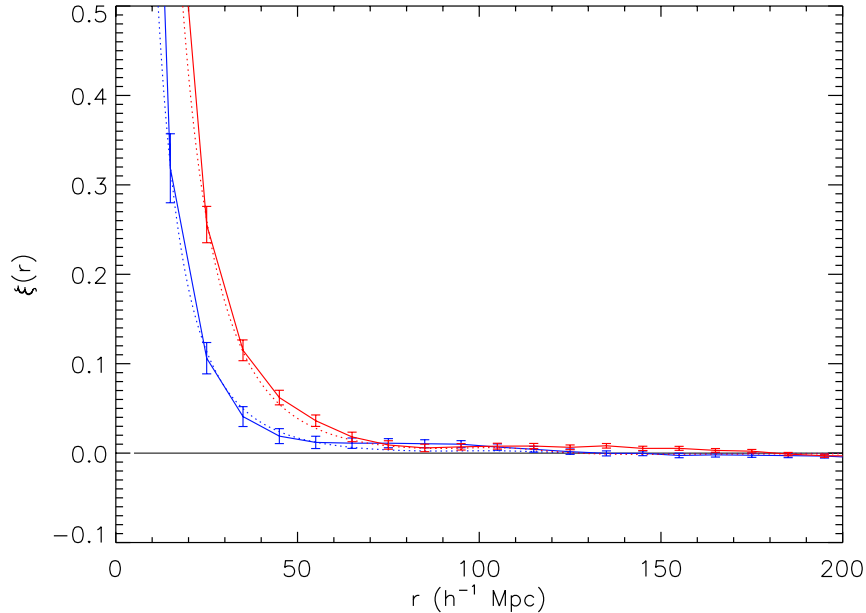


FIGURE 4.1 – Fonctions de corrélation estimées pour les échantillons Main VL (trait plein bleu) et LRG VL (trait plein rouge) pour une cosmologie fiducielle  $\Lambda\text{CDM}$  plate avec  $\Omega_m = 0.27$ , et fonctions de corrélation théoriques  $b^2 \xi_{\Lambda\text{CDM}}$  avec un biais ajusté pour fitter les données. On trouve des valeurs du biais égales à  $b = 1.65$  pour le Main VL et  $b = 2.5$  pour le LRG VL. Les barres d'erreur sont estimées à partir de 2,000 simulations log-normales  $\Lambda\text{CDM}$  avec les paramètres  $\Omega_m = 0.27$ ,  $h = 0.7$ ,  $\Omega_b h^2 = 2.205^{-2}$ ,  $n_s = 1$ ,  $\sigma_8 = 0.8$  et respectivement  $b = 1.65$  pour le Main VL et  $b = 2.5$  pour le LRG VL.

## 4.4 Effet de la contrainte intégrale sur des simulations

### 4.4.1 Effet de la contrainte intégrale sur des processus de segments de Cox

Pour montrer comment la contrainte intégrale peut affecter l'estimation de la fonction de corrélation, on génère des réalisations du processus de segments de Cox (voir [Martínez and Saar \(2002\)](#)). Le champ consiste en des segments de longueur  $l$  distribués aléatoirement dans le volume et des points distribués aléatoirement sur chaque segment. L'intensité du processus ponctuel  $\lambda$  est égale à la longueur moyenne de segments par unité de volume  $L_V$ , multipliée par le nombre moyen de points par unité de longueur  $\lambda_l$ . Ce processus est facile à générer et sa fonction de corrélation est connue analytiquement ([Stoyan et al. 1995](#)) :

$$\xi(r) = \begin{cases} \frac{1}{2\pi r^2 L_V} - \frac{1}{2\pi r l L_V} & \text{pour } r \leq l \\ 0 & \text{pour } r \geq l \end{cases} \quad (4.56)$$

Cette fonction de corrélation est toujours positive, donc la contrainte intégrale force les estimateurs à prendre de fausses valeurs négatives.

Pour nos expériences, on considère le processus avec longueur de segments  $l = 10$  (l'unité est ici arbitraire), une longueur moyenne par unité de volume  $L_V = 0.1$  et un nombre moyen de points par unité de longueur de  $\lambda_l = 1.8$ . On a généré 2000 cubes de taille  $a = 10$ , 2000 cubes de taille  $a = 20$  et 512 cubes de taille  $a = 50$ , sur lesquels on a estimé la fonction de corrélation avec des tailles d'intervalles  $dr = 2$ . On trace sur la figure 4.2 les valeurs empiriques de la moyenne et de l'écart-type :

$$\bar{\xi} = \frac{1}{N} \sum_{k=1}^N \hat{\xi}_k \quad (4.57)$$

$$s(\hat{\xi}) = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (\hat{\xi}_k - \bar{\xi})^2} \quad (4.58)$$

Sur la figure 4.2, on montre aussi la corrélation analytique, discrétisée de façon équivalente au comptage de paires. Pour un intervalle  $[r_i - dr/2, r_i + dr/2]$ , la fonction de corrélation analytique discrétisée est égale à :

$$\xi(r_i) = \frac{\int_{r_i - dr/2}^{r_i + dr/2} \xi(r) r^2 dr}{\int_{r_i - dr/2}^{r_i + dr/2} r^2 dr} \quad (4.59)$$

Ainsi, en l'absence de biais, les estimateurs discrétisés de la fonction de corrélation ont une espérance égale à la fonction analytique discrétisée, i.e.  $\mathbb{E}[\hat{\xi}(r_i)] = \xi(r_i)$ .

Pour illustrer la présence d'un biais dû à la contrainte intégrale, on trace sur la figure 4.3 les moyennes des estimateurs  $\bar{\xi}$  avec l'incertitude sur cette moyenne calculé comme  $\frac{s(\hat{\xi})}{\sqrt{N}}$  dans les cas  $a = 20$  et  $a = 50$ . En effet l'espérance de l'estimateur  $\mathbb{E}[\bar{\xi}]$  est égale à la moyenne  $\bar{\xi}$  à cette incertitude près<sup>4</sup>, et un écart de la moyenne plus grand que cette incertitude par rapport à la vraie corrélation traduit un biais de l'estimateur.

On observe un biais pour tous les estimateurs et toutes les tailles de cube. Comme on pouvait le prévoir, le biais diminue quand le volume augmente, de même que la variance diminue. Le biais est approximativement égal à la moitié de l'écart-type de l'estimateur dans une large région pour  $a = 10$  et  $a = 20$ , alors que le biais devient très petit par rapport à l'écart-type pour  $a = 50$ . Le biais est similaire pour les différents estimateurs pour  $a = 10$  et  $a = 20$ . Alors que pour  $a = 50$ , le biais devient plus petit pour les estimateurs de Hamilton et Landy-Szalay que pour les estimateurs de Peebles-Hausser et Davis-Peebles. Ainsi le biais diminue plus rapidement avec le volume pour les estimateurs de Hamilton et Landy-Szalay, i.e. qu'ils sont moins affectés par la contrainte intégrale.

On montre sur la figure 4.4 les moyennes pondérées des estimateurs  $f(r_i)\bar{\xi}(r_i)$  dans les cas  $a = 20$  et  $a = 50$ . On peut voir que la somme pondérée de  $\bar{\xi}$  s'annule pour les estimateurs et non pour la fonction de corrélation analytique  $\xi$ . Cet effet est clair pour les cas  $a = 20$  et  $a = 10$  (ce dernier cas n'est pas montré car les résultats sont similaires au cas  $a = 20$ ). En particulier la contrainte

4. Pour voir cela, il suffit de voir que la variance sur un estimateur est environ égale à  $s(\hat{\xi})^2$ , si bien que la variance de  $\bar{\xi}$  est égale à  $\frac{1}{N} s(\hat{\xi})^2$ .

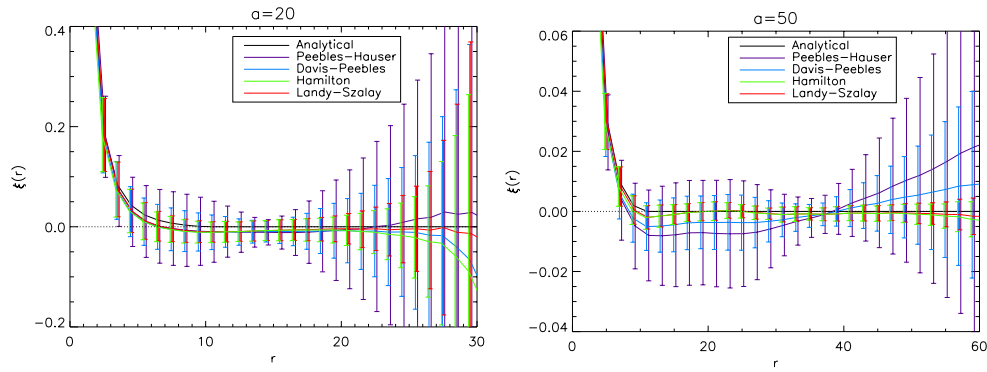


FIGURE 4.2 – A gauche : Moyenne  $\bar{\xi}$  et écart-type  $s(\hat{\xi})$  des différents estimateurs avec  $N = 2000$  réalisations de segments de Cox pour  $a = 20$ . On trace la fonction de corrélation analytique (noir), les estimateurs Peebles-Hauser (violet), Davis-Peebles (bleu clair), Hamilton (vert) et Landy-Szalay (rouge). A droite : Même chose avec  $N = 512$  réalisations de segments de Cox pour  $a = 50$ .

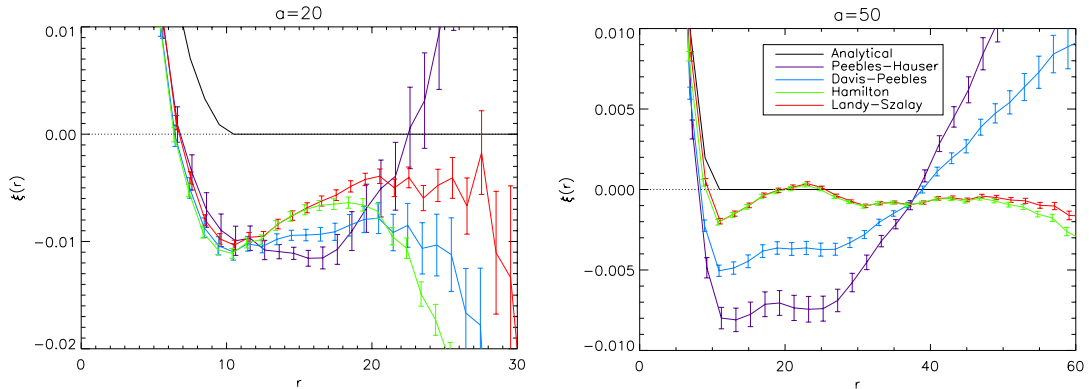


FIGURE 4.3 – A gauche : Moyenne  $\bar{\xi}$  des différents estimateurs et incertitude sur cette moyenne  $\frac{s(\hat{\xi})}{\sqrt{N}}$  avec  $N = 2000$  réalisations de segments de Cox pour  $a = 20$ . A droite : Même chose avec  $N = 512$  réalisations de segments de Cox pour  $a = 50$ .

intégrale force les estimateurs à prendre des valeurs négatives aux échelles  $r$  intermédiaires, i.e. pour les grandes valeurs de  $f(r)$ . Le cas  $a = 50$  est moins clair et le biais des estimateurs ne semble pas venir uniquement de la contrainte intégrale car il peut être positif ou négatif selon l'échelle. Le faible biais toujours présent pour  $a = 50$  pourrait donc venir d'autres effets (e.g. du nombre fini de points dans les catalogues aléatoires).

On montre dans la table 4.1 les valeurs de  $I$  pour la fonction analytique  $\xi$  et  $\hat{I}$  pour les moyennes des différents estimateurs. La contrainte est très bien vérifiée ( $\hat{I} \approx 0$ ), en particulier pour l'estimateur de Peebles-Hauser, même si elle n'est pas bien vérifiée pour la fonction analytique ( $I \gg \hat{I}$ ).

La fonction pondératrice  $f$  somme à 1 (i.e.  $\sum_i f(r_i) = 1$ , voir équation (4.41)). Donc une différence entre  $\sum_i f(r_i)\xi(r_i)$  et  $\sum_i f(r_i)\hat{\xi}(r_i)$  (i.e. entre  $I$  et  $\hat{I}$ ) implique en moyenne une même différence entre  $\xi$  et  $\hat{\xi}$ . Comme des différences positives et négatives entre  $\xi$  et  $\hat{\xi}$  peuvent se compenser dans la somme, cette *ansatz* peut en réalité donner une sous-estimation. Autrement dit, la différence en valeur absolue entre  $\xi$  et  $\hat{\xi}$  peut être en moyenne plus grande que la différence en valeur absolue entre  $I$  et  $\hat{I}$ .

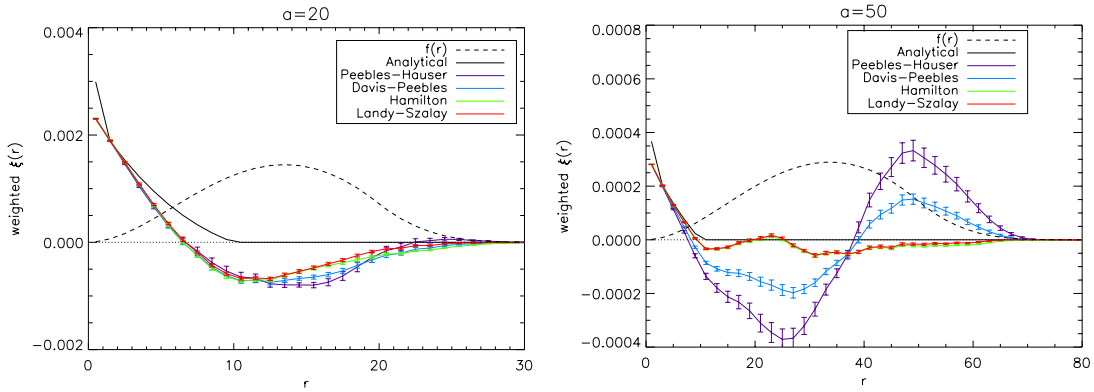


FIGURE 4.4 – A gauche : Moyenne pondérée des différents estimateurs  $f(r_i)\bar{\xi}$  et incertitude  $f(r_i)\frac{s(\hat{\xi})}{\sqrt{N}}$  sur cette moyenne avec  $N = 2000$  réalisations de segments de Cox pour  $a = 20$ . On trace en également la fonction  $f(r)$  (ligne pointillée noire). A droite : Même chose avec  $N = 512$  réalisations de segments de Cox pour  $a = 50$ .

TABLE 4.1 –

	$I$	$\sigma(\mathcal{V})$	$\hat{I}_{PH}$	$\hat{I}_{DP}$	$\hat{I}_H$	$\hat{I}_{LS}$
$a = 10$	0.059	0.24	$2.87 \times 10^{-5}$	$9.24 \times 10^{-5}$	$7.16 \times 10^{-3}$	0.0125
$a = 20$	$9.6 \times 10^{-3}$	0.098	$1.88 \times 10^{-5}$	$-5.54 \times 10^{-4}$	$2.23 \times 10^{-4}$	$1.47 \times 10^{-3}$
$a = 50$	$7.8 \times 10^{-4}$	0.027	$3.8 \times 10^{-6}$	$-7.04 \times 10^{-5}$	$-1.86 \times 10^{-5}$	$1.02 \times 10^{-4}$

NOTES.—Valeurs de  $I$  pour la fonction analytique  $\xi$  et  $\hat{I}$  pour les différents estimateurs et pour  $a = 10$ ,  $a = 20$  et  $a = 50$ .

#### 4.4.2 Prédiction de l'effet de la contrainte intégrale

On vient de voir que le biais sur  $\xi$  dû à la contrainte intégrale est directement relié à la valeur de l'intégrale  $I$  de la fonction analytique  $\xi$ . Ainsi on peut tenter de prédire l'importance du biais à partir de la valeur de  $I$ .

Pour les estimateurs d'Hamilton et Landy-Szalay sur les processus de segments de Cox, la contrainte devient faible entre  $a = 20$  et  $a = 50$ . Ces valeurs de  $a$  correspondent à des valeurs de  $I$  pour la fonction analytique  $\xi$  approximativement égale à 0.01 et 0.001. Une quantité plus intuitive que  $I$  est la variance de masse normalisée dans un domaine  $\mathcal{V}$  :

$$\sigma^2(\mathcal{V}) = \frac{\mathbb{E}[M(\mathcal{V})^2] - \mathbb{E}[M(\mathcal{V})]^2}{\mathbb{E}[M(\mathcal{V})]^2} \quad (4.60)$$

$\sigma(\mathcal{V})$  représente la fluctuation de masse dans le domaine. On peut montrer que  $I$  est égale à  $\sigma^2(\mathcal{V})$  au bruit de Poisson près (*shot noise*, voir Gabrielli et al. (2002)). Ce bruit de Poisson qui est dû au nombre finis de galaxies peut généralement être négligé. Ainsi on peut prédire l'effet de la contrainte à partir de la valeur de  $\sigma(\mathcal{V})$ . Les processus de segments de Cox avec  $a = 20$  et  $a = 50$  correspondent respectivement à  $\sigma(\mathcal{V}) \approx 0.1$  et  $\sigma(\mathcal{V}) \approx 0.03$ . Donc la contrainte affecte encore les estimateurs pour un niveau d'homogénéité de 10% (fluctuation de masse de 10% dans le volume), et devient négligeable pour un niveau d'homogénéité de 3%.

TABLE 4.2 –

	$I$	$\sigma(V)$	$\hat{I}_{PH}$	$\hat{I}_{DP}$	$\hat{I}_H$	$\hat{I}_{LS}$
Main VL	$6.61 \times 10^{-4}$	0.026	$2.39 \times 10^{-4}$	$9.05 \times 10^{-5}$	$1.38 \times 10^{-4}$	$6.99 \times 10^{-4}$
LRG VL	$1.01 \times 10^{-4}$	0.010	$6.17 \times 10^{-5}$	$7.52 \times 10^{-5}$	$9.21 \times 10^{-5}$	$9.54 \times 10^{-5}$

NOTES.—Valeurs de  $I$  et  $\hat{I}$  pour les différents estimateurs.

#### 4.4.3 Effet de la contrainte intégrale sur des simulations Main VL et LRG VL

On s'intéresse maintenant à l'effet de la contrainte intégrale sur des simulations des catalogues Main VL et LRG VL du SDSS-II présentés dans la section 3.2.1.

La contrainte est de la forme  $\int_0^{r^{max}} f(r)\hat{\xi}(r) \approx 0$  avec  $f(r) \approx \frac{4\pi r^2 dr}{V}$  pour  $r$  petit. En supposant que  $\int_0^\infty r^2 \xi(r) dr$  converge, l'intégrale  $\int_0^{r^{max}} f(r)\xi(r)$  tend vers 0 en  $\frac{1}{V}$  pour les volumes larges.

Dans les modèles cosmologiques usuels, le spectre de puissance vérifie  $P(k) \propto k^{n_s}$  pour  $k \ll 1$  avec l'indice spectral  $n_s \approx 1$  (le spectre Harrison Zel'dovich invariant d'échelle correspond à  $n_s = 1$ , voir section 1.3.1). Ainsi le spectre de puissance tend vers 0 pour  $k \rightarrow 0$ , et donc la fonction de corrélation qui est sa transformée de Fourier a une intégrale nulle :

$$\int_0^\infty r^2 \xi(r) dr = 0 \quad (4.61)$$

Ceci facilite la convergence de l'intégrale  $I = \int_0^{r^{max}} f(r)\hat{\xi}(r) \approx 0$  quand le volume augmente. La table 4.2 donne les valeurs de  $I$  pour la fonction de corrélation analytique des simulations  $\xi_{\Lambda\text{CDM}}$  et  $\hat{I}$  pour les différents estimateurs  $\hat{\xi}$ . Pour le catalogue Main VL,  $\hat{I}$  est significativement plus proche de 0 que  $I$ , ce qui veut dire que la contrainte intégrale affecte les estimateurs. En revanche ce n'est pas le cas pour le catalogue LRG VL où la contrainte intégrale ne semble pas affecter les estimateurs.

On rappelle que la valeur de  $I$  donne approximativement la valeur du biais :  $I$  est de l'ordre de  $10^{-3}$  pour le catalogue Main VL et de l'ordre de  $10^{-4}$  pour le catalogue LRG VL. Si l'on compare ceci aux valeurs de  $\xi$  aux échelles qui nous intéressent (i.e. habituellement entre  $50$  et  $200 h^{-1} \text{Mpc}$ ), le biais est important pour le catalogue Main VL et négligeable pour le catalogue LRG VL.

On peut aussi établir une correspondance avec les résultats sur les processus de segments de Cox obtenus section 4.4.1, où la contrainte devient négligeable pour  $\sigma(V) \approx 0.03$ . Le catalogue Main VL a la même valeur de  $\sigma(V) \approx 0.03$  mais l'effet reste présent. Dans le catalogue LRG VL,  $\sigma(V)$  est environ trois fois plus petit, et il n'est donc pas surprenant que le biais soit négligeable.

## 4.5 Comparaison des différents estimateurs sur des simulations Main VL et LRG VL

Dans cette section on veut comparer les différents estimateurs sur des simulations des catalogues Main VL et LRG VL du SDSS-II. On continue à s'intéresser au biais dû à la contrainte intégrale, mais on regarde également la variance des estimateurs. On utilise  $N = 2000$  réalisations log-normales à la fois pour le catalogue Main VL et pour le catalogue LRG VL.

Pour chaque réalisation on utilise 100,000 points dans les catalogues aléatoires pour le calcul des estimateurs (i.e. pour calculer les quantités  $RR(r)$  and  $DR(r)$  introduites section 4.2.1). Ceci garantit que l'erreur due au nombre finis de points aléatoires est petite. Pour chaque réalisation on

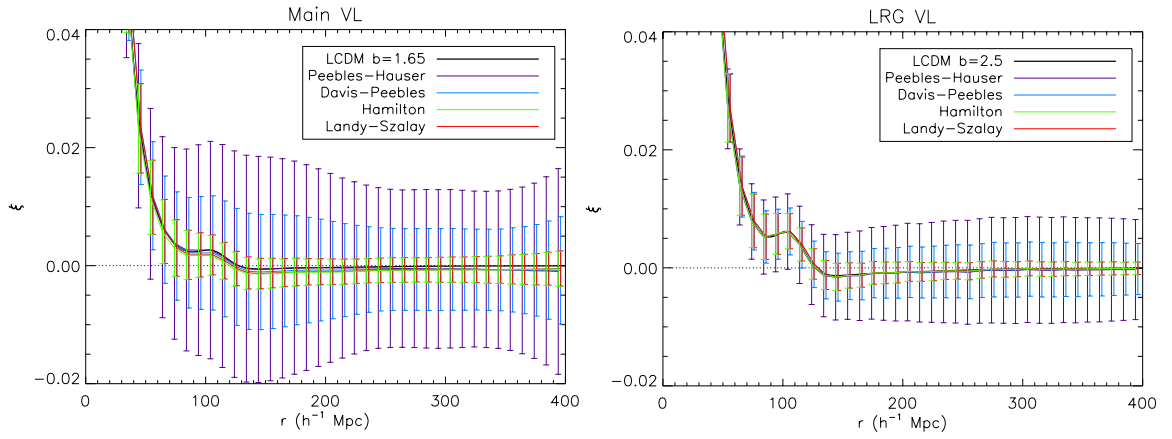


FIGURE 4.5 – A gauche : Moyenne  $\bar{\xi}$  et écart-type  $s(\hat{\xi})$  des différents estimateurs avec  $N = 2000$  réalisations Main VL, Peebles-Hauser (violet), Davis-Peebles (bleu clair), Hamilton (vert), Landy-Szalay (rouge) et modèle  $\Lambda$ CDM à  $z = 0.1$  avec  $b = 1.5$  (noir). A droite : Même chose avec  $N = 2000$  réalisations LRG VL, mais avec un modèle  $\Lambda$ CDM à  $z = 0.3$  avec  $b = 2.5$ .

utilise un catalogue aléatoire différent, et lorsque l'on moyenne sur toutes les réalisations log-normales l'influence du nombre fini de points aléatoires est extrêmement petite. Toutefois pour une réalisation donnée, la fluctuation due au nombre fini de points aléatoires peut accroître quelque peu la variance des estimateurs. La contribution à la variance des estimateurs vient du nombre de points aléatoires, de la géométrie du catalogue, ainsi que de la taille des intervalles pour estimer  $\xi$  (dans tous nos tests on a pris des tailles d'intervalle de  $10h^{-1}\text{Mpc}$ ). De façon plus précise, la variance vient de la précision avec laquelle  $\frac{1}{N_{RR}}RR(r)$  approxime  $\frac{1}{V^2} \int_V d^3\mathbf{x} \int_V d^3\mathbf{y} 1_{|\mathbf{y}-\mathbf{x}| \in [r \pm dr/2]}$ .

On montre sur la figure 4.5 les moyennes et écarts-types des estimateurs comparés à la fonction de corrélation analytique des simulations. De même que pour les processus de Cox, la fonction de corrélation analytique est discrétisée de façon équivalente au comptage de paire. Pour un intervalle  $[r_i - dr/2, r_i + dr/2]$ , la fonction de corrélation analytique discrétisée est égale à :

$$\xi_{\Lambda\text{CDM}}(r_i) = \frac{\int_{r_i-dr/2}^{r_i+dr/2} \xi_{\Lambda\text{CDM}}(r) r^2 dr}{\int_{r_i-dr/2}^{r_i+dr/2} r^2 dr} \quad (4.62)$$

Par soucis de clarté, on a translaté les courbes de  $\pm 1h^{-1}\text{Mpc}$ . On peut voir qu'il y a un biais dans les estimateurs du catalogue Main VL, avec une moyenne qui diffère d'environ la moitié de l'écart-type pour  $r > 90h^{-1}\text{Mpc}$  par rapport à la vraie valeur de  $\xi$ . On peut voir cela plus clairement sur les figures 4.6 qui montrent l'incertitude sur la moyenne empirique des  $N$  simulations, i.e.  $\frac{\hat{\sigma}}{\sqrt{N}}$ . Pour le catalogue LRG VL, les moyennes des estimateurs sont pratiquement indistinguables de la vraie valeur.

Concernant les variances des estimateurs, elles sont bien plus faibles sur le catalogue LRG VL que sur le catalogue Main VL. Ceci vient du fait que le volume est bien plus grand et que le bruit de Poisson (i.e. le nombre fini de paires) reste faible pour un nombre de galaxies  $N_D \approx 34000$  et des intervalles de taille  $10h^{-1}\text{Mpc}$ .

On voit finalement que les estimateurs de Hamilton et Landy-Szalay sont bien meilleurs que les autres estimateurs en terme de variance. Cela est en accord avec des études précédentes (Pons-Bordería et al. 1999; Kerscher et al. 2000) qui montrent une supériorité de ces estimateurs sur des catalogues plus petits. Ceci est aussi en accord avec l'analyse dans Landy and Szalay (1993) qui fait une étude analytique sur les processus de Poisson sans corrélation. Dans ce cas, les estimateurs de Landy-Szalay et de Hamilton ont une variance qui décroît en  $\frac{1}{n}$  (i.e. en  $\frac{1}{V^2}$ , avec  $V$  le volume de l'échantillon), alors que les estimateurs de Peebles-Hauser et de Davis-Peebles ont une variance qui décroît en  $\frac{1}{n}$  (i.e. en  $\frac{1}{V}$ ).

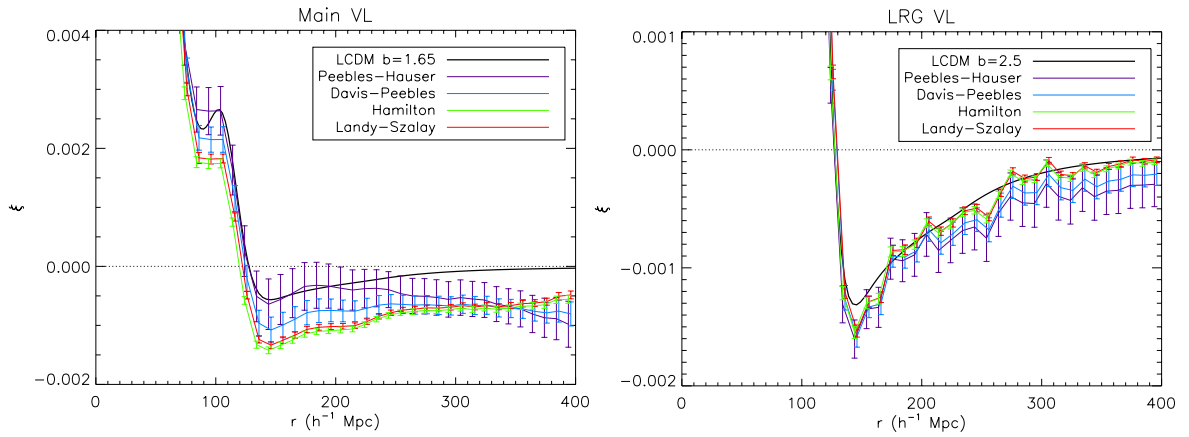


FIGURE 4.6 – A gauche : Moyenne  $\bar{\xi}$  des différents estimateurs et incertitude sur cette moyenne  $\frac{s(\hat{\xi})}{\sqrt{N}}$  avec  $N = 2000$  réalisations Main VL, Peebles-Hauser (violet), Davis-Peebles (bleu clair), Hamilton (vert), Landy-Szalay (rouge) et modèle  $\Lambda$ CDM à  $z = 0.1$  avec  $b = 1.5$  (noir). A droite : Même chose avec  $N = 2000$  réalisations LRG VL, mais avec un modèle  $\Lambda$ CDM à  $z = 0.3$  avec  $b = 2.5$ .

## 4.6 Conclusion

Comme on l'a vu dans ce chapitre, l'incertitude sur  $\bar{n}$  cause un biais dans les estimateurs de la fonction de corrélation, et notamment le biais connu sous le nom de contrainte intégrale. Notre objectif principal était de quantifier ce biais dans le cadre du modèle standard et pour les relevés de galaxies actuels, afin de vérifier la validité de l'analyse de la fonction de corrélation.

En utilisant des simulations log-normales  $\Lambda$ CDM du relevé SDSS DR7, on a montré que ce biais est très faible pour les relevés de galaxies actuels, et qu'il devient d'autant plus faible que la taille du relevé augmente. Pour les processus de segments de Cox, l'effet devient faible lorsque les fluctuations de masse dans la volume deviennent inférieures à 3% ( $\sigma(\mathcal{V}) < 0.03$ ). Ce niveau d'homogénéité est atteint l'échantillon Main VL. Toutefois la contrainte intégrale a toujours un effet pour ce catalogue, puisqu'elle cause un biais d'environ la moitié de l'écart-type  $s(\hat{\xi})$  pour  $r > 90h^{-1}\text{Mpc}$ . Pour le catalogue LRG VL, avec  $\sigma(\mathcal{V}) \approx 0.01$ , nous n'avons pu détecter aucun biais dans les estimateurs.

L'analyse classique de la fonction de corrélation dans le cadre du modèle  $\Lambda$ CDM est donc cohérente et non biaisée pour les relevés de galaxies actuels. Ce résultat montre que le modèle  $\Lambda$ CDM est testable puisqu'il conduit à une estimation non biaisée de la fonction de corrélation.

Au cours de notre étude, nous avons également pu comparer les différents estimateurs. On a vérifié les conclusions provenant d'études précédentes, à savoir que les estimateurs de Landy-Szalay et de Hamilton ont des variances nettement plus faibles que les estimateurs de Peebles-Hauser et de Davis-Peebles. Nous avons donné une raison de cette supériorité, à savoir la sensibilité des différents estimateurs à l'incertitude sur la densité moyenne  $\bar{n}$  (à partir des quantités  $\hat{\delta}$  et  $\Psi(r)$  de la section 4.2.1) : les estimateurs de Peebles-Hauser et de Davis-Peebles dépendent de cette fluctuation au premier ordre, alors que les estimateurs de Hamilton et Landy-Szalay n'ont qu'une dépendance au second ordre.

Mentionnons enfin qu'une étude récente (Vargas Magaña et al. in prep.) se penche sur l'estimation de la fonction de corrélation, en considérant une généralisation des estimateurs que l'on a étudié. L'idée est de considérer toutes les combinaisons possibles des ratios  $DD$ ,  $DR$  et  $RR$  au 1<sup>er</sup> et 2<sup>nd</sup> ordre. La conclusion de cette étude est qu'un tel estimateur optimisé sur toutes ces combinaisons permet une amélioration de 20 à 25% des barres d'erreur par rapport à l'estimateur de Landy-Szalay.

# Chapitre 5

## Détection des Oscillations Baryoniques Acoustiques

### Sommaire

---

<b>5.1</b>	<b>Les différentes méthodes de détection des BAOs</b>	<b>86</b>
5.1.1	Principe de la détection des BAOs	86
5.1.2	Historique de la détection des BAOs dans les grandes structures	89
5.1.3	But du travail dans ce chapitre	90
<b>5.2</b>	<b>Méthode classique pour la détection des BAOs</b>	<b>91</b>
5.2.1	La statistique du $\chi^2$	91
5.2.2	La méthode du $\chi^2$ pour la détection des BAOs	91
5.2.3	Limitations de la méthode du $\chi^2$	93
<b>5.3</b>	<b>Nouvelle méthode pour la détection des BAOs</b>	<b>94</b>
<b>5.4</b>	<b>Modélisation de <math>\hat{\xi}</math> et simulations SDSS DR7-Full</b>	<b>96</b>
5.4.1	Modèles de fonctions de corrélation pour $\mathcal{H}_0$ et $\mathcal{H}_1$	96
5.4.2	Simulations log-normales SDSS DR7-Full utilisées	98
5.4.3	Vérification de la modélisation statistique de $\hat{\xi}$	99
<b>5.5</b>	<b>Détection des BAOs sur des simulations SDSS DR7-Full</b>	<b>101</b>
5.5.1	Détection des BAOs avec une matrice de covariance constante $C$	101
5.5.2	Détection des BAOs avec une matrice de covariance $C_\theta$	108
<b>5.6</b>	<b>Conclusion</b>	<b>110</b>

---

Comme on l'a vu dans la section 1.4, les modèles CDM avec baryons prédisent l'existence d'ondes acoustiques (BAOs) dans le plasma chaud avant la recombinaison, quand les baryons et les photons étaient couplés. Ces ondes sphériques sont générées par la compétition entre la gravitation qui provoque l'effondrement des surdensités et la pression des photons qui s'oppose à cet effondrement. Environ 380,000 ans après le Big Bang, il y a découplage des baryons et des photons, et ces ondes sphériques se 'figent' à la distance de l'horizon sonique  $r_s$ . A cause de leur grande taille, leur empreinte sur la distribution de matière a principalement subie une évolution linéaire et devrait être bien visible dans les grandes structures actuelles.

L'existence des BAOs a d'abord été prévue théoriquement dans [Peebles and Yu \(1970\)](#), et la détection des BAOs dans les grandes structures de l'univers à faible *redshift* est donc une confirmation très forte du paradigme cosmologique. Le fait d'observer le pic BAO à l'échelle attendue dans les grandes structures est plus précisément un argument très fort pour le modèle standard actuel : un modèle CDM avec un évolution linéaire aux grandes échelles à partir d'un *redshift*  $z \approx 1,000$  et avec une composante d'énergie noire qui accélère l'expansion.



## 5.1 Les différentes méthodes de détection des BAOs

### 5.1.1 Principe de la détection des BAOs

Définissons les deux différentes hypothèses :

$$\begin{aligned}\mathcal{H}_0 & : \text{ hypothèse non-BAO} \\ \mathcal{H}_1 & : \text{ hypothèse BAO}\end{aligned}$$

La détection des BAOs est équivalente à ce problème de test d'hypothèse. La procédure habituelle en cosmologie est de choisir une statistique (*test statistic*) pour voir si les observations sont compatibles avec  $\mathcal{H}_0$ . Si la statistique a une valeur très improbable sous  $\mathcal{H}_0$  (plus qu'un seuil donné), l'hypothèse  $\mathcal{H}_0$  est rejetée et l'hypothèse  $\mathcal{H}_1$  est acceptée.

On se restreint ici au cas où l'observation est donnée par la fonction de corrélation discrétisée  $\hat{\xi} = (\hat{\xi}_i)_{1 \leq i \leq n}$ , mais les différentes méthodes que l'on va présenter peuvent aussi s'appliquer au spectre de puissance par exemple. Alors si l'on choisit une statistique  $S(\hat{\xi})$  à partir de l'observation  $\hat{\xi}$ , le test d'hypothèse de taille  $\alpha$  peut être formulé comme suit :

- si  $S(\hat{\xi}) \leq \eta$  alors accepter  $\mathcal{H}_1$  (i.e. rejeter  $\mathcal{H}_0$ )
- si  $S(\hat{\xi}) > \eta$  alors accepter  $\mathcal{H}_0$  (i.e. rejeter  $\mathcal{H}_1$ )

où la taille  $\alpha$  correspond à la probabilité de rejeter  $\mathcal{H}_0$  si  $\mathcal{H}_0$  est vraie :

$$\alpha = P\left(S(\hat{\xi}) \leq \eta \mid \mathcal{H}_0\right) \quad (5.1)$$

En pratique, les tests avec des niveaux de seuil  $\alpha$  prédéfinis sont peu utilisés en cosmologie, et il est plus commun de citer le niveau de significativité correspondant à la valeur obtenue pour la statistique. Pour une réalisation avec une valeur donnée  $S(\hat{\xi}) = x$ , la significativité (donnée comme une valeur  $p$ ) est égale à :

$$\alpha(x) = P\left(S(\hat{\xi}) \leq x \mid \mathcal{H}_0\right) \quad (5.2)$$

Une façon plus répandue d'exprimer la significativité est de la quantifier comme un nombre de  $\sigma$ . Il s'agit simplement du nombre d'écart-types par rapport à la moyenne pour une variable Gaussienne, qui correspond à la valeur  $p$  obtenue :

$$\alpha(x) = P(|X - \mu| > k\sigma) \quad (5.3)$$

où  $X$  est une variable Gaussienne  $X \sim \mathcal{N}(\mu, \sigma)$ , et  $k\sigma$  est l'expression de la significativité comme un nombre de  $\sigma$ .

A notre connaissance, toutes les méthodes qui ont été utilisées pour la détection des BAOs ont cette méthodologie en commun, et fournissent un niveau de significativité soit en termes de nombre de  $\sigma$ , soit en termes de valeur  $p$ . La différence entre les différentes méthodes de détection provient donc simplement des hypothèses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$  ainsi que du choix de la statistique. En réalité, le niveau de détection ne dépend que du choix de  $\mathcal{H}_0$  et de la statistique, mais l'hypothèse  $\mathcal{H}_1$  est souvent utilisée pour optimiser la statistique comme ce sera le cas pour le rapport de vraisemblance généralisé dans la section 5.3.

On va distinguer deux types de méthodes pour la détection des BAOs. Le premier type de méthode basée sur l'analyse en ondelette, est principalement sensible au pic BAO dans la fonction de corrélation. D'un autre côté les méthodes de détection classiques, et notamment la méthode classique du  $\chi^2$ , sont basées sur une modélisation statistique complète de l'estimateur  $\hat{\xi}$  pour les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ .

#### Méthodes basées sur un modélisation complète de $\hat{\xi}$

La plupart des méthodes de détection sont basées sur la modélisation statistique de  $\hat{\xi}$  pour chaque hypothèse  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . Ceci inclut la modélisation de l'espérance  $\mathbb{E}[\hat{\xi}]$  de l'estimateur  $\hat{\xi}$  de la section 2.2.5, mais cela inclut aussi la modélisation de la variance de l'estimateur  $\hat{\xi}$ , et notamment de la matrice de covariance :

$$C_{ij} = \mathbb{E}\left[\left(\hat{\xi}(r_i) - \mathbb{E}[\hat{\xi}(r_i)]\right)\left(\hat{\xi}(r_j) - \mathbb{E}[\hat{\xi}(r_j)]\right)\right] \quad (5.4)$$

La plupart des analyses de la fonction de corrélation font la modélisation simple suivante de la fonction de corrélation, pour un modèle cosmologique avec paramètre  $\theta$  :

$$\hat{\xi} \sim \mathcal{N}(\xi_\theta, C_\theta) \quad (5.5)$$

De plus, les hypothèses BAO et non-BAO incluent une paramétrisation des modèles de fonctions de corrélation, i.e. qu'elles permettent des variations des paramètres cosmologiques :

$$\begin{aligned} \mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO,\theta}, C_{noBAO,\theta}) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO,\theta}, C_{BAO,\theta}) \end{aligned}$$

Idéalement, la modélisation de  $\hat{\xi}$  dans les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  devrait être faite à partir de simulations  $N$ -corps pour chaque valeur de  $\theta$ , ce qui n'imposerait pas la Gaussianité de  $\hat{\xi}$ . On verra toutefois dans la section 5.4.3 que la modélisation Gaussienne de  $\hat{\xi}$  est particulièrement bien vérifiée sur nos simulations log-normales.

La méthode classique du  $\chi^2$  pour la détection des BAOs qui est la plus couramment utilisée, et que l'on va présenter dans la section 5.2, fait encore une simplification importante des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . En effet la méthode considère des hypothèses où la matrice de covariance de l'estimateur  $\hat{\xi}$  est constante pour tous les modèles et les deux hypothèses ( $C_{noBAO,\theta} = C_{BAO,\theta} = C$ ). La raison de cette simplification est la suivante : d'une part il est très difficile d'évaluer la matrice de covariance pour de nombreuses valeurs de  $\theta$ , et d'autre part la méthode classique du  $\chi^2$ , qui requiert que la matrice de covariance soit constante, est particulièrement simple à mettre en oeuvre. Dans la section 5.3, on proposera une extension de la méthode du  $\chi^2$  qui n'impose pas une matrice de covariance constante dans les hypothèses.

La distribution de  $\hat{\xi}$  est entièrement spécifiée pour chaque hypothèse lorsque l'on fixe  $\theta$ , mais ce n'est pas le cas lorsque  $\theta$  peut varier. Dans le premier cas les hypothèses sont dites simples, alors que dans le second cas les hypothèses sont dites composites<sup>1</sup>.

La détection des BAOs a plus de sens lorsqu'on autorise des variations de  $\theta$  dans les hypothèses. En effet, on ne connaît pas avec une précision parfaite les paramètres au sein des modèles cosmologiques actuels (l'un des objectifs majeurs de la cosmologie moderne est en réalité de mieux contraindre ces paramètres). On verra au cours des sections 5.2 et 5.3 qu'il y a des difficultés inhérentes aux tests d'hypothèses composites.

Mentionnons à ce stade qu'un autre type d'approche permet aussi de tester deux hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  incluant un paramètre de variation  $\theta$ . Il s'agit d'une approche bayésienne de comparaison de modèles basée sur le facteur de Bayes.

Dans le cas qui nous intéresse, les probabilités bayésiennes des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  sont données par :

$$P(\mathcal{H}_0 | \hat{\xi}) = \frac{P(\hat{\xi} | \mathcal{H}_0)P(\mathcal{H}_0)}{P(\hat{\xi})} \quad (5.6)$$

$$P(\mathcal{H}_1 | \hat{\xi}) = \frac{P(\hat{\xi} | \mathcal{H}_1)P(\mathcal{H}_1)}{P(\hat{\xi})} \quad (5.7)$$

où les termes  $P(\hat{\xi} | \mathcal{H}_0)$  et  $P(\hat{\xi} | \mathcal{H}_1)$  peuvent être vus comme des vraisemblances.

Le facteur de Bayes est alors construit comme le ratio de ces deux termes :

$$K = \frac{P(\hat{\xi} | \mathcal{H}_1)}{P(\hat{\xi} | \mathcal{H}_0)} = \frac{\int P(\hat{\xi} | \mathcal{H}_1, \theta)P(\theta | \mathcal{H}_1)d\theta}{\int P(\hat{\xi} | \mathcal{H}_0, \theta)P(\theta | \mathcal{H}_0)d\theta} \quad (5.8)$$

Selon la valeur de  $K$ , l'hypothèse  $\mathcal{H}_1$  est privilégiée ou non pour rapport à l'hypothèse  $\mathcal{H}_0$  (avec une confiance d'autant plus importante que  $K$  est grand).

Un des avantages de cette approche est qu'elle permet des paramétrisations différentes  $\theta_0$  et  $\theta_1$  dans les hypothèses (et notamment elle ne force pas une paramétrisation de même dimension). Son

1. En statistiques une hypothèse est dite simple lorsque la distribution de la variable est complètement spécifiée. Elle est dite composite lorsque la distribution n'est pas complètement spécifiée. Par exemple l'union d'hypothèses simples donne une hypothèse composite.

inconvenient majeur est qu'elle oblige un traitement bayésien complet, et donc de choisir des probabilités *a priori*  $P(\theta | \mathcal{H}_0)$  et  $P(\theta | \mathcal{H}_1)$  sur  $\theta$ . Du fait du caractère quelque peu arbitraire de ce choix, nous nous sommes restreint dans notre étude aux tests d'hypothèses classiques, que l'on a détaillés plus haut, et qui sont généralement utilisés en cosmologie.

### Méthodes d'analyse en ondelette

Les BAOs se manifestent comme un pic caractéristique dans la fonction de corrélation à une échelle proche de l'horizon sonique  $r_s$ . Pour détecter ce pic caractéristique de nouvelles méthodes ont été utilisées récemment, et sont basées sur l'analyse en ondelette (Xu et al. 2010; Tian et al. 2011; Arnalte-Mur et al. 2012). La transformée en ondelettes est apparue dans les années 80, et a depuis été utilisée avec succès dans de nombreux domaines, et notamment en traitement d'images (Mallat 2008; Starck et al. 2010). Cette transformation est particulièrement utile pour l'analyse de données à différentes échelles, et pour l'identification de structures caractéristiques.

Dans le cas qui nous intéresse, la structure caractéristique est le pic BAO dans la fonction de corrélation, et l'analyse à différentes échelles est utile pour prendre en compte l'incertitude sur sa forme et sa taille apparentes<sup>2</sup>. En effet, même si le pic BAO est un étalon standard (i.e. que sa taille est bien connue en coordonnées comobiles), sa taille apparente dépendent de la cosmologie fiducielle utilisée pour convertir les *redshifts* en distance. L'incertitude sur la forme du pic BAO provient de possibles erreurs de modélisation pour l'évolution non linéaire (voir section 1.3.4), les distortions de *redshift* (voir section 2.2.3) et le biais dépendant d'échelle (voir section 2.2.2).

Pour ces méthodes, une ondelette  $w = (w(r_i))_{1 \leq i \leq n}$  de dimension  $n$  agit comme un détecteur de pic (*peak finder*) pour détecter le pic BAO dans la fonction de corrélation estimée  $(\hat{\xi}(r_i))_{1 \leq i \leq n}$ . L'ondelette  $w(R, s)$  est paramétrisée par  $R$  et  $s$ , qui contrôlent respectivement l'échelle et la largeur du pic. Une façon simple de procéder consiste à prendre une ondelette mère  $w_0$  et d'effectuer la transformation suivante (voir figure 5.1) :

$$w_i(R, s) = w_0 \left( \frac{r_i - R}{s} \right) \quad (5.9)$$

On obtient alors une version filtrée  $S_w(R, s)$  de la fonction de corrélation pour chaque paire  $(R, s)$  :

$$S_w(R, s) = \left\langle w(R, s), \hat{\xi} \right\rangle = \sum_{i=1}^n w_i(R, s) \hat{\xi}(r_i) \quad (5.10)$$

L'étape suivante consiste à diviser  $S_w(R, s)$  par le bruit  $\sigma_w(R, s)$  sous  $\mathcal{H}_0$  pour chaque paire  $(R, s)$ , afin d'obtenir le  $Z$ -score  $Z_w(R, s)$  (ou signal sur bruit). Les paramètres  $(R_{max}, s_{max})$  qui donnent le  $Z$ -score maximum sont alors utilisés pour estimer l'échelle et la taille du pic BAO :

$$Z_w(R, s) = \frac{S_w(R, s)}{\sigma_w(R, s)} \quad (5.11)$$

$$Z_w^{max} = Z_w(R_{max}, s_{max}) \quad (5.12)$$

Pour rejeter  $\mathcal{H}_0$ , on génère des simulations sans BAOs et l'on calcule avec quelle probabilité on trouve une valeur de  $Z_w^{max}$  aussi forte que pour les observations. Cela donne la significativité de la détection. Il faut noter que les études Arnalte-Mur et al. (2012) et Tian et al. (2011) n'ont utilisé qu'un seul modèle pour l'hypothèse  $\mathcal{H}_0$ , et non une classe de modèles qui tiendrait compte des variations des paramètres cosmologiques. L'idée sous-jacente est que des variations des paramètres cosmologiques n'altéreraient pas beaucoup l'analyse. En effet la paramétrisation de l'ondelette permet notamment d'être peu sensible aux dilatations d'échelles. L'hypothèse  $\mathcal{H}_0$  qui est réellement rejetée est donc plus vaste qu'un simple modèle dans  $\mathcal{H}_0$ , mais l'hypothèse complète est difficile à formuler rigoureusement.

2. On utilise ici le terme de taille apparente, comme la taille d'un objet vu en coordonnées fiducielles. Cette taille apparente est différente de sa vraie taille vue en coordonnées comobiles. En effet l'objet est déformé en coordonnées fiducielles par rapport aux coordonnées comobiles, puisque l'on utilise pas la bonne conversion *redshift*-distance

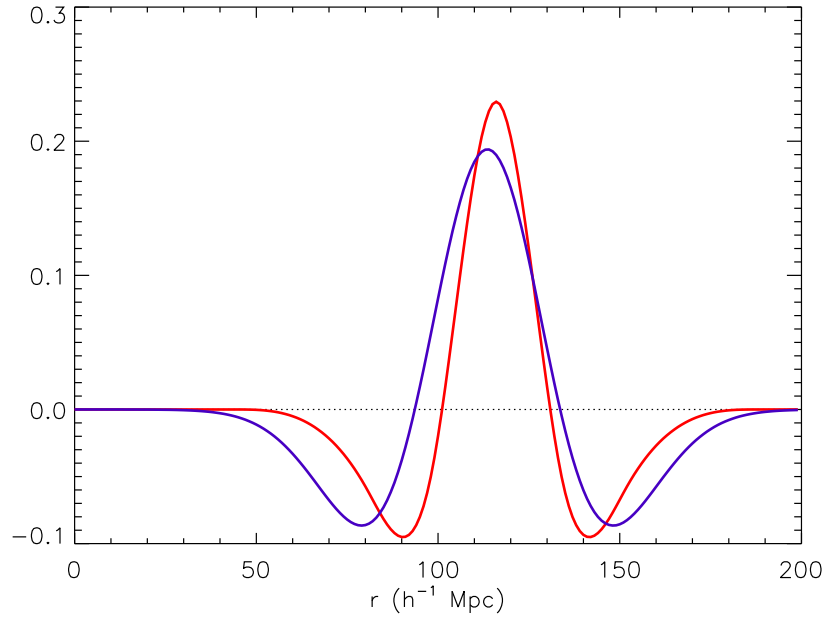


FIGURE 5.1 – Différentes ondelettes mère utilisées pour détecter les BAOs. On montre l’ondelette ‘chapeau mexicain’ avec les paramètres  $R = 113.6 h^{-1}\text{Mpc}$ ,  $s = 20 h^{-1}\text{Mpc}$  (en bleu) de l’étude [Tian et al. \(2011\)](#) et la BAOlet avec les paramètres  $R = 116 h^{-1}\text{Mpc}$ ,  $s = 36 h^{-1}\text{Mpc}$  (en rouge) de l’étude [Arnalte-Mur et al. \(2012\)](#). Ces paramètres correspondent aux réponses maximum  $Z_w^{max}$  dans les études respectives [Tian et al. \(2011\)](#) et [Arnalte-Mur et al. \(2012\)](#).

### Avantages et inconvénients de chaque type de méthodes

L’avantage majeur des méthodes utilisant des ondelettes est qu’elles sont principalement sensibles au pic BAO dans la fonction de corrélation, et non à sa forme globale. Elles permettent aussi de ne pas être affecté par des erreurs de modélisation en autorisant des variations de la largeur du pic BAO. Ainsi les méthodes en ondelettes permettent de détecter le pic BAO tout en étant robustes aux erreurs de modélisation dans la fonction de corrélation.

Le prix à payer est qu’elles sont moins performantes que les autres méthodes lorsque la modélisation de  $\mathcal{H}_0$  et  $\mathcal{H}_1$  est correcte. Comme on va le voir dans la section 5.2, la statistique optimale pour tester deux hypothèses simples  $\mathcal{H}_0$  et  $\mathcal{H}_1$  est le rapport de vraisemblance. Dans le cas qui nous intéresse, les hypothèses sont composites mais on peut quand même considérer un rapport de vraisemblance généralisé comme on le verra dans les section 5.2 et 5.3. On verra dans la section 5.5.2 que l’utilisation du rapport de vraisemblance est effectivement justifiée. Etant donné que les statistiques basées sur les transformées en ondelettes ne sont pas directement liées au rapport de vraisemblance, on s’attend donc à de moins bons résultats.

### 5.1.2 Historique de la détection des BAOs dans les grandes structures

La première détection convaincante des BAOs dans les grandes structures a été faite à un niveau de significativité de  $3.4\sigma$  dans [Eisenstein et al. \(2005\)](#) avec l’analyse de la fonction de corrélation dans l’échantillon LRG du SDSS DR3. Cela a été suivi par de nombreuses autres détections, comme dans [Cole et al. \(2005\)](#) avec l’analyse du spectre de puissance du 2dFGRS à un niveau de  $2.5\sigma$ . L’analyse du spectre de puissance a aussi été utilisée dans [Hütsi \(2006\)](#) avec une détection à  $3.3\sigma$  dans l’échantillon LRG du SDSS DR4. Dans [Percival et al. \(2007\)](#) (respectivement [Percival et al. \(2010\)](#)) l’analyse combinée du spectre de puissance des échantillons LRG et Main du SDSS DR5 (respectivement DR7) ainsi que du 2dFGRS a donné une détection à  $3\sigma$  (respectivement  $3.6\sigma$ ). Dans [Xu et al. \(2012\)](#) la technique de reconstruction du pic BAO (*BAO reconstruction*, pour une revue sur le principe de cette technique on renvoie le lecteur à [Eisenstein et al. \(2007a\)](#)) a permis de passer d’une détection à un niveau de  $3\sigma$  à un niveau de  $4\sigma$  sur l’analyse de la fonction de corrélation de l’échantillon LRG du SDSS DR7.

Des études sur les relevés plus récents à différents *redshifts* ont confirmé la présence des BAOs dans les grandes structures. Dans [Blake et al. \(2011a\)](#) l'analyse du spectre de puissance du *WiggleZ Dark Energy Survey* à un plus haut *redshift* de  $z = 0.6$  a permis une détection à  $3.2\sigma$ . L'étude de la fonction de corrélation dans le relevé 6dFGRS à faible *redshift*  $z = 0.106$  a permis une détection à  $2.4\sigma$  dans [Beutler et al. \(2011\)](#). En combinant les relevés WiggleZ, SDSS LRG et 6dFGRS, [Blake et al. \(2011b\)](#) a obtenu une détection à  $4.9\sigma$  avec la fonction de corrélation. Finalement [Anderson et al. \(2012\)](#) a obtenu une détection à  $5\sigma$  avec la fonction de corrélation du relevé BOSS (i.e. le SDSS DR9).

Ces détections ont toutes été basées sur une modélisation complète de l'estimateur de la fonction de corrélation (ou du spectre de puissance) et correspondent donc au type de méthode présenté dans la section 5.1.1.

D'autres analyses ont utilisées le type de méthodes plus complexes présentées dans la section 5.1.1, i.e. qu'elles sont basées sur l'analyse en ondelette pour détecter le pic de la fonction de corrélation dû aux BAOs. Avec ce type de technique, [Arnalte-Mur et al. \(2012\)](#) obtient une détection à  $4.4\sigma$  à partir de la corrélation entre les échantillons LRG et Main du SDSS DR7, tandis que [Tian et al. \(2011\)](#) obtient une détection à  $4\sigma$  à partir de la fonction de corrélation de l'échantillon Main du SDSS DR7.

On doit garder à l'esprit que ces différents niveaux de détection ne sont en général pas comparables, puisqu'ils ne correspondent pas forcément à la même hypothèse  $\mathcal{H}_0$  rejetée. Par exemple dans [Eisenstein et al. \(2005\)](#) et [Cole et al. \(2005\)](#) l'hypothèse  $\mathcal{H}_0$  est composée de modèles zéro-baryon (modèles CDM purs), alors que dans [Hütsi \(2006\)](#) et [Blake et al. \(2011a\)](#) l'hypothèse  $\mathcal{H}_0$  est composée de modèles 'no wiggles' de [Eisenstein and Hu \(1998\)](#) qui suppriment la signature du pic BAO dans la fonction de corrélation mais conservent la suppression de puissance aux échelles intermédiaires due aux baryons. Dans [Percival et al. \(2007\)](#) et [Percival et al. \(2010\)](#) les modèles  $\mathcal{H}_0$  autorisent un lissage du spectre de puissance par des splines (i.e. qu'il y a marginalisation des spectres de puissance sur l'espace des splines). Concernant les analyses en ondelettes de [Arnalte-Mur et al. \(2012\)](#) et [Tian et al. \(2011\)](#), elles sont basées sur la détection d'une réponse maximum en espace ondelettes, qui traduit l'existence d'un pic de corrélation dû aux BAOs. L'hypothèse  $\mathcal{H}_0$  rejetée par ces méthodes est en réalité difficile à formuler rigoureusement, ce qui rend donc la comparaison avec d'autres méthodes difficile.

Même lorsque les méthodes de détection sont identiques (i.e. basés sur les mêmes hypothèses et utilisant les mêmes statistiques), ce n'est pas parce que les relevés deviennent plus grands que les niveaux de détection doivent nécessairement croître. Le niveau de détection a bien entendu tendance à augmenter, mais la variance d'échantillon provoque une variabilité du niveau de détection, si bien qu'il y a une part d'aléatoire dans cette mesure.

Du fait de cette part d'aléatoire, il faut aussi se méfier des analyses *a posteriori* où l'on construirait une statistique, simplement pour optimiser le niveau de détection sur les données (alors qu'une bonne procédure pour construire une statistique consiste à optimiser le niveau de détection moyen sur des simulations).

### 5.1.3 But du travail dans ce chapitre

Dans ce chapitre on s'intéresse à la question de la détection des BAOs, principalement d'un point de vue statistique. On vient d'exposer la principe général de la détection de ce pic caractéristique (et notamment les différents types de méthodes existantes). On a vu notamment que la détection des BAOs est équivalente à un test d'hypothèses. Il s'ensuit que les différentes méthodes ne sont comparables que pour un même choix d'hypothèses, et qu'il est très important de bien formuler les hypothèses afin que la détection des BAOs ait un sens.

On va maintenant étudier en détails la méthode classique utilisée pour la détection, basée sur la statistique du  $\chi^2$ . On va montrer que cette méthode suppose des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  avec la forme suivante :

$$\begin{aligned}\mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO,\theta}, C) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO,\theta}, C)\end{aligned}$$

On va étudier la validité de cette modélisation statistique de  $\hat{\xi}$ . Pour cela, on va utiliser une nouvelle procédure nous permettant d'obtenir des simulations log-normales sur une grille en  $\theta = (\omega_m, \alpha, B)$  en un temps raisonnable. Le paramètre  $\omega_m = \Omega_m h^2$  correspond à la densité de matière,  $\alpha$  est la

paramètre de dilatation introduit dans la section 2.2.5 et  $B$  est un paramètre d'amplitude constant dans la fonction de corrélation.

Ces simulations vont nous permettre d'obtenir une matrice de covariance réaliste  $C_\theta$  pour  $\hat{\xi}$  en fonction de  $\theta$ . On va d'abord vérifier l'absence de biais dans l'estimateur de Landy-Szalay et pour différents modèles cosmologiques utilisées dans les simulations. On va également vérifier la validité de l'hypothèse de Gaussianité sur  $\hat{\xi}$ . Il ne restera donc qu'à étudier la validité de l'approximation de  $C_\theta$  par une matrice constante  $C$  dans les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  de la méthode classique.

En considérant des hypothèses avec une matrice de covariance constante  $C$ , on verra que la méthode classique du  $\chi^2$  donne effectivement une bonne approximation de la vraie significativité. Toutefois lorsque l'on va considérer des hypothèses avec la matrice de covariance  $C_\theta$  de nos simulations, on verra que la méthode classique du  $\chi^2$  donne des significativités grossièrement incorrectes. Cela va notamment montrer que l'approximation de  $C_\theta$  par une matrice constante  $C$  fausse grandement la procédure de détection des BAOs.

On proposera une nouvelle méthode qui permet d'obtenir des significativités rigoureuses, même dans le cas d'hypothèses réalistes avec la matrice  $C_\theta$ , et qui utilise une statistique  $\Delta l$  donnant de meilleures significativités dans ce cas.

## 5.2 Méthode classique pour la détection des BAOs

La méthode standard pour détecter les BAOs est basée sur la statistique du  $\chi^2$ . A la différence des méthodes basées sur les ondelettes, cette méthode est très dépendante du modèle, et elle est donc surtout utile quand tous les effets intervenant dans la fonction de corrélation sont bien compris et modélisés.

### 5.2.1 La statistique du $\chi^2$

Pour une fonction de corrélation estimée  $\hat{\xi} \sim \mathcal{N}(\xi_m, C)$  de dimension  $n$ , la statistique du  $\chi^2$  s'écrit :

$$\chi^2 = \left\langle \hat{\xi} - \xi_m, C^{-1}(\hat{\xi} - \xi_m) \right\rangle \quad (5.13)$$

$$= \sum_{1 \leq i, j \leq n} \left[ \hat{\xi}(r_i) - \xi_m(r_i) \right] C_{i,j}^{-1} \left[ \hat{\xi}(r_j) - \xi_m(r_j) \right] \quad (5.14)$$

En supposant que le modèle est correct (i.e.  $\hat{\xi} \sim \mathcal{N}(\xi_m, C)$ ), la statistique du  $\chi^2$  suit une loi  $\chi_n^2$ , i.e. une loi du  $\chi^2$  à  $n$  degrés de liberté. La loi  $\chi_n^2$  peut s'interpréter comme celle qui est suivie par le carré de  $n$  variables Gaussiennes centrées réduites. Ainsi si  $X_1, \dots, X_n$  sont de telles variables indépendantes, alors  $\sum_{k=1}^n X_k^2$  suit une loi  $\chi_n^2$ .

### 5.2.2 La méthode du $\chi^2$ pour la détection des BAOs

Montrons comment fonctionne la procédure qui est utilisée classiquement pour la détection des BAOs. On note  $\theta = (\theta_1, \dots, \theta_k) \in \Theta$  les paramètres dont dépendent les fonctions de corrélation avec et sans BAOs,  $\xi_{BAO, \theta}$  et  $\xi_{noBAO, \theta}$ . Les hypothèses sont données par :

$$\begin{aligned} \mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO, \theta}, C) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO, \theta}, C) \end{aligned}$$

La méthode du  $\chi^2$  teste des hypothèses avec une matrice de covariance constante  $C$ . Les paramètres de  $\theta$  ne représentent pas directement les paramètres cosmologiques usuels mais ils leur sont liés. Par exemple dans Eisenstein et al. (2005), ils sont donnés par un paramètre de dilatation  $\alpha$  (pour modéliser l'effet d'une cosmologie fiducielle pour convertir les *redshift* en distances), un paramètre d'amplitude  $B = b^2$  (pour modéliser l'effet du biais des galaxies par rapport à la matière, des *redshift distortions*, et pour le normalisation du spectre de puissance  $\sigma_8$ ), et le paramètre  $\omega_m = \Omega_m h^2$  (qui détermine l'échelle de l'horizon à l'époque de l'égalité matière-radiation, l'amplitude du pic BAO et plus modérément la position du pic). D'autres paramètres ont aussi un impact sur la vraie fonction de corrélation ( $\omega_b = \Omega_b h^2$  et l'indice spectral  $n$ ), mais ils sont mieux contraints par les mesures sur le CMB et on peut les fixer en 1ère approximation. Dans ce chapitre, nous utiliserons la même paramétrisation.

On peut donc définir des statistiques  $\chi^2$  en fonction de  $\theta$  :

$$\chi_{BAO,\theta}^2 = \left\langle \hat{\xi} - \xi_{BAO,\theta}, C^{-1}(\hat{\xi} - \xi_{BAO,\theta}) \right\rangle \quad (5.15)$$

$$\chi_{noBAO,\theta}^2 = \left\langle \hat{\xi} - \xi_{noBAO,\theta}, C^{-1}(\hat{\xi} - \xi_{noBAO,\theta}) \right\rangle \quad (5.16)$$

Pour chaque classe de modèles on peut regarder les valeurs du meilleur *fit* du  $\chi^2$ ,  $\min_{\theta} \chi_{noBAO,\theta}^2$  et  $\min_{\theta} \chi_{BAO,\theta}^2$ . L'approximation habituelle est de considérer que le meilleur *fit* du  $\chi^2$  suit une loi  $\chi_{n-k}^2$  si le vrai modèle fait partie de la classe de modèles utilisée pour le *fitting*. Donc le nombre de degrés de liberté décroît du nombre de paramètres du *fit* par rapport au  $\chi^2$  avec le vrai modèle de l'équation (5.13). Soulignons que ce résultat n'est rigoureusement valide que si l'espace des modèles de fonctions de corrélation est affine (voir annexe B.2)<sup>3</sup>.

Ce résultat peut être utilisé avec  $\min_{\theta} \chi_{BAO,\theta}^2$  pour vérifier que la mesure est compatible avec  $\mathcal{H}_1$ . Plus précisément, on peut tester si  $\min_{\theta} \chi_{BAO,\theta}^2$  est compatible avec la distribution  $\chi_{n-k}^2$  suivie par cette variable si  $\mathcal{H}_1$  est vraie :

$$\min_{\theta} \chi_{BAO,\theta}^2 \sim \chi_{n-k}^2 \text{ si } \mathcal{H}_1 \text{ est vraie} \quad (5.17)$$

Pour rejeter l'hypothèse  $\mathcal{H}_0$ , la procédure habituelle est plus complexe. Pour comprendre cette procédure, on doit ajouter un paramètre artificiel dans le *fit*, afin de réunir les 2 classes de modèles. Par exemple ce paramètre, que l'on va nommer  $\beta$ , peut être une pondération de  $\xi_{BAO,\theta}$  et  $\xi_{noBAO,\theta}$  dans le modèle de fonction de corrélation :

$$\xi_{\beta,\theta} = \beta \xi_{BAO,\theta} + (1 - \beta) \xi_{noBAO,\theta} \quad (5.18)$$

Sous  $\mathcal{H}_0$ , la vraie fonction de corrélation (l'espérance de la fonction de corrélation mesurée  $\mathbb{E}[\hat{\xi}]$  s'il n'y a pas de biais) est de la forme  $\xi_{noBAO,\theta}$  pour la vraie valeur du paramètre  $\theta = \theta_0$ . Mais elle est également de la forme  $\xi_{\beta,\theta}$  avec  $\beta = 0$  et  $\theta = \theta_0$ . Ainsi le meilleur *fit*  $\chi^2$  dans la classe non-BAO suit une loi  $\chi_{n-k}^2$ , et le meilleur *fit*  $\chi^2$  dans la classe étendue suit une loi  $\chi_{n-(k+1)}^2$  (puisque  $\beta$  est un paramètre additionnel).

On est dans le cas de 2 classes de modèles imbriquées, qui contiennent toutes les 2 le vrai modèle. Dans ce cas la différence entre les valeurs de meilleur *fit* suit une loi du chi-carré avec un nombre de degré de liberté égal à la différence du nombre de paramètres entre les 2 classes. A nouveau ce résultat n'est pas rigoureusement valide dans le cas général, mais seulement si les espaces des modèles de fonctions de corrélation sont affines (voir annexe B.2).

Ici il n'y a qu'un paramètre additionnel,  $\beta$ , dans la classe étendue. Ainsi la différence des meilleurs *fit* du  $\chi^2$  suit une loi  $\chi_1^2$  :

$$\Delta\chi_{global}^2 = \min_{\theta} \chi_{noBAO,\theta}^2 - \min_{\beta,\theta} \chi_{\beta,\theta}^2 \sim \chi_1^2 \quad (5.19)$$

Ceci montre qu'ajouter un paramètre additionnel, qui n'est pas requis par le vrai modèle, ne fait que décroître modérément la valeur du meilleur *fit*. Pour rejeter  $\mathcal{H}_0$ , on peut regarder la différence  $\Delta\chi_{global}^2$  et calculer la probabilité d'obtenir une telle valeur sous  $\mathcal{H}_0$ , i.e. pour une loi  $\chi_1^2$ .

En pratique, le meilleur *fit* dans la classe étendue est remplacée par le meilleur *fit* dans la classe BAO (i.e. en restreignant à  $\beta = 1$ ) :

$$\Delta\chi^2 = \min_{\theta} \chi_{noBAO,\theta}^2 - \min_{\theta} \chi_{BAO,\theta}^2 \quad (5.20)$$

Cette différence  $\Delta\chi^2$  est nécessairement inférieure à  $\Delta\chi_{global}^2$  de l'équation (5.19), qui suit une loi  $\chi_1^2$  sous  $\mathcal{H}_0$ . En effet on a  $\min_{\theta} \chi_{BAO,\theta}^2 \geq \min_{\beta,\theta} \chi_{\beta,\theta}^2$  et donc  $\Delta\chi^2 \leq \Delta\chi_{global}^2$ .

Ainsi pour une réalisation avec une valeur donnée  $\Delta\chi^2 = x$ , on a :

$$P(\Delta\chi^2 > x | \mathcal{H}_0) \leq P(\Delta\chi_{global}^2 > x | \mathcal{H}_0) \quad (5.21)$$

3. Rappelons que la fonction de corrélation mesurée  $\hat{\xi}$  et la fonction de corrélation théorique  $\xi_m(\theta)$  sont des vecteurs  $n$ -dimensionnels, i.e. sont des versions discrétisées de leur équivalent continu. Donc l'ensemble des modèles de corrélation  $(\xi_m(\theta))_{\theta \in \Theta}$  constitue un sous-espace de  $\mathbb{R}^n$ , qui a doit être un espace affine de dimension  $k$  pour que le résultat mentionné soit valide.

Une loi  $\chi_1^2$  est simplement la loi suivie par le carré d'une variable normale centrée réduite. Si l'on note  $\Phi$  la fonction cumulative d'une fonction Gaussienne centrée réduite, comme  $\Delta\chi_{global}^2 \sim \chi_1^2$ , on obtient pour  $x \geq 0$  :

$$P(\Delta\chi^2 > x | \mathcal{H}_0) \leq P(\Delta\chi_{global}^2 > x | \mathcal{H}_0) = 2\Phi(-\sqrt{x}) \quad (5.22)$$

ce qui correspond à un nombre de  $\sigma$  égal à  $\sqrt{x}$  pour une loi normale. Ainsi, quand  $\Delta\chi^2 \geq 0$ , on peut évaluer la significativité de la détection des BAOs comme  $\sqrt{\Delta\chi^2} \cdot \sigma$ .

Dans Eisenstein et al. (2005) cette statistique est égale à  $\Delta\chi^2 = 11.7$ , ce qui correspond à une détection à  $3.4\sigma$  avec cette méthode. Dans Percival et al. (2010) elle est égale à  $\Delta\chi^2 = 13.1$ , correspondant à une détection à  $3.6\sigma$ .

À cause de l'inégalité dans l'équation (B.12) la méthode peut sembler conservatrice. Cependant comme les hypothèses de la méthode ne sont pas vérifiées, on verra dans la section 5.5.1 que la méthode peut en fait surestimer la significativité.

### 5.2.3 Limitations de la méthode du $\chi^2$

Le lemme de Neyman-Pearson indique que lorsque l'on effectue un test d'hypothèses entre  $\mathcal{H}_0$  et  $\mathcal{H}_1$ , le test le plus puissant est basé sur le rapport de vraisemblance :

$$\Lambda(\hat{\xi}) = \frac{\mathcal{L}_{\mathcal{H}_0}(\hat{\xi})}{\mathcal{L}_{\mathcal{H}_1}(\hat{\xi})} \quad (5.23)$$

Plus précisément, le test le plus puissant de taille  $\alpha$  est le suivant :

- si  $\Lambda(\hat{\xi}) \leq \eta$  alors accepter  $\mathcal{H}_1$  (i.e. rejeter  $\mathcal{H}_0$ )
- si  $\Lambda(\hat{\xi}) > \eta$  alors accepter  $\mathcal{H}_0$  (i.e. rejeter  $\mathcal{H}_1$ )

avec  $\alpha$  la probabilité de rejeter  $\mathcal{H}_0$  si  $\mathcal{H}_0$  est vraie (risque de première espèce) :

$$\alpha = P\left(\Lambda(\hat{\xi}) \leq \eta | \mathcal{H}_0\right) \quad (5.24)$$

La puissance du test est définie comme la probabilité d'accepter  $\mathcal{H}_1$  si cette hypothèse est vraie. Elle est égale à  $1 - \beta$ , avec  $\beta$  la probabilité de rejeter  $\mathcal{H}_1$  si cette hypothèse est vraie (risque de seconde espèce) :

$$\beta = P\left(\Lambda(\hat{\xi}) > \eta | \mathcal{H}_1\right) \quad (5.25)$$

Comme on l'a vu dans la section 5.1.1, les tests avec des niveaux de seuil prédéfinis sont peu utilisés en cosmologie, et il est plus commun de citer le niveau de significativité correspondant à la valeur obtenue pour la statistique. Pour une réalisation avec une valeur donnée  $\Lambda(\hat{\xi}) = x$ , la significativité (donnée comme une valeur  $p$ ) est égale à :

$$\alpha(x) = P\left(\Lambda(\hat{\xi}) \leq x | \mathcal{H}_0\right) \quad (5.26)$$

Comme on le montre en annexe B.3, le lemme de Neyman-Pearson implique que l'espérance de la significativité obtenue de cette manière sous  $\mathcal{H}_1$ , est meilleure avec  $\Lambda(\hat{\xi})$  qu'avec n'importe quelle autre statistique. Plus précisément, sous  $\mathcal{H}_1$ , l'espérance de la valeur  $p$  de l'équation (5.26) est plus petite et l'espérance du nombre de  $\sigma$  est plus grand pour  $\Lambda(\hat{\xi})$  que pour n'importe quelle autre statistique.

Il faut noter que la statistique  $\Lambda(\hat{\xi})$  est optimale dans ce sens, mais on doit connaître sa distribution sous  $\mathcal{H}_0$  (pour calculer la significativité  $\alpha(x)$  correspondant à une réalisation  $\Lambda(\hat{\xi}) = x$ ). De plus dans le cas d'hypothèses composites, la significativité n'est pas bien définie. L'avantage de la statistique  $\Delta\chi_{global}^2$  est que sa distribution est identique pour tous les modèles inclus dans  $\mathcal{H}_0$  (une distribution  $\chi_1^2$ ). Dans ce cas on est capable de donner une significativité même pour des hypothèses composites.

Toutefois comme on l'a vu plus haut, ce résultat nécessite que les espaces des modèles des fonctions de corrélation soient affines. Comme cela n'a pas de raison d'être vérifié en pratique, on verra avec des simulations dans la section 5.5.1 que la loi de  $\Delta\chi_{global}^2$  est très différente d'une loi  $\chi_1^2$ . Le fait que la méthode du  $\chi^2$  utilise la statistique  $\Delta\chi^2$  et non  $\Delta\chi_{global}^2$  va toutefois compenser cette erreur comme on le verra dans la section 5.5.1.



Cependant la méthode du  $\chi^2$  va atteindre ses limites lorsque l'on considère des hypothèses plus réalistes, avec une matrice de covariance dépendant du modèle :

$$\begin{aligned}\mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO,\theta}, C_{noBAO,\theta}) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO,\theta}, C_{BAO,\theta})\end{aligned}$$

Considérons le cas où les paramètres  $\theta$  sont donnés par  $\theta = (\omega_m, \alpha, B)$ . Pour l'illustration, prenons un modèle simple où la matrice de covariance ne dépend que du biais de normalisation  $B$ . Cela donne un facteur multiplicatif  $B$  dans le modèle de fonction de corrélation. Dans la fonction de corrélation estimée  $\hat{\xi}$ , cela donne aussi un facteur multiplicatif  $B$  en négligeant le bruit de Poisson dû au nombre fini de paires (ceci est d'autant mieux vérifié que les intervalles de distance sont grands). Il s'agit donc en première approximation d'un facteur multiplicatif  $B^2$  dans la matrice de covariance :

$$C_{BAO,\theta} = C_{noBAO,\theta} \times B^2 C \quad (5.27)$$

Dans la méthode du  $\chi^2$  classique,  $\chi_{BAO,\theta}^2$  et  $\chi_{noBAO,\theta}^2$  sont calculés avec une matrice de covariance constante  $C$ . Ainsi pour une réalisation  $B_0 \hat{\xi}$  avec  $A \geq 0$ , on obtient le meilleur *fit* BAO :

$$\begin{aligned}\min_{\theta} \chi_{BAO,\theta}^2(B_0 \hat{\xi}) & \\ & = \min_{\theta} \left\langle B_0 \hat{\xi} - \xi_{BAO,\theta}, C^{-1} (B_0 \hat{\xi} - \xi_{BAO,\theta}) \right\rangle \\ & = B_0^2 \min_{\theta} \left\langle \hat{\xi} - \frac{1}{B_0} \xi_{BAO,\theta}, C^{-1} \left( \hat{\xi} - \frac{1}{B_0} \xi_{BAO,\theta} \right) \right\rangle \\ & = B_0^2 \min_{\theta} \chi_{BAO,\theta}^2(\hat{\xi})\end{aligned} \quad (5.28)$$

La dernière égalité vient du rôle de  $B$ , qui permet n'importe quelle normalisation positive du modèle. Le même raisonnement peut être appliqué pour  $\min_{\theta} \chi_{noBAO,\theta}^2$  qui est multiplié par  $B_0^2$ , et donc la statistique  $\Delta\chi^2$  est aussi multipliée par  $B_0^2$ .

Avec les hypothèses où la matrice de covariance varie comme  $B^2$ , des réalisations de  $\mathcal{H}_0$  avec  $\theta = (\omega_m, \alpha, B_2)$  peuvent être obtenues à partir de réalisations de  $\mathcal{H}_0$  avec  $\theta = (\omega_m h^2, \alpha, B_1)$  avec une multiplication par  $B_2/B_1$ . Ainsi il y a une dilatation d'un facteur  $(B_2/B_1)^2$  de la distribution de  $\Delta\chi^2$  entre les 2 modèles. Cela crée de grandes différences entre les 2 modèles dans  $\mathcal{H}_0$ , de sorte que la significativité donnée par la méthode classique du  $\chi^2$  est fautive. La conclusion est que la méthode classique du  $\chi^2$  ne peut pas être utilisée dans le cas d'une matrice de covariance dépendant du modèle.

### 5.3 Nouvelle méthode pour la détection des BAOs

Dans cette section on propose deux modifications à la méthode classique du  $\chi^2$  pour surmonter ces limitations. Une première modification permet d'obtenir une significativité correcte dans tous les cas. Ainsi, au contraire de la méthode classique du  $\chi^2$ , notre méthode peut être appliquée avec des hypothèses où la matrice de covariance dépend du modèle :

$$\begin{aligned}\mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO,\theta}, C_{noBAO,\theta}) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO,\theta}, C_{BAO,\theta})\end{aligned}$$

La façon de calculer rigoureusement la valeur  $p$  pour une mesure  $\Delta\chi^2 = x$  est de considérer le pire cas possible dans l'hypothèse  $\mathcal{H}_0$  :

$$p(x) = \max_{\theta \in \Theta} P(\Delta\chi^2 \geq x | \mathcal{H}_0, \theta) \quad (5.29)$$

Donc pour chaque modèle dans  $\mathcal{H}_0$ , la valeur  $p$  de la mesure  $\Delta\chi^2 = x$  est inférieure à  $p(x)$ . Si l'on considère la significativité  $s(x)\sigma$  correspondant à cette valeur  $p$ , on obtient que tous les modèles dans  $\mathcal{H}_0$  sont rejetés à au moins  $s(x)\sigma$ . Il faut noter que ceci est la plus grande significativité que l'on peut obtenir si l'on veut rejeter tous les modèles  $\mathcal{H}_0$  simultanément.

Cette façon d'obtenir la significativité ne repose sur aucune hypothèse de régularité, à la différence de la méthode du  $\chi^2$  classique. Cependant cela requiert plus de travail (et engendre plus de difficultés)

pour calculer la distribution de  $\Delta\chi^2$  pour tous les modèles dans  $\mathcal{H}_0$ . On verra précisément comment ceci est possible dans les sections 5.5.1 et 5.5.2, où l'on applique notre méthode sur des simulations.

La seconde modification que l'on propose concerne la statistique elle-même. Comme on a vu dans la section 5.2.3, la statistique optimale pour tester des hypothèses simples  $\mathcal{H}_0$  et  $\mathcal{H}_1$  est le rapport de vraisemblance. Cependant lorsqu'on travaille avec des hypothèses composites, les vraisemblances ne sont pas bien définies. En effet, elles sont définies pour chaque modèle dans  $\mathcal{H}_0$  et  $\mathcal{H}_1$  mais elles ne sont pas définies pour les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  elles-mêmes. L'idée de la statistique  $\Delta\chi^2$  est qu'elle peut être vue comme un rapport de vraisemblance généralisé entre des hypothèses composites. En effet dans le cas spécial d'une matrice de covariance constante, on a :

$$\begin{aligned}\Delta\chi^2 &= \min_{\theta} \chi_{noBAO,\theta}^2 - \min_{\theta} \chi_{BAO,\theta}^2 \\ &= -2 \left[ \max_{\theta} \ln(\mathcal{L}_{noBAO,\theta}) - \max_{\theta} \ln(\mathcal{L}_{BAO,\theta}) \right] \\ &= -2 \ln \left[ \frac{\max_{\theta} \mathcal{L}_{noBAO,\theta}}{\max_{\theta} \mathcal{L}_{BAO,\theta}} \right]\end{aligned}\quad (5.30)$$

où on a utilisé la relation suivante entre  $\chi^2$  et la vraisemblance dans le cas d'une matrice de covariance constante :

$$\mathcal{L}_{\theta} = \frac{1}{|C|^{1/2}} e^{-\frac{1}{2} \langle \hat{\xi} - \xi_{\theta}, C^{-1} (\hat{\xi} - \xi_{\theta}) \rangle} = \frac{1}{|C|^{1/2}} e^{-\frac{1}{2} \chi_{\theta}^2} \quad (5.31)$$

$$-2 \ln \mathcal{L}_{\theta} = \chi_{\theta}^2 + const \quad (5.32)$$

Ceci est seulement valide pour une matrice de covariance constante. Pour étendre cette idée dans le cas d'une matrice de covariance qui dépend du modèle, on considère simplement la statistique :

$$\Delta l = -2 \left[ \max_{\theta} \ln(\mathcal{L}_{noBAO,\theta}) - \max_{\theta} \ln(\mathcal{L}_{BAO,\theta}) \right] \quad (5.33)$$

On utilise la notation  $\Delta l$  parce que cela désigne une différence de log-vraisemblances (i.e. le logarithme d'un rapport de vraisemblance). C'est un léger abus de notation car ne n'est pas une différence de log-vraisemblances au sens strict (mais plutôt ce que l'on a appelé précédemment un rapport de vraisemblance généralisé). A la différence de  $\Delta\chi^2$ ,  $\Delta l$  reste un rapport de vraisemblance généralisé pour des matrices de covariance dépendant du modèle. Cela devrait donc donner de meilleurs résultats dans ce cas, comme on le vérifiera dans la section 5.5.2.

Dans le cas d'une matrice de covariance dépendant du modèle, les vraisemblances sont données par :

$$\mathcal{L}_{BAO,\theta} \propto |C_{BAO,\theta}|^{-1/2} e^{-\frac{1}{2} \chi_{BAO,\theta}^2} \quad (5.34)$$

$$\mathcal{L}_{noBAO,\theta} \propto |C_{noBAO,\theta}|^{-1/2} e^{-\frac{1}{2} \chi_{noBAO,\theta}^2} \quad (5.35)$$

$$(5.36)$$

où  $\chi_{BAO,\theta}^2$  et  $\chi_{noBAO,\theta}^2$  sont calculés en prenant en compte les variations de la matrice de covariance :

$$\chi_{BAO,\theta}^2 = \left\langle \hat{\xi} - \xi_{BAO,\theta}, C_{BAO,\theta}^{-1} (\hat{\xi} - \xi_{BAO,\theta}) \right\rangle \quad (5.37)$$

$$\chi_{noBAO,\theta}^2 = \left\langle \hat{\xi} - \xi_{noBAO,\theta}, C_{noBAO,\theta}^{-1} (\hat{\xi} - \xi_{noBAO,\theta}) \right\rangle \quad (5.38)$$

Si l'on note  $l_{noBAO,\theta}$  et  $l_{BAO,\theta}$  pour  $-2 \ln(\mathcal{L}_{noBAO,\theta})$  et  $-2 \ln(\mathcal{L}_{BAO,\theta})$ , on obtient :

$$\Delta l = \min_{\theta} l_{noBAO,\theta} - \min_{\theta} l_{BAO,\theta} \quad (5.39)$$

$$l_{BAO,\theta} = \chi_{BAO,\theta}^2 + \ln |C_{BAO,\theta}| + const \quad (5.40)$$

$$l_{noBAO,\theta} = \chi_{noBAO,\theta}^2 + \ln |C_{noBAO,\theta}| + const \quad (5.41)$$

avec la même constante additive pour  $l_{BAO,\theta}$  et  $l_{noBAO,\theta}$ , qui peut être prise égale à 0.

Il faut noter que  $\Delta l$  n'est pas équivalent à  $\Delta\chi^2$  même si  $\chi_{BAO,\theta}^2$  et  $\chi_{noBAO,\theta}^2$  sont calculés en utilisant les équations (5.37) et (5.38). En effet dans le cas de matrices de covariance dépendant du modèle, il faut prendre en compte les variations du déterminant de la matrice dans les équations (5.40) et (5.41).

On désignera comme méthode  $\Delta l$  notre méthode modifiée comme présenté ci-dessus, i.e. avec remplacement de  $\Delta\chi^2$  par  $\Delta l$  et avec le changement de la procédure pour obtenir une significativité correcte.

## 5.4 Modélisation de $\hat{\xi}$ et simulations SDSS DR7-Full

### 5.4.1 Modèles de fonctions de corrélation pour $\mathcal{H}_0$ et $\mathcal{H}_1$

Comme on l'a vu dans la section 2.2.5, la fonction de corrélation linéaire ne dépend (à un facteur d'amplitude près) que de la densité de matière  $\omega_m = \Omega_m h^2$ , de la densité de baryons  $\omega_b = \Omega_b h^2$  et de l'indice spectral  $n_s$ . Dans notre analyse on va négliger la dépendance par rapport aux paramètres  $\omega_b$  et  $n_s$ , qui sont bien contraints par les données WMAP7 du CMB (Komatsu et al. 2011). On fixe donc ces paramètres aux valeurs de vraisemblance maximum de WMAP7, qui sont  $\omega_b = 2.227 \times 10^{-2}$  et  $n_s = 0.966$  (on va aussi fixer le paramètre  $\sigma_8 = 0.81$  pour la normalisation du spectre de puissance linéaire). Le seul paramètre que l'on varie dans la fonction de corrélation linéaire est donc  $\omega_m$ .

On utilise un modèle simple pour les corrections non linéaires de l'évolution du champ de matière. Plus précisément on utilise la procédure décrite dans la section 1.3.4, avec une correction invariante d'échelle donnée par HALOFIT (Smith et al. 2003) et une dégradation non linéaire des BAOs. A partir du spectre de puissance linéaire on modélise la dégradation non linéaire du pic BAO par un lissage Gaussien, appliqué seulement sur la signature BAO.

Pour isoler la signature BAO dans le spectre de puissance, on se sert du spectre de puissance *no wiggles* de Eisenstein and Hu (1998), qui est un spectre lisse avec la même forme que le spectre de puissance linéaire mais sans la signature BAO. On applique donc le lissage Gaussien à la différence entre le spectre linéaire et le spectre *no wiggles*,  $P_L(k) - P_{nw,L}(k)$ , et l'on restaure la forme générale du spectre en rajoutant  $P_{nw,L}(k)$  après le lissage :

$$P_{dw,L}(k) = P_{nw,L}(k) + e^{-a^2 k^2 / 2} [P_L(k) - P_{nw,L}(k)] \quad (5.42)$$

Pour prendre en compte les corrections non linéaires invariantes d'échelle, on applique à  $P_{dw,L}$  les corrections non linéaires de HALOFIT correspondant au spectre invariant d'échelle  $P_{nw,L}(k)$  :

$$P_{dw,NL}(k) = \frac{P_{nw,NL}(k)}{P_{nw,L}(k)} P_{dw,L}(k) \quad (5.43)$$

Pour calculer ces spectres au *redshift* moyen  $z = 0.3$  du catalogue, on utilise la librairie IDL *iCosmo* (Refregier et al. 2011), qui donne les spectres de puissances  $P_L$  et  $P_{nw,L}$  avec l'approximation de Eisenstein and Hu (1998).

Pour obtenir la fonction de corrélation de la distribution de galaxies en espace *redshift* et en coordonnées comobiles en fonction de  $\omega_m$ , il reste à choisir la valeur de  $a$  dans la formule (6.11) et à modéliser le biais dépendant d'échelle des galaxies de l'échantillon par rapport à la matière. Pour ce faire, on utilise les catalogues *Large Suite of Dark Matter Simulations* (LasDamas, McBride et al., in prep.) présentés dans la section 3.4.4.

On utilise la version gamma des simulations LasDamas, et plus précisément l'ensemble de simulations Oriana qui est disponible en ligne<sup>4</sup>. Cet ensemble est composé de 40 simulations  $N$ -corps, avec chaque simulation pouvant reproduire deux fois la couverture angulaire Nord+Sud du SDSS. L'ensemble produit donc un total de 80 réalisations de l'échantillon SDSS. Les paramètres cosmologiques de la simulations sont  $\Omega_m = 0.25$ ,  $\Omega_\Lambda = 0.75$ ,  $\Omega_b = 0.04$ ,  $h = 0.7$ ,  $\sigma_8 = 0.8$  et  $n_s = 1$ . On utilise les catalogues contenant les galaxies LRG avec une magnitude absolue dans la bande  $g$ ,  $-23.2 < M_g < -21.2$ , comme pour l'échantillon SDSS DR7-Full. Comme ces catalogues sont presque *volume-limited*, leur extension en *redshift* ( $0.16 < z < 0.36$ ) est plus faible que pour l'échantillon SDSS DR7-Full.

4. <http://lss.phy.vanderbilt.edu/lasdamas/mocks/>

Pour chaque réalisation, on estime la fonction de corrélation avec l'estimateur de Landy-Szalay (voir section 2.2.4). Comme le modèle HOD utilisé n'évolue pas avec le *redshift* et que les galaxies sont observées sur un cône de lumière, il y a une légère décroissance de la densité moyenne  $\bar{n}(z)$  en fonction du *redshift*. Pour tenir compte de cet effet, la fonction de corrélation est estimée pour chaque simulation,  $\hat{\xi}_k$ , avec le catalogue aléatoire de Poisson fourni par l'équipe LasDamas, qui a le même profil de décroissance de la densité. On calcule ensuite la moyenne  $\bar{\xi}$  des fonctions de corrélation estimées pour chaque simulation. On compare cette fonction de corrélation avec la fonction de corrélation  $\xi_{dw,NL}$ , obtenue par transformée de Hankel du spectre  $P_{dw,NL}$  de la formule (6.12) avec les paramètres cosmologiques des simulations LasDamas.

On ajuste tout d'abord le facteur  $a$  de la formule (6.11) pour reproduire la dégradation non linéaire observée dans la fonction de corrélation des simulations LasDamas  $\bar{\xi}$ , et l'on trouve que la valeur  $a = 9.5h^{-1}\text{Mpc}$  donne un bon résultat. Finalement on ajuste le biais dépendant d'échelle des galaxies  $B(r)$  qui intervient surtout aux petites échelles (voir section 2.2.2), en divisant  $\bar{\xi}$  par la fonction de corrélation non linéaire du champ de matière,  $\xi_{dw,NL}$ . On trouve une correction dépendante d'échelle d'environ 10% pour  $r = 5h^{-1}\text{Mpc}$  et qui décroît jusqu'à  $r = 55h^{-1}\text{Mpc}$ .

On obtient finalement la fonction de corrélation des galaxies en espace *redshift* et en coordonnées comobiles, à partir du biais  $B(r)$  et des formules (6.11) et (6.12) avec le choix  $a = 9.5h^{-1}\text{Mpc}$  :

$$\xi_{galaxies,\omega_m}(r) = B(r) \xi_{dw,NL}(r) \quad (5.44)$$

Comme on garde les paramètres  $a$  et  $B(r)$  fixes dans l'analyse, la fonction de corrélation  $\xi_{galaxies,\omega_m}$  ne dépend que des spectres linéaires  $P_L$  et  $P_{nw,L}$  dans la formule (6.11). Comme on l'a déjà précisé, on ne fait varier que le paramètre  $\omega_m$  dans le spectre linéaire, si bien que la fonction  $\xi_{galaxies,\omega_m}$  ne dépend que de  $\omega_m$ .

On modélise maintenant les effets géométriques introduits par la cosmologie fiducielle. Comme on l'a vu dans la section section 2.2.5, l'effet de la cosmologie fiducielle peut être approximé comme un facteur de dilatation  $\alpha$  dans le monopole de la fonction de corrélation avec  $\alpha = D_V(\langle z \rangle) / D_V^{fid}(\langle z \rangle)$ , où  $\langle z \rangle = 0.3$  est le *redshift* moyen de l'échantillon et  $D_V(z)$  est l'échelle de dilatation donnée par :

$$D_V(z) = \left[ (1+z)^2 D_A(z)^2 \frac{cz}{H(z)} \right]^{1/3} \quad (5.45)$$

Notre choix d'une cosmologie fiducielle  $\Lambda\text{CDM}$  plate avec  $\Omega_m = 0.25$  pour étudier l'échantillon SDSS DR7-Full donne  $D_V^{fid}(z = 0.3) = 1180 \text{ Mpc}$ .

On introduit finalement un facteur multiplicatif constant  $b$  pour modéliser les variations de  $\sigma_8$ , les distortions de *redshift* linéaires et le régime constant du biais des galaxies par rapport à la matière. On obtient donc le modèle final de corrélation en fonction de  $\omega_m, \alpha$  et  $B = b^2$  :

$$\xi_{\omega_m,\alpha,B}(r) = b^2 \xi_{galaxies,\omega_m}(\alpha r) \quad (5.46)$$

Il ne reste plus qu'à discrétiser ce modèle continu de façon équivalente au comptage de paires lors de l'estimation de la fonction de corrélation. Pour un intervalle  $[r_i - dr/2, r_i + dr/2]$ , la fonction de corrélation discrétisée est donnée par :

$$\xi_{\omega_m,\alpha,B}(r_i) = \frac{\int_{r_i-dr/2}^{r_i+dr/2} \xi_{\omega_m h^2, \alpha, B}(r) r^2 dr}{\int_{r_i-dr/2}^{r_i+dr/2} r^2 dr} \quad (5.47)$$

Dans toute cette étude, on utilise des intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  entre les distances  $r = 20h^{-1}\text{Mpc}$  et  $r = 200h^{-1}\text{Mpc}$ , ce qui correspond à  $n = 18$  intervalles, i.e. un vecteur  $(\xi_{\omega_m,\alpha,B}(r_i))$  de dimension  $n = 18$ .

Toute cette procédure permet d'obtenir les modèles de fonctions de corrélation  $(\xi_{\omega_m,\alpha,B})$  pour l'hypothèse  $\mathcal{H}_1$ . Pour obtenir les modèles de l'hypothèse  $\mathcal{H}_0$ , on utilise une procédure similaire. Plus précisément, on part simplement de la fonction de corrélation *no wiggles*  $\xi_{nw,NL}$  au lieu de la fonction de corrélation *dewiggled*  $\xi_{dw,NL}$  dans l'équation (5.44). On applique ensuite la même dilatation par le paramètre  $\alpha$ , le même biais dépendant d'échelle  $B(r)$  et la même discrétisation. Cette hypothèse  $\mathcal{H}_0$  correspond à une hypothèse *no wiggles*. A certains endroits, on considérera plutôt une hypothèse  $\mathcal{H}_0$  zéro-baryon. Pour cela, on remplacera simplement la fonction de corrélation  $\xi_{nw,NL}$  par la fonction de corrélation  $\xi_{NL}$  avec  $\omega_b = 0$ .

### 5.4.2 Simulations log-normales SDSS DR7-Full utilisées

On détaille ici les simulations log-normales que l'on va utiliser dans notre analyse. Dans la section 5.4.3 on va se servir de ces simulations pour vérifier l'absence de biais dans les estimateurs ainsi que leur Gaussianité, i.e. que les hypothèses suivantes sont valides :

$$\mathcal{H}_0 : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO,\theta}, C_{noBAO,\theta})$$

$$\mathcal{H}_1 : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO,\theta}, C_{BAO,\theta})$$

D'autre part ces simulations vont permettre d'estimer les matrices de covariance dépendantes du modèle  $C_{noBAO,\theta}$  et  $C_{BAO,\theta}$ . On fera en outre l'approximation<sup>5</sup> que ces matrices de dépendent que de  $\theta$ , i.e.  $C_{noBAO,\theta} = C_{BAO,\theta} = C_\theta$ . On rappelle la paramétrisation tridimensionnelle  $\theta = (\omega_m, \alpha, B)$  que l'on utilise pour modéliser les fonctions de corrélation  $\xi_{noBAO,\theta}$  et  $\xi_{BAO,\theta}$ .

Nos simulations reproduisent le catalogue SDSS DR7-Full présenté dans la section 3.2.2, i.e. que l'on utilise la même couverture angulaire du ciel et le même densité moyenne de galaxies  $\bar{n}(r)$  dans nos simulations que dans le catalogue SDSS DR7-Full.

Pour chaque simulation on utilise l'estimateur de Landy-Szalay avec des poids constants pour toutes les galaxies<sup>6</sup>. Pour des raisons de temps de calcul, on n'estime pas la fonction de corrélation  $\hat{\xi}$  à partir de l'échantillon complet, mais en séparant l'hémisphère nord en coordonnées galactiques (*Northern Galactic Cap*),  $\hat{\xi}_{NGC}$ , et l'hémisphère sud en coordonnées galactiques (*Southern Galactic Cap*),  $\hat{\xi}_{SGC}$ , qui peuvent être considérés comme indépendants (la distance minimum entre une galaxie de l'hémisphère nord et une galaxie de l'hémisphère sud étant de plus de  $350h^{-1}$  Mpc). Egalement pour des raisons de temps de calcul, on ne peut pas utiliser des catalogues aléatoires trop grands, et l'on va donc utiliser des catalogues aléatoires avec la même densité que l'échantillon SDSS DR7-Full.

A partir des mesures de  $\hat{\xi}_{NGC}$  et  $\hat{\xi}_{SGC}$  sur de nombreuses simulations et pour différentes valeurs de  $\theta$ , on obtient les matrices de covariances dépendant du modèle  $C_{NGC,\theta}$ , and  $C_{SGC,\theta}$ , comme les matrices de covariance empiriques des simulations. La moyenne et la covariance empirique d'un ensemble de  $N$  réalisations sont données par :

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^N [\hat{\xi}_k(r_i) - \bar{\xi}(r_i)][\hat{\xi}_k(r_j) - \bar{\xi}(r_j)] \quad (5.48)$$

$$\bar{\xi} = \frac{1}{N} \sum_{k=1}^N \hat{\xi}_k \quad (5.49)$$

Pour chaque valeur de  $\theta$ , on estime la fonction de corrélation de l'échantillon complet  $\hat{\xi}$ , en prenant la combinaison linéaire optimale de  $\hat{\xi}_{NGC}$  et  $\hat{\xi}_{SGC}$  comme dans [White et al. \(2011\)](#) (voir annexe C) :

$$\hat{\xi} = C_\theta \left[ C_{NGC,\theta}^{-1} \hat{\xi}_{NGC} + C_{SGC,\theta}^{-1} \hat{\xi}_{SGC} \right] \quad (5.50)$$

$$C_\theta = \left( C_{NGC,\theta}^{-1} + C_{SGC,\theta}^{-1} \right)^{-1} \quad (5.51)$$

avec  $C_\theta$  la matrice de covariance de l'estimateur complet  $\hat{\xi}$  résultant de ce choix.

Les trois paramètres de  $\theta = (\omega_m, \alpha, B)$  sont pris en compte de façon différente dans nos simulations. Le paramètre  $\omega_m$  change la forme globale de la fonction de corrélation d'une façon complexe, et l'on n'a pas d'autre choix que de générer de nouveaux ensembles de simulations log-normales pour chaque valeur de  $\omega_m$ . On choisit de considérer 5 valeurs différentes,  $\omega_m = 0.08, 0.105, 0.13, 0.155, 0.18$  afin d'estimer une matrice de covariance pour chacune de ces valeurs. Puis on va simplement interpoler linéairement la matrice de covariance pour les valeurs intermédiaires de  $\omega_m$  (l'interpolation étant faite dans le sens classique, i.e. coefficient par coefficient).

D'un autre côté, le paramètre  $\alpha$  peut être pris en compte de façon moins contraignante avec la procédure détaillée dans la section 3.4.3. On utilise les valeurs minimum  $\alpha_{min} = 0.8$  et maximum

5. Cette approximation est très forte, mais cela n'est pas si grave puisque notre but est seulement de comparer les différentes méthodes de détection des BAOs.

6. Ces poids ne sont pas optimaux puisque la densité  $\bar{n}(z)$  varie en fonction du *redshift*, mais comme on l'a vu dans la section 3.2.2, l'estimateur de la fonction de corrélation dépend très peu de ce choix.

$\alpha_{max} = 0.8$ . Etant donné que  $D_V^{fid}(z = 0.3) = 1180 \text{Mpc}$ , les valeurs minimum et maximum de  $\alpha$  correspondent à un intervalle  $D_V(0.3) \in [944 \text{ Mpc}, 1416 \text{ Mpc}]$  pour nos simulations.

Etudions finalement comment modéliser l'effet de  $B = b^2$ , qui change aussi la vraie distribution en coordonnées comobiles, de la même façon que  $\omega_m$ . Mais parce que c'est simplement un facteur multiplicatif constant  $B$ , cela devrait donner approximativement un facteur multiplicatif  $B^2$  dans la matrice de covariance de  $\hat{\xi}$ . On rappelle qu'il y a deux sources de bruit dans l'estimateur  $\hat{\xi}$  (voir section 2.2.4) :

- Le bruit d'échantillon dû au volume fini du catalogue
- Le bruit de Poisson dû au nombre fini de galaxies

L'approximation d'une matrice de covariance proportionnelle à  $B^2$  est valide lorsque l'on peut négliger la contribution du bruit de Poisson par rapport au bruit d'échantillon. Cette approximation sera donc mieux vérifiée pour des valeurs grandes de  $B$ . On vérifiera dans la section 6.3 que cette approximation est bonne pour des valeurs raisonnables de  $B$ , et que l'approximation  $C_B \propto B^2 C$  est bien plus proche de la vraie matrice de covariance que l'approximation d'une matrice constante  $C_B = C$ . Ce paramètre peut donc être traité sans avoir besoin de simulations additionnelles.

La procédure de la section 3.4.3 permet d'estimer la fonction de corrélation pour toutes les valeurs de  $\alpha$  simultanément avec un temps seulement multiplié par un facteur 4 par rapport à une unique valeur de  $\alpha$ . Donc tout notre ensemble de simulations sur la grille en  $\theta$  va demander seulement un temps de calcul multiplié par un facteur 20 par rapport à des simulations pour une unique valeur de  $\theta$ .

Notre ensemble principal de simulations est généré avec une valeur  $B = 2.5^2$  (notons que l'on considère ici le biais par rapport à la corrélation en espace réel et non en espace *redshift*, i.e. sans le facteur de Kaiser de l'équation (2.30)). Dans la section 6.3, on vérifiera la validité de l'approximation  $C_B \propto B^2 C$  en comparant la matrice de covariance obtenue sur l'ensemble avec  $B = 2.5^2$  à la matrice de covariance obtenue sur un ensemble avec  $B = 3.0^2$ .

Pour chaque valeur de  $(\omega_m, \alpha)$  on génère  $N = 2,000$  simulations log-normales pour estimer la matrice de covariance dépendant du modèle  $C_{\omega_m, \alpha}$ , de sorte qu'il y a très peu de bruit dans l'estimation.

### 5.4.3 Verification de la modélisation statistique de $\hat{\xi}$

Dans cette section on vérifie que les hypothèses suivantes sont valides :

$$\begin{aligned} \mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO, \theta}, C_{noBAO, \theta}) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO, \theta}, C_{BAO, \theta}) \end{aligned}$$

Cela implique de vérifier d'une part que les estimateurs ne sont pas biaisés, et d'autre part qu'ils sont bien Gaussiens.

#### Absence de biais dans $\hat{\xi}$

Pour vérifier l'absence de biais dans les estimateurs, on utilise les simulations log-normales de la section 5.4.2 et l'on calcule la moyenne des estimateurs de la fonction de corrélation  $\hat{\xi}_k$  pour les paramètres  $\alpha = 1$ ,  $B = 2.5^2$  et pour différentes valeurs de  $\omega_m = 0.08, 0.105, 0.13, 0.155, 0.18$ , avec à chaque fois  $N = 2000$  réalisations :

$$\bar{\xi}_{\omega_m} = \frac{1}{N} \sum_{k=1}^N \hat{\xi}_{k, \omega_m} \quad (5.52)$$

On montre sur la figure 5.2 les moyennes des estimateurs  $\bar{\xi}_{\omega_m}$  sur 2,000 réalisations, comparées aux fonctions de corrélation attendues  $\xi_{\omega_m}$  donnée en entrée des simulations log-normales. La figure 5.2 montre l'excellent accord entre ces fonctions, i.e. que les estimateurs ont un biais négligeable.

#### Verification de la Gaussianité de $\hat{\xi}$

Dans cette section on vérifie la Gaussianité de l'estimateur  $\hat{\xi}$  pour les  $N = 2,000$  réalisations log-normales de la section 5.4.2 pour la valeur  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1, 2.5^2)$ . On estime la matrice

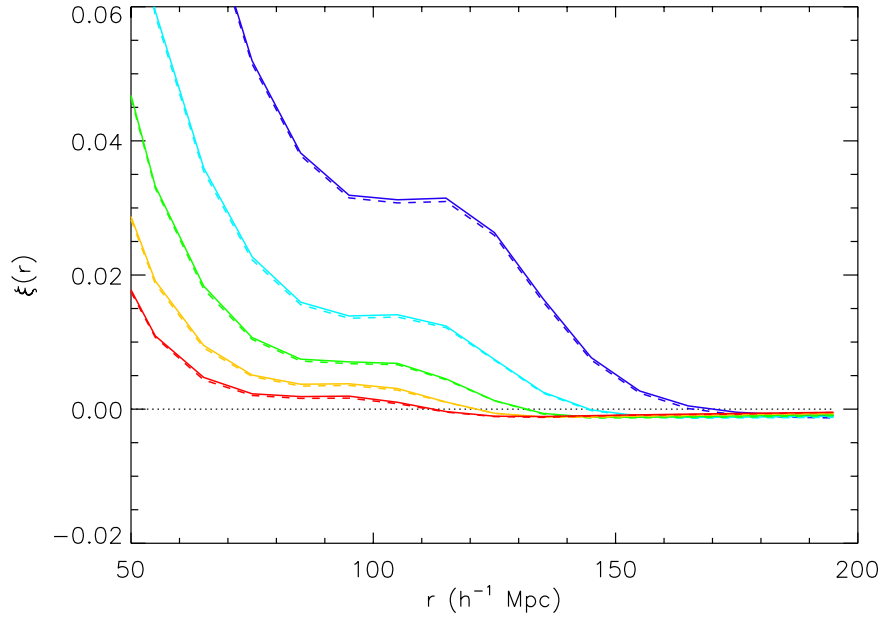


FIGURE 5.2 – Moyennes des estimateurs  $\bar{\xi}_{\omega_m}$  en tirets comparées aux fonctions de corrélation attendues  $\xi_{\omega_m}$  en traits pleins pour  $\alpha = 1$ ,  $B = 2.5^2$  et pour  $\omega_m = 0.08$  (violet), 0.105 (bleu clair), 0.13 (vert), 0.155 (jaune), 0.18 (rouge).

de covariance empirique sur ces  $N = 2,000$  réalisations :

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^N [\hat{\xi}_k(r_i) - \bar{\xi}(r_i)][\hat{\xi}_k(r_j) - \bar{\xi}(r_j)] \quad (5.53)$$

$$\bar{\xi} = \frac{1}{N} \sum_{k=1}^N \hat{\xi}_k \quad (5.54)$$

On regarde d'abord la statistique  $\chi_{\theta_0}^2$  :

$$\chi_{\theta_0}^2 = \left\langle \hat{\xi} - \xi_{\theta_0}, C^{-1}(\hat{\xi} - \xi_{\theta_0}) \right\rangle \quad (5.55)$$

$$= \sum_{1 \leq i, j \leq n} \left[ \hat{\xi}(r_i) - \xi_{\theta_0}(r_i) \right] C_{i,j}^{-1} \left[ \hat{\xi}(r_j) - \xi_{\theta_0}(r_j) \right] \quad (5.56)$$

Avec l'hypothèse Gaussienne  $\hat{\xi} \sim \mathcal{N}(\xi_{\theta_0}, C)$ ,  $\chi_{\theta_0}^2$  suit une loi du  $\chi^2$  avec  $n$  degrés de liberté :

$$\chi_{\theta_0}^2 \sim \chi_n^2 \quad (5.57)$$

On rappelle que l'on utilise des intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  entre les distances  $r = 20h^{-1}\text{Mpc}$  et  $r = 200h^{-1}\text{Mpc}$ , ce qui correspond à  $n = 18$  intervalles. On compare donc l'histogramme de  $\chi_{\theta_0}^2$  sur nos simulations log-normales avec la densité de probabilité (pdf pour *probability density function*) pour une loi  $\chi_{18}^2$ . On montre figure 5.3 le très bon accord entre l'histogramme et la distribution théorique.

Pour les tests suivants, on utilise les méthodes en ondelettes de la section 5.1.1. Dans ces méthodes, on obtient une réponse de l'ondelette  $S_w(R, s)$  et un  $Z$ -score  $Z_w(R, s)$  pour chaque paramètre  $(R, s)$  :

$$S_w(R, s) = \left\langle w(R, s), \hat{\xi} \right\rangle = \sum_{i=1}^n w_i(R, s) \hat{\xi}(r_i) \quad (5.58)$$

$$Z_w(R, s) = \frac{S_w(R, s)}{\sigma_w(R, s)} \quad (5.59)$$

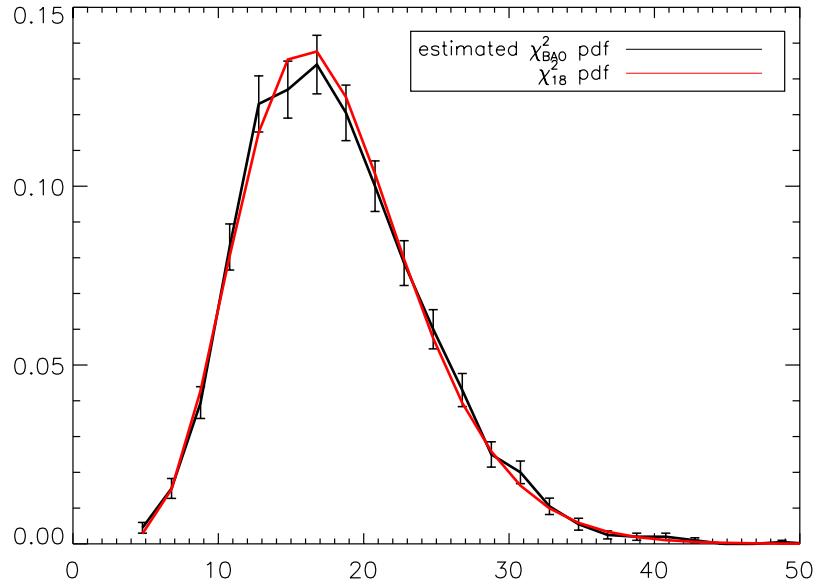


FIGURE 5.3 – pdf estimée de  $\chi_{\theta_0}^2$  (noir) en utilisant l’histogramme des 2,000 simulations log-normales et pdf d’une loi  $\chi_{18}^2$  (rouge). Les barres d’erreur donnent l’erreur de Poisson dans l’estimation de la pdf due au nombre fini de simulations.

On considère seulement le filtre du chapeau mexicain avec les paramètres  $R = 113.6 h^{-1}\text{Mpc}$ ,  $s = 20 h^{-1}\text{Mpc}$ , et le filtre BAOlet avec les paramètres  $R = 116 h^{-1}\text{Mpc}$ ,  $s = 36 h^{-1}\text{Mpc}$ . Ces paramètres maximisent le  $Z$ -score obtenu sur la fonction de corrélation estimée dans les études respectives de [Tian et al. \(2011\)](#) et [Arnalte-Mur et al. \(2012\)](#). Afin d’obtenir  $Z_w$  sur nos simulations, on calcule le bruit  $\sigma_w$  (i.e. l’écart-type) sur  $S_w$  en utilisant la matrice de covariance des simulations :

$$\sigma_w = \sqrt{\langle w, Cw \rangle} \quad (5.60)$$

Avec l’hypothèse Gaussienne  $\hat{\xi} \sim \mathcal{N}(\xi_{\theta_0}, C)$ ,  $Z_w$  est une variable Gaussienne d’espérance  $\mathbb{E}[Z_w] = \langle w, \xi_{\theta_0} \rangle / \sigma_w(R, s)$  et d’écart-type égal à 1. On trace sur les figures 5.4 et 5.5 l’histogramme de  $Z_w$  sur nos simulations log-normales, respectivement pour le chapeau mexicain et pour le filtre BAOlet.

On trouve à nouveau un très bon accord entre le résultat sur les simulations et la prédiction Gaussienne.

## 5.5 Détection des BAOs sur des simulations SDSS DR7-Full

### 5.5.1 Détection des BAOs avec une matrice de covariance constante $C$

**Erreur de la méthode du  $\chi^2$  classique pour la réjection de modèles  $\mathcal{H}_0$  particuliers**

Les hypothèses testées sont :

$$\begin{aligned} \mathcal{H}_0 &: \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO, \theta}, C) \\ \mathcal{H}_1 &: \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO, \theta}, C) \end{aligned}$$

où l’on prend comme matrice de covariance constante  $C$  la matrice de covariance des  $N = 2,000$  simulations log-normales pour la valeur  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1, 2.5^2)$ .

L’hypothèse de Gaussianité des estimateurs est bien valide (au moins pour nos simulations log-normales) comme on l’a vu dans la section 5.4.3. Ici on fait toutefois l’approximation que la matrice de



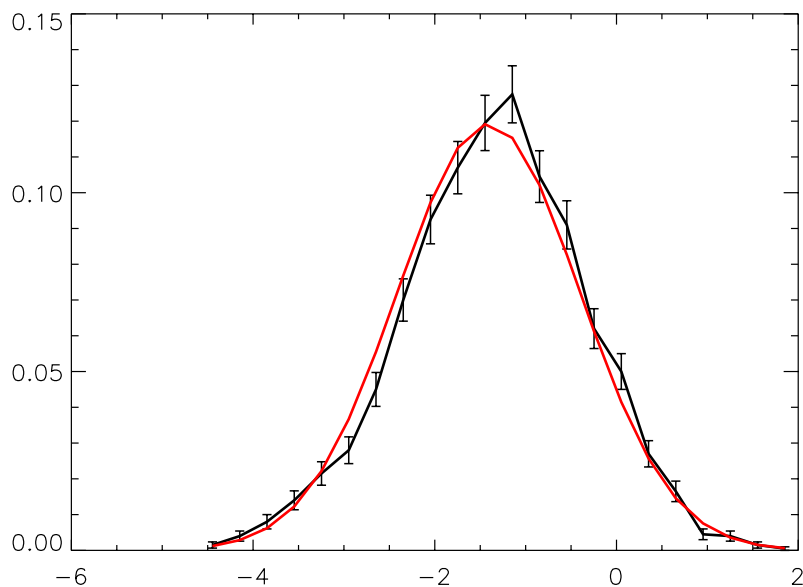


FIGURE 5.4 – pdf estimée de  $Z_w$  pour le chapeau mexicain avec les paramètres  $R = 113.6 h^{-1}\text{Mpc}$ ,  $s = 20 h^{-1}\text{Mpc}$  en utilisant l'histogramme sur 2,000 simulations log-normales (noir) et pdf d'une loi normale centrée sur  $\langle w, \xi_{\theta_0} \rangle / \sigma_w(R, s)$  et d'écart-type 1. Les barres d'erreur donnent l'erreur de Poisson sur l'histogramme due au nombre fini de simulations.

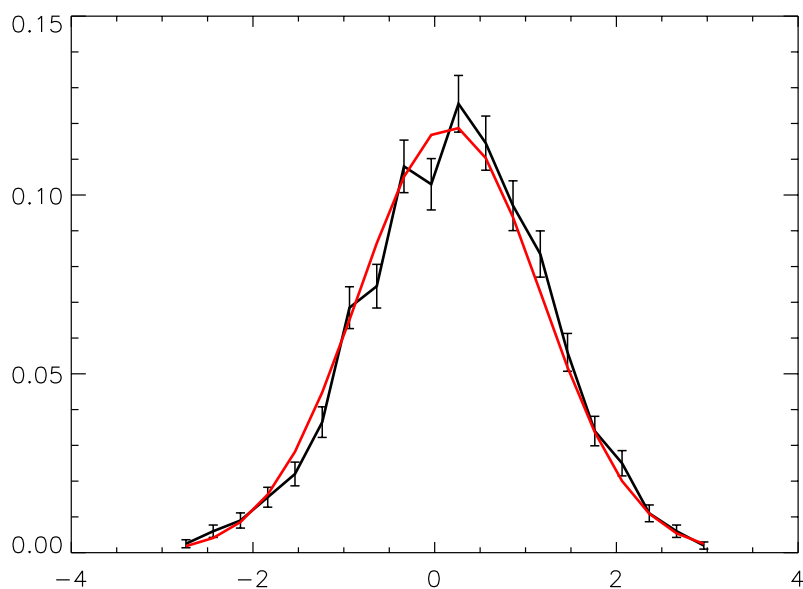


FIGURE 5.5 – pdf estimée de  $Z_w$  pour le filtre BAOlet avec les paramètres  $R = 116 h^{-1}\text{Mpc}$ ,  $s = 36 h^{-1}\text{Mpc}$  en utilisant l'histogramme sur 2,000 simulations log-normales (noir) et pdf d'une loi normale centrée sur  $\langle w, \xi_{\theta_0} \rangle / \sigma_w(R, s)$  et d'écart-type 1. Les barres d'erreur donnent l'erreur de Poisson sur l'histogramme due au nombre fini de simulations.

covariance est constante, alors qu'elle dépend fortement de  $\theta$  comme on le montre dans la section 6.3. Ceci va toutefois nous permettre de tester si méthode du  $\chi^2$  classique donne la bonne significativité dans le cadre de cette approximation. Les statistiques chi-carré dépendent de  $\theta$  :

$$\chi_{BAO,\theta}^2 = \left\langle \hat{\xi} - \xi_{BAO,\theta}, C^{-1}(\hat{\xi} - \xi_{BAO,\theta}) \right\rangle \quad (5.61)$$

$$\chi_{noBAO,\theta}^2 = \left\langle \hat{\xi} - \xi_{noBAO,\theta}, C^{-1}(\hat{\xi} - \xi_{noBAO,\theta}) \right\rangle \quad (5.62)$$

Un modèle étendu de fonctions de corrélation est implicitement défini, pour intégrer à la fois les modèles BAO et non-BAO, par exemple :

$$\xi_{\beta,\theta} = \beta \xi_{BAO,\theta} + (1 - \beta) \xi_{noBAO,\theta} \quad (5.63)$$

Notons comme précédemment :

$$\Delta\chi^2 = \min_{\theta} \chi_{noBAO,\theta}^2 - \min_{\theta} \chi_{BAO,\theta}^2 \quad (5.64)$$

$$\Delta\chi_{global}^2 = \min_{\theta} \chi_{noBAO,\theta}^2 - \min_{\beta,\theta} \chi_{\beta,\theta}^2 \quad (5.65)$$

La méthode suppose que  $\Delta\chi_{global}^2$  suit une loi  $\chi_1^2$  sous  $\mathcal{H}_0$ . Comme on a  $\Delta\chi^2 \leq \Delta\chi_{global}^2$  par construction, une valeur conservatrice de la significativité est donnée par  $\sqrt{\Delta\chi^2} \cdot \sigma$  lorsque  $\Delta\chi^2 \geq 0$

On rappelle que  $\Delta\chi_{global}^2 \sim \chi_1^2$  est soumis à l'hypothèse de régularité que les espaces de modèles de fonctions de corrélation  $(\xi_{noBAO,\theta})_{\theta \in \Theta}$  et  $(\xi_{\beta,\theta})_{\beta \in \mathbb{R}, \theta \in \Theta}$  sont affines. Comme cela n'est pas facile à vérifier (il est également difficile de quantifier l'écart par rapport à cette hypothèse), nous allons simplement tester sur nos simulations si l'approximation  $\Delta\chi_{global}^2 \sim \chi_1^2$  sous  $\mathcal{H}_0$  est bien vérifiée.

Pour un modèle  $\theta$  dans  $\mathcal{H}_0$ , on génère des réalisations de la façon suivante :

$$\hat{\xi} = C^{1/2}g + \xi_{noBAO,\theta} \quad (5.66)$$

où  $g$  est une variable Gaussienne multivariée standard (i.e.  $g \sim \mathcal{N}(0, I_n)$ ). Pour chaque réalisation on calcule le meilleur *fit* en testant toutes les valeurs de  $\omega_m$  et  $\alpha$  sur une grille. Les paramètres restants, qui sont l'amplitude  $B$  et le paramètre  $\beta$ , sont trouvées analytiquement pour des valeurs de  $\omega_m$  et  $\alpha$  données.

Pour la grille  $(\omega_m, \alpha)$ , on prend  $\omega_m \in [0.10, 0.16]$  avec un pas de grille égal à 0.001, et  $\alpha \in [0.8, 1.2]$  avec un pas de grille égal à 0.04. On autorise n'importe quelle valeur de  $B = b^2 \geq 0$  et  $\beta$ . On teste deux modèles dans  $\mathcal{H}_0$  avec valeurs  $\omega_m = 0.13$ ,  $\alpha = 1.0$ ,  $B = 2.5^2$  et avec des valeurs différentes  $\omega_m = 0.115$ ,  $\alpha = 0.9$ ,  $B = 2.5^2$ . A chaque fois, on génère 10,000 réalisations en utilisant l'équation (5.66) pour estimer la distribution de  $\Delta\chi_{global}^2$  et  $\Delta\chi^2$ . On montre dans les tables 5.1 et 5.2 pour différents seuils  $t$ , la valeur  $p$  et la significativité correspondante pour une loi  $\chi_1^2$ , pour  $\Delta\chi_{global}^2$ , et pour  $\Delta\chi^2$ . Nos résultats montrent que l'approximation  $\Delta\chi_{global}^2 \sim \chi_1^2$  est très mauvaise. En particulier, la moyenne de  $\Delta\chi_{global}^2$  pour les deux différents modèles est respectivement de 1.91 et 1.68, alors que l'espérance d'une loi  $\chi_1^2$  est égale à 1.

Dans les deux cas, on surestime grossièrement la significativité en identifiant  $\Delta\chi_{global}^2$  à une loi  $\chi_1^2$ . Le fait que la méthode du  $\chi^2$  classique soit conservatrice car elle utilise la valeur de  $\Delta\chi^2$  au lieu de  $\Delta\chi_{global}^2$  peut compenser cette surestimation. Pour le modèle  $\mathcal{H}_0$  de la table 5.1, identifier  $\Delta\chi^2$  avec une loi  $\chi_1^2$  peut encore donner une très légère surestimation de la significativité. Pour le modèle  $\mathcal{H}_0$  de la table 5.2, cela donne par contre une légère sous-estimation de la significativité.

Soulignons que ces significativités ne sont pas à strictement parler les significativités de la détection des BAOs, mais seulement de modèles particuliers dans  $\mathcal{H}_0$ . En effet la détection des BAOs consiste à rejeter tous les modèles  $\mathcal{H}_0$  simultanément, par rapport aux modèles  $\mathcal{H}_1$ .

Nous avons également fait ce type de tests avec les modèles zéro-baryon pour  $\mathcal{H}_0$ , et nous avons trouvé les mêmes résultats qualitatifs. On a trouvé que la distribution de  $\Delta\chi_{global}^2$  est très différente de celle d'une variable  $\chi_1^2$ , et que l'identification de  $\Delta\chi^2$  avec une loi du  $\chi_1^2$  compense cette surestimation, bien que cela puisse mener à de légères sous-estimations ou surestimation de la significativité.

L'approximation  $\Delta\chi_{global}^2 \sim \chi_1^2$  n'est pas valide parce que les espaces de modèles de fonctions de corrélation  $(\xi_{noBAO,\theta})_{\theta \in \Theta}$  et  $(\xi_{\beta,\theta})_{\beta \in \mathbb{R}, \theta \in \Theta}$  ne peuvent pas être considérés comme affines. Il est facile

TABLE 5.1 –

	$\chi_1^2$	$\Delta\chi_{global}^2$	$\Delta\chi^2$
$P(X \geq 1.0)$	0.32 ( $1\sigma$ )	0.58 ( $0.55\sigma$ )	0.18 ( $1.34\sigma$ )
$P(X \geq 2.25)$	0.13 ( $1.5\sigma$ )	0.31 ( $1.02\sigma$ )	0.11 ( $1.60\sigma$ )
$P(X \geq 4.0)$	$4.5 \times 10^{-2}$ ( $2\sigma$ )	0.13 ( $1.51\sigma$ )	$5.2 \times 10^{-2}$ ( $1.94\sigma$ )
$P(X \geq 6.25)$	$1.2 \times 10^{-2}$ ( $2.5\sigma$ )	$4.1 \times 10^{-2}$ ( $2.04\sigma$ )	$1.6 \times 10^{-2}$ ( $2.41\sigma$ )
$P(X \geq 9.0)$	$2.7 \times 10^{-3}$ ( $3\sigma$ )	$1.0 \times 10^{-2}$ ( $2.57\sigma$ )	$3.2 \times 10^{-3}$ ( $2.95\sigma$ )

NOTES.—Valeurs  $p$  et significativités correspondantes pour différentes distributions et pour la réjection d'un modèle particulier dans  $\mathcal{H}_0$  avec  $\omega_m = 0.13$ ,  $\alpha = 1.0$ ,  $B = 2.5^2$ . On regarde la distribution  $\chi_1^2$  et les distributions  $\Delta\chi_{global}^2$ ,  $\Delta\chi^2$ . L'approximation  $\Delta\chi_{global}^2 \sim \chi_1^2$  est mauvaise avec une significativité qui est grossièrement surestimée. Le fait que la méthode du  $\chi^2$  utilise  $\Delta\chi^2$  au lieu de  $\Delta\chi_{global}^2$  compense cette surestimation. Pour la réjection de ce modèle  $\mathcal{H}_0$  en particulier, il peut encore y avoir une très légère surestimation de la significativité si l'on identifie  $\Delta\chi^2$  à une variable  $\chi_1^2$ .

TABLE 5.2 –

	$\chi_1^2$	$\Delta\chi_{global}^2$	$\Delta\chi^2$
$P(X \geq 1.0)$	0.32 ( $1\sigma$ )	0.50 ( $0.67\sigma$ )	$4.2 \times 10^{-2}$ ( $2.03\sigma$ )
$P(X \geq 2.25)$	0.13 ( $1.5\sigma$ )	0.27 ( $1.11\sigma$ )	$2.9 \times 10^{-2}$ ( $2.18\sigma$ )
$P(X \geq 4.0)$	$4.5 \times 10^{-2}$ ( $2\sigma$ )	0.11 ( $1.60\sigma$ )	$1.6 \times 10^{-2}$ ( $2.41\sigma$ )
$P(X \geq 6.25)$	$1.2 \times 10^{-2}$ ( $2.5\sigma$ )	$3.7 \times 10^{-2}$ ( $2.09\sigma$ )	$6.6 \times 10^{-3}$ ( $2.72\sigma$ )
$P(X \geq 9.0)$	$2.7 \times 10^{-3}$ ( $3\sigma$ )	$8.4 \times 10^{-3}$ ( $2.64\sigma$ )	$1.7 \times 10^{-3}$ ( $3.14\sigma$ )

NOTES.—Même chose que table 5.1 pour la réjection d'un modèle particulier dans  $\mathcal{H}_0$  avec  $\omega_m = 0.115$ ,  $\alpha = 0.9$ ,  $B = 2.5^2$ . A nouveau l'approximation  $\Delta\chi_{global}^2 \sim \chi_1^2$  est mauvaise avec une significativité qui est grossièrement surestimée. Le fait que la méthode du  $\chi^2$  utilise  $\Delta\chi^2$  au lieu de  $\Delta\chi_{global}^2$  compense cette surestimation. Pour la réjection de ce modèle  $\mathcal{H}_0$  en particulier, la significativité devient légèrement sous-estimée si l'on identifie  $\Delta\chi^2$  à une variable  $\chi_1^2$ .

de voir par exemple que les fonctions de corrélation non-BAO sont plus dégénérées par rapport à leurs 3 paramètres que les fonctions de corrélation BAO<sup>7</sup>. Ainsi pour un même espace des paramètres, l'espace des fonctions de corrélation BAO est plus grand que l'espace des fonctions de corrélation non-BAO. Ceci favorise la minimisation de  $\chi_{BAO}^2$  par rapport à  $\chi_{noBAO}^2$ , i.e. tend à augmenter la valeur de  $\Delta\chi_{global}^2$  et  $\Delta\chi^2$ . Dans ce cas, les espaces de fonction de corrélation ne sont pas affines simplement à cause de la limitation de l'espace des paramètres.

Comme on l'a vu dans la section 5.3, on a besoin de considérer le pire scénario pour le modèle  $\mathcal{H}_0$  afin d'obtenir une significativité pour la réjection de tous les modèles  $\mathcal{H}_0$  simultanément. Pour une réalisation  $\Delta\chi^2 = x$  la valeur  $p$  est donnée par :

$$p(x) = \max_{\theta \in \Theta} P(\Delta\chi^2 \geq x | \mathcal{H}_0, \theta) \quad (5.67)$$

On a vu précédemment (tables 5.1 et 5.2) que la significativité peut être légèrement surestimée ou sous-estimée pour la réjection d'un modèle particulier dans  $\mathcal{H}_0$ , lorsque l'on identifie  $\Delta\chi^2$  avec une variable  $\chi_1^2$ . Supposons qu'elle soit légèrement surestimée pour un modèle particulier dans  $\mathcal{H}_0$  avec paramètre  $\theta$ , i.e. que  $\Delta\chi^2$  peut être minorée par une variable  $\chi_1^2$ , i.e. pour  $x \geq 0$  :

$$P(\Delta\chi^2 \geq x | \mathcal{H}_0, \theta) > p_{\chi_1^2}(x) = 2\Phi(-\sqrt{x}) \quad (5.68)$$

La conséquence est que la méthode du  $\chi^2$  classique peut légèrement surestimer la significativité pour la détection des BAOs. En effet en utilisant l'équation (5.67) pour déterminer la significativité de la réjection de tous les modèles  $\mathcal{H}_0$  simultanément (i.e. la réjection de  $\mathcal{H}_0$ ), on obtient :

$$p(x) > p_{\chi_1^2}(x) = 2\Phi(-\sqrt{x}) \quad (5.69)$$

On verra dans la section 5.5.1 que la significativité  $\sqrt{\Delta\chi^2}\sigma$  donnée par la méthode classique du  $\chi^2$  donne quand même une bonne approximation de la vraie significativité dans le cas d'une matrice de covariance constante  $C$ . Toutefois on verra dans la section 5.5.2, que lorsque l'on considère des hypothèses avec une matrice de covariance qui dépend du modèle  $C_\theta$ , la significativité donnée par la méthode classique du  $\chi^2$  devient grossièrement fausse.

### Erreur de la méthode du $\chi^2$ classique pour la réjection de tous les modèles $\mathcal{H}_0$ simultanément (i.e. pour la détection des BAOs)

Dans cette section, on teste la modification de la procédure pour calculer la significativité que l'on a introduit dans la section 5.3. On considère toujours une matrice de covariance constante, i.e. les hypothèses :

$$\begin{aligned} \mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO, \theta}, C) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO, \theta}, C) \end{aligned}$$

où l'on prend encore comme matrice de covariance constante  $C$  la matrice de covariance des  $N = 2,000$  simulations log-normales pour la valeur  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1, 2.5^2)$ .

Comme nous nous plaçons toujours dans le cas d'une matrice de covariance constante, la statistique  $\Delta\chi^2$  est encore un rapport de vraisemblance généralisé. On test le changement de la procédure pour calculer la significativité afin d'obtenir une valeur correcte comme dans l'équation (5.29) ou (5.67). Pour une réalisation  $\Delta l = x$  la valeur  $p$  est donnée par :

$$p(x) = \max_{\theta \in \Theta} P(\Delta l \geq x | \mathcal{H}_0, \theta) \quad (5.70)$$

La méthode requiert de précalculer  $p(x)$  afin d'obtenir la valeur  $p$  pour une réalisation donnée  $\Delta l = x$ . Ici on considère une grille de paramètres  $\omega_m \in [0.10, 0.16]$  avec un pas de grille 0.01, et  $\alpha \in [0.8, 1.2]$  avec un pas de grille 0.05,  $B = b^2 \in [4, 9]$  avec un pas de grille 0.25. Cette grille n'est pas très fine parce que le calcul de  $p(x)$  est proportionnel au carré de la taille de la grille, i.e. qu'il augmente très rapidement pour 3 paramètres. En effet il faut calculer la distribution de  $\Delta l$  pour tous

7. En effet les fonctions de corrélation sans BAOs sont lisses et presque invariantes d'échelles, alors que le pic BAO introduit une caractéristique supplémentaire dans les fonctions de corrélation avec BAOs.

les modèles de la grille sous  $\mathcal{H}_0$ . Et pour toutes les réalisations d'un modèle  $\mathcal{H}_0$  particulier, le calcul de  $\Delta l$  requiert aussi une minimisation sur tous les paramètres de la grille. On a testé de raffiner la grille et on a trouvé un bon accord à quelques pourcents près, ce qui est suffisant pour notre objectif d'illustration (ce serait toutefois important pour améliorer la précision d'une analyse sur de vraies données).

Pour chaque modèle de la grille sous  $\mathcal{H}_0$ , on génère 50,000 réalisations avec l'équation (5.66) comme précédemment, et on calcule la statistique  $\Delta l$ . Avec ces réalisations pour tous les modèles  $\mathcal{H}_0$ , on peut donc calculer la fonction  $p(x)$  de l'équation (5.70).

Le fait que l'on utilise seulement 50,000 réalisations pour chaque modèle ne permet pas de bien quantifier des valeurs  $p$  plus petites qu'environ  $2 \times 10^{-5}$ , i.e. des significativités de plus de  $4.25\sigma$ . Notons qu'à la différence de la méthode classique du  $\chi^2$ , cette imprécision est seulement computationnelle et n'intervient que pour des niveaux de détection élevés. Donc si l'on veut quantifier la détection des BAOs avec une mesure sur des données  $\hat{\xi}$ , l'imprécision n'a lieu que lorsque la significativité est élevée et la détection déjà claire.

On effectue la même procédure de simulation pour  $\mathcal{H}_1$ , i.e. qu'on génère 50,000 réalisations pour chaque modèle de la grille sous  $\mathcal{H}_1$  avec la formule suivante :

$$\hat{\xi} = C^{1/2}g + \xi_{BAO,\theta} \quad (5.71)$$

avec  $g \sim \mathcal{N}(0, I_n)$ .

Comme  $\mathcal{H}_1$  est une hypothèse composite, la distribution de  $\Delta l$  n'est pas bien définie sous  $\mathcal{H}_1$ . Par exemple on ne peut pas parler d'espérance sous  $\mathcal{H}_1$ . Ici on va simplement s'intéresser à l'espérance de la significativité pour chaque modèle  $\mathcal{H}_1$ , que l'on va moyenner sur tous les modèles (on appellera cela la significativité moyenne)<sup>8</sup>.

On montre nos résultats dans la table 5.3. On obtient une significativité moyenne de  $2.75\sigma$  avec cette procédure sous  $\mathcal{H}_1$ . D'un autre côté lorsqu'on calcule la significativité avec  $\sqrt{\Delta\chi^2}\sigma$  comme dans la méthode classique du  $\chi^2$ , on obtient une significativité moyenne de  $3.15\sigma$  (avec la convention que  $\Delta\chi^2 \leq 0$  correspond à une significativité de  $0\sigma$ ). Regardons maintenant l'effet de l'imprécision aux significativités élevées en ne considérant que les réalisations sous la limite des  $4.25\sigma$ . Dans ce cas on obtient une significativité moyenne de  $2.43\sigma$  avec la procédure rigoureuse et une moyenne de  $\sqrt{\Delta\chi^2}$  égale à 2.48. On voit donc que les moyennes obtenues avec la méthode classique du  $\chi^2$  et notre méthode rigoureuse sont très similaires.

Finalement on effectue les mêmes calculs pour des modèles  $\mathcal{H}_0$  avec zéro baryon et l'on montre nos résultats dans la table 5.4. Si l'on utilise la procédure rigoureuse, on obtient une significativité moyenne de  $2.88\sigma$  et la moyenne de  $\sqrt{\Delta\chi^2}$  est égale à 3.88. Si l'on se restreint aux réalisations sous la limite de  $4.25\sigma$ , on obtient une significativité moyenne de  $2.36\sigma$  avec notre procédure et une moyenne de  $\sqrt{\Delta\chi^2}$  égale à 2.5. A nouveau, la méthode classique du  $\chi^2$  donne une bonne approximation de la significativité moyenne.

A partir des tables 5.3 et 5.4, on a vu que la méthode du  $\chi^2$  classique donne une bonne estimation de la significativité moyenne. Pour comprendre si cette estimation est bonne de façon générale, on compare pour toutes les valeurs possibles de la statistique  $\Delta\chi^2$  les estimations de la significativité  $\sqrt{\Delta\chi^2}\sigma$  par la méthode du  $\chi^2$  classique et  $\sigma_{\text{reel}}(\Delta\chi^2)$  pour notre méthode rigoureuse. On montre cette comparaison sur la figure 5.6 construit à partir de toutes les valeurs de  $\Delta\chi^2$  obtenues sur nos simulations  $\mathcal{H}_1$ . La sous-figure de gauche montre le cas d'une hypothèse  $\mathcal{H}_0$  *no wiggles* et la sous-figure de droite le cas d'une hypothèse  $\mathcal{H}_0$  zéro baryon. On voit que la significativité donnée par  $\sqrt{\Delta\chi^2}\sigma$  n'est pas tout à fait correcte. Elle est sous-estimée pour les faibles significativités et devient surestimée aux hautes significativités. Toutefois dans l'ensemble on peut dire que la significativité  $\sqrt{\Delta\chi^2}\sigma$  donnée par la méthode classique est une approximation très correcte de la vraie significativité dans le cas d'une matrice de covariance constante  $C$ .

8. Ceci est en fait équivalent à l'espérance de la significativité sous  $\mathcal{H}_1$ , si l'on introduit une probabilité a priori  $p(\theta)$  constante pour les paramètres  $\theta$  de  $\mathcal{H}_1$ .

TABLE 5.3 –

	$\overline{\sqrt{\Delta\chi^2}}$	$\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$
Toutes simus $\mathcal{H}_1$	3.15 $\sigma$	2.75 $\sigma$
Simus $\mathcal{H}_1$ avec $\sigma_{\text{reel}}(\Delta\chi^2) < 4.25\sigma$	2.48 $\sigma$	2.43 $\sigma$

NOTES.— Significativités moyennes obtenues sous  $\mathcal{H}_1$  avec la méthode classique  $\sqrt{\Delta\chi^2}$  (avec la convention  $\sqrt{\Delta\chi^2} = 0$  pour  $\Delta\chi^2 < 0$ ) et avec notre méthode rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$ , avec une hypothèse  $\mathcal{H}_0$  *no wiggles*. On montre les moyennes  $\overline{\sqrt{\Delta\chi^2}}$  et  $\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$  obtenues avec toutes les simulations  $\mathcal{H}_1$  et en se limitant au seuil de significativité de notre méthode  $\sigma_{\text{reel}}(\Delta\chi^2) < 4.25\sigma$ . En utilisant toutes les simulations  $\mathcal{H}_1$ , on a l'impression que la méthode classique du  $\chi^2$  surestime la significativité en moyenne. Toutefois en prenant en compte le seuil de significativité, la méthode classique du  $\chi^2$  et notre méthode rigoureuse donnent des significativités moyennes similaires.

TABLE 5.4 –

	$\overline{\sqrt{\Delta\chi^2}}$	$\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$
Toutes simus $\mathcal{H}_1$	3.88 $\sigma$	2.88 $\sigma$
Simus $\mathcal{H}_1$ avec $\sigma_{\text{reel}}(\Delta\chi^2) < 4.25\sigma$	2.51 $\sigma$	2.36 $\sigma$

NOTES.— Significativités moyennes obtenues sous  $\mathcal{H}_1$  avec la méthode classique  $\sqrt{\Delta\chi^2}$  (avec la convention  $\sqrt{\Delta\chi^2} = 0$  pour  $\Delta\chi^2 < 0$ ) et avec notre méthode rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$ , avec une hypothèse  $\mathcal{H}_0$  zéro baryon. On montre les moyennes  $\overline{\sqrt{\Delta\chi^2}}$  et  $\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$  obtenues avec toutes les simulations  $\mathcal{H}_1$  et en se limitant au seuil de significativité de notre méthode  $\sigma_{\text{reel}}(\Delta\chi^2) < 4.25\sigma$ . En utilisant toutes les simulations  $\mathcal{H}_1$ , on a l'impression que la méthode classique du  $\chi^2$  surestime la significativité en moyenne. Toutefois en prenant en compte le seuil de significativité, la méthode classique du  $\chi^2$  et notre méthode rigoureuse donnent des significativités moyennes similaires.

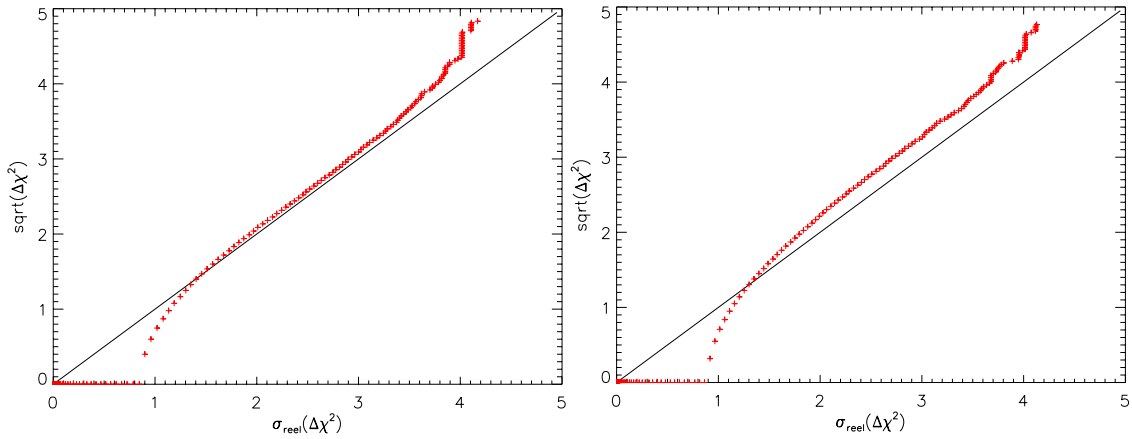


FIGURE 5.6 – A gauche : Significativité de la détection des BAOs obtenue avec la méthode classique  $\sqrt{\Delta\chi^2}$  (avec la convention  $\sqrt{\Delta\chi^2} = 0$  pour  $\Delta\chi^2 < 0$ ) en fonction de la significativité rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$ , avec une hypothèse  $\mathcal{H}_0$  *no wiggles*. La courbe est obtenue pour 50,000 réalisations qui prennent différentes valeurs de  $\Delta\chi^2$ . On voit que pour  $\Delta\chi^2 < 0$  et pour des valeurs faibles de  $\Delta\chi^2$ , la méthode classique sous-estime la significativité. En revanche aux plus grandes valeurs de  $\Delta\chi^2$ , la significativité devient surestimée. Dans l'ensemble, on peut quand même dire que la significativité  $\sqrt{\Delta\chi^2}\sigma$  donnée par la méthode classique est une approximation très correcte de la vraie significativité. A droite : Même chose pour une hypothèse  $\mathcal{H}_0$  zéro baryon.

### 5.5.2 Détection des BAOs avec une matrice de covariance $C_\theta$

On considère finalement le cas général où la matrice de covariance dépend de  $\theta$ , ce qui constitue des hypothèses plus réalistes :

$$\mathcal{H}_0 : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{\text{noBAO},\theta}, C_{\text{noBAO},\theta})$$

$$\mathcal{H}_1 : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{\text{BAO},\theta}, C_{\text{BAO},\theta})$$

Pour cela on utilise la matrice de covariance complète  $C_\theta$  obtenue à partir des simulations de la section 5.4.2. On montre la dépendance de la matrice de covariance  $C_\theta$  en fonction des 3 paramètres de  $\theta = (\omega_m, \alpha, B)$ .

On applique ici la méthode  $\Delta l$  avec la même procédure que celle décrite dans la section 5.5.1, sauf que l'on considère ici la statistique  $\Delta l$  au lieu de  $\Delta\chi^2$ . En effet dans le cas d'une matrice de covariance  $C_\theta$ , la statistique  $\Delta\chi^2$  n'est plus un rapport de vraisemblance généralisé. On génère des réalisations sous  $\mathcal{H}_0$  et  $\mathcal{H}_1$  avec les formules suivantes :

$$\hat{\xi} = C_\theta^{1/2} g + \xi_{\text{noBAO},\theta} \text{ sous } \mathcal{H}_0 \quad (5.72)$$

$$\hat{\xi} = C_\theta^{1/2} g + \xi_{\text{BAO},\theta} \text{ sous } \mathcal{H}_1 \quad (5.73)$$

avec  $g \sim \mathcal{N}(0, I_n)$ .

La statistique  $\Delta l$  est différente de la statistique  $\Delta\chi^2$ , et est calculée avec les équations (5.37), (5.38), (5.39), (5.40) et (5.41).

On montre nos résultats avec une hypothèse  $\mathcal{H}_0$  *no wiggles* dans la table 5.5. Cette fois on obtient une significativité moyenne de  $2.53\sigma$  sous  $\mathcal{H}_1$ , ce qui est un peu plus faible que pour une matrice de covariance constante ( $2.88\sigma$ ). D'un autre côté si l'on utilise la statistique  $\Delta\chi^2$  au lieu de  $\Delta l$  on obtient une significativité moyenne de  $1.87\sigma$  sous  $\mathcal{H}_1$  avec notre procédure rigoureuse. Cela justifie notre choix de remplacer la statistique  $\Delta\chi^2$  par  $\Delta l$ , qui reste un rapport de vraisemblance généralisé dans ce cas (voir la section 5.3). En utilisant la méthode classique du  $\chi^2$ , on obtient une moyenne de  $\sqrt{\Delta\chi^2}$  égale à 3.10. Dans ce cas la significativité donnée par la méthode classique est très loin de la vraie significativité moyenne égale à  $1.87\sigma$ . Comme on l'a déjà mentionné dans la section 5.2.3, la méthode classique du  $\chi^2$  ne peut pas être utilisée dans le cas où la matrice de covariance dépend du modèle.

TABLE 5.5 –

	$\overline{\sqrt{\Delta\chi^2}}$	$\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$	$\overline{\sigma_{\text{reel}}(\Delta l)}$
Toutes simus $\mathcal{H}_1$	$3.10\sigma$	$1.87\sigma$	$2.53\sigma$
Simus $\mathcal{H}_1$ avec $\sigma_{\text{reel}} < 4.25\sigma$	$3.06\sigma$	$1.86\sigma$	$2.53\sigma$

NOTES.— Significativités moyennes obtenues sous  $\mathcal{H}_1$  avec la méthode classique  $\sqrt{\Delta\chi^2}$  (avec la convention  $\sqrt{\Delta\chi^2} = 0$  pour  $\Delta\chi^2 < 0$ ), avec notre méthode rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$  et avec notre méthode rigoureuse pour le rapport de vraisemblance généralisé  $\sigma_{\text{reel}}(\Delta l)$ , avec une hypothèse  $\mathcal{H}_0$  *no wiggles*. On montre les moyennes  $\overline{\sqrt{\Delta\chi^2}}$ ,  $\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$  et  $\overline{\sigma_{\text{reel}}(\Delta l)}$  obtenues avec toutes les simulations  $\mathcal{H}_1$  et en se limitant au seuil de significativité de notre méthode  $\sigma_{\text{reel}}(\Delta\chi^2) < 4.25\sigma$ . Ici la méthode classique du  $\chi^2$  surestime grossièrement la significativité. On voit aussi que le rapport de vraisemblance généralisé  $\Delta l$  donne de meilleurs résultats que la statistique  $\Delta\chi^2$ .

TABLE 5.6 –

	$\overline{\sqrt{\Delta\chi^2}}$	$\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$	$\overline{\sigma_{\text{reel}}(\Delta l)}$
Toutes simus $\mathcal{H}_1$	$3.88\sigma$	$1.46\sigma$	$2.71\sigma$
Simus $\mathcal{H}_1$ avec $\sigma_{\text{reel}} < 4.25\sigma$	$3.81\sigma$	$1.45\sigma$	$2.54\sigma$

NOTES.— Significativités moyennes obtenues sous  $\mathcal{H}_1$  avec la méthode classique  $\sqrt{\Delta\chi^2}$  (avec la convention  $\sqrt{\Delta\chi^2} = 0$  pour  $\Delta\chi^2 < 0$ ), avec notre méthode rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$  et avec notre méthode rigoureuse pour le rapport de vraisemblance généralisé  $\sigma_{\text{reel}}(\Delta l)$ , avec une hypothèse  $\mathcal{H}_0$  zéro baryon. On montre les moyennes  $\overline{\sqrt{\Delta\chi^2}}$ ,  $\overline{\sigma_{\text{reel}}(\Delta\chi^2)}$  et  $\overline{\sigma_{\text{reel}}(\Delta l)}$  obtenues avec toutes les simulations  $\mathcal{H}_1$  et en se limitant au seuil de significativité de notre méthode  $\sigma_{\text{reel}}(\Delta\chi^2) < 4.25\sigma$ . Ici la méthode classique du  $\chi^2$  surestime grossièrement la significativité. On voit aussi que le rapport de vraisemblance généralisé  $\Delta l$  donne de meilleurs résultats que la statistique  $\Delta\chi^2$ .

On vérifie que ces conclusions ne sont pas dues à l'imprécision de notre procédure pour les significativités élevées. En effet lorsque l'on considère seulement les réalisations sous la limite  $4.25\sigma$ , notre procédure donne une significativité moyenne de  $2.53\sigma$  avec  $\Delta l$  et de  $1.86\sigma$  pour  $\Delta\chi^2$ , alors que la moyenne de  $\sqrt{\Delta\chi^2}$  est de  $3.06$ .

On effectue les mêmes calculs pour une hypothèse  $\mathcal{H}_0$  zéro baryon et l'on montre nos résultats dans la table 5.6. On obtient une significativité moyenne de  $2.71\sigma$  avec  $\Delta l$  contre  $1.46\sigma$  avec  $\Delta\chi^2$ . La significativité moyenne donnée par la méthode du  $\chi^2$  classique de  $3.88\sigma$  est à nouveau grossièrement fautive. De plus ces résultats ne sont pas dus à l'imprécision de notre méthode aux hautes significativités.

A ce point on peut noter que les significativités moyennes obtenues avec la statistique  $\Delta\chi^2$  sont bien plus faibles que pour une matrice de covariance constante  $C$ . On en déduit que l'approximation de  $C_\theta$  par une matrice constante  $C$  affecte de façon importante la procédure de détection des BAOs.

A nouveau on peut comprendre plus en détails les différences de significativités données par les différentes méthodes. Sur la figure 5.7 on montre la significativité  $\sqrt{\Delta\chi^2}\sigma$  donnée par la méthode classique du  $\chi^2$  en fonction de la significativité rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$ . La sous-figure de gauche montre le cas d'une hypothèse  $\mathcal{H}_0$  *no wiggles* et la sous-figure de droite le cas d'une hypothèse  $\mathcal{H}_0$  zéro baryon. Contrairement au cas d'une matrice de covariance constante dans la section 5.5.1, on voit que la significativité donnée par la méthode classique du  $\chi^2$  est grossièrement fautive. Elle est sous-estimée



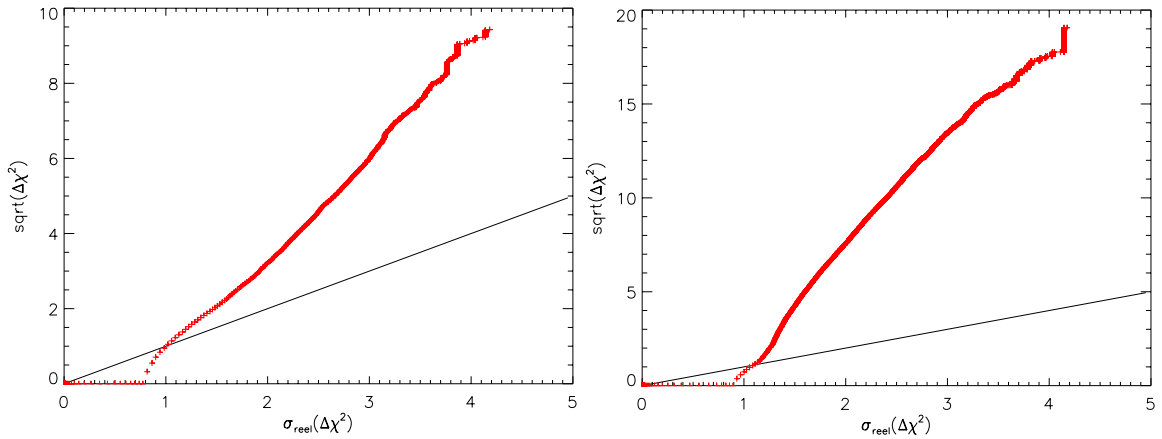


FIGURE 5.7 – A gauche : Significativité de la détection des BAOs obtenue avec la méthode classique  $\sqrt{\Delta\chi^2}$  (avec la convention  $\sqrt{\Delta\chi^2} = 0$  pour  $\Delta\chi^2 < 0$ ) en fonction de la significativité rigoureuse  $\sigma_{\text{reel}}(\Delta\chi^2)$ , avec une hypothèse  $\mathcal{H}_0$  *no wiggles*. La courbe est obtenue pour 50,000 réalisations qui prennent différentes valeurs de  $\Delta\chi^2$ . On voit que pour  $\Delta\chi^2 < 0$  et pour des valeurs faibles de  $\Delta\chi^2$ , la méthode classique sous-estime la significativité. En revanche aux plus grandes valeurs de  $\Delta\chi^2$ , la significativité devient grossièrement surestimée, si bien que la méthode classique est grossièrement fautive dans le cas d’une matrice de covariance  $C_\theta$ . A droite : Même chose pour une hypothèse  $\mathcal{H}_0$  zéro baryon.

pour les faibles significativités, mais devient grossièrement surestimée aux plus hautes significativités (en particulier pour une hypothèse  $\mathcal{H}_0$  zéro baryon). La méthode classique du  $\chi^2$  est donc à proscrire dans le cas d’une matrice de covariance dépendant du modèle  $C_\theta$ .

Enfin on montre sur la figure 5.8 une comparaison de la significativité rigoureuse obtenue avec  $\Delta l$  en fonction de la significativité rigoureuse obtenue avec  $\Delta\chi^2$ . A nouveau, la sous-figure de gauche montre le cas d’une hypothèse  $\mathcal{H}_0$  *no wiggles* alors que la sous-figure de droite montre le cas d’une hypothèse  $\mathcal{H}_0$  zéro baryon. Pour des raisons de visibilité, on montre seulement 1,000 réalisations avec un modèle pris à chaque fois aléatoirement dans tous les modèles  $\mathcal{H}_1$ . On voit que pour la majorité des réalisations, la significativité obtenue est plus grande avec  $\Delta l$  qu’avec  $\Delta\chi^2$ . Ceci justifie notre recommandation de continuer à utiliser un rapport de vraisemblance généralisé dans le cas d’une matrice  $C_\theta$ .

Donnons un exemple concret où l’on obtiendrait une valeur  $\Delta\chi^2 = 25$  qui s’interpréterait comme une détection à un niveau  $5\sigma$  avec la méthode classique du  $\chi^2$ . Pour une hypothèse  $\mathcal{H}_0$  *no wiggles*, on voit avec la figure 5.7 que cela correspond en fait à une vraie significativité égale à  $\sigma_{\text{reel}}(\Delta\chi^2) = 2.66\sigma$ . En revanche si l’on utilise la statistique  $\Delta l$ , une valeur  $\Delta\chi^2 = 25$  correspond en moyenne à une significativité  $\sigma_{\text{reel}}(\Delta l) = 3.51\sigma$  obtenue avec la statistique  $\Delta l$ .

## 5.6 Conclusion

Dans ce chapitre, on s’est focalisé sur différentes questions statistiques concernant la détection des BAOs. On a étudié en détails la méthode classique du  $\chi^2$ , qui est la méthode habituellement utilisée pour la détection des BAOs. On a vu que cette méthode est basée sur des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  de la forme suivante :

$$\begin{aligned}\mathcal{H}_0 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{\text{noBAO},\theta}, C) \\ \mathcal{H}_1 & : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{\text{BAO},\theta}, C)\end{aligned}$$

On a aussi vu que cette méthode faisait aussi l’approximation que les espaces de fonctions de corrélation  $(\xi_{\text{noBAO},\theta})_{\theta \in \Theta}$  et  $(\xi_{\text{BAO},\theta})_{\theta \in \Theta}$  sont des espaces affines.

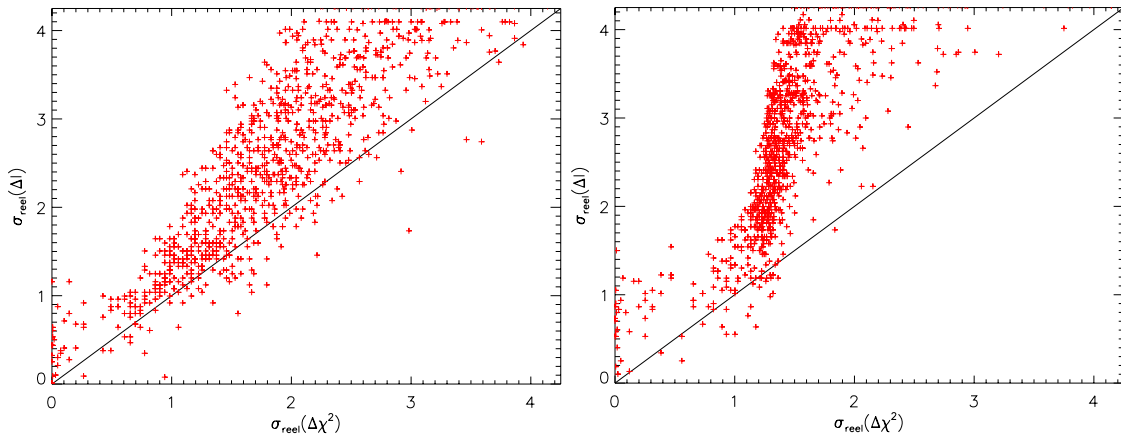


FIGURE 5.8 – A gauche : Significativité de la détection des BAOs obtenue sous  $\mathcal{H}_1$  avec  $\Delta l$  en fonction de la significativité obtenue avec  $\Delta\chi^2$ , avec une hypothèse  $\mathcal{H}_0$  *no wiggles*. Pour des raisons de visibilité, on montre seulement 1,000 réalisations avec un modèle pris à chaque fois aléatoirement dans tous les modèles  $\mathcal{H}_1$ . On voit que pour la majorité des réalisations, la significativité obtenue est plus grande avec  $\Delta l$  qu’avec  $\Delta\chi^2$ . Ceci justifie notre recommandation de continuer à utiliser un rapport de vraisemblance généralisé dans le cas d’une matrice  $C_\theta$ . A droite : Même chose avec une hypothèse  $\mathcal{H}_0$  zéro baryon.

On a étudié la modélisation statistique de  $\hat{\xi}$ . Pour cela, on a utilisé une nouvelle procédure nous permettant d’obtenir des simulations log-normales sur une grille en  $\theta = (\omega_m, \alpha, B)$  en un temps raisonnable. Ces simulations nous ont permis d’obtenir une matrice de covariance réaliste  $C_\theta$  pour  $\hat{\xi}$  en fonction de  $\theta$ . On a vérifié l’absence de biais dans l’estimateur de Landy-Szalay et pour différents modèles cosmologiques utilisées dans les simulations. On a également vérifié la validité de l’hypothèse de Gaussianité sur  $\hat{\xi}$ .

On a ensuite testé la méthode classique du  $\chi^2$  pour différents types d’hypothèses. On a vu que si l’on considère une matrice de covariance constante  $C$  dans les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ , alors la méthode classique du  $\chi^2$  donne une significativité  $\sqrt{\Delta\chi^2}\sigma$  qui est une bonne approximation de la vraie significativité. Cependant l’approximation de  $C_\theta$  par une matrice constante  $C$  est en réalité très mauvaise pour la détection des BAOs. En effet, en utilisant la matrice de covariance réaliste  $C_\theta$  obtenue avec nos simulations, on a vu que les résultats changeaient beaucoup par rapport à une matrice constante  $C$ . On a vu notamment que dans le cas où les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  sont plus réalistes et utilisent la matrice  $C_\theta$ , la méthode classique du  $\chi^2$  donne des niveaux de significativité grossièrement faux.

On a proposé une nouvelle méthode pour estimer la significativité de façon rigoureuse dans tous les cas. Cette méthode peut être assez couteuse en temps de calcul si l’on veut quantifier des hauts niveaux de significativité. Mais ce problème n’est pas si grave, puisqu’il n’intervient que lorsque la significativité est grande et la détection des BAOs déjà sûre. Finalement on a proposé une nouvelle statistique  $\Delta l$  qui reste un rapport de vraisemblance généralisé dans le cas d’une matrice de covariance  $C_\theta$ . On a montré que cette statistique surpasse la statistique habituelle  $\Delta\chi^2$  dans ce cas.



# Chapitre 6

## Effet de $C_\theta$ pour contraindre les paramètres cosmologiques

### Sommaire

---

<b>6.1</b>	<b>Utiliser les BAOs pour contraindre les paramètres cosmologiques</b>	<b>114</b>
6.1.1	Approche basée seulement sur la position du pic BAO	114
6.1.2	Approche basée sur la forme globale de la fonction de corrélation	115
6.1.3	But du travail dans ce chapitre	116
<b>6.2</b>	<b>Modélisation de <math>\hat{\xi}</math> et simulations SDSS DR7-Full</b>	<b>117</b>
6.2.1	Modèles de fonctions de corrélation	117
6.2.2	Simulations log-normales SDSS DR7-Full utilisées	118
<b>6.3</b>	<b>Dépendance de <math>C_\theta</math> par rapport à <math>\omega_m</math>, <math>\alpha</math> et <math>B</math></b>	<b>118</b>
<b>6.4</b>	<b>Effet de <math>C_\theta</math> pour contraindre les paramètres cosmologiques</b>	<b>119</b>
6.4.1	Erreur de modélisation pour la mesure $\hat{\xi}$ du SDSS DR7-Full	122
6.4.2	Quantification de l'erreur de modélisation sur des simulations du SDSS DR7-Full	124
6.4.3	Quantification de l'erreur de modélisation pour les relevés futurs	125
<b>6.5</b>	<b>Conclusion</b>	<b>128</b>

---

La notion d'étalon standard est bien connue dans la vie courante, puisque l'on peut identifier la distance d'un objet à partir de sa taille apparente lorsqu'on connaît sa vraie taille. En cosmologie, le même principe peut être appliqué. En effet, nous observons les objets en espace *redshift*, et si l'on connaît la vraie taille de ces objets on peut contraindre la relation *redshift*-distance pour qu'ils nous apparaissent avec la bonne taille.

A partir du modèle physique décrit dans la section 1.4, on voit que l'échelle  $r_s$  du BAO fournit un étalon standard pour étudier la géométrie de l'univers (Eisenstein et al. 1998; Blake and Glazebrook 2003; Seo and Eisenstein 2005). C'est un étalon standard de choix puisque l'échelle imprimée dans la fonction de corrélation reste fixe en coordonnées comobiles par évolution linéaire (qui ne change que l'amplitude globale du spectre de puissance et donc de la fonction de corrélation, voir section 1.3.2).

Notons que la relation *redshift*-distance à faible *redshift* fournit des informations importantes sur l'énergie noire, qui est dominante aux époques récentes. Pour ces différentes raisons, le rapport du *Dark Energy Task Force* (Albrecht et al. 2006), qui est un groupe de travail conçu pour comparer les différentes approches d'étude de l'énergie sombre, conclut que l'approche basées sur les BAOs est celle qui est la moins soumise à des effets systématiques.

## 6.1 Utiliser les BAOs pour contraindre les paramètres cosmologiques

On montre dans cette section comment on peut utiliser en pratique la propriété d'étalon standard des BAOs afin de contraindre les paramètres cosmologiques (et notamment les paramètres d'énergie sombre).

Notons d'abord que le pic BAO n'est pas un étalon standard au sens strict car il y a une légère dépendance de l'horizon sonique  $r_s$  en fonction des paramètres cosmologiques (voir équation (1.112)). On peut aussi voir cette dépendance sur la figure 1.7 où la position du pic BAO dans la fonction de corrélation varie légèrement en fonction du modèle. Il y a également certaines subtilités à prendre en compte pour une analyse rigoureuse. Un point très important est que la position du maximum local de la fonction de corrélation théorique ne coïncide pas avec l'horizon sonique  $r_s$ , même dans le régime linéaire (Sánchez et al. 2008). Enfin l'évolution non linéaire cause aussi un décalage du pic BAO dans la fonction de corrélation (voir section 1.3.4).

Pour comprendre comment utiliser la propriété d'étalon standard, on rappelle comment on peut étudier la géométrie de l'univers en mesurant la fonction de corrélation (voir section 2.2.5). La cosmologie fiducielle utilisée pour convertir les *redshifts* en distances introduit des distortions du relevé de galaxies, qui sont différentes suivant la ligne de visée et la direction transverse. Ceci se traduit comme une dilatation de la fonction de corrélation bidimensionnelle. Si l'on considère l'espérance de l'estimateur de la fonction de corrélation estimée en coordonnées fiducielles, on obtient :

$$\mathbb{E} \left[ \hat{\xi}(\pi, \sigma) \right] = \xi \left( \frac{H^{fid}(z)}{H(z)} \pi, \frac{D_A(z)}{D_A^{fid}(z)} \sigma \right) \quad (6.1)$$

où  $\xi(\pi, \sigma)$  est la vraie fonction de corrélation en espace *redshift* 2D. Si l'on ne s'intéresse qu'au monopole de la fonction de corrélation  $\xi_0(r)$ , on peut approximer l'effet géométrique comme une simple dilatation de la fonction de corrélation par un facteur  $\alpha = D_V(\langle z \rangle) / D_V^{fid}(\langle z \rangle)$ , où  $\langle z \rangle$  est le *redshift* moyen du relevé et  $D_V$  est l'échelle de dilatation (*dilation scale*) donnée par :

$$D_V(z) = \left[ (1+z)^2 D_A(z)^2 \frac{cz}{H(z)} \right]^{1/3} \propto D_A(z)^{2/3} \frac{1}{H(z)^{1/3}} \quad (6.2)$$

L'espérance du monopole de la fonction de corrélation en coordonnées fiducielles (quasiment toujours notée  $\xi(r)$  et appelée fonction de corrélation, même si cela constitue un abus de langage) est donc reliée à la vraie fonction de corrélation par la relation :

$$\mathbb{E} \left[ \hat{\xi}_0(s) \right] = \xi(\alpha s) \quad (6.3)$$

Montrons maintenant comment la propriété d'étalon standard à travers l'étude de la fonction de corrélation. Il existe principalement 2 types d'approches (des approches équivalentes peuvent être utilisées pour le spectre de puissance) :

- Un premier type d'approche consiste à simplement utiliser la relation entre le pic BAO dans la fonction de corrélation et la taille de l'horizon sonique  $r_s$  (Kazin et al. 2010; Mehta et al. 2012)
- Le second type d'approche consiste à utiliser la forme globale de la fonction de corrélation (incluant implicitement la position du pic BAO) pour contraindre les paramètres cosmologiques (Eisenstein et al. 2005; Sánchez et al. 2009; Beutler et al. 2011; Blake et al. 2011a,b)

### 6.1.1 Approche basée seulement sur la position du pic BAO

Considérons d'abord la première approche où l'on ne s'intéresse qu'à la position du pic BAO dans la fonction de corrélation. Pour montrer le principe général, on suit l'analyse de Kazin et al. (2010) qui se base sur la relation suivante :

$$\frac{D_V(\langle z \rangle)}{r_s} = \frac{D_V^{fid}(\langle z \rangle)}{\gamma_{s_p}^{fid}(\langle z \rangle)} \quad (6.4)$$

avec  $D_V(\langle z \rangle)$ ,  $r_s$  respectivement la vraie échelle de dilatation et la vraie taille de l'horizon sonique,  $D_V^{fid}(\langle z \rangle)$ ,  $s_p^{fid}(\langle z \rangle)$  respectivement l'échelle de dilatation et la position du pic BAO (i.e. le maximum local dans la fonction de corrélation) en coordonnées fiducielles, et  $\gamma$  un facteur de proportionnalité entre  $r_s$  et  $s_p$ . Cela suppose qu'il existe une relation de proportionnalité  $r_s = \gamma s_p$  qui serait valable pour tous les modèles cosmologiques, et qui peut être mesurée sur des simulations. En réalité cette relation n'a pas de raison d'être valable pour tous les modèles, mais elle est valable au premier ordre. Avec cette relation de proportionnalité, la relation (6.4) se réduit simplement à l'hypothèse de dilatation de la fonction de corrélation suivant le paramètre  $\alpha = D_V(\langle z \rangle)/D_V^{fid}(\langle z \rangle)$  :

$$\frac{D_V(\langle z \rangle)}{s_p} = \frac{D_V^{fid}(\langle z \rangle)}{s_p^{fid}(\langle z \rangle)} \quad (6.5)$$

On peut facilement calculer  $D_V^{fid}(\langle z \rangle)$  pour la cosmologie fiducielle choisie.  $r_s$  est bien contraint par les mesures du CMB (Komatsu et al. 2009) et  $s_p^{fid}(\langle z \rangle)$  est mesuré à une incertitude près à partir de la fonction de corrélation estimée en coordonnées fiducielles. On peut donc se servir de la relation (6.4) pour contraindre le paramètre  $D_V(\langle z \rangle)$ .

### 6.1.2 Approche basée sur la forme globale de la fonction de corrélation

Considérons maintenant le second type d'approche qui utilise la forme globale de la fonction de corrélation. Pour cela on fait l'hypothèse suivante pour l'estimateur  $\hat{\xi}$  de la fonction de corrélation :

$$\exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_\theta, C_\theta)$$

i.e. qu'il existe un modèle  $\theta$  correct, tel que l'espérance de la fonction de corrélation estimée est égale à la fonction de corrélation du modèle,  $\mathbb{E}[\hat{\xi}] = \xi_\theta$ , et la matrice de covariance de l'estimateur est égale à  $C_\theta$ .  $\theta$  comprend tous les paramètres dont peut dépendre la fonction de corrélation (e.g. tous les paramètres du modèle  $\Lambda$ CDM).

L'hypothèse de la Gaussianité de  $\hat{\xi}$  est très courante, et on a vérifié qu'elle est justifiée dans la section 5.4.3 (elle a aussi été justifiée par d'autres études, notamment dans Manera et al. (2012)).

On veut contraindre le paramètre  $\theta$  à partir de l'estimateur  $\hat{\xi}$ , et pour cela on doit adopter un point de vue Bayésien et supposer une probabilité a priori  $p(\theta)$ . La postérieure sur  $\theta$  est alors donnée par le théorème de Bayes :

$$p(\theta | \hat{\xi}) \propto p(\theta) \mathcal{L}_\theta(\hat{\xi}) \quad (6.6)$$

avec  $\mathcal{L}_\theta(\hat{\xi})$  la vraisemblance de la mesure pour le modèle  $\theta$ . Avec l'hypothèse Gaussienne, la vraisemblance est donnée par :

$$\mathcal{L}_\theta(\hat{\xi}) \propto |C_\theta|^{-1/2} e^{-\frac{1}{2} \langle \hat{\xi} - \xi_\theta, C_\theta^{-1} (\hat{\xi} - \xi_\theta) \rangle} \quad (6.7)$$

La question du choix de  $p(\theta)$  peut être délicate puisqu'il y a une part de subjectivité. Toutefois comme on s'y attend, plus la mesure  $\hat{\xi}$  est précise et moins le choix de la probabilité a priori  $p(\theta)$  influence la postérieure  $p(\theta | \hat{\xi})$ .

A partir de la postérieure  $p(\theta | \hat{\xi})$ , on peut obtenir les contraintes unidimensionnelles sur les paramètres  $\theta_1, \dots, \theta_n$  de  $\theta = (\theta_1, \dots, \theta_n)$  en marginalisant par rapport aux autres paramètres :

$$p(\theta_i | \hat{\xi}) = \int p(\theta_1, \dots, \theta_{i-1}, \theta_i, \theta_{i+1}, \dots, \theta_n | \hat{\xi}) d\theta_1 \dots d\theta_{i-1} d\theta_{i+1} \dots d\theta_n \quad (6.8)$$

Cette approche basée sur la forme globale de la fonction de corrélation possède à la fois des avantages et des inconvénients par rapport à l'approche basée seulement sur la position du pic BAO. Le principal avantage est qu'elle utilise la totalité de l'information contenue dans l'estimateur de la fonction de corrélation  $\hat{\xi}$  et permet donc d'obtenir des meilleures contraintes. Le principal inconvénient est qu'elle requiert une très bonne modélisation de la fonction de corrélation, notamment des effets non linéaires et du biais dépendant d'échelle des galaxies (voir sections 1.3.4 et 2.2.2)). En effet, des erreurs de modélisation peuvent mener à des contraintes biaisées sur  $\theta$ . Une façon de réduire ce risque est de considérer des échelles  $r$  suffisamment grandes, où l'on est confiant par rapport à la modélisation de ces effets.

Le paramètre  $\theta$  peut avoir une dimension importante (6 paramètres dans un modèle  $\Lambda$ CDM), et il devient alors quasiment impossible de calculer la valeur de  $p(\theta|\hat{\xi})$  sur une grille en  $\theta$ . L'approche habituelle dans ce cas est d'utiliser des méthodes MCMC (pour *Markov Chain Monte Carlo*), qui utilisent des chaînes de Markov dont la distribution à l'équilibre est égale à la probabilité recherchée  $p(\theta|\hat{\xi})$ . Avec un nombre suffisant de chaînes on peut donc avoir un échantillon de points représentatifs de la distribution  $p(\theta|\hat{\xi})$  et en déduire des contraintes unidimensionnelles  $p(\theta_i|\hat{\xi})$ .

Notons finalement que cette approche n'utilise pas de façon explicite la propriété d'étalon standard des BAOs, mais elle l'utilise de façon naturelle à travers la dépendance de  $\xi_\theta$  par rapport au paramètre  $\alpha$ . En gardant les autres paramètres constants,  $\alpha$  induit une dilatation de la fonction de corrélation. Ainsi la position du pic BAO permet de naturellement contraindre ce paramètre  $\alpha$  dans la probabilité postérieure.

Dans la suite on utilisera un modèle simple avec seulement trois paramètres dans  $\xi_\theta$  :  $\theta = (\omega_m, \alpha, B)$ , avec  $\omega_m = \Omega_m h^2$  la densité de matière,  $\alpha$  le paramètre de dilatation et  $B = b^2$  un simple facteur d'amplitude dans la fonction de corrélation. Ceci est nécessaire pour rendre possible notre analyse : estimer une matrice de covariance dépendant du modèle  $C_\theta$  et estimer les erreurs induites dans la postérieure  $p(\theta|\hat{\xi})$  lorsqu'on approxime la matrice de covariance dépendant du modèle  $C_\theta$  par une matrice constante  $C$ .

### 6.1.3 But du travail dans ce chapitre

On a vu dans la section 6.1 qu'une approche classique pour contraindre les paramètres cosmologiques avec les BAOs, est de mesurer la fonction de corrélation et de faire une analyse de vraisemblance  $\mathcal{L}_\theta(\hat{\xi})$ , à partir d'un modèle statistique de  $\hat{\xi}$  en fonction de  $\theta$ . Un modèle habituel est de considérer que  $\hat{\xi}$  est simplement une variable Gaussienne, centrée sur une fonction de corrélation modèle  $\xi_\theta$  avec une matrice de covariance constante  $C$ .

L'hypothèse de Gaussianité a été vérifiée dans la section 5.4.3. Toutefois l'approximation d'une matrice de covariance constante  $C$  n'a pas été beaucoup étudiée, probablement parce qu'il est difficile d'estimer une matrice de covariance dépendant du modèle  $C_\theta$ . En effet la procédure habituelle est de calculer la matrice de covariance empirique sur un grand nombre de simulations :

$$C_{ij} = \frac{1}{N-1} \sum_{k=1}^N [\hat{\xi}_k(r_i) - \bar{\xi}(r_i)][\hat{\xi}_k(r_j) - \bar{\xi}(r_j)] \quad (6.9)$$

$$\bar{\xi} = \frac{1}{N} \sum_{k=1}^N \hat{\xi}_k \quad (6.10)$$

Cette procédure peut déjà être longue pour une seule valeur de  $\theta$ , et elle semble infaisable pour une grille multidimensionnelle de valeurs de  $\theta$ .

Une alternative consiste à utiliser des formules analytiques pour estimer la matrice de covariance de  $\hat{\xi}$  en fonction de  $\theta$ . Une récente tentative a été faite dans [Xu et al. \(2012\)](#). Cette étude commence par calculer analytiquement une matrice de covariance dépendant de  $\theta$  dans le cas d'un champ de galaxies Gaussien. La matrice de covariance est ensuite modifiée pour être en accord avec une matrice de covariance empirique estimée sur des simulations. Cette procédure est très intéressante puisqu'elle fournit une matrice de covariance  $C_\theta$  de façon très rapide. Toutefois elle n'est pas totalement aveugle, puisqu'elle requiert d'ajuster certains paramètres pour être en accord avec les simulations. En outre il n'a pas encore été montré que l'estimation était correcte pour des valeurs de  $\theta$  différentes de la valeur utilisée pour l'ajustement des paramètres.

Dans ce chapitre nous n'allons pas étudier cette question de modéliser de façon analytique la matrice de covariance. On va plutôt regarder si cette modélisation est requise, i.e. si l'approximation habituelle d'une matrice de covariance constante  $C$  au lieu de  $C_\theta$  engendre une erreur de modélisation (e.g. en changeant les régions de confiance).

Pour que notre analyse soit possible on va seulement considérer un paramètre tridimensionnel  $\theta = (\omega_m, \alpha, B)$  comme dans le chapitre 5. On se servira des simulations détaillées dans la section 5.4.2 afin d'obtenir une matrice de covariance  $C_\theta$ . Comme on l'a déjà vu, les paramètres  $\alpha$  et  $B$  peuvent être pris en compte de façon optimisée, i.e. sans ajouter beaucoup de temps de calcul.

On va d'abord étudier de façon précise la dépendance de  $C_\theta$  en fonction des différents paramètres  $\omega_m$ ,  $\alpha$  et  $B$ . On va ensuite regarder l'erreur de modélisation lorsqu'on approxime  $C_\theta$  par une matrice

constante  $C$ , pour la fonction de corrélation estimée du catalogue SDSS DR7-Full. Finalement on va quantifier l'erreur de modélisation sur des simulations du SDSS DR7-Full et sur des simulations correspondant à des relevés futurs.

## 6.2 Modélisation de $\hat{\xi}$ et simulations SDSS DR7-Full

### 6.2.1 Modèles de fonctions de corrélation

On utilise ici la même modélisation de la fonction de corrélation que dans le chapitre 5. La seule différence est que l'on ne considère ici que l'hypothèse  $\mathcal{H}_1$ , i.e. seulement les modèles de fonctions de corrélation avec BAOs. En effet, notre but est d'obtenir des contraintes cosmologiques dans la cadre du modèle cosmologique actuel, et non de tester l'existence des BAOs comme c'était le cas dans le chapitre 5. Résumons brièvement les différentes étapes pour obtenir les modèles de fonctions de corrélation  $\xi_\theta$  avec le paramètre tridimensionnel  $\theta = (\omega_m, \alpha, B)$ .

On considère d'abord les spectres de puissance linéaires  $P_L$  en fonction de  $\omega_m$  (i.e. en fixant les paramètres  $\omega_b = 2.227 \times 10^{-2}$  et  $n_s = 0.966$  aux valeurs du maximum de vraisemblance WMAP7) et avec le choix  $\sigma_8 = 0.81$ .

On corrige ensuite l'évolution non linéaire avec la procédure décrite dans la section 1.3.4. Plus précisément, on applique un lissage Gaussien à la différence entre le spectre linéaire et le spectre *no wiggles* de Eisenstein and Hu (1998),  $P_L(k) - P_{nw,L}(k)$ , et l'on restaure la forme générale du spectre en rajoutant  $P_{nw,L}(k)$  après le lissage :

$$P_{dw,L}(k) = P_{nw,L}(k) + e^{-a^2 k^2 / 2} [P_L(k) - P_{nw,L}(k)] \quad (6.11)$$

Pour prendre en compte les corrections non linéaires invariantes d'échelle, on applique à  $P_{dw,L}$  les corrections non linéaires de HALOFIT correspondant au spectre invariant d'échelle  $P_{nw,L}(k)$  :

$$P_{dw,NL}(k) = \frac{P_{nw,NL}(k)}{P_{nw,L}(k)} P_{dw,L}(k) \quad (6.12)$$

Tous ces spectres sont calculés au *redshift* moyen  $z = 0.3$  du catalogue. La fonction de corrélation de la distribution de galaxies, qui dépend de  $\omega_m$ , est obtenue en fixant la valeur de  $a$  dans la formule (6.11) et en modélisant le biais dépendant d'échelle  $B(r)$  des galaxies de l'échantillon par rapport à la matière :

$$\xi_{galaxies, \omega_m}(r) = B(r) \xi_{dw,NL}(r) \quad (6.13)$$

Pour ce faire, on utilise les catalogues LasDamas, en choisissant les paramètres  $a$  et  $B(r)$  afin d'ajuster la fonction de corrélation LasDamas par rapport à notre modèle.

On modélise ensuite les effets géométriques introduits par la cosmologie fiducielle en introduisant le paramètre  $\alpha = D_V(\langle z \rangle) / D_V^{fid}(\langle z \rangle)$ , où  $\langle z \rangle = 0.3$  est le *redshift* moyen de l'échantillon et  $D_V(z)$  est l'échelle de dilatation.

On introduit finalement un facteur multiplicatif constant  $B = b^2$  et l'on obtient le modèle final de corrélation en fonction de  $\omega_m, \alpha$  et  $B = b^2$  :

$$\xi_{\omega_m, \alpha, B}(r) = b^2 \xi_{galaxies, \omega_m}(\alpha r) \quad (6.14)$$

Il ne reste plus qu'à discrétiser ce modèle continu de façon équivalente au comptage de paires lors de l'estimation de la fonction de corrélation. Pour un intervalle  $[r_i - dr/2, r_i + dr/2]$ , la fonction de corrélation discrétisée est donnée par :

$$\xi_{\omega_m, \alpha, B}(r_i) = \frac{\int_{r_i - dr/2}^{r_i + dr/2} \xi_{\omega_m, \alpha, B}(r) r^2 dr}{\int_{r_i - dr/2}^{r_i + dr/2} r^2 dr} \quad (6.15)$$

Dans toute cette étude, on utilise des intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  entre les distances  $r = 20h^{-1}\text{Mpc}$  et  $r = 200h^{-1}\text{Mpc}$ , ce qui correspond à  $n = 18$  intervalles, i.e. un vecteur  $(\xi_{\omega_m, \alpha, B}(r_i))$  de dimension  $n = 18$ .



### 6.2.2 Simulations log-normales SDSS DR7-Full utilisées

On utilise ici les mêmes simulations log-normales du catalogue DR7-Full que dans le chapitre 5. Ces simulations ont d'abord comme but de vérifier l'absence de biais dans les estimateurs ainsi que leur Gaussianité, i.e. que l'hypothèse suivante est valide :

$$\exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO,\theta}, C_{BAO,\theta})$$

où l'on rappelle la paramétrisation tridimensionnelle  $\theta = (\omega_m, \alpha, B)$  que l'on utilise. Nous avons fait cette vérification dans la section 5.4.3.

Ces simulations ont ensuite servi à estimer la matrice de covariance dépendant du modèle  $C_\theta$ . En effet, notre but dans ce chapitre est de quantifier l'erreur de modélisation due à l'approximation de  $C_\theta$  par une matrice constante  $C$ . Il est donc important de pouvoir bien estimer la matrice de covariance  $C_\theta$ , et de comprendre sa dépendance en fonction de  $\theta$ .

Résumons brièvement notre procédure pour générer des simulations SDSS DR7-Full (voir la section 3.2.2 pour une description de ce catalogue) sur une grille de valeurs de  $\theta$ .

Nos simulations utilisent la même couverture angulaire du ciel et le même densité moyenne de galaxies  $\bar{n}(r)$  que le catalogue SDSS DR7-Full. Pour chaque simulation on utilise l'estimateur de Landy-Szalay avec des poids constants pour toutes les galaxies. Pour des raisons de temps de calcul on estime séparément la fonction de corrélation dans l'hémisphère galactique sud et l'hémisphère galactique nord. On réunit les deux estimateurs (respectivement  $\hat{\xi}_{NGC}$  et  $\hat{\xi}_{SGC}$ ) avec la combinaison linéaire optimale (voir annexe C) :

$$\hat{\xi} = C_\theta \left[ C_{NGC,\theta}^{-1} \hat{\xi}_{NGC} + C_{SGC,\theta}^{-1} \hat{\xi}_{SGC} \right] \quad (6.16)$$

$$C_\theta = \left( C_{NGC,\theta}^{-1} + C_{SGC,\theta}^{-1} \right)^{-1} \quad (6.17)$$

avec  $C_\theta$  la matrice de covariance de l'estimateur complet  $\hat{\xi}$  résultant de ce choix.

Les trois paramètres de  $\theta = (\omega_m, \alpha, B)$  sont pris en compte de façon différente dans les simulations. Pour le paramètre  $\omega_m$ , on n'a pas d'autre choix que de générer de nouveaux ensembles de simulations log-normales pour chaque valeur de  $\omega_m$ . On choisit de considérer 5 valeurs différentes,  $\omega_m = 0.08, 0.105, 0.13, 0.155, 0.18$  et l'on va simplement interpoler linéairement la matrice de covariance pour les valeurs intermédiaires (l'interpolation est faite dans le sens classique, i.e. coefficient par coefficient).

D'un autre côté, le paramètre  $\alpha$  peut être pris en compte de façon moins contraignante avec la procédure détaillée dans la section 3.4.3. On utilise les valeurs minimum  $\alpha_{min} = 0.8$  et maximum  $\alpha_{max} = 1.2$ . Etant donné que  $D_V^{fid}(z = 0.3) = 1180 \text{ Mpc}$ , les valeurs minimum et maximum de  $\alpha$  correspondent à un intervalle  $D_V(0.3) \in [944 \text{ Mpc}, 1416 \text{ Mpc}]$  pour nos simulations.

Finalement il reste à modéliser l'effet de  $B = b^2$ , qui change la vraie distribution en coordonnées comobiles, de la même façon que  $\omega_m$ . Mais parce que c'est simplement un facteur multiplicatif constant  $B$ , cela devrait donner approximativement un facteur multiplicatif  $B^2$  dans la matrice de covariance de  $\hat{\xi}$ . On vérifiera dans la section 6.3 que cette approximation est bonne pour des valeurs raisonnables de  $b$ , et que l'approximation  $C_B \propto B^2 C$  est bien plus proche de la vraie matrice de covariance que l'approximation d'une matrice constante  $C_B = C$ . Ce paramètre peut donc être traité sans avoir besoin de simulations additionnelles.

Notre ensemble principal de simulations est généré avec une valeur  $B = b^2 = 2.5^2$  (notons que l'on considère ici le biais par rapport à la corrélation en espace réel et non en espace *redshift*, i.e. sans le facteur de Kaiser de l'équation (2.30)). On vérifiera la validité de l'approximation  $C_B \propto B^2 C$  en comparant la matrice de covariance obtenue sur l'ensemble avec  $B = 2.5^2$  à la matrice de covariance obtenue sur un ensemble avec  $B = 3.0^2$ .

Pour chaque valeur de  $(\omega_m, \alpha)$  on génère  $N = 2,000$  simulations log-normales pour estimer la matrice de covariance dépendant du modèle  $C_{\omega_m, \alpha}$ , de sorte qu'il y a très peu de bruit dans l'estimation.

### 6.3 Dépendance de $C_\theta$ par rapport à $\omega_m$ , $\alpha$ et $B$

On décrit ici la dépendance de  $C_\theta$  (obtenue avec l'ensemble de simulations log-normales présenté dans la section 6.2.2) par rapport à  $\omega_m$ ,  $\alpha$  et  $B$ . Nous avons différé cette analyse que l'on aurait pu

faire dans le chapitre 5, puisqu'il s'agit du but de ce chapitre de comprendre l'erreur de modélisation induite par l'approximation de  $C_\theta$  par une matrice constante  $C$ , et donc de comprendre la dépendance de  $C_\theta$  par rapport à  $\theta$ .

On vérifie d'abord que la dépendance de  $C_\theta$  par rapport à  $B = b^2$  peut être effectivement approximée comme  $C_{\omega_m, \alpha, B} \propto B^2 C_{\omega_m, \alpha}$ . Pour cela on compare la matrice de covariance  $C_1 = C_{\omega_m, \alpha, B_1}$  avec la matrice de covariance  $C_2 = C_{\omega_m, \alpha, B_2}$ , chacune étant obtenue avec 2,000 simulations log-normales, respectivement avec  $\omega_m = 0.13$ ,  $\alpha = 1$ ,  $B_1 = 2.5^2$  et  $\omega_m = 0.13$ ,  $\alpha = 1$ ,  $B_2 = 3.0^2$ .

On calcule la distance L2 entre  $(B_2/B_1)^2 C_1$  et  $C_2$ , que l'on compare à la distance L2 entre  $C_1$  et  $C_2$  :

$$\frac{\|(B_2/B_1)^2 C_1 - C_2\|_2}{\|C_1 - C_2\|_2} = 0.22 \quad (6.18)$$

On obtient donc que l'approximation  $C_{\omega_m, \alpha, B} \propto B^2 C_{\omega_m, \alpha}$  est 5 fois plus précise que l'approximation d'une matrice de covariance constante, ce qui justifie notre modélisation.

On s'intéresse maintenant à la dépendance de  $C_\theta$  par rapport aux 2 autres paramètres  $\omega_m$  et  $\alpha$ . On commence par analyser la dépendance de  $C_\theta$  par rapport à  $\omega_m$  dans le cas  $\alpha = 1$  et  $B = 2.5^2$ . On montre sur les figures 6.1 et 6.2 les variations de  $C_\theta$  pour  $\omega_m = 0.08, 0.105, 0.13, 0.155, 0.18$ . Par souci de clarté on distingue la matrice de corrélation (i.e. la matrice de covariance normalisée par ses éléments diagonaux)  $\rho_\theta$  de la formule (6.19) et la variance diagonale  $\sigma_\theta = (\sqrt{C_{\theta, ii}})$ , qui donnent une description complète de la matrice de covariance lorsqu'on les combine.

$$\rho_{\theta, ij} = \frac{C_{\theta, ij}}{\sqrt{C_{\theta, ii} C_{\theta, jj}}} = \frac{1}{\sigma_{\theta, i} \sigma_{\theta, j}} C_{\theta, ij} \quad (6.19)$$

On rappelle que la fonction de corrélation possède  $n = 18$  intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  entre  $20h^{-1}\text{Mpc}$  et  $200h^{-1}\text{Mpc}$ . On constate une forte dépendance à la fois de  $\rho_\theta$  et  $\sigma_\theta$  par rapport à  $\omega_m$ .

Finalement on étudie la dépendance de  $C_\theta$  par rapport à  $\alpha$  dans le cas  $\omega_m = 0.13$  and  $B = 2.5^2$ . On montre sur les figures 6.3 and 6.4 les variations de  $C_\theta$  pour  $\alpha = 0.8, 0.9, 1.0, 1.1, 1.2$ , à nouveau en séparant la matrice de corrélation  $\rho_\theta$  et la variance diagonale  $\sigma_\theta$ . On trouve également une dépendance à la fois de  $\rho_\theta$  et  $\sigma_\theta$  par rapport à  $\alpha$ , mais cette dépendance n'est pas aussi forte que pour  $\omega_m$ . Notons que cette conclusion dépend clairement de la plage de variation des paramètres, mais on a considéré ici des variations habituelles.

## 6.4 Effet de $C_\theta$ pour contraindre les paramètres cosmologiques

Pour obtenir des contraintes de paramètres cosmologiques avec les BAOs, on étudie généralement la vraisemblance de la fonction de corrélation globale (Eisenstein et al. 2005; Sánchez et al. 2009; Beutler et al. 2011; Blake et al. 2011a,b) ou du spectre de puissance global (Cole et al. 2005; Tegmark et al. 2006; Padmanabhan et al. 2007; Reid et al. 2010; Ho et al. 2012), bien que certaines études se concentrent uniquement sur la position de pic BAO (Percival et al. 2007, 2010; Kazin et al. 2010; Mehta et al. 2012). Il s'agit des deux types de méthodes que l'on a présenté respectivement dans les section 6.1.2 et 6.1.1.

Ici on va suivre l'approche basée sur la vraisemblance de la fonction de corrélation globale. On suppose que l'hypothèse suivante est correcte, et l'on veut contraindre le paramètre  $\theta$  :

$$\exists \theta \in \Theta \text{ s.t. } \hat{\xi} \sim \mathcal{N}(\xi_\theta, C_\theta)$$

Pour obtenir une information a posteriori sur  $\theta$  (i.e. après la mesure de  $\hat{\xi}$ ), on a besoin d'un point de vue bayésien en supposant une probabilité a priori  $p(\theta)$  (ou probabilité antérieure). Alors la probabilité a posteriori (ou probabilité postérieure) sur  $\theta$  est donnée par le théorème de Bayes :

$$p(\theta | \hat{\xi}) \propto p(\theta) \mathcal{L}_\theta(\hat{\xi}) \quad (6.20)$$

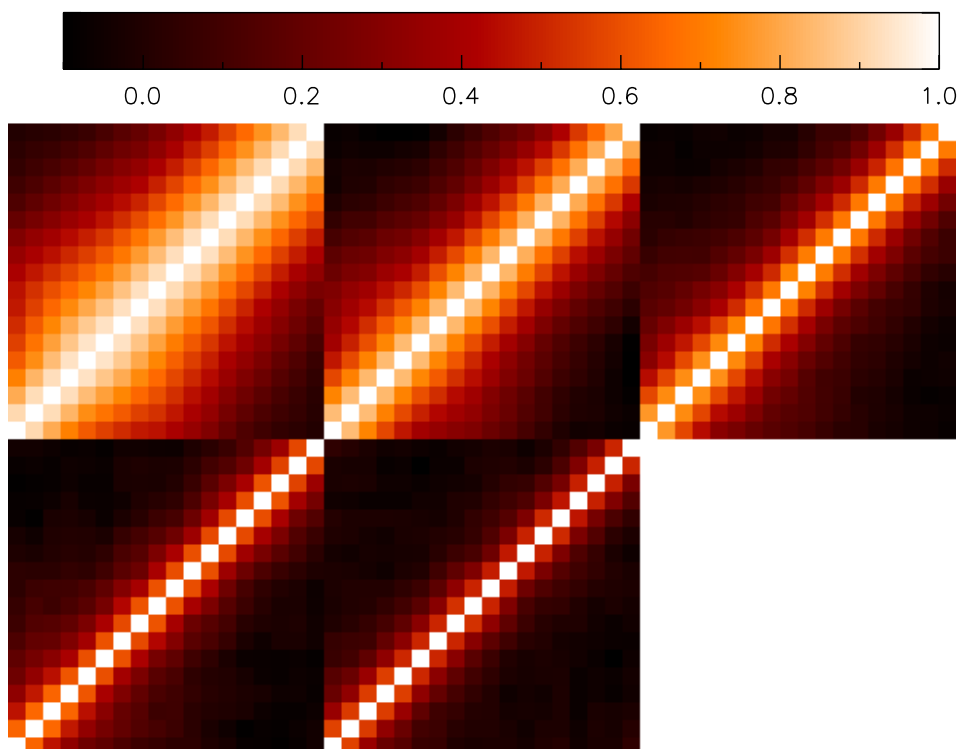


FIGURE 6.1 – Dépendance de  $\rho_\theta$  par rapport à  $\omega_m$  dans le cas  $\alpha = 1$  et  $B = 2.5^2$ . On montre  $\rho_\theta$  pour les valeurs  $\omega_m = 0.08$  (en haut à gauche),  $0.105$  (en haut au milieu),  $0.13$  (en haut à droite),  $0.155$  (en bas à gauche),  $0.18$  (en bas au milieu). La corrélation entre les différents intervalles diminue fortement quand  $\omega_m$  augmente. On montre les  $n = 18$  intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  de  $20h^{-1}\text{Mpc}$  à  $200h^{-1}\text{Mpc}$ .

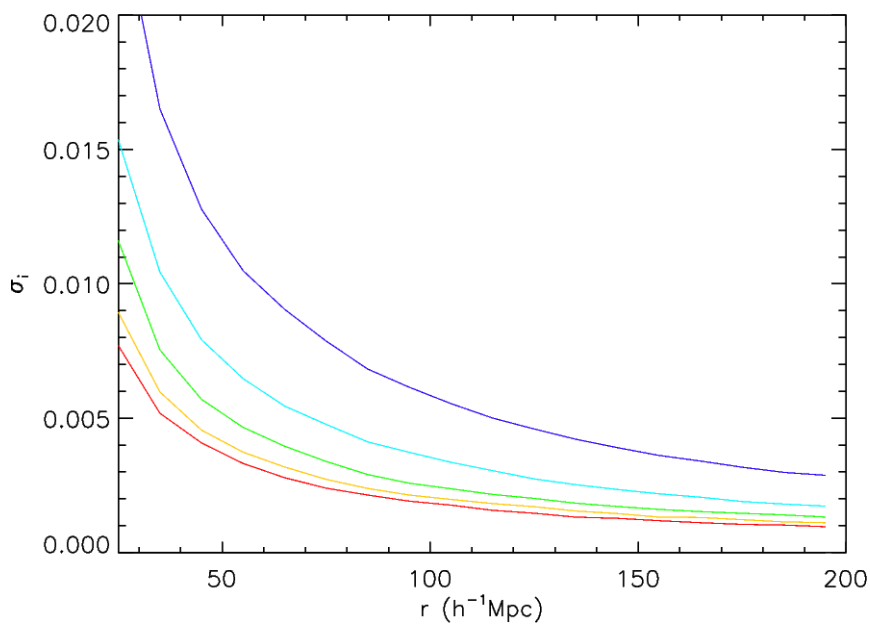


FIGURE 6.2 – Dépendance de  $\sigma_\theta = (\sqrt{C_{\theta,ii}})$  par rapport à  $\omega_m$  dans le cas  $\alpha = 1$  et  $B = 2.5^2$ . On montre  $\sigma_\theta$  pour les valeurs  $\omega_m = 0.08$  (violet),  $0.105$  (bleu clair),  $0.13$  (vert),  $0.155$  (jaune),  $0.18$  (rouge). La variance diagonale diminue fortement quand  $\omega_m$  augmente.

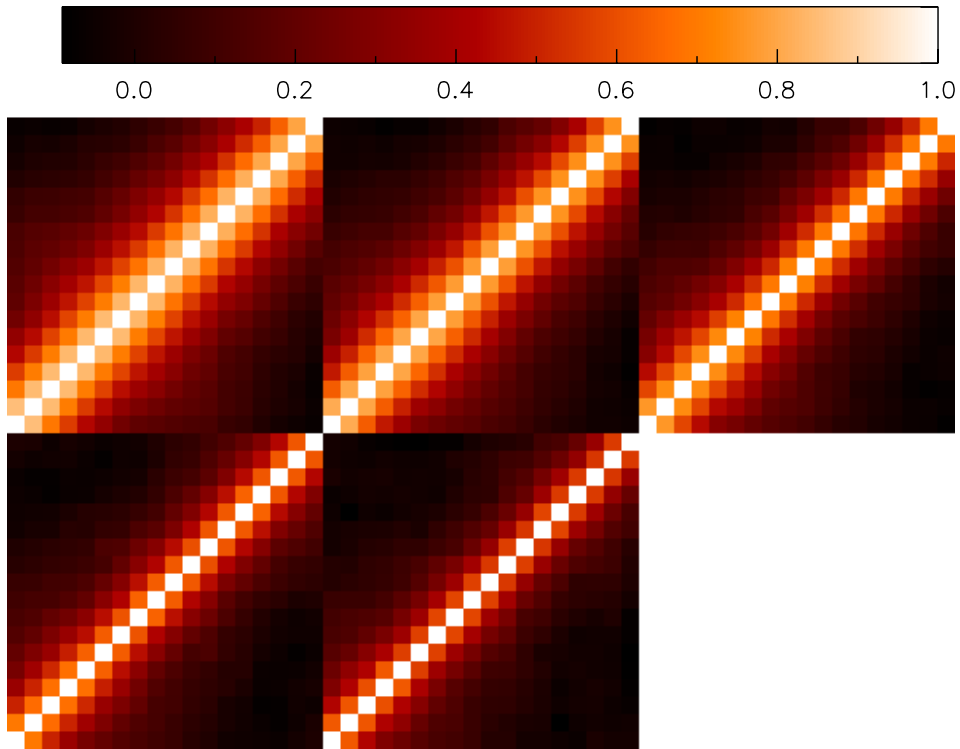


FIGURE 6.3 – Dépendance de  $\rho_\theta$  par rapport à  $\alpha$  dans le cas  $\omega_m = 0.13$  et  $B = 2.5^2$ . On montre  $\rho_\theta$  pour les valeurs  $\alpha = 0.8$  (en haut à gauche),  $0.9$  (en haut au milieu),  $1.0$  (en haut à droite),  $1.1$  (en bas à gauche),  $1.2$  (en bas au milieu). La corrélation entre les différents intervalles diminue quand  $\alpha$  augmente. On montre les  $n = 18$  intervalles de taille  $dr = 10h^{-1}\text{Mpc}$  de  $20h^{-1}\text{Mpc}$  à  $200h^{-1}\text{Mpc}$ .

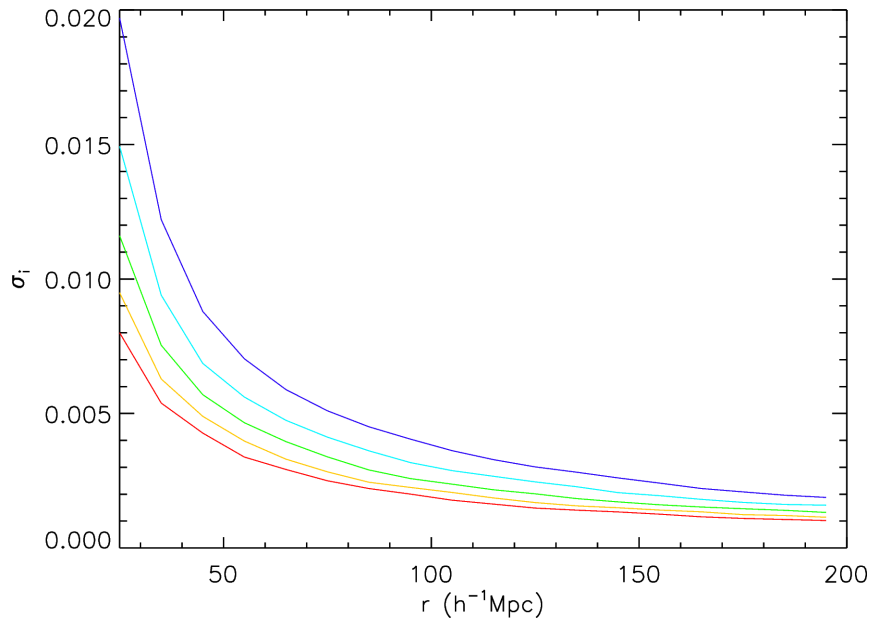


FIGURE 6.4 – Dépendance de  $\sigma_\theta = (\sqrt{C_{\theta,ii}})$  par rapport à  $\alpha$  dans le cas  $\omega_m = 0.13$  et  $B = 2.5^2$ . On montre  $\sigma_\theta$  pour les valeurs  $\alpha = 0.8$  (violet),  $0.9$  (bleu clair),  $1.0$  (vert),  $1.1$  (jaune),  $1.2$  (rouge). La variance diagonale diminue quand  $\alpha$  augmente.

L'association de la mesure  $\hat{\xi}$  avec d'autres observations indépendantes peut être faite 'à l'intérieur' de la probabilité a priori. Par exemple pour combiner les données CMB avec la mesure de  $\xi$  :

$$p(\theta | \text{CMB}, \hat{\xi}) \propto p(\theta, \text{CMB}, \hat{\xi}) \quad (6.21)$$

$$\propto p(\theta, \text{CMB}) p(\hat{\xi} | \theta, \text{CMB}) \quad (6.22)$$

$$\propto p(\theta | \text{CMB}) \mathcal{L}_\theta(\hat{\xi}) \quad (6.23)$$

où l'on a utilisé l'indépendance de  $\hat{\xi}$  avec les données CMB. Ajouter la connaissance des données CMB est donc équivalent à utiliser une probabilité a priori  $p(\theta) = p(\theta | \text{CMB})$ .

Pour contraindre  $\theta$  seulement à partir de la mesure  $\hat{\xi}$ , la question de choisir une probabilité à priori  $p(\theta)$  est délicate. Dans cette étude on prend une probabilité a priori  $p(\theta)$  constante, mais il faut noter que ce choix est arbitraire. Alors la probabilité postérieure est équivalente à la vraisemblance :

$$\mathcal{L}_\theta(\hat{\xi}) \propto |C_\theta|^{-1/2} e^{-\frac{1}{2} \langle \hat{\xi} - \xi_\theta, C_\theta^{-1} (\hat{\xi} - \xi_\theta) \rangle} \quad (6.24)$$

Dans toute la suite on compare la probabilité postérieure obtenue avec notre matrice de covariance dépendant du modèle  $C_\theta$  avec une matrice de covariance constante  $C = C_{\theta_0}$  pour la valeur particulière  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1.0, 2.5^2)$ . On va seulement montrer les probabilités postérieures 2D  $p(\omega_m, D_V(0.3) | \hat{\xi})$  (on rappelle la relation simple entre  $\alpha$  et  $D_V(0.3)$ ,  $\alpha = D_V(0.3)/D_V^{fid}(0.3)$ ), i.e. après marginalisation sur le paramètre  $B = b^2$  :

$$p(\omega_m, D_V(0.3) | \hat{\xi}) = \int_B p(\omega_m, D_V(0.3), B | \hat{\xi}) dB \quad (6.25)$$

On va toujours considérer la grille de paramètres suivantes :  $B \in [4.0, 9.0]$  avec un pas  $dB = 0.01$ ,  $\omega_m \in [0.08, 0.18]$  avec un pas  $d\omega_m = 0.00025$  et  $\alpha \in [0.8, 1.2]$  avec un pas  $d\alpha = 0.001$ . Cette grille en  $\alpha$  correspond à une grille  $D_V(0.3) \in [944 \text{ Mpc}, 1416 \text{ Mpc}]$  avec un pas  $dD_V(0.3) = 1.18 \text{ Mpc}$ . Parce que l'on utilise un modèle simple pour la fonction de corrélation avec seulement 3 paramètres, cette étude n'a pas pour but d'améliorer la précision des contraintes cosmologiques obtenues avec le SDSS DR7. On va simplement chercher à quantifier l'erreur de modélisation due à l'approximation de la matrice de covariance dépendant du modèle  $C_\theta$  par une matrice constante  $C$ .

#### 6.4.1 Erreur de modélisation pour la mesure $\hat{\xi}$ du SDSS DR7-Full

On travaille ici avec la fonction de corrélation estimée  $\hat{\xi}$  sur le SDSS DR7-Full (voir figure 3.7). On montre sur les figures 6.5 et 6.6 les probabilités postérieures  $p(\omega_m, D_V(0.3) | \hat{\xi})$ , respectivement pour une matrice de covariance constante  $C = C_{\theta_0}$  avec  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1.0, 2.5^2)$ , et pour notre matrice de covariance dépendant du modèle  $C_\theta$ .

On peut d'abord remarquer que la postérieure  $p(\omega_m, D_V(0.3) | \hat{\xi})$  est moins régulière et plus bruitée pour notre matrice de covariance dépendant du modèle  $C_\theta$ . Cela peut être facilement expliqué par le bruit sur l'estimation de  $C_\theta$ .

On peut aussi noter que les postérieures bidimensionnelles ne sont pas si bien approximées par des Gaussiennes bidimensionnelles (caractérisées notamment par des contours elliptiques), en particulier dans le cas d'une covariance constante  $C$ . On explique ceci par le comportement de la fonction de corrélation modèle  $\xi_\theta$  aux grands  $\omega_m$  et faibles  $\alpha$  (en bas à droite sur la figure 6.5).

A partir des postérieures bidimensionnelles, on calcule les postérieures unidimensionnelles de  $\omega_m$  et  $D_V(0.3)$ , en marginalisant sur l'autre paramètre. On peut alors calculer les contraintes unidimensionnelles sur chaque paramètre comme l'intervalle de confiance à 68% de probabilité (intervalle à  $1\sigma$ ) autour du maximum de la postérieure unidimensionnelle.

Dans le cas d'une matrice de covariance constante  $C$ , on obtient les contraintes  $\omega_m = 0.144 \pm 0.013$  (précision de 9.0%) et  $D_V(0.3) = 1104 \pm 83 \text{ Mpc}$  (précision de 7.5%). Alors que dans le cas de notre matrice de covariance dépendant du modèle  $C_\theta$ , on obtient les contraintes  $\omega_m = 0.140 \pm 0.011$  (précision de 7.9%) et  $D_V(0.3) = 1114 \pm 74 \text{ Mpc}$  (précision de 6.7%). En terme de  $\alpha$ , cela donne respectivement  $\alpha = 0.935 \pm 0.070$  pour une matrice constante  $C$ , et  $\alpha = 0.944 \pm 0.063$  pour la matrice dépendant du modèle  $C_\theta$ .

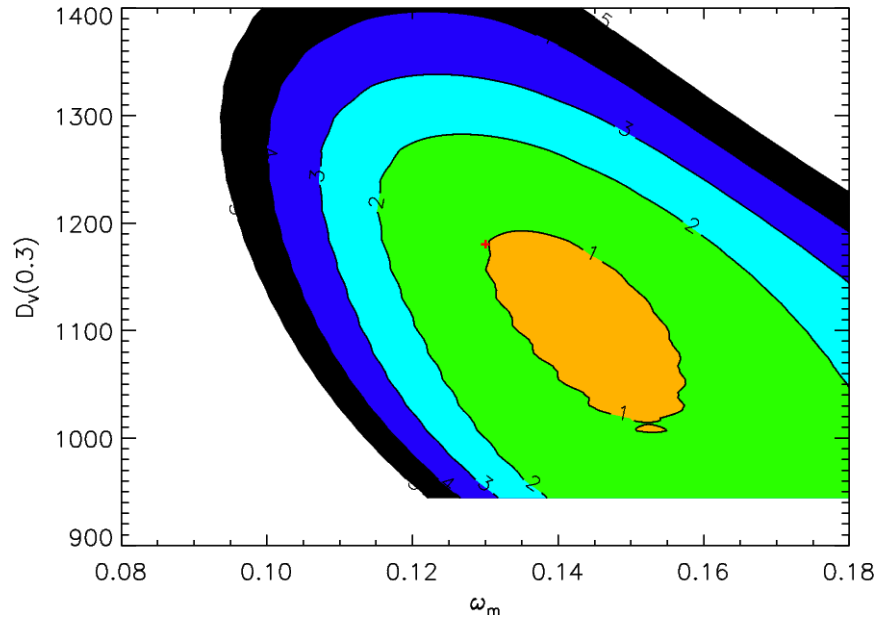


FIGURE 6.5 – Postérieure  $p(\omega_m, D_V(0.3) | \hat{\xi})$  dans le cas d’une matrice de covariance constante  $C = C_{\theta_0}$  avec  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1.0, 2.5^2)$  (position de la croix rouge sur la figure) pour la mesure  $\hat{\xi}$  sur le SDSS DR7-Full. On montre les régions de confiance de  $1\sigma$  à  $5\sigma$  avec l’approximation que la postérieure est un Gaussienne bidimensionnelle. Les contours correspondent à des valeurs  $-2 \ln(p) = -2 \ln(p_{max}) + 2.29, 6.16, 11.81, 19.32, 28.74$  (voir la section ‘Confidence Limits on Estimated Model Parameters’ dans [Press et al. \(2007\)](#)). On obtient les contraintes unidimensionnelles  $\omega_m = 0.144 \pm 0.013$  (précision de 9.0%) et  $D_V(0.3) = 1104 \pm 83$  Mpc (précision de 7.5%).

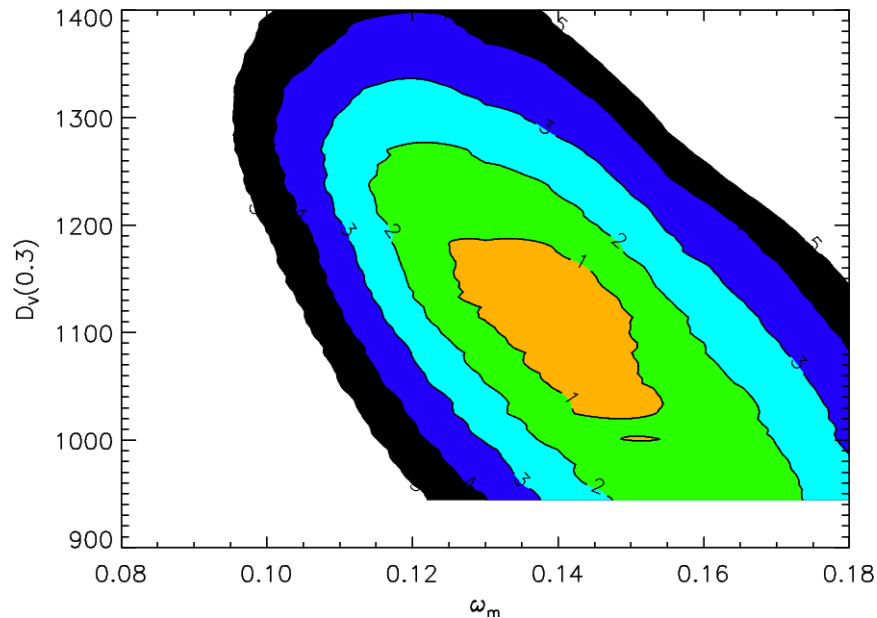


FIGURE 6.6 – Postérieure  $p(\omega_m, D_V(0.3) | \hat{\xi})$  dans le cas de notre matrice de covariance dépendant du modèle  $C_\theta$  pour la mesure  $\hat{\xi}$  sur le SDSS DR7-Full. On obtient les contraintes unidimensionnelles  $\omega_m = 0.140 \pm 0.011$  (précision de 7.9%) et  $D_V(0.3) = 1114 \pm 74$  Mpc (précision de 6.7%). Il y a seulement un petit déplacement de la position du maximum de la postérieure et les régions de confiance deviennent aussi un peu plus petites dans le cas de la matrice de covariance  $C_\theta$ .

Comme on peut le voir en comparant les figures 6.5 et 6.6, l'erreur de modélisation due à l'approximation de la matrice de covariance  $C_\theta$  par une constante  $C$  est assez faible. Il y a seulement un petit déplacement de la position du maximum de la postérieure et les régions de confiance deviennent aussi un peu plus petites dans le cas de la matrice de covariance  $C_\theta$ . On verra toutefois dans la section 6.4.2 que cette réduction de la taille des régions de confiance n'est pas systématique.

### 6.4.2 Quantification de l'erreur de modélisation sur des simulations du SDSS DR7-Full

On étudie finalement l'effet général de l'approximation de  $C_\theta$  par une constante  $C$ , parce que cet effet dépend de la réalisation particulière  $\hat{\xi}$ . On va quantifier cet effet sur les contraintes cosmologiques avec un grand nombre de simulations du SDSS DR7-Full :

$$\hat{\xi} \sim \mathcal{N}(\xi_{\theta_0}, C_{\theta_0}) \quad (6.26)$$

à nouveau avec  $\theta_0 = (\omega_m, \alpha, B) = (0.13, 1.0, 2.5^2)$ . Notons ici que le choix d'un modèle particulier  $\theta_0$  est un peu réducteur. Mais ce choix est raisonnable et doit permettre d'avoir une bonne idée de l'erreur de modélisation.

Pour chaque simulation on calcule la postérieure 2D  $p(\omega_m, \alpha | \hat{\xi})$  dans le cas d'une matrice de covariance constante  $C$  et d'une matrice de covariance  $C_\theta$ . Nous allons nous intéresser à 2 types particuliers d'erreur de modélisation :

- Erreur sur la position des maxima des postérieures,  $\omega_m^{\max}$  et  $\alpha^{\max}$
- Erreur sur la taille des intervalles de confiance à  $1\sigma$ ,  $\sigma_{\omega_m}$  et  $\sigma_\alpha$

On adopte les notations suivantes :

$$\delta\omega_m^{\max} = (\omega_m^{\max, C} - \omega_m^{\max, C_\theta}) / \sigma_{\omega_m}^C \quad (6.27)$$

$$\delta\sigma_{\omega_m} = (\sigma_{\omega_m}^C - \sigma_{\omega_m}^{C_\theta}) / \sigma_{\omega_m}^C \quad (6.28)$$

$$\delta\alpha^{\max} = (\alpha^{\max, C} - \alpha^{\max, C_\theta}) / \sigma_\alpha^C \quad (6.29)$$

$$\delta\sigma_\alpha = (\sigma_\alpha^C - \sigma_\alpha^{C_\theta}) / \sigma_\alpha^C \quad (6.30)$$

On génère 2,000 réalisations suivant le modèle de la formule (6.26) et l'on s'intéresse aux différentes quantités  $\delta\omega_m^{\max}$ ,  $\delta\sigma_{\omega_m}$ ,  $\delta\alpha^{\max}$  et  $\delta\sigma_\alpha$ , qui caractérisent l'erreur de modélisation due à une matrice de covariance incorrecte. Chaque quantité est divisée par la taille de l'intervalle à  $1\sigma$  (l'incertitude statistique) dans les équations 6.27), (6.28), (6.29) and (6.30), pour comparer l'erreur de modélisation à l'incertitude statistique.

On calcule les valeurs moyennes  $\langle \delta\omega_m^{\max} \rangle$ ,  $\langle \delta\sigma_{\omega_m} \rangle$ ,  $\langle \delta\alpha^{\max} \rangle$  et  $\langle \delta\sigma_\alpha \rangle$  pour quantifier un éventuel déplacement systématique de la position du maximum et un changement systématique de la taille des intervalles à  $1\sigma$ . On trouve toutefois que ces moyennes sont négligeables ( $\approx 2\%$ ).

On calcule ensuite les moyennes en valeur absolue  $\langle |\delta\omega_m^{\max}| \rangle$ ,  $\langle |\delta\sigma_{\omega_m}| \rangle$ ,  $\langle |\delta\alpha^{\max}| \rangle$  et  $\langle |\delta\sigma_\alpha| \rangle$ .  $\langle |\delta\omega_m^{\max}| \rangle$  et  $\langle |\delta\alpha^{\max}| \rangle$  correspondent aux erreurs moyennes sur la position du maximum de la postérieure comparées à la taille de l'intervalle à  $1\sigma$ . De même,  $\langle |\delta\sigma_{\omega_m}| \rangle$  et  $\langle |\delta\sigma_\alpha| \rangle$  donnent l'erreur relative moyenne sur la taille de l'intervalle à  $1\sigma$ . On montre les résultats que l'on obtient dans la table 6.1.

Comme on le voit dans la table 6.1, il y a en une erreur de modélisation moyenne de 21% à 28% pour la position du maximum de la postérieure et de 7.5% à 10% pour la taille de l'intervalle à  $1\sigma$ . La position du maximum de la postérieure est donc bien plus affectée par l'erreur de modélisation que l'intervalle à  $1\sigma$ . Cependant l'erreur moyenne reste modérée par rapport à la taille de l'intervalle à  $1\sigma$ .

A partir de la table 6.1, on voit que l'erreur sur les extrémités de l'intervalle à  $1\sigma$  est généralement plus faible que  $21\% + 7.5\% = 28.5\%$  pour  $\omega_m$  et  $28\% + 10\% = 38\%$  pour  $\alpha$ . Une façon de prendre en compte l'erreur de modélisation (bien qu'on ne puisse pas la prendre en compte de façon certaine, puisqu'elle dépend de la réalisation  $\hat{\xi}$ ) est de multiplier la taille de l'intervalle à  $1\sigma$  obtenu avec une matrice constante  $C$  par un facteur  $\approx 1.3$  pour  $\omega_m$  et  $\approx 1.4$  pour  $\alpha$ . Ainsi, les intervalles à  $1\sigma$  étendus vont très probablement recouvrir une grosse partie des vrais intervalles à  $1\sigma$  (i.e. ceux obtenus avec la matrice de covariance dépendant du modèle  $C_\theta$ ).

TABLE 6.1 –

$\langle  \delta\omega_m^{\max}  \rangle$	21%
$\langle  \delta\sigma_{\omega_m}  \rangle$	7.5%
$\langle  \delta\alpha^{\max}  \rangle$	28%
$\langle  \delta\sigma_\alpha  \rangle$	10%

NOTES.—Importance de l’erreur de modélisation comparée à la taille de l’intervalle à  $1\sigma$ , à la fois pour la position du maximum de la postérieure et pour la taille de l’intervalle  $1\sigma$  lui-même. Il y a en moyenne une petite erreur de modélisation, principalement sur la position du maximum de la postérieure. Cependant cette erreur moyenne reste modérée par rapport à la taille de l’intervalle à  $1\sigma$ .

Pour illustrer comment l’erreur de modélisation peut varier en fonction de la réalisation  $\hat{\xi}$ , on montre sur la figure 6.8 pour chaque quantité  $\delta\omega_m^{\max}$ ,  $\delta\sigma_{\omega_m}$ ,  $\delta\alpha^{\max}$  et  $\delta\sigma_\alpha$ , la pdf estimée en utilisant leur histogramme sur 2,000 réalisations. On voit que l’erreur de modélisation dépend clairement de la réalisation  $\hat{\xi}$ .

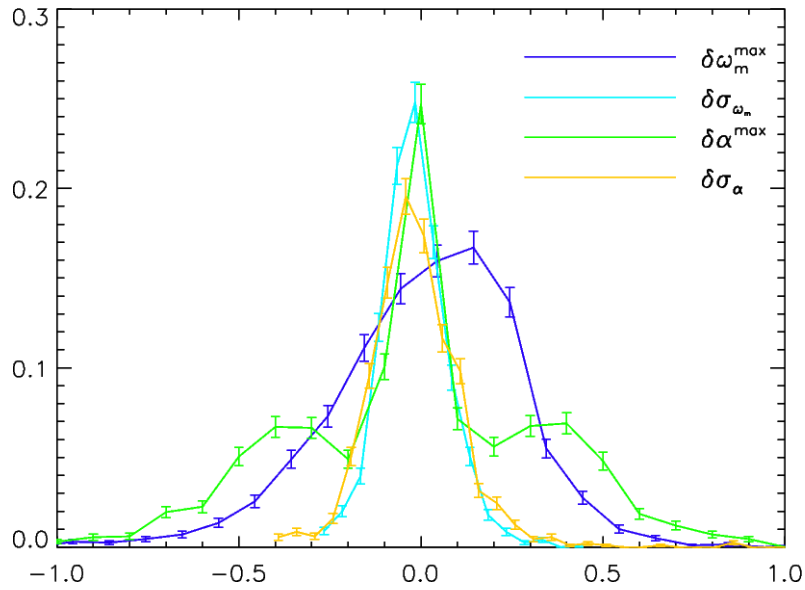


FIGURE 6.7 – pdf estimée de  $\delta\omega_m^{\max}$ ,  $\delta\sigma_{\omega_m}$ ,  $\delta\alpha^{\max}$  et  $\delta\sigma_\alpha$  en utilisant leur histogramme sur 2,000 réalisations de la formule (6.26). Les barres d’erreur donnent l’incertitude de Poisson sur l’estimation de la pdf due au nombre fini de réalisations.

Finalement on fait une étude visuelle des postérieures 2D  $p(\omega_m, D_V(0.3) | \hat{\xi})$ , à la fois dans le cas d’une matrice constante  $C$  et de la matrice dépendant du modèle  $C_\theta$ . Comme dans la section 6.4.1, on constate des déviations par rapport à une Gaussienne bidimensionnelle pour la majorité des réalisations  $\hat{\xi}$ . Ces déviations sont localisées aux grands  $\omega_m$  et aux faibles  $\alpha$ , à la fois dans le cas d’une matrice constante  $C$  et de la matrice dépendant du modèle  $C_\theta$ . Ces déviations sont donc simplement dues au comportement de la fonction de corrélation modèle  $\xi_\theta$  dans cette région.

### 6.4.3 Quantification de l’erreur de modélisation pour les relevés futurs

Finalement on quantifie cette erreur de modélisation pour les prochaines générations de relevés de galaxies. Pour ce faire, on divise simplement les matrices de covariance  $C$  et  $C_\theta$  par un facteur



TABLE 6.2 –

	$c = 2$	$c = 4$
$\langle  \delta\omega_m^{\max}  \rangle$	16%	13%
$\langle  \delta\sigma_{\omega_m}  \rangle$	6.3%	5.1%
$\langle  \delta\alpha^{\max}  \rangle$	23%	20%
$\langle  \delta\sigma_\alpha  \rangle$	8.5%	6.9%

NOTES.—Importance de l’erreur de modélisation comparée à la taille de l’intervalle à  $1\sigma$ , à la fois pour la position du maximum de la postérieure et pour la taille de l’intervalle  $1\sigma$  lui-même, pour des matrices de covariance  $\frac{1}{c}C$  et  $\frac{1}{c}C_\theta$  avec  $c = 2$  et  $c = 4$ . Il y a en moyenne une petite erreur de modélisation, principalement sur la position du maximum de la postérieure. Cependant cette erreur moyenne reste modérée par rapport à la taille de l’intervalle à  $1\sigma$ . De plus l’erreur moyenne est plus faible ici que dans le cas du SDSS DR7-Full, et elle décroît avec la taille du relevé.

constant  $c$  avec  $c = 2$  et  $c = 4$ , et l’on répète la procédure de la section 6.4.2. Pour illustrer ce que cela représente en terme de taille de relevé, considérons la matrice de covariance de  $\xi_{12}$  estimée sur la réunion ’1+2’ de 2 relevés de même taille :

$$C \left[ \hat{\xi}_{12} \right] \approx C \left[ \frac{1}{2}\hat{\xi}_1 + \frac{1}{2}\hat{\xi}_2 \right] \approx \frac{1}{4}C \left[ \hat{\xi}_1 \right] + \frac{1}{4}C \left[ \hat{\xi}_2 \right] \quad (6.31)$$

$$\approx \frac{1}{2}C \left[ \hat{\xi}_1 \right] \quad (6.32)$$

parce que la fonction de corrélation estimée  $\hat{\xi}_{12}$  pour le relevé ’1+2’ est approximativement la moyenne de  $\hat{\xi}_1$  et  $\hat{\xi}_2$  pour des relevés assez grands. Une division de la matrice de covariance par un facteur 2 équivaut donc à un doublement de la taille du relevé. De façon similaire, une division de la matrice de covariance par un facteur 4 équivaut à un quadruplement de la taille du relevé.

On génère désormais des réalisations suivant le modèle suivant :

$$\hat{\xi} \sim \mathcal{N} \left( \xi_{\theta_0}, \frac{1}{c}C_{\theta_0} \right) \quad (6.33)$$

La vraisemblance approximée (avec une matrice de covariance constante  $C$ ), et la vraisemblance réelle (avec la matrice de covariance dépendant du modèle  $C_\theta$ ) sont maintenant données par :

$$\mathcal{L}_\theta^C(\hat{\xi}) \propto e^{-\frac{c}{2}\langle \hat{\xi} - \xi_\theta, C^{-1}(\hat{\xi} - \xi_\theta) \rangle} \quad (6.34)$$

$$\mathcal{L}_\theta^{C_\theta}(\hat{\xi}) \propto |C_\theta|^{-1/2} e^{-\frac{c}{2}\langle \hat{\xi} - \xi_\theta, C_\theta^{-1}(\hat{\xi} - \xi_\theta) \rangle} \quad (6.35)$$

On répète l’analyse de la table 6.1 avec 2,000 réalisations de la formule (6.33), pour chacun des cas  $c = 2$  et  $c = 4$ . On montre les résultats dans la table 6.2.

A partir de la table 6.2, on voit à nouveau qu’il y a en moyenne une petite erreur de modélisation, principalement sur la position du maximum de la postérieure. Cependant cette erreur moyenne reste modérée par rapport à la taille de l’intervalle à  $1\sigma$ . De plus l’erreur moyenne est plus faible ici que dans le cas du SDSS DR7-Full, et elle décroît avec la taille du relevé.

Notre conclusion est que l’erreur de modélisation due à l’approximation de  $C_\theta$  par une matrice constante  $C$  est relativement faible. A nouveau on peut prendre en compte cette erreur de modélisation en multipliant la taille des intervalles à  $1\sigma$  obtenus avec une matrice de covariance constante  $C$  par un facteur  $\approx 1.3$ . On doit toutefois souligner que notre étude n’est pas complète puisque l’on a considéré seulement la dépendance en fonction de trois paramètres :  $\theta = (\omega_m, \alpha, B)$ .

Cette conclusion est un peu surprenante puisque l'on a trouvé une forte dépendance de  $C_\theta$  par rapport à  $\theta$ . Il est toutefois facile de constater qu'il y a deux effets opposés intervenant dans la vraisemblance. On rappelle l'expression de la vraisemblance pour une matrice de covariance  $C_\theta$  :

$$\mathcal{L}_\theta(\hat{\xi}) \propto |C_\theta|^{-1/2} e^{-\frac{1}{2} \langle \hat{\xi} - \xi_\theta, C_\theta^{-1} (\hat{\xi} - \xi_\theta) \rangle} \quad (6.36)$$

Si l'on multiplie par exemple la matrice de covariance par un facteur constant  $c$ , les termes  $|C_\theta|^{-1/2}$  et  $e^{-\frac{1}{2} \langle \hat{\xi} - \xi_\theta, C_\theta^{-1} (\hat{\xi} - \xi_\theta) \rangle}$  vont avoir des effets opposés sur la vraisemblance, ce qui va tendre à diminuer l'effet global sur la vraisemblance. On a notamment vérifié lors de nos tests que le terme  $|C_\theta|^{-1/2}$  a un rôle important : si ce terme est oublié, on observe des variations de la vraisemblance bien plus importantes.

On peut aussi voir cet effet dans un cas particulier. Regardons d'abord l'expression de  $-2 \ln \mathcal{L}_\theta(\hat{\xi})$  :

$$-2 \ln \mathcal{L}_\theta(\hat{\xi}) = \ln |C_\theta| + \langle \hat{\xi} - \xi_\theta, C_\theta^{-1} (\hat{\xi} - \xi_\theta) \rangle + const \quad (6.37)$$

Effectuons maintenant la transformation  $|C_\theta| \rightarrow \frac{1}{c} |C_\theta|$ , notons  $\chi_\theta^2 = \langle \hat{\xi} - \xi_\theta, C_\theta^{-1} (\hat{\xi} - \xi_\theta) \rangle$  et regardons la dépendance de  $-2 \ln \mathcal{L}_\theta(\hat{\xi})$  par rapport à  $c$  :

$$-2 \ln \mathcal{L}_\theta(\hat{\xi}) = \ln \left[ \frac{1}{c^n} |C_\theta| \right] + c \chi_\theta^2 + const \quad (6.38)$$

$$= -n \ln c + c \chi_\theta^2 + const \quad (6.39)$$

avec  $n$  le nombre d'intervalles de discrétisation dans la fonction de corrélation. Prenons le cas particulier de l'étude précédente avec  $n = 18$  et prenons la valeur particulière  $\chi_\theta^2 = 18$ . C'est une valeur raisonnable puisque  $\mathbb{E}[\chi_\theta^2] = n$  si le modèle  $\theta$  est correct. On compare sur la figure 6.8 les courbes  $c \chi_\theta^2$ ,  $-n \ln c$  et  $-n \ln c + c \chi_\theta^2$  en fonction de  $c$ . On voit que les 2 termes de la vraisemblance se compensent nettement, ce qui tend à atténuer les variations de la vraisemblance comparées aux variations de chacun des termes (i.e. comparées aux variations de la matrice de covariance).

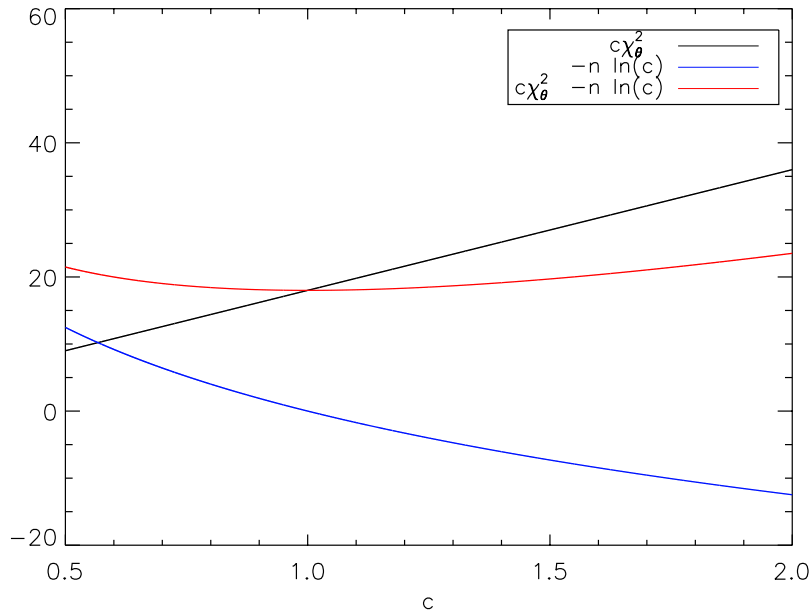


FIGURE 6.8 – Importance des différents termes intervenant dans la log-vraisemblance  $-2 \ln \mathcal{L}_\theta(\hat{\xi})$  lorsque l'on fait varier un paramètre d'amplitude  $c$  dans la matrice de covariance  $\frac{1}{c} C_\theta$ , et dans le cas particulier  $n = 18$  et  $\chi_\theta^2 = 18$ . On voit que les 2 termes de la vraisemblance se compensent nettement, ce qui tend à atténuer les variations de la vraisemblance comparées aux variations de chacun des termes (i.e. comparées aux variations de la matrice de covariance).

## 6.5 Conclusion

L'objectif de ce chapitre était d'étudier l'effet de l'approximation de  $C_\theta$  par une matrice constante  $C$  sur les contraintes cosmologiques.

On a commencé par étudier en détails la dépendance de  $C_\theta$  par rapport à  $\omega_m$ ,  $\alpha$  et  $B = b^2$ . On a vu que le paramètre d'amplitude  $B$  peut être bien approximé comme un facteur  $B^2$  dans la matrice de covariance (pour  $b$  assez grand, i.e.  $b > 2$ ). Pour les deux autres paramètres  $\omega_m$  et  $\alpha$ , on a vu que leurs variations affecte la globalité de la matrice de covariance de façon non triviale. On a aussi vu que le paramètre  $\omega_m$  a un effet plus important que  $\alpha$  pour les plages de variations habituellement considérées.

On s'est ensuite focalisé sur l'effet de l'approximation de  $C_\theta$  par une matrice constante  $C$  sur les contraintes cosmologiques. Plus précisément, on a comparé les contraintes obtenues avec la matrice  $C_\theta$  par rapport aux contraintes obtenues avec l'approximation d'une matrice constante  $C = C_{\theta_0}$  pour la valeur particulière  $\theta_0 = (\omega_m, \alpha, b^2) = (0.13, 1.0, 2.5^2)$ .

Pour la fonction de corrélation estimée du SDSS DR7-Full, on obtient les contraintes  $\omega_m = 0.145 \pm 0.016$  (10.8% de précision) et  $D_V(0.3) = 1104 \pm 105$  Mpc (9.5% de précision) pour une matrice  $C$ , alors qu'on obtient  $\omega_m = 0.140 \pm 0.011$  (7.9% de précision) et  $D_V(0.3) = 1114 \pm 74$  Mpc (6.7% de précision) pour notre matrice  $C_\theta$ . L'approximation d'une matrice de covariance constante cause donc un léger décalage du maximum de la postérieure, et un léger élargissement des intervalles de confiance à  $1\sigma$ .

Toutefois c'est effet n'est pas systématique et dépend de la réalisation particulière  $\hat{\xi}$ . Autrement dit, l'approximation de  $C_\theta$  par une matrice constante  $C$  crée une erreur de modélisation à la fois sur la position du maximum de la postérieure et sur la taille des intervalles à  $1\sigma$ , qui dépend de la réalisation  $\hat{\xi}$ . On a quantifié cette erreur de modélisation en utilisant de nombreuses simulations du catalogue SDSS DR7-Full :

$$\hat{\xi} \sim \mathcal{N}(\xi_{\theta_0}, C_{\theta_0}) \quad (6.40)$$

Pour les deux paramètres,  $\omega_m$  et  $\alpha$ , on a étudié l'erreur sur la position du maximum de la postérieure et sur la taille de l'intervalle  $1\sigma$ . On a trouvé qu'il y a en moyenne une erreur de modélisation sur la position du maximum de vraisemblance égale à 20 – 30% de la taille des intervalles  $1\sigma$ . L'erreur sur la taille des intervalles  $1\sigma$  est plus faible et est en moyenne approximativement de 10%.

Finalement on a fait la même analyse pour des catalogues futurs, simplement en divisant les matrices de covariance par un facteur  $c$ , avec  $c = 2$  et  $c = 4$  :

$$\hat{\xi} \sim \mathcal{N}\left(\xi_{\theta_0}, \frac{1}{c} C_{\theta_0}\right) \quad (6.41)$$

On a également trouvé une assez faible erreur en moyenne, à la fois sur la position du maximum de la postérieure et sur la taille des intervalles  $1\sigma$ . On a vu que cette erreur de modélisation diminue quand la taille des catalogues augmente. Plus précisément, pour des catalogues d'environ 4 fois la taille du SDSS DR7-Full, l'erreur de modélisation moyenne sur la position du maximum de la postérieure atteint  $\approx 20\%$  de la taille des intervalles  $1\sigma$ , alors que l'erreur de modélisation sur la taille des intervalles  $1\sigma$  atteint  $\approx 6\%$ .

Notre conclusion est donc que l'erreur de modélisation due à l'approximation de  $C_\theta$  par une matrice constante  $C$  est relativement faible. Toutefois pour que l'analyse soit plus sûre, nous recommandons d'élargir d'un facteur  $\approx 1.3$  les intervalles à  $1\sigma$  obtenus avec une matrice de covariance constante  $C$ . De cette façon on est quasiment sûr que les intervalles obtenus recouvrent en grosse partie les vrais intervalles à  $1\sigma$ .

Cette conclusion peut sembler surprenante puisque l'on a constaté une forte dépendance de  $C_\theta$  par rapport à  $\theta$ . Toutefois il y a un effet de compensation dans la vraisemblance  $\mathcal{L}_\theta(\xi)$  qui tend à atténuer les variations d'amplitude de la covariance  $C_\theta$ .

L'estimation d'une matrice  $C_\theta$  avec un paramètre  $\theta$  de plus grande dimension semble très difficile. L'approche proposée par Xu et al. (2012) d'une modélisation semi-analytique de  $C_\theta$  semble être une possible alternative. Toutefois cette approche requiert un ajustement de certains paramètres de façon *ad hoc*. Afin de faire l'ajustement de ces paramètres, nos simulations pour un paramètre  $\theta$  tridimensionnel pourraient être très utiles. Une telle analyse permettrait de voir si nos conclusions sont toujours valables lorsque l'on prend en compte les variations de l'ensemble des paramètres cosmologiques.

# Conclusion

L'étude des grandes structures de l'univers fournit des informations cruciales concernant leur formation et leur évolution, qui peuvent être utilisées pour tester les modèles cosmologiques. Le bon accord trouvé entre les observations et le modèle standard  $\Lambda$ CDM permet d'espérer que ce modèle soit une fondation durable pour la cosmologie moderne. En outre, ces observations ont pu être utilisées pour faire de la cosmologie de précision, i.e. pour contraindre les paramètres du modèle standard.

Les grandes structures de l'univers sont principalement étudiées à partir des relevés de galaxies, qui permettent de cartographier de très grands volumes de l'univers. Les grands relevés de galaxies récents comme le SDSS ou le 2dF ont notamment permis la première détection des BAOs dans les grandes structures, ouvrant un domaine très prometteur dans la cosmologie de précision.

Les BAOs peuvent être étudiés à partir des statistiques du second ordre (la fonction de corrélation ou le spectre de puissance) dans la distribution des galaxies. Dans cette thèse, on s'est restreint au cas de la fonction de corrélation. Les BAOs apparaissent dans la fonction de corrélation comme un faible pic à l'échelle de l'horizon sonique  $r_s$ . Il y a deux applications principales des BAOs : la détection des BAOs et la contraintes des paramètres cosmologiques grâce à la propriété d'étalon standard. La détection du pic BAO à une échelle proche de l'échelle attendue permet de confirmer le modèle cosmologique actuel. Quant à la contrainte des paramètres cosmologiques, permettant d'étudier l'énergie sombre, il s'agit d'un objectif majeur de la cosmologie moderne.

Les relevés de galaxies futurs (e.g. BOSS, DES, BigBOSS, EUCLID) vont cartographier des volumes de plus en plus grands et continuer à améliorer notre compréhension des modèles cosmologiques. Alors que les observations vont augmenter en taille et devenir plus précises, il va devenir très important de les analyser rigoureusement. En particulier, certains effets systématiques, qui étaient auparavant négligeables devant l'incertitude statistique, devront être pris en compte.

Des effets systématiques peuvent notamment venir d'une modélisation statistique incorrecte des données. Dans le cas de la fonction de corrélation des galaxies, et notamment des BAOs, on modélise généralement la fonction de corrélation estimée  $\hat{\xi}$  par une variable Gaussienne centrée sur une fonction de corrélation modèle  $\xi_\theta$  avec une matrice de covariance  $C_\theta$ , généralement approximée par une matrice constante  $C$  (i.e. indépendante de  $\theta$ ) :

$$\exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_\theta, C) \quad (6.42)$$

On a également besoin d'une telle modélisation pour la détection des BAOs, à la fois pour l'hypothèse  $\mathcal{H}_0$  sans BAOs et pour l'hypothèse  $\mathcal{H}_1$  avec BAOs :

$$\mathcal{H}_0 : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{noBAO, \theta}, C) \quad (6.43)$$

$$\mathcal{H}_1 : \exists \theta \in \Theta \text{ t.q. } \hat{\xi} \sim \mathcal{N}(\xi_{BAO, \theta}, C) \quad (6.44)$$

Dans cette thèse, on s'est intéressé à la validité de cette modélisation statistique, l'effet des approximations qu'elle comporte, et les possibilités pour l'améliorer. Pour ce faire, nous avons utilisé les données et des simulations provenant du relevé SDSS-II. En particulier nous avons développé une nouvelle procédure pour générer des simulations log-normales sur une grille en  $\theta = (\Omega_m h^2, \alpha, B)$ , qui sont les trois paramètres avec la plus grande influence sur la fonction de corrélation. Cette procédure nous a notamment permis d'estimer une matrice de covariance réaliste  $C_\theta$  en fonction de  $\theta$ .

On récapitule les principales conclusions des études menées dans les chapitres 4, 5 et 6 :

- ★ Dans le chapitre 4, nous avons étudié l'effet de l'incertitude sur la moyenne  $\bar{n}$  du champ de galaxies sur les estimateurs  $\hat{\xi}$ , et notamment le biais connu sous le nom de contrainte intégrale

qui en résulte. On a montré que ce biais est très faible pour les relevés de galaxies actuels et qu'il devient d'autant plus faible que la taille des relevés augmente. Ce résultat montre que le modèle  $\Lambda$ CDM est testable puisqu'il conduit à une estimation non biaisée de la fonction de corrélation. Nous avons également pu comparer la qualité des différents estimateurs de la fonction de corrélation. En accord avec les études précédentes, nous avons constaté une supériorité des estimateurs de Landy-Szalay et de Hamilton, à la fois en terme de variance et de biais.

- ★ Au cours des chapitres 5 et 6, nous avons continué à étudier la modélisation statistique de  $\hat{\xi}$ . Dans le chapitre 5, nous avons notamment vérifié à nouveau l'absence de biais dans l'estimateur de Landy-Szalay et pour différents modèles cosmologiques utilisées dans les simulations. Nous avons également vérifié la validité de l'hypothèse de Gaussianité sur  $\hat{\xi}$ . La dernière étape pour étudier la validité de la modélisation statistique de  $\hat{\xi}$  des formules (6.42), (6.43) et (6.44) est donc de regarder la validité de l'approximation de  $C_\theta$  par une matrice constante  $C$ . Pour cela, nous avons étudié dans le chapitre 6 la dépendance de  $C_\theta$  en fonction des différents paramètres de  $\theta$ . On a trouvé que le paramètre d'amplitude  $B = b^2$  dans la fonction de corrélation peut être bien approximé par un facteur  $B^2$  dans la matrice de covariance. Nous avons aussi trouvé que les variations des deux autres paramètres,  $\Omega_m h^2$  et  $\alpha$  ont un effet significatif sur la matrice de covariance  $C_\theta$ . La conclusion est donc que l'approximation de  $C_\theta$  par une matrice constante  $C$  peut causer des effets systématiques importants.
- ★ Dans le chapitre 5, nous avons étudié différentes questions statistiques concernant la détection des BAOs, et notamment l'effet de cette approximation de  $C_\theta$  par une matrice constante  $C$ . On a vu que la méthode classique du  $\chi^2$  pour la détection des BAOs requiert cette forte approximation. En effet dans le cas d'hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  avec une matrice de covariance constante  $C$ , la significativité de la détection des BAOs donnée par la méthode classique est une bonne approximation de la significativité réelle. Toutefois lorsque l'on considère des hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$  plus réalistes avec la matrice de covariance  $C_\theta$ , la méthode classique donne des niveaux de significativité grossièrement faux. Cela veut dire que l'approximation de  $C_\theta$  par une constante  $C$  affecte de façon très importante la détection des BAOs.

Nous avons proposé deux modifications à la méthode classique du  $\chi^2$  : une nouvelle méthode pour estimer de façon rigoureuse la significativité dans tous les cas, et l'utilisation d'une nouvelle statistique  $\Delta l$  qui reste un rapport de vraisemblance généralisé dans le cas d'une matrice de covariance  $C_\theta$ . Nous avons montré que cette statistique surpasse la statistique habituelle  $\Delta\chi^2$  dans ce cas.

- ★ Dans le chapitre 6, nous avons à nouveau étudié l'effet de l'approximation de  $C_\theta$  par une matrice constante  $C$ , mais cette fois sur les contraintes cosmologiques. On a vu que cette approximation cause un léger décalage du maximum de la postérieure et un léger élargissement des régions de confiance pour la fonction de corrélation estimée  $\hat{\xi}$  du catalogue SDSS DR7-Full. L'erreur de modélisation due à l'approximation de  $C_\theta$  par une matrice constante  $C$  dépend en réalité de la réalisation  $\hat{\xi}$ . Nous avons quantifié cette erreur avec de nombreuses simulations. Nous avons trouvé une erreur moyenne sur la position du maximum de la postérieure égale à  $\approx 20 - 30\%$  de la taille des intervalles  $1\sigma$ . L'erreur sur la taille des intervalles  $1\sigma$  est plus faible, et est en moyenne  $\approx 10\%$ . De plus, cette erreur tend à diminuer par rapport à la taille des intervalles  $1\sigma$  lorsque les relevés grandissent.

Notre conclusion est donc que l'erreur de modélisation due à l'approximation de  $C_\theta$  par une matrice constante  $C$  est relativement faible. Toutefois pour que l'analyse soit plus sûre, nous recommandons d'élargir d'un facteur  $\approx 1.3$  les intervalles à  $1\sigma$  obtenus avec une matrice de covariance constante  $C$ . De cette façon on est quasiment sûr que les intervalles à  $1\sigma$  obtenus recouvrent en grosse partie les vrais intervalles à  $1\sigma$ .

\*  
\* \*

Dans notre étude, nous avons mis en lumière un effet systématique important, provenant de l'ap-

proximation de la matrice  $C_\theta$  par une matrice constante  $C$ . Voyons quelles sont les solutions possibles pour prendre en compte cet effet, à la fois pour la détection des BAOs et pour les contraintes cosmologiques. Notons que la question de la détection des BAOs est moins critique, puisque l'existence de cette caractéristique devient de plus en plus difficile à remettre en question, après sa détection dans de nombreux relevés de galaxies (SDSS, 2dF, WiggleZ, 6dF, BOSS). Mais la détection des BAOs reste tout de même un moyen intéressant pour établir un niveau de détection (i.e. un niveau de confirmation du modèle standard) pour un relevé de galaxies donné.

Pour la détection des BAOs, nous avons vu que les niveaux de détection données par la méthode classique du  $\chi^2$  n'étaient pas fiables, en particulier pour les hauts niveaux de détection attendus pour les relevés de galaxies futurs (e.g. BOSS, DES, BigBOSS, EUCLID). Pour que les niveaux de détection aient un sens, il faudra que ces études considèrent une matrice de covariance dépendant du modèle  $C_\theta$ , et appliquent une procédure rigoureuse pour calculer la significativité. Nous soulignons à nouveau que la statistique  $\Delta l$  que l'on a introduit permet d'obtenir des niveaux de significativité à la fois rigoureux et plus élevés que  $\Delta\chi^2$ .

En ce qui concerne la contrainte des paramètres cosmologiques, nous voyons deux possibilités pour prendre en compte cet effet systématique. Une première possibilité est de considérer effectivement la matrice de covariance dépendant du modèle  $C_\theta$  au lieu de l'approximation d'une constante  $C$ . La deuxième possibilité est de simplement étendre les intervalles à  $1\sigma$  obtenus avec une matrice de covariance constante  $C$ . En effet, nous avons vu sur nos simulations du chapitre 6 qu'en étendant les intervalles à  $1\sigma$  par un facteur  $\approx 1.3$ , on était quasiment sûrs que les intervalles à  $1\sigma$  obtenus recouvraient en grande partie les vrais intervalles à  $1\sigma$ .

Il reste toutefois du travail pour que ces méthodologies puissent être appliquées de façon rigoureuse sur de vraies données. D'abord il paraît impossible d'estimer une matrice de covariance  $C_\theta$  avec des simulations sur une grille en  $\theta$ , qui tienne compte de tous les paramètres cosmologiques envisagés (e.g. pour un modèle  $\Lambda$ CDM). En effet, nous nous sommes arrêtés à une simple paramétrisation tridimensionnelle pour estimer  $C_\theta$ .

L'approche proposée par [Xu et al. \(2012\)](#) d'une modélisation semi-analytique de  $C_\theta$  serait une alternative possible. Toutefois cette approche requiert encore du travail. En ajustant de façon *ad hoc* certains paramètres, cette approche permet de bien approximer la matrice de covariance pour un modèle donné. Mais il n'a pas encore été montré que la matrice de covariance est également bien approximée pour des modèles différents du modèle utilisé pour l'ajustement.

Une seconde approche consisterait à ne considérer la dépendance de  $C_\theta$  que par rapport aux principaux paramètres affectant la fonction de corrélation. De cette façon on pourrait être ramené à des simulations sur une grille tridimensionnelle ou quadri-dimensionnelle, ce qui reste envisageable. Notons que cela donnerait forcément des résultats plus précis que l'approximation d'une matrice de covariance constante  $C$ , à la fois pour la détection des BAOs et pour les contraintes de paramètres cosmologiques.

Concernant l'approche d'extension des intervalles à  $1\sigma$  obtenus avec une matrice constante  $C$ , il reste aussi du travail à effectuer, notamment pour trouver le facteur d'extension. En effet, nous avons vu que ce facteur diminue quand la taille des relevés augmente. Nous pensons qu'il existe une valeur asymptotique pour ce paramètre, qui pourrait être trouvé de façon analytique à partir de la dépendance de  $C_\theta$  par rapport à  $\theta$ . En tout état de cause, il faudrait faire cet étalonnage en considérant pour  $\theta$  les variations de tous les paramètres cosmologiques envisagés.

Enfin signalons qu'il serait aussi intéressant d'étudier cet effet systématique dans les contraintes cosmologiques pour d'autres types d'observations, comme par exemple pour les effets de lentille gravitationnelle faibles.



## Annexe A

# Espérance et covariance d'un champ log-normal

### A.1 Espérance

Si  $X$  est une variable aléatoire Gaussienne d'espérance  $\mu$  et de variance  $\sigma_X^2$ , on obtient pour  $Y = e^X$  :

$$\mathbb{E}[Y] = e^\mu \mathbb{E}[e^{X-\mu}] \quad (\text{A.1})$$

Or pour une variable Gaussienne centrée comme l'est  $X - \mu$ , on a :

$$\mathbb{E}[e^X] = \int \frac{1}{\sqrt{2\pi\sigma_X}} e^x e^{-\frac{x^2}{2\sigma_X^2}} dx \quad (\text{A.2})$$

$$= \frac{1}{\sqrt{2\pi\sigma_X}} \int e^{-\frac{(x-\sigma_X^2)^2}{2\sigma_X^2}} e^{\frac{\sigma_X^2}{2}} dx \quad (\text{A.3})$$

$$= e^{\frac{\sigma_X^2}{2}} \quad (\text{A.4})$$

d'où :

$$\mathbb{E}[Y] = e^\mu \mathbb{E}[e^{X-\mu}] = e^\mu e^{\frac{\sigma_X^2}{2}} \quad (\text{A.5})$$

Dans le cas où le champ Gaussien est d'espérance  $\mathbb{E}[X] = 0$  et de variance  $\sigma_X^2$ , on obtient un champ log-normal d'espérance 1 en normalisant de la façon suivante :

$$Y = e^{X - \frac{\sigma_X^2}{2}} \quad (\text{A.6})$$

### A.2 Covariance du champ log-normal

Soit  $Y$  un champ log-normal défini à partir d'un champ Gaussien  $X$  par la relation  $Y = e^{X - \frac{\sigma_X^2}{2}}$ . Pour 2 points  $i$  et  $j$  on veut calculer la covariance entre  $Y_i$  et  $Y_j$  qui est aussi la fonction de corrélation à 2 points du champ  $Y$  dans ce cas où  $\mathbb{E}[Y] = 1$  (voir section 2.1.2) :

$$\mathbb{E}[Y_i Y_j] = e^{-\sigma_X^2} \mathbb{E}[e^{X_i} e^{X_j}] \quad (\text{A.7})$$

On peut écrire :

$$X_j = \rho(X_i, X_j) \frac{\sigma(X_j)}{\sigma(X_i)} X_i + \sqrt{1 - \rho(X_i, X_j)^2} \tilde{X}_j \quad (\text{A.8})$$

$$= \rho(X_i, X_j) X_i + \sqrt{1 - \rho(X_i, X_j)^2} \tilde{X}_j \quad (\text{A.9})$$

avec  $\tilde{X}_j$  indépendant de  $X_i$ ,  $\sigma(\tilde{X}_j) = \sigma(X_j)$ , et  $\rho(X_i, X_j)$  la corrélation mathématique entre les 2 variables  $X_i$  et  $X_j$  :

$$\rho(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sigma(X_i)\sigma(X_j)} = \frac{\text{Cov}(X_i, X_j)}{\sigma_X^2} \quad (\text{A.10})$$



Donc on a :

$$\mathbb{E} [e^{X_i} e^{X_j}] = \mathbb{E} [e^{(1+\rho)X_i} e^{\sqrt{1-\rho^2}X_j}] \quad (\text{A.11})$$

Les 2 variables étant indépendantes, l'espérance est le produit des 2 espérances, et d'après le calcul précédent on obtient :

$$\mathbb{E} [e^{X_i} e^{X_j}] = e^{(1+\rho)^2 \frac{\sigma_X^2}{2}} e^{(1-\rho^2) \frac{\sigma_X^2}{2}} \quad (\text{A.12})$$

$$= e^{(2+2\rho) \frac{\sigma_X^2}{2}} \quad (\text{A.13})$$

$$= e^{(1+\rho)\sigma_X^2} \quad (\text{A.14})$$

Donc en utilisant la formule (A.7) :

$$\mathbb{E} [Y_i Y_j] = e^{-\sigma_X^2} e^{(1+\rho)\sigma_X^2} \quad (\text{A.15})$$

$$= e^{\rho\sigma_X^2} \quad (\text{A.16})$$

$$= e^{\text{Cov}(X_i, X_j)} \quad (\text{A.17})$$

Finalement on obtient :

$$\text{Cov}(Y_i, Y_j) = \mathbb{E} [Y_i Y_j] - \mathbb{E} [Y_i] \mathbb{E} [Y_j] = e^{\text{Cov}(X_i, X_j)} - 1 \quad (\text{A.18})$$

ou encore  $\text{Cov}(X_i, X_j) = \ln [\text{Cov}(Y_i, Y_j) + 1]$ . En particulier, on obtient la variance en chaque point du champ log-normal en prenant  $X_j = X_i$  :

$$\sigma_Y^2 = e^{\sigma_X^2} - 1 \quad (\text{A.19})$$

# Annexe B

## Tests d'hypothèses

### B.1 Meilleur fit $\chi^2$

On considère une classe de modèles de fonctions de corrélation  $\xi_\theta = (\xi_\theta(r_i))_{1 \leq i \leq n}$ , avec un paramètre  $k$ -dimensionnel  $\theta = (\theta_1, \dots, \theta_k) \in \Theta$ . On suppose que l'estimateur  $\hat{\xi}$  de la fonction de corrélation est Gaussien avec une espérance dans la classe de modèles (i.e.  $\exists \theta_0$  tel que  $\hat{\xi} \sim \mathcal{N}(\xi_{\theta_0}, C)$ ) et une matrice de covariance  $C$ . On regarde la statistique  $\chi_\theta^2$  en fonction du paramètre  $\theta$  :

$$\chi_\theta^2 = \sum_{1 \leq i, j \leq n} [\hat{\xi}(r_i) - \xi_\theta(r_i)] C_{i,j}^{-1} [\hat{\xi}(r_j) - \xi_\theta(r_j)] \quad (\text{B.1})$$

On fait maintenant une hypothèse importante, à savoir que l'espace des modèles de fonctions de corrélation  $(\xi_\theta)_{\theta \in \Theta}$  est un sous-espace affine de dimension  $k$  de  $\mathbb{R}^n$ . Alors la statistique  $\min_\theta \chi_\theta^2$  suit une loi du  $\chi^2$  avec  $n - k$  degrés de liberté (la dimension de la mesure moins la dimension de l'espace des modèles) :

$$\min_\theta \chi_\theta^2 \sim \chi_{n-k}^2 \quad (\text{B.2})$$

Comme  $C$  est une matrice définie positive, on peut considérer  $C^{-1/2}$ . Notons  $\hat{X} = C^{-1/2}(\hat{\xi} - \xi_{\theta_0})$  et  $X_\theta = C^{-1/2}(\xi_\theta - \xi_{\theta_0})$ , de sorte qu'on puisse réécrire  $\chi_\theta^2$  de la façon suivante :

$$\chi_\theta^2 = \|\hat{X} - X_\theta\|^2 \quad (\text{B.3})$$

Ceci représente la transformée de Karhunen-Loève qui transforme le vecteur mesuré  $\hat{\xi}$  en un bruit blanc. Cela veut dire que le vecteur résultant  $\hat{X}$  est une variable Gaussienne multivariée d'espérance 0 et de matrice de covariance égale à l'identité. En effet la matrice de covariance de  $\hat{X}$  est égale à :

$$\mathbb{E}[\hat{X} \hat{X}^T] = C^{-1/2} \mathbb{E}[(\hat{\xi} - \xi_{\theta_0})(\hat{\xi} - \xi_{\theta_0})^T] C^{-1/2} \quad (\text{B.4})$$

$$= C^{-1/2} C C^{-1/2} = I_n \quad (\text{B.5})$$

avec  $I_n$  la matrice identité  $n \times n$ . Ainsi, dans n'importe quelle base orthogonale de  $\mathbb{R}^n$ , les  $n$  composantes de  $\hat{X}$  sont des variables Gaussiennes standards et indépendantes. Notons  $F_\Theta = (X_\theta)_{\theta \in \Theta}$ , qui est un espace vectoriel de dimension  $k$ , et  $F_\Theta^\perp$  son complémentaire orthogonal de dimension  $n - k$ . Notons  $(\hat{Y}_1, \dots, \hat{Y}_{k+1}, \dots, \hat{Y}_n)$  les composantes de  $\hat{X}$  dans une base orthonormée qui a ses  $k$  premiers vecteurs dans  $F_\Theta$ , et ses  $n - k$  vecteurs suivants dans  $F_\Theta^\perp$ . Alors les composantes  $\hat{Y}_i$  sont des variables Gaussiennes standard et indépendantes. De plus  $\chi_\theta^2$  est minimisé lorsque  $X_\theta$  est la projection de  $\hat{X}$  sur  $F_\Theta$ , et est égal à :

$$\min_\theta \chi_\theta^2 = \|\hat{X} - \hat{X}_{F_\Theta}\|^2 = \|\hat{X}_{F_\Theta^\perp}\|^2 \quad (\text{B.6})$$

$$= \sum_{i=k+1}^n \hat{Y}_i^2 \quad (\text{B.7})$$

Cela montre que  $\min_\theta \chi_\theta^2$  suit une loi du  $\chi^2$  avec  $n - k$  degrés de liberté, i.e.  $\min_\theta \chi_\theta^2 \sim \chi_{n-k}^2$ .

## B.2 Différence de meilleurs *fits* $\chi^2$ pour des modèles imbriqués

On considère ici 2 classes de modèles imbriquées de fonctions de corrélation,  $\xi_\theta$  avec  $\theta \in \Theta_1$  pour la première classe et  $\theta \in \Theta_2$  pour la seconde classe. On suppose que  $\Theta_1$  est  $k$ -dimensionnel et  $\Theta_2$  est  $(k+l)$ -dimensionnel avec  $\Theta_1 \subset \Theta_2$ .

On suppose à nouveau que l'estimateur  $\hat{\xi}$  est Gaussien avec une espérance dans la classe restreinte de modèles (i.e.  $\exists \theta_0 \in \Theta_1$  tel que  $\hat{\xi} \sim \mathcal{N}(\xi_{\theta_0}, C)$ ) et une matrice de covariance  $C$ . On fait à nouveau l'hypothèse importante que les espaces de modèles de fonctions de corrélation  $(\xi_\theta)_{\theta \in \Theta_1}$  et  $(\xi_\theta)_{\theta \in \Theta_2}$  sont des sous-espaces affines de  $\mathbb{R}^n$ , de dimensions respectives  $k$  et  $k+l$ . Alors la différence des meilleurs *fits*  $\chi^2$  entre les 2 classes  $\min_{\theta \in \Theta_1} \chi_\theta^2 - \min_{\theta \in \Theta_2} \chi_\theta^2$  suit une loi du  $\chi^2$  avec  $l$  degrés de liberté (la différence de dimension des espaces de modèles) :

$$\min_{\theta \in \Theta_1} \chi_\theta^2 - \min_{\theta \in \Theta_2} \chi_\theta^2 \sim \chi_l^2 \quad (\text{B.8})$$

Cela découle facilement de l'annexe B.1. On considère à nouveau les transformées de Karhunen-Loève  $\hat{X} = C^{-1/2}(\hat{\xi} - \xi_{\theta_0})$  et  $X_\theta = C^{-1/2}(\xi_\theta - \xi_{\theta_0})$ . Notons les espaces de modèles  $F_{\Theta_1} = (X_\theta)_{\theta \in \Theta_1}$  et  $F_{\Theta_2} = (X_\theta)_{\theta \in \Theta_2}$ , et leurs compléments orthogonaux  $F_{\Theta_1}^\perp$  et  $F_{\Theta_2}^\perp$ . Notons  $(\hat{Y}_1, \dots, \hat{Y}_{k+1}, \dots, \hat{Y}_{k+l+1}, \dots, \hat{Y}_n)$  les composantes de  $\hat{X}$  dans une base orthonormée qui a ses  $k$  premiers vecteurs dans  $F_{\Theta_1} \cap F_{\Theta_2}$ , ses  $l$  vecteurs suivants dans  $F_{\Theta_1}^\perp \cap F_{\Theta_2}$ , et ses  $n - (k+l)$  derniers vecteurs dans  $F_{\Theta_1}^\perp \cap F_{\Theta_2}^\perp$ . Pour chaque classe de modèles,  $\chi_\theta^2$  est minimisé lorsque  $X_\theta$  est la projection de  $\hat{X}$  sur l'espace de modèles  $F_\Theta$  :

$$\min_{\theta \in \Theta_1} \chi_\theta^2 = \|X_{F_{\Theta_1}^\perp}\|^2 = \sum_{i=k+1}^n \hat{Y}_i^2 \quad (\text{B.9})$$

$$\min_{\theta \in \Theta_2} \chi_\theta^2 = \|X_{F_{\Theta_2}^\perp}\|^2 = \sum_{i=k+l+1}^n \hat{Y}_i^2 \quad (\text{B.10})$$

Donc la différence des meilleurs  $\chi^2$  entre les 2 classes est égale à :

$$\min_{\theta \in \Theta_1} \chi_\theta^2 - \min_{\theta \in \Theta_2} \chi_\theta^2 = \sum_{i=k+1}^{k+l} \hat{Y}_i^2 \quad (\text{B.11})$$

Cela montre que  $\min_{\theta \in \Theta_1} \chi_\theta^2 - \min_{\theta \in \Theta_2} \chi_\theta^2$  suit une loi de  $\chi^2$  avec  $l$  degrés de liberté, i.e.  $\min_{\theta \in \Theta_1} \chi_\theta^2 - \min_{\theta \in \Theta_2} \chi_\theta^2 \sim \chi_l^2$ .

## B.3 Optimalité du rapport de vraisemblance

Considérons le rapport de vraisemblance  $\Lambda(\hat{\xi}) = \mathcal{L}_{\mathcal{H}_0}(\hat{\xi})/\mathcal{L}_{\mathcal{H}_1}(\hat{\xi})$  ainsi qu'une autre statistique quelconque  $S(\hat{\xi})$  pour tester les hypothèses  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . On suppose que  $\mathcal{H}_1$  est préférée par rapport à  $\mathcal{H}_0$  pour des valeurs faibles de  $S(\hat{\xi})$ , comme c'est le cas pour le rapport de vraisemblance (on peut toujours se ramener à ce cas en considérant  $-S$ ).

On considère d'abord des tests statistiques de significativité  $\alpha$  donnée. Le test basé sur le rapport de vraisemblance est donné par :

- si  $\Lambda(\hat{\xi}) \leq \eta_\Lambda$  alors accepter  $\mathcal{H}_1$
- si  $\Lambda(\hat{\xi}) > \eta_\Lambda$  alors accepter  $\mathcal{H}_0$

Le test basé sur la statistique  $S$  est donné par :

- if  $S(\hat{\xi}) \leq \eta_S$  alors accepter  $\mathcal{H}_1$
- if  $S(\hat{\xi}) > \eta_S$  alors accepter  $\mathcal{H}_0$

avec  $\alpha = P(\Lambda(\hat{\xi}) \leq \eta_\Lambda | \mathcal{H}_0)$  et  $\alpha = P(S(\hat{\xi}) \leq \eta_S | \mathcal{H}_0)$ . D'après le Lemme de Neyman-Pearson, la puissance du test basé sur le rapport de vraisemblance est plus grande que pour n'importe quel

autre test. Cela veut dire que la probabilité d'accepter  $\mathcal{H}_1$  si l'hypothèse est vraie est plus grande pour le test de rapport de vraisemblance.

$$P\left(\Lambda(\hat{\xi}) \leq \eta_\Lambda \mid \mathcal{H}_1\right) \geq P\left(S(\hat{\xi}) \leq \eta_S \mid \mathcal{H}_1\right) \quad (\text{B.12})$$

On considère maintenant la significativité correspondant à des réalisations  $\Lambda(\hat{\xi}) = x$  et  $S(\hat{\xi}) = y$ , que l'on note respectivement  $\alpha_\Lambda$  et  $\alpha_S$  :

$$\alpha_\Lambda(x) = P(\Lambda(\hat{\xi}) \leq x \mid \mathcal{H}_0) \quad (\text{B.13})$$

$$\alpha_S(y) = P(S(\hat{\xi}) \leq y \mid \mathcal{H}_0) \quad (\text{B.14})$$

On a  $\alpha_\Lambda(\eta_\Lambda) = \alpha$  et  $\alpha_S(\eta_S) = \alpha$ . Comme  $\alpha_\Lambda$  et  $\alpha_S$  sont des fonctions croissantes, les conditions  $\Lambda(\hat{\xi}) \leq \eta_\Lambda$  et  $S(\hat{\xi}) \leq \eta_S$  sont respectivement équivalentes à  $\alpha_\Lambda(\Lambda(\hat{\xi})) \leq \alpha$  et  $\alpha_S(S(\hat{\xi})) \leq \alpha$ . Si l'on simplifie les notations et que l'on note  $\alpha_\Lambda(\hat{\xi})$  pour  $\alpha_\Lambda(\Lambda(\hat{\xi}))$  et  $\alpha_S(\hat{\xi})$  pour  $\alpha_S(S(\hat{\xi}))$ , on obtient à partir de l'équation (B.12) que pour tout  $\alpha$  :

$$P\left(\alpha_\Lambda(\hat{\xi}) \leq \alpha \mid \mathcal{H}_1\right) \geq P\left(\alpha_S(\hat{\xi}) \leq \alpha \mid \mathcal{H}_1\right) \quad (\text{B.15})$$

Nous allons montrer que cela implique l'inégalité suivante :

$$\mathbb{E}\left[\alpha_\Lambda(\hat{\xi}) \mid \mathcal{H}_1\right] \leq \mathbb{E}\left[\alpha_S(\hat{\xi}) \mid \mathcal{H}_1\right] \quad (\text{B.16})$$

Dans la suite, on va toujours considérer les probabilités et espérances avec la condition  $\mathcal{H}_1$ , donc on va omettre cette condition pour simplifier les notations. On note  $F_\Lambda$  et  $F_S$  les fonctions de répartition données par  $F_\Lambda(\alpha) = P(\alpha_\Lambda(\hat{\xi}) \leq \alpha)$  et  $F_S(\alpha) = P(\alpha_S(\hat{\xi}) \leq \alpha)$ . L'équation (B.15) implique que pour tout  $\alpha$  et tout  $p$  :

$$F_\Lambda(\alpha) \geq F_S(\alpha) \quad (\text{B.17})$$

$$F_\Lambda^{-1}(p) \leq F_S^{-1}(p) \quad (\text{B.18})$$

L'espérance de  $\alpha_\Lambda(\hat{\xi})$  est donnée par :

$$\mathbb{E}[\alpha_\Lambda(\hat{\xi})] = \int \alpha \, dF_\Lambda(\alpha) = \int_0^1 F_\Lambda^{-1}(p) \, dp \quad (\text{B.19})$$

où l'on a effectué le changement de variable  $p = F_\Lambda(\alpha)$ . Le même calcul peut être effectué pour  $\mathbb{E}[\alpha_S(\hat{\xi})]$ , et puisque  $F_\Lambda^{-1}(p) \leq F_S^{-1}(p)$  on obtient  $\mathbb{E}[\alpha_\Lambda(\hat{\xi})] \leq \mathbb{E}[\alpha_S(\hat{\xi})]$ . L'espérance de la significativité donnée en terme de valeur  $p$  est minimisée pour le rapport de vraisemblance.

Si l'on mesure la significativité en terme de nombre de  $\sigma$  au lieu de valeur  $p$ , les 2 inégalités (B.17) et (B.18) sont inversées. L'inégalité sur l'espérance est donc également inversée, i.e. que l'espérance du nombre de  $\sigma$  est maximisée pour le rapport de vraisemblance.



## Annexe C

# Combinaison linéaire optimale d'estimateurs

Dans cette section on suppose que l'on connaît 2 estimateurs non biaisés, indépendants et Gaussiens  $X_1, X_2$  (de dimension  $n$ ) de  $X_0$ , avec matrices de covariance respectives  $C_1$  et  $C_2$  :

$$X_1 \sim \mathcal{N}(X_0, C_1) \quad (\text{C.1})$$

$$X_2 \sim \mathcal{N}(X_0, C_2) \quad (\text{C.2})$$

Considérons un estimateur non biaisé  $X$  de  $X_0$  comme combinaison linéaire de  $X_1$  et  $X_2$  :

$$X = AX_1 + (Id - A)X_2 \quad (\text{C.3})$$

avec  $A$  une matrice carré  $n \times n$ . La matrice de covariance de  $X$  est donnée par :

$$C = \mathbb{E}[XX^T] = AC_1A^T + (Id - A)C_2(Id - A)^T \quad (\text{C.4})$$

où l'on a utilisé le fait que  $X_1$  et  $X_2$  sont indépendants. On va montrer que le choix suivant de la matrice  $A$  donne un extremum de  $\det(C)$  :

$$A = (C_1^{-1} + C_2^{-1})^{-1} C_1^{-1} \quad (\text{C.5})$$

$$Id - A = (C_1^{-1} + C_2^{-1})^{-1} C_2^{-1} \quad (\text{C.6})$$

Pour cela on utilise la formule de dérivation suivante, avec  $B$  une matrice symétrique  $n \times n$  :

$$\frac{\partial \det(C)}{\partial C} = \det(C)C^{-T} \quad (\text{C.7})$$

$$\frac{\partial (ABA^T)}{\partial A_{ji}} = ABJ^{ij} + J^{ji}BA^T = ABJ^{ij} + (ABJ^{ij})^T \quad (\text{C.8})$$

avec  $(J^{ij})_{kl} = \delta_{ik}\delta_{jl}$ . En dérivant  $\det(C)$  par rapport à  $A$ , on obtient :

$$\frac{\partial \det(C)}{\partial A_{ji}} = \sum_{kl} \det(C)C_{kl}^{-T} \frac{\partial C_{kl}}{\partial A_{ji}} \quad (\text{C.9})$$

Il est suffisant d'avoir pour tout  $i, j$  que  $\frac{\partial C}{\partial A_{ji}} = 0$  :

$$\frac{\partial C}{\partial A_{ji}} = \frac{\partial (AC_1A^T)}{\partial A_{ji}} + \frac{\partial ((Id - A)C_2(Id - A)^T)}{\partial A_{ji}} \quad (\text{C.10})$$

$$= (AC_1 - (Id - A)C_2)J^{ij} + [(AC_1 - (Id - A)C_2)J^{ij}]^T \quad (\text{C.11})$$

A nouveau il est suffisant d'avoir  $AC_1 - (Id - A)C_2 = 0$ , ce qui donne :

$$A(C_1 + C_2) = C_2 \quad (\text{C.12})$$

$$A = C_2(C_1 + C_2)^{-1} \quad (\text{C.13})$$

$$A = C_2C_2^{-1}(C_1^{-1} + C_2^{-1})^{-1}C_1^{-1} \quad (\text{C.14})$$

Donc on obtient la solution donnée par les équations (C.6) et (C.6) :

$$A = (C_1^{-1} + C_2^{-1})^{-1} C_1^{-1} \quad (\text{C.15})$$

$$Id - A = (C_1^{-1} + C_2^{-1})^{-1} C_2^{-1} \quad (\text{C.16})$$

Finalement en utilisation cette expression de  $A$  dans l'équation (C.4), on obtient :

$$C = \mathbb{E}[XX^T] = (C_1^{-1} + C_2^{-1})^{-1} \quad (\text{C.17})$$

# Annexe D

## Mes contributions

### Articles de journaux

1. **Uncertainty in 2-point correlation function estimators and baryon acoustic oscillation detection in galaxy surveys**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey, P. Arnalte-Mur*, *Statistical Methodology*, 9, 85-100, 2012
2. **Detecting Baryon Acoustic Oscillations**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey*, *The Astrophysical Journal*, 746, 172, 2012
3. **Effect of model-dependent covariance matrix for studying Baryon Acoustic Oscillations**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey*, *The Astrophysical Journal*, accepté
4. **Wavelet analysis of baryon acoustic structures in the galaxy distribution**, *P. Arnalte-Mur, A. Labatie, N. Clerc, V. J. Martínez, J.-L. Starck, M. Lachièze-Rey, E. Saar, S. Paredes*, *Astronomy and Astrophysics*, 542, A34, 2012
5. **The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey : Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample**, *Anderson et al.*, arXiv :1203.6594, 2012
6. **The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey : cosmological implications of the large-scale two-point correlation function**, *Sánchez et al.*, *Monthly Notices of the Royal Astronomical Society*, 425, 415-437, 2012

### Communications

1. **Uncertainty in 2-point correlation function estimators**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey, P. Arnalte-Mur*, *Astronomical Data Analysis VI*, Monastir, Tunisie, Mai 2010
2. **A new method for detecting Baryon Acoustic Oscillations**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey*, *Astronomical Data Analysis VII*, Cargèse, Corse, Mai 2012
3. **Effect of model-dependent covariance matrix for studying Baryon Acoustic Oscillations**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey*, *Astronomical Data Analysis VII*, Cargèse, Corse, Mai 2012
4. **BAO detection methods**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey*, workshop Cosmology and Sparsity, Nice, Juin 2011



## Chapitre de livre

1. **Statistical Study of the Galaxy Distribution**, *A. Labatie, J.-L. Starck, M. Lachièze-Rey*, dans *Advances in Modern Cosmology*, Adnan Ghribi (Ed.), ISBN : 978-953-307-423-8, InTech, 2011

# Bibliographie

- K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, and et al. The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182 :543, June 2009. 53
- A. Albrecht, G. Bernstein, R. Cahn, W. L. Freedman, J. Hewitt, W. Hu, J. Huth, M. Kamionkowski, E. W. Kolb, L. Knox, J. C. Mather, S. Staggs, and N. B. Suntzeff. Report of the Dark Energy Task Force. *ArXiv Astrophysics e-prints*, September 2006. 113
- R. Amanullah, C. Lidman, D. Rubin, G. Aldering, P. Astier, K. Barbary, M. S. Burns, A. Conley, K. S. Dawson, S. E. Deustua, M. Doi, S. Fabbro, L. Faccioli, H. K. Fakhouri, G. Folatelli, A. S. Fruchter, H. Furusawa, G. Garavini, G. Goldhaber, A. Goobar, D. E. Groom, I. Hook, D. A. Howell, N. Kashikawa, A. G. Kim, R. A. Knop, M. Kowalski, E. Linder, J. Meyers, T. Morokuma, S. Nobili, J. Nordin, P. E. Nugent, L. Östman, R. Pain, N. Panagia, S. Perlmutter, J. Raux, P. Ruiz-Lapuente, A. L. Spadafora, M. Strovink, N. Suzuki, L. Wang, W. M. Wood-Vasey, N. Yasuda, and T. Supernova Cosmology Project. Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at 0.511<math>z</math>1.12 and the Union2 Compilation. *ApJ*, 716 :712–738, 2010. 13
- L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, A. J. Cuesta, L. N. A. da Costa, K. S. Dawson, R. de Putter, D. J. Eisenstein, J. E. Gunn, H. Guo, J.-C. Hamilton, P. Harding, S. Ho, K. Honscheid, E. Kazin, D. Kirkby, J.-P. Kneib, A. Labatie, C. Loomis, R. H. Lupton, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, C. K. McBride, K. T. Mehta, O. Mena, F. Montesano, D. Muna, R. C. Nichol, S. E. Nuza, M. D. Olmstead, D. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. Parejko, I. Paris, W. J. Percival, P. Petitjean, F. Prada, B. Reid, N. A. Roe, A. J. Ross, N. P. Ross, L. Samushia, A. G. Sanchez, D. J. S. D. P. Schneider, C. G. Scoccola, H.-J. Seo, E. S. Sheldon, A. Simmons, R. A. Skibba, M. A. Strauss, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, M. Vargas Magana, L. Verde, C. Wagner, D. A. Wake, B. A. Weaver, D. H. Weinberg, M. White, X. Xu, C. Yeche, I. Zehavi, and G.-B. Zhao. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey : Baryon Acoustic Oscillations in the Data Release 9 Spectroscopic Galaxy Sample. *ArXiv e-prints*, March 2012. XIV, 60, 61, 62, 90
- P. Arnalte-Mur, A. Labatie, N. Clerc, V. J. Martínez, J.-L. Starck, M. Lachièze-Rey, E. Saar, and S. Paredes. Wavelet analysis of baryon acoustic structures in the galaxy distribution. *A&A*, 542 : A34, June 2012. XV, 88, 89, 90, 101
- J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *ApJ*, 304 :15–61, May 1986. 36
- A. A. Berlind and D. H. Weinberg. The Halo Occupation Distribution : Toward an Empirical Determination of the Relation between Galaxies and Mass. *ApJ*, 575 :587–616, August 2002. 37
- F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro. Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, 367 :1–248, 2002. 19, 20
- F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson. The 6dF Galaxy Survey : baryon acoustic oscillations and the local Hubble constant. *MNRAS*, 416 :3017–3032, October 2011. 27, 59, 64, 71, 90, 114, 119
- C. Blake and K. Glazebrook. Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler. *ApJ*, 594 :665–673, September 2003. 113

- C. Blake, T. Davis, G. B. Poole, D. Parkinson, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I.-H. Li, B. Madore, D. C. Martin, K. Pimbblet, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey : testing the cosmological model with baryon acoustic oscillations at  $z=0.6$ . *MNRAS*, 415 :2892–2909, August 2011a. 25, 37, 59, 64, 71, 90, 114, 119
- C. Blake, E. A. Kazin, F. Beutler, T. M. Davis, D. Parkinson, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I.-H. Li, B. Madore, D. C. Martin, K. Pimbblet, G. B. Poole, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey : mapping the distance-redshift relation with baryon acoustic oscillations. *MNRAS*, 418 :1707–1724, December 2011b. XIII, 13, 25, 27, 53, 64, 71, 90, 114, 119
- M. R. Blanton, D. W. Hogg, N. A. Bahcall, J. Brinkmann, M. Britton, A. J. Connolly, I. Csabai, M. Fukugita, J. Loveday, A. Meiksin, J. A. Munn, R. C. Nichol, S. Okamura, T. Quinn, D. P. Schneider, K. Shimasaku, M. A. Strauss, M. Tegmark, M. S. Vogeley, and D. H. Weinberg. The Galaxy Luminosity Function and Luminosity Density at Redshift  $z = 0.1$ . *ApJ*, 592 :819–838, August 2003. 54
- M. R. Blanton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, D. Finkbeiner, M. Fukugita, J. E. Gunn, D. W. Hogg, Ž. Ivezić, G. R. Knapp, R. H. Lupton, J. A. Munn, D. P. Schneider, M. Tegmark, and I. Zehavi. New York University Value-Added Galaxy Catalog : A Galaxy Catalog Based on New Public Surveys. *AJ*, 129 :2562–2578, June 2005. 54
- J. R. Bond and G. Efstathiou. Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *ApJ*, 285 :L45–L48, October 1984. 22
- J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380 :603–606, April 1996. 33
- S. M. Carroll. The Cosmological Constant. *Living Reviews in Relativity*, 4 :1, 2001. XI, 12, 18
- S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey : power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362 :505–534, September 2005. 52, 71, 89, 90, 119
- P. Coles and B. Jones. A lognormal model for the cosmological mass distribution. *MNRAS*, 248 :1–13, January 1991. 63, 64
- A. Cooray and R. Sheth. Halo models of large scale structure. *Phys. Rep.*, 372 :1–129, December 2002. 25
- M. Crocce and R. Scoccimarro. Renormalized cosmological perturbation theory. *Phys. Rev. D*, 73 (6) :063519, March 2006. 25
- M. Crocce and R. Scoccimarro. Nonlinear evolution of baryon acoustic oscillations. *Phys. Rev. D*, 77 (2) :023533, January 2008. 25
- L. N. da Costa, P. S. Pellegrini, M. Davis, A. Meiksin, W. L. W. Sargent, and J. L. Tonry. Southern Sky Redshift Survey - The catalog. *ApJS*, 75 :935–964, April 1991. 51
- M. Davis. Is the Universe Homogeneous on Large Scales? In *Critical Dialogues in Cosmology*, pages 13–+, 1997. 3, 47
- M. Davis and P. J. E. Peebles. A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, 267 :465–482, April 1983. 40, 74
- S. Dodelson. *Modern Cosmology*. Academic Press, 1 edition, March 2003. URL <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0122191412>. 20

- D. J. Eisenstein and W. Hu. Baryonic Features in the Matter Transfer Function. *ApJ*, 496 :605, March 1998. XII, XIV, 22, 23, 25, 29, 66, 67, 78, 90, 96, 117
- D. J. Eisenstein and W. Hu. Power Spectra for Cold Dark Matter and Its Variants. *ApJ*, 511 :5–15, January 1999. 22, 23
- D. J. Eisenstein, W. Hu, J. Silk, and A. S. Szalay. Can Baryonic Features Produce the Observed 100 H -1 MPC Clustering? *ApJ*, 494 :L1, February 1998. 113
- D. J. Eisenstein, J. Annis, J. E. Gunn, A. S. Szalay, A. J. Connolly, R. C. Nichol, N. A. Bahcall, M. Bernardi, S. Burles, F. J. Castander, M. Fukugita, D. W. Hogg, Ž. Ivezić, G. R. Knapp, R. H. Lupton, V. Narayanan, M. Postman, D. E. Reichart, M. Richmond, D. P. Schneider, D. J. Schlegel, M. A. Strauss, M. SubbaRao, D. L. Tucker, D. Vanden Berk, M. S. Vogeley, D. H. Weinberg, and B. Yanny. Spectroscopic Target Selection for the Sloan Digital Sky Survey : The Luminous Red Galaxy Sample. *AJ*, 122 :2267–2280, November 2001. 53, 54, 57
- D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633 :560–574, November 2005. XI, 16, 17, 25, 37, 43, 52, 71, 89, 90, 91, 93, 114, 119
- D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel. Improving Cosmological Distance Measurements by Reconstruction of the Baryon Acoustic Peak. *ApJ*, 664 :675–679, August 2007a. 59, 89
- D. J. Eisenstein, H.-J. Seo, and M. White. On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter. *ApJ*, 664 :660–674, August 2007b. XII, 25, 27, 28, 29
- D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. Allende Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot, and et al. SDSS-III : Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, 142 :72, September 2011. 60
- K. Falconer. *Fractal Geometry*. Mathematical Foundations and Applications. John Wiley & Sons Ltd., New-York, 1990. XIII, 44, 46
- J. Feder. *Fractals*. Plenum Press, New York, 1988. XIII, 45
- H. A. Feldman, N. Kaiser, and J. A. Peacock. Power-spectrum analysis of three-dimensional redshift surveys. *ApJ*, 426 :23–37, May 1994. 61
- D. J. Fixsen. The Temperature of the Cosmic Microwave Background. *ApJ*, 707 :916–920, December 2009. 41
- J. N. Fry and E. Gaztanaga. Biasing and hierarchical statistics in large-scale structure. *ApJ*, 413 :447–452, August 1993. 36
- A. Gabrielli, M. Joyce, and F. Sylos Labini. Glass-like universe : Real-space correlation properties of standard cosmological models. *Phys. Rev. D*, 65(8) :083523, April 2002. 81
- A. J. S. Hamilton. Toward Better Ways to Measure the Galaxy Correlation Function. *ApJ*, 417 :19, November 1993. 40, 41, 61, 74
- A. J. S. Hamilton and M. Tegmark. A scheme to deal accurately and efficiently with complex angular masks in galaxy surveys. *MNRAS*, 349 :115–128, March 2004. 54

- S. Ho, A. Cuesta, H.-J. Seo, R. de Putter, A. J. Ross, M. White, N. Padmanabhan, S. Saito, D. J. Schlegel, E. Schlafly, U. Seljak, C. Hernandez-Monteagudo, A. G. Sanchez, W. J. Percival, M. Blanton, R. Skibba, D. Schneider, B. Reid, O. Mena, M. Viel, D. J. Eisenstein, F. Prada, B. Weaver, N. Bahcall, D. Bizyaev, H. Brewinton, J. Brinkman, L. Nicolaci da Costa, J. R. Gott, E. Malanushenko, V. Malanushenko, B. Nichol, D. Oravetz, K. Pan, N. Palanque-Delabrouille, N. P. Ross, A. Simmons, F. de Simoni, S. Snedden, and C. Yeche. Clustering of Sloan Digital Sky Survey III Photometric Luminous Galaxies : The Measurement, Systematics and Cosmological Implications. *ArXiv e-prints*, January 2012. 71, 119
- J. Hogan. Unseen Universe : Welcome to the dark side. *Nature*, 448 :240–245, July 2007. 12
- D. W. Hogg, I. K. Baldry, M. R. Blanton, and D. J. Eisenstein. The K correction. *ArXiv Astrophysics e-prints*, October 2002. 54
- D. W. Hogg, D. J. Eisenstein, M. R. Blanton, N. A. Bahcall, J. Brinkmann, J. E. Gunn, and D. P. Schneider. Cosmic Homogeneity Demonstrated with Luminous Red Galaxies. *ApJ*, 624 :54–58, May 2005. 48
- E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15 :168–173, March 1929. 7
- E. Hubble. The Distribution of Extra-Galactic Nebulae. *ApJ*, 79 :8, January 1934. 64
- J. Huchra, M. Davis, D. Latham, and J. Tonry. A survey of galaxy redshifts. IV - The data. *ApJS*, 52 :89–119, June 1983. 51
- G. Hütsi. Acoustic oscillations in the SDSS DR4 luminous red galaxy sample power spectrum. *A&A*, 449 :891–902, April 2006. 89, 90
- N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Sky Maps, Systematic Errors, and Basic Results. *ApJS*, 192 :14, February 2011. XI, 14, 15
- B. J. Jones, V. J. Martínez, E. Saar, and V. Trimble. Scaling laws in the distribution of galaxies. *Reviews of Modern Physics*, 76 :1211–1266, October 2004. XIII, 52
- N. Kaiser. On the spatial correlations of Abell clusters. *ApJ*, 284 :L9–L12, September 1984. 36
- N. Kaiser. Clustering in real space and in redshift space. *MNRAS*, 227 :1–21, July 1987. 37, 38
- I. Kayo, A. Taruya, and Y. Suto. Probability Distribution Function of Cosmological Density Fluctuations from a Gaussian Initial Condition : Comparison of One-Point and Two-Point Lognormal Model Predictions with N-Body Simulations. *ApJ*, 561 :22–34, November 2001. 64
- E. A. Kazin, M. R. Blanton, R. Scoccimarro, C. K. McBride, A. A. Berlind, N. A. Bahcall, J. Brinkmann, P. Czarapata, J. A. Frieman, S. M. Kent, D. P. Schneider, and A. S. Szalay. The Baryonic Acoustic Feature and Large-Scale Clustering in the Sloan Digital Sky Survey Luminous Red Galaxy Sample. *ApJ*, 710 :1444–1461, February 2010. XIII, 57, 58, 59, 60, 71, 114, 119
- M. Kerscher, I. Szapudi, and A. S. Szalay. A Comparison of Estimators for the Two-Point Correlation Function. *ApJ*, 535 :L13–L16, May 2000. 83
- F.-S. Kitaura, J. Jasche, and R. B. Metcalf. Recovering the non-linear density field from the galaxy distribution with a Poisson-lognormal filter. *MNRAS*, 403 :589–604, April 2010. 64
- E. Komatsu, J. Dunkley, M. R. Nolta, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. Limon, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, S. S. Meyer, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations : Cosmological Interpretation. *ApJS*, 180 :330–376, February 2009. 16, 115

- E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Cosmological Interpretation. *ApJS*, 192 :18, 2011. XII, XIV, 13, 15, 24, 66, 67, 96
- M. Kowalski, D. Rubin, G. Aldering, R. J. Agostinho, A. Amadon, R. Amanullah, C. Balland, K. Barbary, G. Blanc, P. J. Challis, A. Conley, N. V. Connolly, R. Covarrubias, K. S. Dawson, S. E. Deustua, R. Ellis, S. Fabbro, V. Fadeyev, X. Fan, B. Farris, G. Folatelli, B. L. Frye, G. Garavini, E. L. Gates, L. Germany, G. Goldhaber, B. Goldman, A. Goobar, D. E. Groom, J. Haissinski, D. Hardin, I. Hook, S. Kent, A. G. Kim, R. A. Knop, C. Lidman, E. V. Linder, J. Mendez, J. Meyers, G. J. Miller, M. Moniez, A. M. Mourão, H. Newberg, S. Nobili, P. E. Nugent, R. Pain, O. Perdureau, S. Perlmutter, M. M. Phillips, V. Prasad, R. Quimby, N. Regnault, J. Rich, E. P. Rubenstein, P. Ruiz-Lapuente, F. D. Santos, B. E. Schaefer, R. A. Schommer, R. C. Smith, A. M. Soderberg, A. L. Spadafora, L.-G. Strolger, M. Strovink, N. B. Suntzeff, N. Suzuki, R. C. Thomas, N. A. Walton, L. Wang, W. M. Wood-Vasey, J. L. Yun, and Supernova Cosmology Project. Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets. *ApJ*, 686 :749–778, 2008. XI, 13, 14
- A. Labatie, J. L. Starck, and M. Lachièze-Rey. Detecting Baryon Acoustic Oscillations. *ApJ*, 746 : 172, February 2012a. 64
- A. Labatie, J. L. Starck, and M. Lachièze-Rey. Effect of model-dependent covariance matrix for studying Baryon Acoustic Oscillations. *ApJ*, 2012b. 64
- A. Labatie, J. L. Starck, M. Lachièze-Rey, and P. Arnalte-Mur. Uncertainty in 2-point correlation function estimators and baryon acoustic oscillation detection in galaxy surveys. *Statistical Methodology*, 9 :85 – 100, 2012c. 64
- O. Lahav, P. B. Lilje, J. R. Primack, and M. J. Rees. Dynamical effects of the cosmological constant. *MNRAS*, 251 :128–136, 1991. 21
- S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412 :64–71, July 1993. 40, 74, 83
- D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Power Spectra and WMAP-derived Parameters. *ApJS*, 192 :16, February 2011. XI, 15
- A. Lewis, A. Challinor, and A. Lasenby. Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *ApJ*, 538 :473–476, August 2000. 22
- E. V. Linder. Exploring the Expansion History of the Universe. *Physical Review Letters*, 90(9) : 091301, March 2003. 13
- R. Maartens. Is the Universe homogeneous? *Royal Society of London Philosophical Transactions Series A*, 369 :5115–5137, December 2011. 50
- S. Mallat. *A Wavelet Tour of Signal Processing, Third Edition : The Sparse Way*. Academic Press, 3rd edition, 2008. ISBN 0123743702, 9780123743701. 88
- B. B. Mandelbrot. *The Fractal Geometry of Nature*. 1982. 44
- M. Manera, R. Scoccimarro, W. J. Percival, L. Samushia, C. K. McBride, A. J. Ross, R. K. Sheth, M. White, B. A. Reid, A. G. Sánchez, R. de Putter, X. Xu, A. A. Berlind, J. Brinkmann, B. Nichol, F. Montesano, N. Padmanabhan, R. A. Skibba, R. Tojeiro, and B. A. Weaver. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey : a large sample of mock galaxy catalogues. *ArXiv e-prints*, March 2012. XIV, 62, 63, 115

- V. J. Martínez and E. Saar. *Statistics of the Galaxy Distribution*. Chapman & Hall/CRC, 2002. 19, 35, 79
- V. J. Martínez, J.-L. Starck, E. Saar, D. L. Donoho, S. C. Reynolds, P. de la Cruz, and S. Paredes. Morphology of the Galaxy Distribution from Wavelet Denoising. *ApJ*, 634 :744–755, November 2005. 33
- V. J. Martínez, P. Arnalte-Mur, E. Saar, P. de la Cruz, M. J. Pons-Bordería, S. Paredes, A. Fernández-Soto, and E. Tempel. Reliability of the Detection of the Baryon Acoustic Peak. *ApJ*, 696 :L93–L97, May 2009. 54, 56
- K. L. Masters, C. Maraston, R. C. Nichol, D. Thomas, A. Beifiori, K. Bundy, E. M. Edmondson, T. D. Higgs, A. Leauthaud, R. Mandelbaum, J. Pforr, A. J. Ross, N. P. Ross, D. P. Schneider, R. Skibba, J. Tinker, R. Tojeiro, D. A. Wake, J. Brinkmann, and B. A. Weaver. The morphology of galaxies in the Baryon Oscillation Spectroscopic Survey. *MNRAS*, 418 :1055–1070, December 2011. 60
- T. Matsubara. Stochasticity of Bias and Nonlocality of Galaxy Formation : Linear Scales. *ApJ*, 525 : 543–553, November 1999. 36
- T. Matsubara. Resumming cosmological perturbations via the Lagrangian picture : One-loop results in real space and in redshift space. *Phys. Rev. D*, 77(6) :063530, March 2008a. 25
- T. Matsubara. Nonlinear perturbation theory with halo bias and redshift-space distortions via the Lagrangian picture. *Phys. Rev. D*, 78(8) :083519, October 2008b. 25
- K. T. Mehta, A. J. Cuesta, X. Xu, D. J. Eisenstein, and N. Padmanabhan. A 2% Distance to  $z = 0.35$  by Reconstructing Baryon Acoustic Oscillations - III : Cosmological Measurements and Interpretation. *ArXiv e-prints*, January 2012. 71, 114, 119
- H. J. Mo, Y. P. Jing, and S. D. M. White. High-order correlations of peaks and haloes : a step towards understanding galaxy biasing. *MNRAS*, 284 :189–201, January 1997. 36
- J. H. Oort. The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems. *Bull. Astron. Inst. Netherlands*, 6 :249, August 1932. 11
- N. Padmanabhan and M. White. Constraining anisotropic baryon oscillations. *Phys. Rev. D*, 77(12) : 123540, June 2008. 17, 43, 44
- N. Padmanabhan, D. J. Schlegel, U. Seljak, A. Makarov, N. A. Bahcall, M. R. Blanton, J. Brinkmann, D. J. Eisenstein, D. P. Finkbeiner, J. E. Gunn, D. W. Hogg, Ž. Ivezić, G. R. Knapp, J. Loveday, R. H. Lupton, R. C. Nichol, D. P. Schneider, M. A. Strauss, M. Tegmark, and D. G. York. The clustering of luminous red galaxies in the Sloan Digital Sky Survey imaging data. *MNRAS*, 378 : 852–872, July 2007. 71, 119
- N. Padmanabhan, X. Xu, D. J. Eisenstein, R. Scalzo, A. J. Cuesta, K. T. Mehta, and E. Kazin. A 2% Distance to  $z=0.35$  by Reconstructing Baryon Acoustic Oscillations - I : Methods and Application to the Sloan Digital Sky Survey. *ArXiv e-prints*, January 2012. 59
- J. A. Peacock and R. E. Smith. Halo occupation numbers and galaxy bias. *MNRAS*, 318 :1144–1156, November 2000. 25
- J. A. Peacock, S. Cole, P. Norberg, C. M. Baugh, J. Bland-Hawthorn, T. Bridges, R. D. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, W. J. Percival, B. A. Peterson, I. Price, W. Sutherland, and K. Taylor. A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, 410 :169–173, March 2001. XII, 38
- P. J. E. Peebles and M. G. Hauser. Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs. *ApJS*, 28 :19, November 1974. 40, 74
- P. J. E. Peebles and J. T. Yu. Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, 162 : 815, December 1970. 85

- W. J. Percival, S. Cole, D. J. Eisenstein, R. C. Nichol, J. A. Peacock, A. C. Pope, and A. S. Szalay. Measuring the Baryon Acoustic Oscillation scale using the Sloan Digital Sky Survey and 2dF Galaxy Redshift Survey. *MNRAS*, 381 :1053–1066, November 2007. 64, 71, 89, 90, 119
- W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, J. Loveday, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, D. N. Spergel, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *MNRAS*, 401 :2148–2168, February 2010. 59, 64, 71, 89, 90, 93, 119
- S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and Supernova Cosmology Project. Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517 :565–586, June 1999. 7, 12, 16
- L. Pietronero. The fractal structure of the universe : Correlations of galaxies and clusters and the average mass density. *Physica A Statistical Mechanics and its Applications*, 144 :257–284, 1987. 72
- M.-J. Pons-Bordería, V. J. Martínez, D. Stoyan, H. Stoyan, and E. Saar. Comparing Estimators of the Galaxy Correlation Function. *ApJ*, 523 :480–491, October 1999. 83
- William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition : The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007. XVI, 123
- S. Räsänen. Backreaction : directions of progress. *Classical and Quantum Gravity*, 28(16) :164008, August 2011. 5, 12
- A. Refregier, A. Amara, T. D. Kitching, and A. Rassat. iCosmo : an interactive cosmology package. *A&A*, 528 :A33, April 2011. 78, 96
- B. A. Reid, W. J. Percival, D. J. Eisenstein, L. Verde, D. N. Spergel, R. A. Skibba, N. A. Bahcall, T. Budavari, J. A. Frieman, M. Fukugita, J. R. Gott, J. E. Gunn, Ž. Ivezić, G. R. Knapp, R. G. Kron, R. H. Lupton, T. A. McKay, A. Meiksin, R. C. Nichol, A. C. Pope, D. J. Schlegel, D. P. Schneider, C. Stoughton, M. A. Strauss, A. S. Szalay, M. Tegmark, M. S. Vogeley, D. H. Weinberg, D. G. York, and I. Zehavi. Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. *MNRAS*, 404 :60–85, May 2010. 25, 71, 119
- A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116 :1009–1038, 1998. 7, 12, 16
- M. Robberto, A. Cimatti, and Space Science Team. An overview of the SPACE mission proposal. *Nuovo Cimento B Serie*, 122 :1467–1471, December 2007. 25
- A. G. Sánchez, C. M. Baugh, and R. Angulo. What is the best way to measure baryonic acoustic oscillations? *MNRAS*, 390 :1470–1490, November 2008. XII, 23, 25, 26, 27, 36, 37, 114
- A. G. Sánchez, M. Crocce, A. Cabré, C. M. Baugh, and E. Gaztañaga. Cosmological parameter constraints from SDSS luminous red galaxies : a new treatment of large-scale clustering. *MNRAS*, 400 :1643–1664, December 2009. 27, 71, 114, 119
- A. G. Sánchez, C. G. Scóccola, A. J. Ross, W. Percival, M. Manera, F. Montesano, X. Mazzalay, A. J. Cuesta, D. J. Eisenstein, E. Kazin, C. K. McBride, K. Mehta, A. D. Montero-Dorta, N. Padmanabhan, F. Prada, J. A. Rubiño-Martín, R. Tojeiro, X. Xu, M. V. Magaña, E. Aubourg, N. A. Bahcall, S. Bailey, D. Bizyaev, A. S. Bolton, H. Brewington, J. Brinkmann, J. R. Brownstein,



- J. R. Gott, J. C. Hamilton, S. Ho, K. Honscheid, A. Labatie, E. Malanushenko, V. Malanushenko, C. Maraston, D. Muna, R. C. Nichol, D. Oravetz, K. Pan, N. P. Ross, N. A. Roe, B. A. Reid, D. J. Schlegel, A. Shelden, D. P. Schneider, A. Simmons, R. Skibba, S. Snedden, D. Thomas, J. Tinker, D. A. Wake, B. A. Weaver, D. H. Weinberg, M. White, I. Zehavi, and G. Zhao. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey : cosmological implications of the large-scale two-point correlation function. *MNRAS*, page 3424, July 2012. XIV, 13, 59, 60, 62, 63
- P. Sarkar, J. Yadav, B. Pandey, and S. Bharadwaj. The scale of homogeneity of the galaxy distribution in SDSS DR6. *MNRAS*, 399 :L128–L131, 2009. XIII, 48
- W. Saunders, W. J. Sutherland, S. J. Maddox, O. Keeble, S. J. Oliver, M. Rowan-Robinson, R. G. McMahon, G. P. Efstathiou, H. Tadros, S. D. M. White, C. S. Frenk, A. Carramiñana, and M. R. S. Hawkins. The PSCz catalogue. *MNRAS*, 317 :55–63, September 2000. 51
- R. Scoccimarro and R. K. Sheth. PTHALOS : a fast method for generating mock galaxy distributions. *MNRAS*, 329 :629–640, January 2002. 61
- R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain. How Many Galaxies Fit in a Halo ? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *ApJ*, 546 :20–34, January 2001. 37
- M. Scrimgeour, T. Davis, C. Blake, J. B. James, G. Poole, L. Staveley-Smith, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladsters, K. Glazebrook, B. Jelliffe, R. J. Jurek, I. Li, B. Madore, C. Martin, K. Pimblett, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. Wyder, and H. Yee. The WiggleZ Dark Energy Survey : the transition to large-scale cosmic homogeneity. *ArXiv e-prints*, May 2012. XIII, 48, 49
- U. Seljak. Analytic model for galaxy and dark matter clustering. *MNRAS*, 318 :203–213, October 2000. 25
- U. Seljak and M. Zaldarriaga. A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469 :437, October 1996. 22
- H.-J. Seo and D. J. Eisenstein. Baryonic Acoustic Oscillations in Simulated Galaxy Redshift Surveys. *ApJ*, 633 :575–588, November 2005. 113
- S. A. Shectman, S. D. Landy, A. Oemler, D. L. Tucker, H. Lin, R. P. Kirshner, and P. L. Schechter. The Las Campanas Redshift Survey. *ApJ*, 470 :172, October 1996. 51
- R. K. Sheth, H. J. Mo, and G. Tormen. Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes. *MNRAS*, 323 :1–12, May 2001. 36
- R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman. Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, 341 :1311–1332, June 2003. XII, 24, 25, 77, 96
- T. Sousbie. Le squelette de l’univers : Un outil d’analyse topologique des grandes structures. 2007. URL <http://tel.archives-ouvertes.fr/tel-00162559/en/>. XI, 9
- V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435 :629–636, June 2005. XII, 33, 34
- J.-L. Starck, F. Murtagh, and J. Fadili. *Sparse Image and Signal Processing : Wavelets, Curvelets, Morphological Diversity*. Cambridge University Press, 2010. ISBN 9780521119139. 88
- D. Stoyan, W. S. Kendall, and J. Mecke. *Stochastic geometry and its applications*. John Wiley & Sons, Chichester, 1995. 35, 63, 79
- M. A. Strauss, D. H. Weinberg, R. H. Lupton, V. K. Narayanan, J. Annis, M. Bernardi, M. Blanton, S. Burles, A. J. Connolly, J. Dalcanton, M. Doi, D. Eisenstein, J. A. Frieman, M. Fukugita, J. E. Gunn, Ž. Ivezić, S. Kent, R. S. J. Kim, G. R. Knapp, R. G. Kron, J. A. Munn, H. J. Newberg, R. C.

- Nichol, S. Okamura, T. R. Quinn, M. W. Richmond, D. J. Schlegel, K. Shimasaku, M. SubbaRao, A. S. Szalay, D. Vanden Berk, M. S. Vogeley, B. Yanny, N. Yasuda, D. G. York, and I. Zehavi. Spectroscopic Target Selection in the Sloan Digital Sky Survey : The Main Galaxy Sample. *AJ*, 124 :1810–1824, September 2002. 53
- N. Suzuki, D. Rubin, C. Lidman, G. Aldering, R. Amanullah, K. Barbary, L. F. Barrientos, J. Botyanszki, M. Brodwin, N. Connolly, K. S. Dawson, A. Dey, M. Doi, M. Donahue, S. Deustua, P. Eisenhardt, E. Ellingson, L. Faccioli, V. Fadeyev, H. K. Fakhouri, A. S. Fruchter, D. G. Gilbank, M. D. Gladders, G. Goldhaber, A. H. Gonzalez, A. Goobar, A. Gude, T. Hattori, H. Hoekstra, E. Hsiao, X. Huang, Y. Ihara, M. J. Jee, D. Johnston, N. Kashikawa, B. Koester, K. Konishi, M. Kowalski, E. V. Linder, L. Lubin, J. Melbourne, J. Meyers, T. Morokuma, F. Munshi, C. Mullis, T. Oda, N. Panagia, S. Perlmutter, M. Postman, T. Pritchard, J. Rhodes, P. Riposte, P. Rosati, D. J. Schlegel, A. Spadafora, S. A. Stanford, V. Stanishev, D. Stern, M. Strovink, N. Takahashi, K. Tokita, M. Wagner, L. Wang, N. Yasuda, H. K. C. Yee, and T. Supernova Cosmology Project. The Hubble Space Telescope Cluster Supernova Survey. V. Improving the Dark-energy Constraints above  $z \sim 1$  and Building an Early-type-hosted Supernova Sample. *ApJ*, 746 :85, February 2012. XI, 16
- F. Sylos Labini, M. Montuori, and L. Pietronero. Scale-invariance of galaxy clustering. *Physics Reports*, 293 :61–226, 1998. 73
- F. Sylos Labini, N. L. Vasilyev, Y. V. Baryshev, and M. López-Corredoira. Absence of anti-correlations and of baryon acoustic oscillations in the galaxy correlation function from the Sloan Digital Sky Survey data release 7. *A&A*, 505 :981–990, October 2009. 73
- A. Taruya, T. Nishimichi, S. Saito, and T. Hiramatsu. Nonlinear evolution of baryon acoustic oscillations from improved perturbation theory in real and redshift spaces. *Phys. Rev. D*, 80(12) :123503, December 2009. 25
- M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton, J. A. Frieman, M. Fukugita, J. E. Gunn, A. J. S. Hamilton, G. R. Knapp, R. C. Nichol, J. P. Ostriker, N. Padmanabhan, W. J. Percival, D. J. Schlegel, D. P. Schneider, R. Scoccimarro, U. Seljak, H.-J. Seo, M. Swanson, A. S. Szalay, M. S. Vogeley, J. Yoo, I. Zehavi, K. Abazajian, S. F. Anderson, J. Annis, N. A. Bahcall, B. Bassett, A. Berlind, J. Brinkmann, T. Budavari, F. Castander, A. Connolly, I. Csabai, M. Doi, D. P. Finkbeiner, B. Gillespie, K. Glazebrook, G. S. Hennessy, D. W. Hogg, Ž. Ivezić, B. Jain, D. Johnston, S. Kent, D. Q. Lamb, B. C. Lee, H. Lin, J. Loveday, R. H. Lupton, J. A. Munn, K. Pan, C. Park, J. Peoples, J. R. Pier, A. Pope, M. Richmond, C. Rockosi, R. Scranton, R. K. Sheth, A. Stebbins, C. Stoughton, I. Szapudi, D. L. Tucker, D. E. vanden Berk, B. Yanny, and D. G. York. Cosmological constraints from the SDSS luminous red galaxies. *Phys. Rev. D*, 74(12) :123507, December 2006. 25, 71, 119
- H. J. Tian, M. C. Neyrinck, T. Budavári, and A. S. Szalay. Redshift-space Enhancement of Line-of-sight Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Main-galaxy Sample. *ApJ*, 728 :34, February 2011. XV, 88, 89, 90, 101
- H. Tomita. *Dynamics and Statistics of Patterns, Vol.1, (World Scientific)*, 113, volume 1, pages 113–157. World Scientific, 1990. 33
- C. G. Tsagas. Peculiar motions, accelerated expansion, and the cosmological axis. *Phys. Rev. D*, 84 (6) :063503, September 2011. 12
- M. Vargas Magaña, J. Bautista, J.-Ch. Hamilton, N.G. Busca, E. Aubourg, A. Labatie, and J.-M. Le Goff. An optimized correlation function estimator for galaxy surveys. *A&A*, in prep. 84
- M. White, M. Blanton, A. Bolton, D. Schlegel, J. Tinker, A. Berlind, L. da Costa, E. Kazin, Y.-T. Lin, M. Maia, C. K. McBride, N. Padmanabhan, J. Parejko, W. Percival, F. Prada, B. Ramos, E. Sheldon, F. de Simoni, R. Skibba, D. Thomas, D. Wake, I. Zehavi, Z. Zheng, R. Nichol, D. P. Schneider, M. A. Strauss, B. A. Weaver, and D. H. Weinberg. The Clustering of Massive Galaxies at  $z \sim 0.5$  from the First Semester of BOSS Data. *ApJ*, 728 :126, February 2011. 60, 98

- X. Xu, M. White, N. Padmanabhan, D. J. Eisenstein, J. Eckel, K. Mehta, M. Metchnik, P. Pinto, and H.-J. Seo. A New Statistic for Analyzing Baryon Acoustic Oscillations. *ApJ*, 718 :1224–1234, August 2010. 88
- X. Xu, N. Padmanabhan, D. J. Eisenstein, K. T. Mehta, and A. J. Cuesta. A 2% Distance to  $z=0.35$  by Reconstructing Baryon Acoustic Oscillations - II : Fitting Techniques. *ArXiv e-prints*, January 2012. 59, 89, 116, 128, 131
- J. Yadav, S. Bharadwaj, B. Pandey, and T. R. Seshadri. Testing homogeneity on large scales in the Sloan Digital Sky Survey Data Release One. *MNRAS*, 364 :601–606, 2005. 48
- I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth, M. A. Strauss, I. Kayo, Y. Suto, M. Fukugita, O. Nakamura, N. A. Bahcall, J. Brinkmann, J. E. Gunn, G. S. Hennessy, Ž. Ivezić, G. R. Knapp, J. Loveday, A. Meiksin, D. J. Schlegel, D. P. Schneider, I. Szapudi, M. Tegmark, M. S. Vogeley, D. G. York, and SDSS Collaboration. The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, 630 :1–27, September 2005. 54
- Y. B. Zel'dovich. Gravitational instability : An approximate theory for large density perturbations. *A&A*, 5 :84–89, March 1970. 23
- Z. Zheng, I. Zehavi, D. J. Eisenstein, D. H. Weinberg, and Y. P. Jing. Halo Occupation Distribution Modeling of Clustering of Luminous Red Galaxies. *ApJ*, 707 :554–572, December 2009. 53
- F. Zwicky. On the Masses of Nebulae and of Clusters of Nebulae. *AJ*, 86 :23, 1937. 11