

1

Développement des mocks

Dans ce chapitre, nous présentons la construction des *mocks* : des spectres de quasars simulés, dont les forêts Ly α sont corrélées entre elles et avec le champ de quasars sous-jacent. Ces mocks visent à reproduire les données d’eBOSS et de DESI. Ils sont nommés **SaclayMocks** et présentés dans ?. Le code est écrit en Python¹ et se trouve en accès libre sur GitHub². L’utilisation de ces mocks et leur validation seront présentés dans les chapitres suivants.

1 Objectifs des mocks

Contrairement à ce qu’on appelle les simulations, les mocks ne contiennent pas de physique à proprement parler : ils ne sont pas utilisés afin de déduire des paramètres astrophysiques. Certaines simulations, les simulations hydrodynamiques, permettent de mesurer des effets astrophysiques, comme par exemple le biais de l’hydrogène ou du Ly α . Mais ces simulations sont très coûteuses car elles nécessitent de modéliser les effets physiques qui affectent les paramètres mesurés. Les mocks, quant à eux, sont conçus afin de répliquer rapidement un jeu de données, dans le but de tester l’analyse qui sera appliquée sur ces données. Les mocks sont donc utilisés afin

- de vérifier la mesure des paramètres $\alpha_{\parallel}, \alpha_{\perp}$: cette mesure est-elle non biaisée ?
- d’identifier les potentielles systématiques : la présence de métaux et d’HCD dans les données est-elle bien modélisée ? Affecte-t-elle la mesure de $\alpha_{\parallel}, \alpha_{\perp}$?
- de tester la matrice de distorsion : la distorsion de la fonction de corrélation due à l’ajustement du continuum du quasar est-elle correctement prise en compte par la matrice de distorsion ?
- de vérifier l’estimation de la matrice de covariance : la matrice de covariance, calculée à partir des données, est-elle bien estimée ?

La production et l’analyse d’un grand nombre de mocks permet de répondre précisément à ces questions. Ces mocks sont donc nécessaires pour pouvoir valider l’analyse menée sur les données.

Les mocks décrits dans ce manuscrit s’inscrivent dans les projets eBOSS et DESI. Ils sont utilisés dans l’analyse Ly α des données complète d’eBOSS (?), et seront utilisés dans l’analyse Ly α de DESI. L’objectif de ces mocks est donc de répliquer au mieux les données Ly α d’eBOSS et de DESI. Ces relevés couvrent un volume de plusieurs dizaines de Gpc³, et les échelles sondées grâce au Ly α descendent jusqu’à la centaine de kpc. Les mocks nécessitent donc de reproduire un volume immense, avec une bonne résolution. Les simulations dites *N-corps* sont des simulations qui traitent le problème à N corps. **Elles sont initialisées à un redshift élevé ($z > 100$), avec une distribution de matière noire représentée par des macro-particules. Généralement, environ $(10\,000)^3$ macro-particules sont présentes dans ces simulations. La masse de ces particules dépend du volume simulé. Par exemple, pour un volume d’environ $(4\,\text{Gpc})^3$, les particules possèdent une masse $\sim 10^9 M_{\odot}$ (Heitmann et al., 2019).** Puis, à chaque pas de temps, ces macro-particules sont déplacées en considérant uniquement les interactions gravitationnelles. Le champ de matière initial évolue ainsi jusqu’à $z = 0$. Ces simulations sont très utiles pour étudier les effets de la gravité à grande échelle. Cependant elles ne sont pas adaptées à notre utilisation : afin d’avoir la résolution et le volume requis, la simulation nécessiterait beaucoup trop de macro-particules pour être réalisable dans un temps raisonnable.

Les simulations hydrodynamiques fonctionnent de la même manière que les simulations N-corps. Elles incluent, en plus des macro-particules de matière noire, la physique baryonique présente dans le milieu galactique. Les baryons sont aussi représentés par des macro-particules. Afin de résoudre l’intérieur des galaxies, les macro-particules utilisées possèdent une masse plus faible que dans le cas

1. <https://www.python.org/>

2. <https://github.com/igmhub/SaclayMocks>

des simulations N-corps. En contrepartie, le volume simulé est plus petit. Dans le cas des simulations hydrodynamiques, la densité, la pression et la température sont tracées dans chaque cellule. Certains effets astrophysiques, comme les supernovae ou les AGN peuvent aussi être présents. Ces simulations nécessitent encore plus de temps de calcul que les simulations à N-corps. Elles ne sont donc pas non plus adaptées à notre utilisation.

Ainsi, seuls les *champs aléatoires gaussiens* (GRF pour Gaussian Random Field) permettent de générer un grand volume avec une grande résolution. Ce sont des champs qui en chaque point prennent une valeur aléatoire selon une statistique gaussienne. Une fois générés, il est possible de donner à ces champs n'importe quelle fonction de corrélation, en utilisant la transformation de Fourier. Ces champs sont utilisés notamment dans les simulations à N-corps, afin de fournir les conditions. Cependant, l'utilisation des GRF ne donne pas accès aux non linéarités qui peuvent émerger dans l'évolution des simulations N-corps et hydrodynamiques. La seule information provient de la fonction de corrélation que l'on applique au GRF. Mais cela est entièrement suffisant pour l'utilisation que nous en avons dans ce manuscrit : nous générons un champ gaussien destiné à simuler le champ d'absorption Ly α et le champ de quasars, dont les fonctions d'auto corrélation et de corrélation croisée sont choisies afin de correspondre à ce qui est observé dans les données.

2 Construction des mocks

Dans cette section, nous détaillons comment les mocks sont générés. Nous présentons d'abord la génération des champs de densité, puis comment les quasars sont tirés à partir de ces champs de densité. Nous expliquons ensuite comment nous calculons la densité le long de la ligne de visée de chaque quasar. Enfin, nous présentons comment cette densité est transformée en fraction de flux transmis, et comment nous tirons les HCD.

2.1 Les champs de densité

La première étape dans la création des mocks est de générer les boîtes qui contiennent le champ de densité δ . D'abord, un GRF est généré dans une boîte de $2560 \times 2560 \times 1536$ voxels, chaque voxel faisant $d_{cell}^3 = (2,19 h^{-1} \text{ Mpc})^3$. Afin que le champ δ possède la bonne fonction de corrélation, une transformation de Fourier 3D¹ est appliquée sur la boîte, puis la boîte δ_k ainsi obtenue est multipliée par

$$\sqrt{\frac{P(k)}{d_{cell}^3}}, \quad (1.1)$$

où $P(k)$ est le spectre de puissance désiré. Il est ensuite possible d'obtenir la boîte δ dans l'espace réel grâce à une transformation de Fourier inverse de la boîte δ_k . Ce procédé garanti que le champ δ suive le spectre de puissance $P(k)$. Le GRF pourrait être tiré directement dans l'espace k . Dans ce cas, il nous faut tirer deux champs gaussiens : un pour la partie réelle, et un autre pour la partie imaginaire. La transformation de Fourier prenant moins de temps que la génération du champ aléatoire, nous préférons générer le champ dans l'espace réel plutôt que dans l'espace k . Dans la suite nous décrivons les différentes boîtes nécessaires à la construction des mocks : les boîtes champs utilisées pour tirer les quasars, la boîte utilisée pour créer l'absorption Ly α , ainsi que les boîtes de vitesse et de gradient de vitesse. Afin de garantir leur corrélation, toutes ces boîtes sont construites à partir de la même boîte initial δ_k .

1. Nous utilisons la librairie pyFFTW (<https://github.com/pyFFTW/pyFFTW>), une adaptation python de la librairie FFTW (<http://www.fftw.org/>).

Les quasars

Afin de construire un relevé de quasars corrélés, nous tirons les quasars selon le champ dans l'espace réel δ_{QSO} , construit à partir de δ_k . Une première solution serait de tirer les quasars dans les cellules dont le champ δ_{QSO} est supérieur à un certain seuil. Cette solution produit une fonction de corrélation correcte aux grandes échelles, mais pas aux petites. En effet, comme montré dans ?, les objets tirés aux endroits où la densité est supérieure à un certain seuil suivent la fonction de corrélation de la densité sous-jacente, avec un biais qui dépend du seuil choisi, si et seulement si la fonction de corrélation est petite devant 1. La fonction de corrélation ainsi obtenue est correcte, sauf pour les petites échelles pour lesquelles la fonction de corrélation est importante. Une solution alternative consiste à considérer que les quasars suivent une distribution log-normale. Ceci permet d'obtenir une meilleure corrélation aux petites échelles. Ce choix est souvent fait pour simuler des relevés de galaxies (Agrawal et al., 2017), et est en accord avec ce qui est observé dans les données (Clerkin et al., 2016). Ainsi, dans chaque voxel, les quasars sont tirés avec une probabilité

$$P \propto e^{\delta_q}, \quad (1.2)$$

où δ_q est le champ de densité dans le voxel considéré. Comme montré par Coles and Jones (1991), afin que les quasars suivent la fonction de corrélation $\xi(r)$, le champ δ_q doit suivre la fonction de corrélation

$$\xi_q(r) = \ln(1 + \xi(r)). \quad (1.3)$$

Nous verrons section 2.2 que, de manière à obtenir un relevé synthétique de quasars dont le biais dépend de z , nous utilisons trois boîtes qui suivent des distributions log-normales, à des redshifts différents. La probabilité pour tirer les quasars dépend de l'interpolation de ces 3 boîtes. Pour chacune des boîtes, nous partons du spectre de puissance de la matière $P_{\text{matière}}(k)$ à $z = 0$, fournit par Camb (Lewis et al., 1999). Nous multiplions ensuite ce spectre de puissance par $(b_{\text{QSO}}(z_i)G(z_i))^2$, où $i \in [1, 2, 3]$. À l'aide de la transformation de Fourier, nous calculons la fonction de corrélation $\xi_i(r)$. Puis, nous déterminons le spectre de puissance $P_{\text{QSO},i}(k)$, à appliquer à la boîte δ_k , comme la transformée de Fourier de $\xi_{\text{QSO},i}(r) = \ln(1 + \xi_i(r))$ (équation ??). Une fois les trois spectres de puissances $P_{\text{QSO},i}(k)$ obtenus, nous construisons 3 boîtes

$$\delta_{k,i}(k) = \delta_k(k) \sqrt{\frac{P_{\text{QSO},i}(k)}{V_{\text{cell}}}}, \quad (1.4)$$

où δ_k est le GRF dans l'espace de Fourier. Une fois ces 3 boîtes construites, nous appliquons à chacune d'entre elle une transformation de Fourier inverse afin d'obtenir les boîtes $\delta_{\text{QSO},i}$. Ces boîtes seront interpolées en z , puis les quasars seront ensuite tirés avec une probabilité $\propto \exp(\delta_{\text{QSO}}(z))$, où δ_{QSO} est la boîte interpolée. Nous expliquons cette étape dans la section 2.2.

Le champ Ly α

Afin de construire le champ d'absorption Ly α , nous avons besoin du champ de densité de l'hydrogène neutre. Comme expliqué dans la section ??, la fraction de flux transmis F est reliée à la profondeur optique τ par

$$F = \exp(-\tau). \quad (1.5)$$

De plus, la formule FGPA (Fluctuating Gunn Peterson Approximation) permet de relier la profondeur optique τ au contraste de densité δ à $z = 0$:

$$\tau(z) = a(z)e^{b(z)G(z)\delta} \quad (1.6)$$

Les paramètres a et b sont des paramètres à ajuster afin d'obtenir le bon biais du Ly α et la bonne transmission moyenne \overline{F} . Leur détermination est décrite dans la section 4. Le facteur de croissance G prend en compte l'évolution avec le redshift du champ de densité δ . Ainsi il nous suffit de construire un GRF qui suit la fonction de corrélation à $z = 0$ pour simuler le champ d'absorption du Ly α . Pour ce faire, nous partons de la même boîte δ_k utilisée pour construire les 3 boîtes log-normales des quasars. Ceci garanti la corrélation croisée entre les quasars et le champ d'absorption Ly α . Le spectre de puissance de la matière $P_{matière}(k)$ à $z = 0$ est ensuite appliqué à la boîte δ_k . Enfin, nous obtenons la boîte de densité $\delta_{matière}$ à $z = 0$ qui servira au calcul du champ d'absorption du Ly α en effectuant la transformée de Fourier de la boîte

$$\delta_{k,matière}(\vec{k}) = \delta_k(\vec{k}) \sqrt{\frac{P_{matière}(\vec{k})}{V_{cell}}} . \quad (1.7)$$

Les champs des vitesses

Afin d'inclure les RSD dans nos mocks, nous simulons aussi le champ des vitesses. A l'ordre linéaire, le champ des vitesses $v_{k,n}$ dans l'espace k selon la direction \vec{u}_n , avec $n \in [X, Y, Z]$, est relié au champ de densité δ_k par la relation (#prov mettre la démo? voir dodelson)

$$v_{k,n}(\vec{k}) = \frac{ik_n}{k^2} \delta_k(\vec{k}) . \quad (1.8)$$

Il est fréquent (#prov approximation réaliste? mettre une source; j'ai demandé a Etienne : theorie lineaire pas de biais, c'est ce qui est utilisé pour la reconstruction. Y a des gens qui regardent dans les simu nbody si y a un biais des vitesses qui apparait, dans les simus eBOSS y a rien de convainquant, mais le papier est tres peu fournit a ce sujet) de considérer que le champ de vitesse des traceurs est le même que celui de la matière sous-jacente. Autrement dit, le champ de vitesse des traceurs est non biaisé. En ce qui concerne les quasars, nous simulons les RSD en déplaçant chaque quasar proportionnellement à sa vitesse le long de la ligne de visée (équation ??). Dans ce but, nous calculons les trois boîtes de vitesses $v_{k,x}$, $v_{k,y}$ et $v_{k,z}$ à $z = 0$, comme

$$v_{k,n}(\vec{k}) = \frac{-ik_n}{k^2} H_0 \frac{dG}{dz} \delta_{k,matière}(\vec{k}) ; \quad n \in [X, Y, Z] . \quad (1.9)$$

Cette équation est équivalente à l'équation 1.8 pour $z = 0$. Comme précédemment, la boîte $\delta_{k,matière}$ est la même que celle utilisée pour construire les boîtes $\delta_{QSO,i}$ des quasars et la boîte $\delta_{matière}$ utilisée pour le Ly α , ceci afin de garantir la corrélation entre la densité des traceurs et leurs vitesses particulières. A l'aide d'une transformation de Fourier inverse, nous obtenons les trois boîtes de vitesse à $z = 0$ dans l'espace réel v_x , v_y et v_z . Le calcul pour obtenir la vitesse parallèle $v_{||}$ à un redshift z est décrit dans la section 2.2.

Concernant le champ d'absorption Ly α , les RSD sont prises en compte par une modification de la formule FGPA. Pour ce faire, nous avons besoin du gradient de vitesse $\eta_{||}$ à $z = 0$. Le gradient η_{nm} selon la direction \vec{u}_m de la vitesse v_n est défini comme

$$\eta_{nm}(\vec{k}) = \frac{k_n k_m}{k^2} \delta_k(\vec{k}) . \quad (1.10)$$

Cette équation permet de retrouver la formule de kaiser :

$$\begin{aligned} \delta_k^s(\vec{k}) &= \delta_k(\vec{k}) + \eta_{||}(\vec{k}) , \\ &= (1 + f \mu_k^2) \delta_k(\vec{k}) . \end{aligned} \quad (1.11)$$

La boîte δ_k utilisée est le GRF initial, afin de garantir les corrélations entre les différents champs. A l'aide d'une transformation de Fourier, nous obtenons¹ les 6 boîtes de gradients de vitesses à $z = 0$ dans l'espace réel η_{xx} , η_{yy} , η_{zz} , η_{xy} , η_{yz} et η_{xz} .

2.2 Le relevé de quasars

Une fois toutes ces boîtes construites, nous définissons la géométrie du relevé. Les boîtes, d'une taille $2560 \times 2560 \times 1536$ selon les axes x , y et z respectivement, sont placées à un redshift central $z_0 = 1.71$, et à une ascension droite α_0 et une déclinaison δ_0 (voir équation 1.18). Leurs dimensions permettent de couvrir les redshifts $1.3 < z < 3.6$. L'observateur est considéré être à $z = 0$, au centre du plan (x, y) .

Afin d'obtenir un biais des quasars qui dépend du redshift, nous pouvons utiliser deux champs log-normaux. Considérons $\delta_{\text{QSO},1}$ et $\delta_{\text{QSO},2}$, deux champs log-normaux à $z_1 = 2,1$ et $z_2 = 3,5$. Pour les redshifts $z < z_1$ et $z > z_2$, nous pouvons extrapoler les champs $\delta_{\text{QSO},1}$ et $\delta_{\text{QSO},2}$, en calculant

$$\hat{\delta}_{\text{QSO},i}(z) = \exp\left(\delta_{\text{QSO},i} \frac{b_{\text{QSO}}(z)(1+z_i)}{b_{\text{QSO}}(z_i)(1+z)}\right), \quad i \in [1;2], \quad (1.12)$$

où b_{QSO} est le biais des quasars. Le redshift dans chaque voxel est calculé en utilisant l'équation ???. Toutes les distances sont comobiles. L'équation précédente nous donne donc le champ $\hat{\delta}_{\text{QSO},1}$ à utiliser pour tirer les quasars lorsque $z \leq z_1$, et le champ $\hat{\delta}_{\text{QSO},2}$ à utiliser lorsque $z \geq z_2$. Pour les redshifts $z_1 < z < z_2$, nous pouvons interpoler les deux champs $\delta_{\text{QSO},1}$ et $\delta_{\text{QSO},2}$ en calculant

$$\hat{\delta}_{\text{QSO},12}(z) = \hat{\delta}_{\text{QSO},1}(z) \frac{z_2 - z}{z_2 - z_1} + \hat{\delta}_{\text{QSO},2}(z) \frac{z - z_1}{z_2 - z_1}. \quad (1.13)$$

Nous pouvons alors tirer les quasars dans chaque cellule proportionnellement à l'interpolation ou l'extrapolation de ces deux champs. La figure 1.1 présente les fonctions de corrélation des champs $\hat{\delta}_{\text{QSO},12}$ à $z = 2,6$ (vert) et à $z = 2,9$ (violet), $\hat{\delta}_{\text{QSO},1}$ à $z = 1,9$ (bleu) et $\hat{\delta}_{\text{QSO},2}$ à $z = 3,6$ (rouge), et $\delta_{\text{QSO},1}$ et $\delta_{\text{QSO},2}$ à $z_1 = 2,1$ et $z_2 = 3,5$ (orange). Ces fonctions de corrélations sont corrigées de leur dépendance en redshift en divisant ξ par

$$\left(\frac{b_{\text{QSO}}(z)(1+z_0)}{b_{\text{QSO}}(z_0)(1+z)}\right)^2, \quad (1.14)$$

avec $z_0 = 2.3$. Sur le graphique de droite de la figure 1.1, nous voyons que les fonctions de corrélation obtenues par interpolation possèdent une amplitude trop grande, et celles obtenues par extrapolation une amplitude trop petite. De plus, nous pouvons voir sur la figure 1.2 que l'effet dû à l'interpolation (en bleu) est similaire et opposé à celui dû à l'extrapolation (orange). Ainsi, en combinant une interpolation et une extrapolation, nous obtenons une approximation bien plus satisfaisante (vert). Plutôt que d'utiliser deux champs, nous utilisons donc trois champs log-normaux à $z_1 = 1,9$, $z_2 = 2,75$, et $z_3 = 3,6$. Ces champs sont stockés dans les boîtes construites dans la section 2.1. Nous construisons les trois boîtes $\hat{\delta}_{\text{QSO},i}$ définies selon l'équation 1.12, puis, nous créons les deux boîtes qui contiennent les champs interpolés $\hat{\delta}_{\text{QSO},12}$ et $\hat{\delta}_{\text{QSO},23}$ (équation 1.13). Enfin, nous construisons la boîte

$$\hat{\delta}_{\text{QSO}}(z) = K(z) \left(\hat{\delta}_{\text{QSO},12}(z) - \hat{\delta}_{\text{QSO},23}(z) \right) + \hat{\delta}_{\text{QSO},23}(z), \quad (1.15)$$

où $K(z)$ est un coefficient qui varie entre 0 et 1. Il est représenté sur la figure ??. Pour chaque z , $K(z)$ est déterminé tel que $K(z)(r_{12} - r_{23}) + r_{23}$ vale 1 à $r = 5 h^{-1}$ Mpc, où les rapports r_{12} et r_{23} sont

1. Lors de la construction des 6 boîtes de gradients de vitesses, nous omettons volontairement le facteur $f(z=0)$ de l'équation 1.10. Ce facteur f manquant sera pris en compte lors de l'ajout de la dépendance en redshift (voir section 2.4)

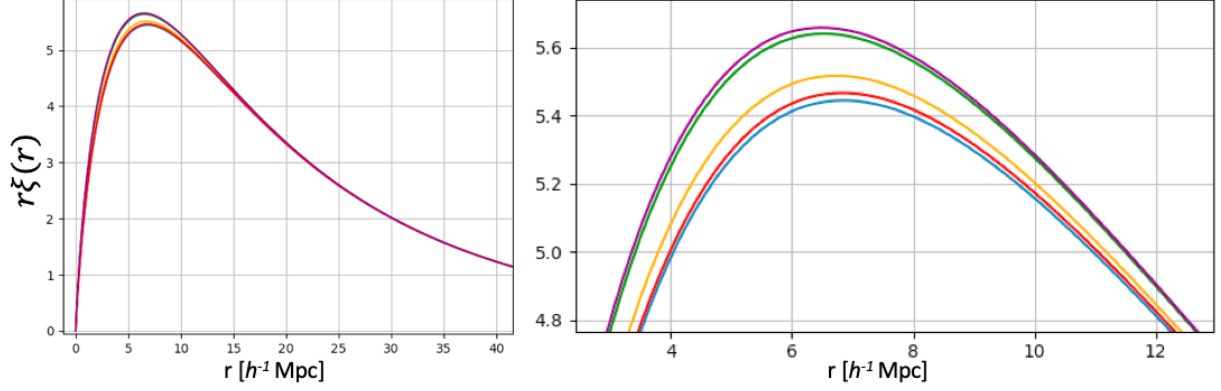


FIGURE 1.1 – Fonctions de corrélation des champs $\hat{\delta}_{\text{QSO},12}$ à $z = 2,6$ (vert) et à $z = 2,9$ (violet), $\hat{\delta}_{\text{QSO},1}$ à $z = 1,9$ (bleu) et $\hat{\delta}_{\text{QSO},2}$ à $z = 3,6$ (rouge), et $\delta_{\text{QSO},1}$ et $\delta_{\text{QSO},2}$ à $z_1 = 2,1$ et $z_2 = 3,5$ (orange) corrigées de leur dépendance en redshift (voir texte). Le graphique de droite présente un zoom de celui de gauche.

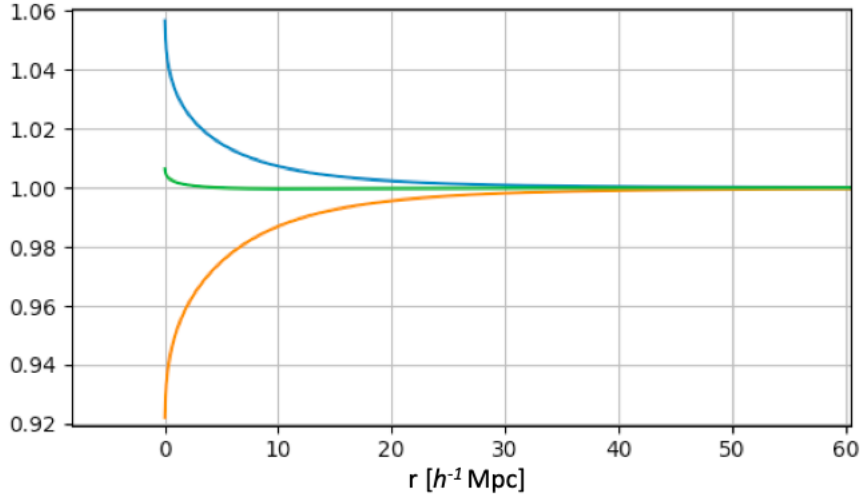
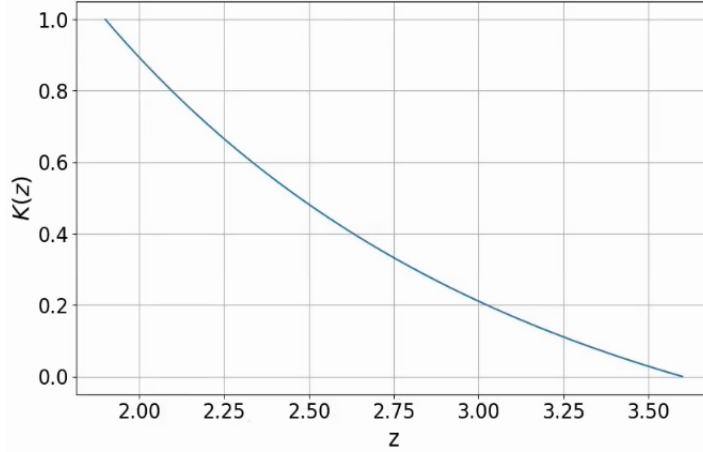


FIGURE 1.2 – Rapports r_{12} (bleu), r_{23} (orange) et $K(z)(r_{12} - r_{23}) + r_{23}$ (vert) à $z = 2,3$ (voir texte). L'effet sur la fonction de corrélation dû à interpolation (bleu) compense celui dû à l'extrapolation (orange). L'ajout des deux donne la courbe verte.


 FIGURE 1.3 – Coefficient $K(z)$ défini dans l'équation 1.15.

définis comme

$$r_{12}(z) = \frac{\xi_{12}(r)}{\xi_1} \left(\frac{b_1(1+z)}{b_{QSO}(z)(1+z_1)} \right)^2, \quad (1.16)$$

$$r_{23}(z) = \frac{\xi_{23}(r)}{\xi_1} \left(\frac{b_1(1+z)}{b_{QSO}(z)(1+z_1)} \right)^2. \quad (1.17)$$

Ces rapports donnent la déviation de la fonction de corrélation due à l'interpolation et à l'extrapolation. La figure 1.2 donne r_{12} (bleu), r_{23} (orange) et $K(z)(r_{12} - r_{23}) + r_{23}$ (vert) à $z = 2, 3$. La courbe verte montre que la fonction de corrélation obtenue avec le champ $\hat{\delta}_{QSO}$ ne dévie pas de plus de 0,6 % de la fonction de corrélation visée à chaque redshift. Les quasars sont ensuite tirés dans chaque voxel, avec une probabilité $P \propto \hat{\delta}_{QSO}$. Pour ce faire, nous générons une variable ϕ aléatoire uniforme entre 0 et 1 dans chaque voxel. Les voxels pour lesquelles $\phi < N(z)\hat{\delta}_{QSO}$ hébergent un quasar. **Le facteur $N(z)$ contient l'évolution avec le redshift du nombre de quasars par degré carré. Il contient aussi un facteur de normalisation.** Les quasars dont le redshift est en dehors de l'intervalle $[1,8; 3,6]$ sont écartés. Les quasars dont l'ascension droite et la déclinaison sont en dehors des intervalles $[-\Delta\alpha; \Delta\alpha]$ et $[-\Delta\delta; \Delta\delta]$ sont aussi écartés. L'ascension droite α et la déclinaison δ du point (x, y, z) sont définies comme

$$\alpha = \arctan\left(\frac{\cos(\alpha_0)X - \sin(\delta_0)\sin(\alpha_0)Y + \cos(\delta_0)\sin(\alpha_0)Z}{-\sin(\alpha_0)X - \sin(\delta_0)\cos(\alpha_0)Y + \cos(\delta_0)\cos(\alpha_0)Z}\right), \quad (1.18)$$

$$\delta = \arcsin\left(\frac{\cos(\delta_0)Y + \sin(\delta_0)Z}{\sqrt{X^2 + Y^2 + Z^2}}\right). \quad (1.19)$$

Enfin, grâce au facteur $N(z)$, les quasars sont tirés selon la distribution en z prédite pour DESI. Cette distribution est présentée sur la figure ???. Cependant, nous dirons environ deux fois plus de quasars, afin de pouvoir simuler, entre autre, la sélection des cibles à l'aide du code quickquasars (présenté dans la section ??). De plus, cela permet d'utiliser les mocks pour d'autres projets, comme WEAVE¹, qui possèdent une densité de cible plus élevée que DESI (#prov pas tellement en fait : 350 000 sur 6000 sq deg ca fait 60 par sq deg). A la fin, nous obtenons environ 100 quasars à $z > 2,1$ par degré carré.

1. WEAVE est un spectrographe à fibre multi objet. Il possède 1000 fibres avec un champ de vue de $3,1 \text{ deg}^2$. Le relevé WEAVE-QSO prévoit d'observer 350 000 quasars à grand redshift, sur un relevé de 6000 deg^2 (Pieri et al., 2016).

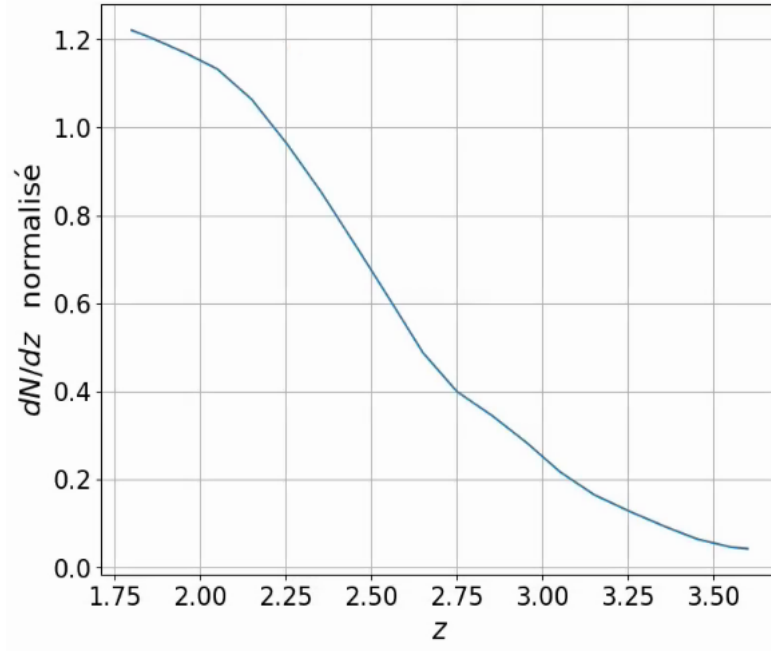


FIGURE 1.4 – Distribution normalisée en redshift des quasars tirés dans les mocks.

Une fois les quasars tirés, nous les déplaçons proportionnellement à leur vitesse v_{\parallel} le long de la ligne de visée. Celle-ci est définie comme

$$v_{\parallel} = \frac{v_x X + v_y Y + v_z Z}{\sqrt{X^2 + Y^2 + Z^2}}. \quad (1.20)$$

Ainsi, un quasar situé à une distance R sera déplacé le long de la ligne de visée à une distance

$$R \rightarrow R + \frac{1}{H(z)} v_{\parallel}(z), \quad (1.21)$$

avec

$$v_{\parallel}(z) = (1+z) \frac{H(z)}{H_0} \frac{dG/dz}{dG/dz(z=0)} v_{\parallel}. \quad (1.22)$$

Le facteur $(1+z)$ vient de la conversion des distances en distances comobiles. Une fois tous les quasars déplacés, leur redshift est recalculé, puis ils sont stockés dans un catalogue. Pour chaque quasar, le catalogue contient leur position dans le ciel (α, δ) , leur redshift avec et sans RSD, ainsi qu'un identifiant unique.

2.3 Création des lignes de visée

A cette étape, nous disposons d'un catalogue de quasars, corrélé avec le champ de densité $\delta_{matière}$ qui sera utilisé pour construire l'absorption Ly α . Nous pouvons donc créer les lignes de visées à partir de chaque quasar, et interpoler la boîte contenant le champ de densité, le long de ces lignes de visée. Dans un premier temps, nous commençons par créer le vecteur en longueurs d'onde observées, sur lequel sera interpolé la boîte de densité. Nous choisissons une taille de pixel $d_{pix} = 0,2 h^{-1}$ Mpc. Les limites $1,8 < z < 3,6$ en redshift se traduisent par des limites $3403,876 < \lambda < 5592,082 \text{ \AA}$ sur la longueur d'onde observée pour le Ly α . Nous ajoutons la limite basse des spectrographes de DESI : $\lambda_{min} = 3530 \text{ \AA}$, que nous réduisons afin d'inclure certains métaux dans les forêts (cela sera expliqué

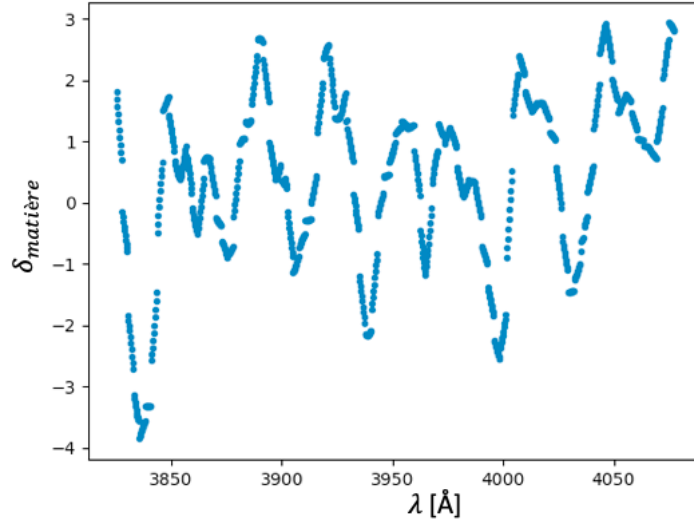


FIGURE 1.5 – Exemple de champ interpolé en utilisant les voxels à $\pm 1\sigma$ du voxel central. Le champ ainsi interpolé possède de nombreuses discontinuités. Ces discontinuités produisent de la puissance supplémentaire aux petites échelles (effet de crénelage ou d'*aliasing*).

dans la section ??). Les longueurs d'onde observées couvrent donc $3476,1877 < \lambda < 5591,566 \text{ \AA}$ à l'aide de 6524 pixels¹.

Une fois ce vecteur en longueur d'onde obtenu, nous le positionnons dans les boîtes afin de construire la ligne de visée à partir de chaque quasar. D'abord, le pixel $\lambda_{\text{QSO}} = (1 + z_{\text{QSO}})\lambda_{\text{Ly}\alpha}$ est placé à la position $(x_{\text{QSO}}, y_{\text{QSO}}, z_{\text{QSO}})$ du quasar, et le vecteur est dirigé vers l'observateur. Puis, pour chaque pixel i entre $\lambda_{\text{min}} = 3476,1877$ et λ_{QSO} , la position (x_i, y_i, z_i) du pixel est déterminée. Une moyenne pondérée avec un lissage gaussien est alors appliquée à la boîte. Ce lissage est nécessaire afin d'éviter le crénelage (*aliasing*) aux petites échelles. Sans lissage, les spectres interpolés possèdent des discontinuités (voir figure 1.5). Ces discontinuités rajoutent de la puissance parasite aux petites échelles lors du calcul du spectre de puissance. Nous appliquons donc un lissage gaussien, de largeur $\sigma = d_{\text{cell}}$. Pour chaque pixel, nous considérons les voxels appartenant au cube de 7 voxels de côté, centré sur le voxel dans lequel se trouve le pixel. Ce cube correspond à considérer les voxels compris à $\pm 3\sigma$ du voxel central. Ceci représente donc, pour chaque pixel, une moyenne sur 343 voxels. La figure 1.5 présente le champ interpolé en considérant uniquement les voxels à $\pm 1\sigma$. Trop peu de voxels sont considérés pour le calcul de chaque pixel, et le champ obtenu contient des discontinuités. La figure 1.6 présente le champ interpolé en considérant les voxels à $\pm 2\sigma$ (orange) et $\pm 3\sigma$ (bleu). La différence entre les deux est faible, cependant certaines discontinuités subsistent pour le champ calculé avec 2σ , comme le montre le zoom sur le graphique de droite. Enfin, nous avons vérifié que les différences entre les champs obtenus avec 3 et 4σ restent faibles. Nous nous limitons donc aux voxels compris à $\pm 3\sigma$ du voxel central pour limiter le temps CPU. Le champ dans le pixel i est donc donné par

$$\delta_i = \frac{\sum_{j=0}^{342} \delta_j w_{ij}}{\sum_{j=0}^{342} w_{ij}}, \quad (1.23)$$

1. Ces limites sont aussi choisies afin de garantir un nombre entier de pixels.

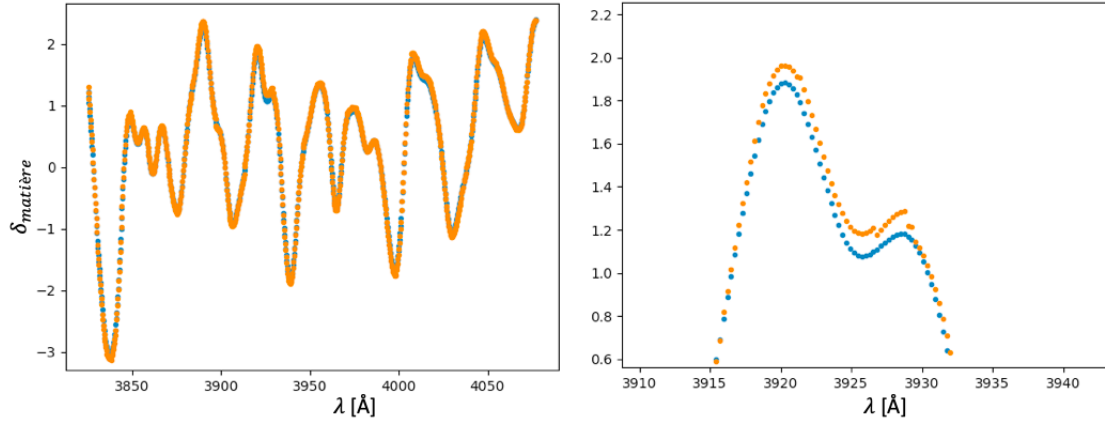


FIGURE 1.6 – Exemple de champ interpolé en utilisant les voxels à $\pm 2\sigma$ (orange) et à $\pm 3\sigma$ (bleu) du voxel central. Les discontinuités sont beaucoup moins nombreuses que pour le champ interpolé avec $\pm 1\sigma$ (figure 1.5). Cependant, certaines sont encore visibles sur le zoom, présenté sur le graphique de droite.

avec

$$w_{ij} = \exp\left(\frac{-(\vec{r}_j - \vec{r}_i)^2}{2\sigma^2}\right), \quad (1.24)$$

où \vec{r}_j est la position du centre du voxel j , \vec{r}_i celle du pixel i , δ est la boîte à interpoler. Ce calcul est effectué pour tous les pixels qui vérifient $\lambda_i < \lambda_{\text{QSO}}$, pour chaque quasar. Les boîtes interpolées sont la boîte $\delta_{\text{matière}}$ utilisée pour construire l'absorption Ly α , les trois boîtes de vitesse utilisées pour ajouter les RSD aux HCD tirés dans chaque ligne de visée (voir section 2.5), et les six boîtes de gradient de vitesse afin d'ajouter les RSD au champ Ly α .

2.4 De la densité à l'absorption

Une fois les lignes de visées interpolées, nous pouvons transformer le champ de densité en absorption Ly α . Ceci est fait via la formule FGPA :

$$F = \exp[-a(z) \exp(b(z)G(z)\delta_{\text{matière}})] . \quad (1.25)$$

Les petites échelles

La boîte $\delta_{\text{matière}}$, utilisée dans l'équation 1.25, contient le champ de matière à grand échelle. Elle est construite grâce à la transformation de Fourier de la boîte δ_k . Cependant, ce champ est construit sur une grille de taille $d_{\text{cell}} = 2,19 h^{-1} \text{ Mpc}$. Par conséquent, la plus petite échelle accessible est

$$k_N = \frac{2\pi}{d_{\text{cell}}} \sim 2,87 h \text{ Mpc}^{-1} . \quad (1.26)$$

Nous manquons donc toutes les fluctuations pour lesquelles $k > k_N$. Sans ces fluctuations le spectre de puissance à une dimension, défini comme

$$P^{1D}(k_{\parallel}) = \frac{1}{2\pi} \int_{k_{\parallel}}^{\infty} k_{\perp} P(k_{\parallel}, k_{\perp}) dk_{\perp} , \quad (1.27)$$

ne possède pas la bonne amplitude. De plus, ce sont ces fluctuations aux petites échelles qui contribuent principalement à la variance du champ. Le champ F construit ne possède donc pas le bon niveau de bruit. Pour palier ce problème, nous rajoutons indépendamment sur chaque ligne de visée un champ

δ_s (*small scales* : petites échelles) au champ $\delta_{matière}$ que nous appelons désormais δ_l (*large scales* : grandes échelles) :

$$F = \exp[-a(z) \exp(b(z)G(z)(\delta_l + \delta_s))]; \quad (1.28)$$

Du fait que ce champ ne soit pas corrélé d'une ligne de visée à une autre, il ne participe pas à la fonction de corrélation à trois dimensions¹. Afin d'ajouter la bonne quantité de fluctuations aux petites échelles, pour chaque ligne de visée nous générons un GRF à une dimension $\delta_{k,s}$, de la taille de la forêt. Puis, nous multiplions $\delta_{k,s}$ par

$$\sqrt{\frac{P_{miss}(k)}{d_{pix}}}, \quad (1.29)$$

où P_{miss} est le spectre de puissance qu'il faut appliquer à $\delta_{k,s}$ afin que F possède le bon P^{1D} . Le calcul de P_{miss} est détaillé dans la section 4. Enfin, nous obtenons δ_s à l'aide de la transformation de Fourier de $\delta_{k,s}$.

Les RSD

Une fois les petites échelles ajoutées, nous obtenons un champ d'absorption F qui possède le bon spectre de puissance à 3 dimension pour les échelles $k_N < k < k_{max}$, avec

$$k_{max} = \frac{2\pi}{1536d_{cell}} \sim 1,9 \times 10^{-3} h \text{ Mpc}^{-1}, \quad (1.30)$$

(#prov y a un facteur 2 en trop pour kny?) ainsi que le bon spectre de puissance à une dimension. Cependant, le champ d'absorption F ne possède pas de RSD pour l'instant, car il est construit à partir du spectre de puissance $P_{matière}(k)$ qui est isotrope. Initialement, nous pensions ajouter les RSD au niveau du spectre de puissance : multiplier le GRF intial par $(1 + \beta\mu^2)P_{matière}(k, \mu)$, avec $\mu = k_z/k$. Cependant, du fait que les lignes de visées ne sont pas parrallèles (elles l'étaient pour les mocks précédents, développés pour BOSS), nous ne pouvons pas confondre l'axe k_z avec l'axe de la ligne de visée $k_{||}$. Nous avons alors choisi d'utilisé le champ de gradient de vitesse $\eta_{||}$, présenté dans la section 2.1. Plusieurs essais (#prov les détailler?) ont été menés afin de savoir comment inclure correctement le champ $\eta_{||}$ dans FGPA. Nous présentons dans les lignes qui suivent la solution retenue. Le champ $\eta_{||}$ est ajouté, en plus du champ δ_s , au champ δ_l . Ceci nous permet de retrouver la formule de Kaiser (équation 1.11). Afin de gérer la quantité de RSD, nous ajoutons un coefficient c , qui dépend de z . L'ajustement de ce paramètre nous permet d'obtenir la bonne dépendance en z pour $\beta_{Ly\alpha}$. La formule FGPA devient donc

$$F = \exp\left[-a(z) \exp\left(b(z)G(z)(\delta_l + \delta_s + c(z)\eta_{||})\right)\right]; \quad (1.31)$$

Les champs δ_l , δ_s et $\eta_{||}$ sont calculés à $z = 0$, la dépendance en z étant prise en compte par le facteur $G(z)$. De plus, le facteur f que nous avons laissé de coté dans la section 2.1 n'est pas explicité ici : pour les redshifts $z > 2$, l'univers est dominé par la matière et donc, en bonne approximation, nous avons $f(z) \sim 1$. Les paramètres $a(z)$, $b(z)$, $c(z)$, ainsi que $P_{miss}(z)$ sont ajustés afin d'obtenir le bon $b_{Ly\alpha}(z)$, $\beta_{Ly\alpha}(z)$, $\overline{F}(z)$ et $P_{Ly\alpha}^{1D}(z)$. L'ajustement est décrit dans la section 4.

La prédiction

Il aurait été possible d'implémenter les RSD différemment. Une solution serait, par exemple, de déplacer chaque pixel d'absorption proportionnellement à la vitesse particulière du gaz dans la cellule

1. Lors du calcul de la fonction de corrélation à 3 dimensions, nous ne considérons pas les paires de pixels issues de la même forêt (voir ??)

considéré, puis de modifier l'absorption en fonction de gradient de vitesse dans cette cellule. En effet, si le gradient de vitesse est non nul, le gaz se retrouve comprimé par endroit, et détendu dans d'autres, modifiant l'absorption dans chaque cellule. Cette méthode pour ajouter les RSD dans des mocks Ly α est la méthode choisie par Farr et al. (2019). La méthode que nous utilisons, décrite dans la section précédente, a l'avantage d'avoir une fonction de corrélation prédictible. C'est pour cela que nous avons fait le choix de cette méthode. Dans les lignes qui suivent, nous expliquons comment calculer la prédiction de la fonction de corrélation. Comme décrit par Font-Ribera et al. (2012), il est possible de relier la fonction de corrélation ξ_F du champ F à la fonction de corrélation ξ_g du champ δ_g . Ces deux fonctions de corrélations sont reliées par

$$\xi_F(r_{12}) = \int_{-\infty}^{\infty} d\delta_{g1} \int_{-\infty}^{\infty} d\delta_{g2} \frac{\exp\left[-\frac{\delta_{g1}^2 + \delta_{g2}^2 - 2\delta_{g1}\delta_{g2}\xi_g(r_{12})}{2(1-\xi_g^2(r_{12}))}\right]}{2\pi\sqrt{1-\xi_g^2(r_{12})}} \delta_F(\delta_{g1})\delta_F(\delta_{g2}), \quad (1.32)$$

où δ_g est un GRF de variance 1, δ_F est le champ d'absorption calculé à partir du champ gaussien, et r_{12} est la distance qui sépare les deux points où sont évalués les champs δ_g et δ_F . Dans notre cas, le champ δ_g représente le champ $bG(z)(\delta_l + \delta_s + c(z)\eta_{||})$. Ce champ est un champ gaussien, de valeur moyenne nulle et de variance σ_g^2 . Nous compensons le fait que $\sigma_g^2 \neq 1$ en remplaçant ξ_g par ξ_g/σ_g^2 dans l'équation précédente. L'équation 1.32 ne dépendant que de la valeur de ξ_g , nous construisons une table qui permet de relier chaque valeur de $\xi_g \in [-1; 1]$ à la valeur ξ_F correspondante. De plus, nous pouvons relier la fonction de corrélation dans l'espace des redshifts $\xi_g(r, \mu)$, à la fonction de corrélation $\xi(r)$ que suit le champ δ_l :

$$\xi(r, \mu) = \xi_0(r) + \frac{1}{2}(3\mu^2 - 1)\xi_2(r) + \frac{1}{8}(35\mu^4 - 30\mu^2 + 3)\xi_4(r), \quad (1.33)$$

avec

$$\xi_0(r) = \left(1 + \frac{2}{3}f + \frac{1}{5}f^2\right)\xi(r), \quad (1.34)$$

$$\xi_2(r) = \left(\frac{4}{3}f + \frac{4}{7}f^2\right)\left[\xi(r) - \bar{\xi}(r)\right], \quad (1.35)$$

$$\xi_4(r) = \frac{8}{35}f^2\left[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)\right], \quad (1.36)$$

et

$$\bar{\xi}(r) = 3r^{-3} \int_0^r \xi(s)s^2 ds, \quad (1.37)$$

$$\bar{\bar{\xi}}(r) = 5r^{-5} \int_0^r \xi(s)s^4 ds. \quad (1.38)$$

Ces équations sont décrites dans Hamilton (1992). Afin d'obtenir la prédiction $\xi_F^{pred}(r, \mu)$, nous commençons par calculer le spectre de puissance que suit la boîte δ_l :

$$P(k) = W^2(k)P_{matière}(z=0), \quad (1.39)$$

où W est le terme représentant le lissage gaussien appliqué à la boîte δ_l (voir section 2.3) :

$$W(k) = e^{-\frac{k^2 d_{cell}^2}{2}}. \quad (1.40)$$

A l'aide d'une transformation de Fourier, nous obtenons la fonction de corrélation $\xi(r)$ que suit la boîte interpolée δ_l . Puis, nous calculons la fonction de corrélation dans l'espace des redshifts $\xi_g(r, \mu)$ que suit le champ δ_g grâce à l'équation 1.33. Enfin, pour tous les couples (r, μ) nécessaires, nous obtenons la fonction de corrélation $\xi_F(r, \mu)$ du champ F comme la valeur correspondante à la valeur tabulée $\xi_g(r, \mu)/\sigma_g^2$ pour ξ_g . (#prov un plot de la prédiction?)

2.5 Ajout des HCD

Les HCD ont un effet important dans les fonctions de corrélation, nous simulons donc aussi leur présence. De manière à avoir une corrélation entre les HCD et les autres traceurs des mocks, nous utilisons la boîte de densité δ_l pour tirer les HCD. Nous ne considérons pas la somme $\delta_l + \delta_s$ car les HCD sont des surdensités à grandes échelles : une résolution de $2,19 h^{-1}$ Mpc est suffisante. De plus, l'ajout de δ_s bruyerait la corrélation entre les HCD et les autres traceurs.

Contrairement aux quasars, les HCD sont tirés proportionnellement au champ δ_l . Nous identifions les cellules dans lesquelles δ_l est au dessus d'un certain seuil, puis les HCD sont tirés dans ces cellules selon une loi de Poisson. Le seuil ν est défini en fonction du biais souhaité pour les HCD. Pour un seuil ν , le biais obtenu est donné par

$$b_\nu = \frac{pdf(\nu)}{cdf(-\nu)}, \quad (1.41)$$

où $pdf(\nu)$ donne la densité de probabilité de ν , et $cdf(-\nu)$ est la fonction de répartition : la probabilité d'être au dessus du seuil ν . Ainsi, pour avoir un biais de 2, il faut que la probabilité que le champ prenne la valeur ν soit 2 fois plus grande que la probabilité que le champ soit au dessus du seuil ν . Afin d'obtenir le seuil pour un biais donné, nous calculons b_ν pour une large gamme de seuils ν puis nous interpolons b_ν sur ν . Ainsi, pour un biais b , nous connaissons le seuil $\nu(b)$ à choisir. Dans notre cas, le champ δ_l suit une densité de probabilité gaussienne. Cependant, sa variance n'est pas égale à 1. De plus, le champ δ_l interpolé le long des lignes de visée correspond au champ de matière à $z = 0$. Ainsi, pour obtenir un biais b_{HCD} , pour chaque redshift nous calculons le seuil ν comme si nous visions un biais $b = b_{\text{HCD}}\sigma_l G(z)$. Le terme σ_l prend en compte la variance du champ δ_l , et $G(z)$ le fait que δ_l soit construit à $z = 0$. Une fois les cellules pouvant héberger un HCD identifiées, nous tirons dans chacune d'entre elles les HCD avec une loi de poisson de paramètre

$$\lambda(z) = \frac{N(z)}{cdf(-\nu(z))}, \quad (1.42)$$

où $N(z)$ donne le nombre moyen de HCD attendu par cellule et $\nu(z)$ le seuil au redshift z . Le nombre de HCD attendu est donné par la librairie `pyigm`¹. La distribution en redshift des HCD est présenté sur le graphique de gauche de la figure 1.7. Une fois tous les HCD tirés, nous leur assignons une densité de colonne dans la gamme $17,2 < \log(n_{\text{HI}}) < 22,5$, selon la distribution donnée par `pyigm`. Cette distribution est présentée sur le graphique de droite de la figure 1.7. Enfin, nous ajoutons les RSD aux HCD tirés. Chaque HCD tiré est déplacé le long de la ligne de visée proportionnellement à la vitesse

$$v_{\parallel} = \frac{v_x X + v_y Y + v_z Z}{\sqrt{X^2 + Y^2 + Z^2}}, \quad (1.43)$$

où v_x , v_y et v_z sont les boîtes de vitesse interpolées le long de la ligne de visée. Ainsi, un HCD à un redshift z sera déplacé à un redshift

$$z \rightarrow z + (1+z)H(z) \frac{dG}{dz} \frac{1}{H_0 \frac{dG}{dz}(z=0)} v_{\parallel}. \quad (1.44)$$

Dans les mocks que nous décrivons ici, le profil d'absorption des HCD n'est pas ajouté dans les forêts. Nous produisons uniquement un catalogue qui regroupe tous les HCD tirés. Le profil d'absorption est ajouté au spectre de chaque quasar par le code `quickquasars`, qui utilise le catalogue de HCD que nous produisons.

1. <https://github.com/pyigm>

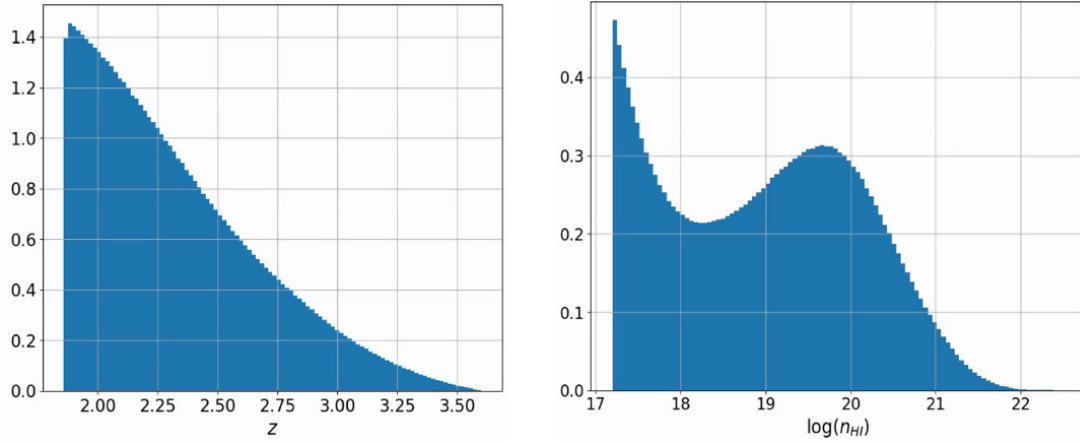


FIGURE 1.7 – Gauche : distribution normalisée en redshift des HCD. Droite : distribution normalisée de $\log(n_{HI})$ des HCD. Ces distributions proviennent de la librairie `pyigm`.

3 Production des mocks

Comme expliqué au début de ce chapitre, l’objectif des mocks est de reproduire les données d’eBOSS et de DESI. Etant donné que le relevé d’eBOSS est contenu dans le relevé de DESI, nous simulons directement le relevé DESI. Ainsi, lorsque nous avons besoin de simuler le relevé d’eBOSS, nous retirons les quasars qui ne sont pas contenu dans ce relevé. La taille des boîtes choisie ($2560 \times 2560 \times 1536$ voxels) et leur résolution ($2,19 h^{-1}$ Mpc) ne suffisent pas à couvrir les 14 000 degrés carrés de DESI. Pour palier ce problème, nous construisons sept *chunks* indépendants, que nous assemblons pour former le relevé de DESI. Le découpage du relevé en sept chunks est montré sur la figure 1.8. Ce choix d’assembler sept boîtes de densité plutôt que d’en utiliser une seule a été contraint par la mémoire maximale des noeuds. Les mocks ont été produits grâce au centre de calcul NERSC¹, avec la machine Cori. **Sur cette machine, nous avons utilisé les noeuds “Haswell”. Ces noeuds possèdent 128Go de mémoire vive. De manière à créer nos boîtes de densité, via la transformation de Fourier à trois dimensions, il faut, pour chaque boîte, que l’intégralité de son contenu soit accessible depuis un même endroit. Nous pourrions distribuer la mémoire et effectuer la transformation de Fourier sur plusieurs noeuds à l’aide de la librairie MPI. Cependant nous n’avons pas l’expertise nécessaire. Nous effectuons donc les transformations de Fourier sur un seul noeud, ce qui explique le découpage du relevé en sept chunks indépendants. Nous profitons néanmoins des 32 cœurs par noeud pour paralléliser notre code. Chaque cœur possède 2 hyper-threads, ce qui permet de gérer 64 tâches simultanément sur un même noeud. #prov taille des boîtes : 37,5 Go (box) + 37,5/2 Go (boxk) + 37,5/2 Go (k) : 75 Go. On est loin des 128 Go.**

Dans les lignes qui suivent, nous détaillons les différents éléments du code. Le schéma 1.9 résume la situation. Le première module, `interpolate_pk.py` permet de calculer puis d’interpoler les quatres spectres de puissance $P_{QSO,i}$ et $P_{matière}$ sur la grille $2560 \times 2560 \times 1536$ dans l’espace k . Le code est lancé séparément sur 16 morceaux de la boîte, puis le code `merge_pk.py` permet de rassembler des 16 morceaux des spectres de puissance interpolés et de les sauver au format FITS (Flexible Image Transport System). Cette étape est effectuée une seule fois, car les spectres de puissance sont communs à toutes les réalisations. Le module suivant est `make_boxes.py`. **Ce code lit les spectres de puissance interpolés puis construit les trois boîtes de densité relatives aux quasars, la**

1. National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

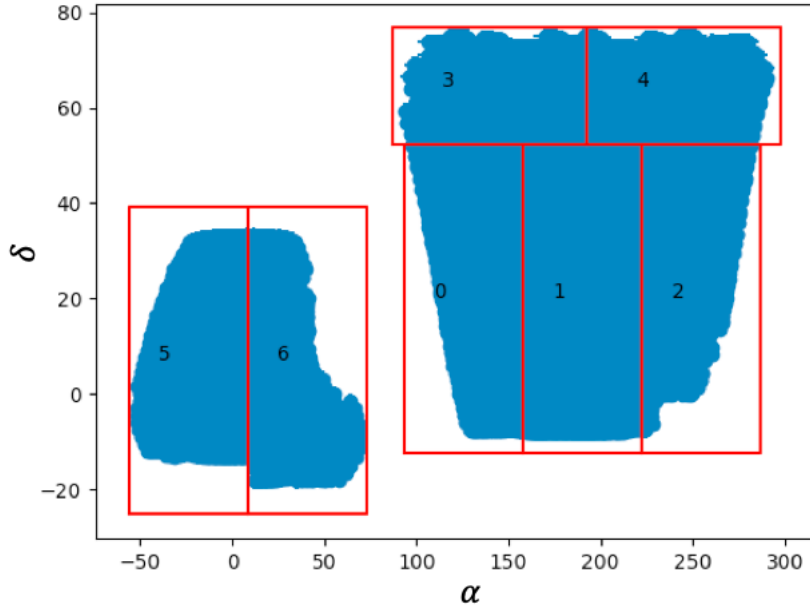


FIGURE 1.8 – Découpage du relevé de DESI en 7 chunks.

boîte de densité relative au $\text{Ly}\alpha$, les trois boîtes de vitesse et les six boîtes de gradient de vitesse, décrites dans la section 2.1. Au total, 14 transformations de Fourier inverses sont effectuées. Ce code est lancé sept fois, afin de produire les boîtes pour les sept chunks. Une fois toutes les boîtes produites, les quasars sont tirés (section 2.2) grâce au code `draw_qso.py`. Afin d’accélérer la production des mocks, les boîtes sont partagées en 512 tranches selon l’axe x . Ces tranches, de taille $5 \times 2560 \times 1536$, sont traitées séparément. Ce code est donc tourné en parallèle 512 fois, sur 16 noeuds \times 32 threads. Une fois les quasars tirés, les lignes de visée sont interpolées (section 2.3) avec le code `make_spectra.py`. De la même manière que le code précédent, il tourne en parallèle 512 fois, sur 16 noeuds \times 32 threads. Chaque instance du code interpole les densités, vitesses et gradients de vitesse le long de chaque ligne de visée présente dans la tranche traitée. Ceci produit, dans chaque tranche, des morceaux de lignes de visée relatifs aux quasars tirés précédemment. Une fois tous les quasars traités, les morceaux de ligne de visée sont mis bout à bout grâce au code `merge_spectra.py`. Encore une fois, le code tourne en parallèle 512 fois : chaque instance du code traite tous les quasars situés dans une même tranche. Le code lit alors tous les morceaux de spectre relatifs à ces quasars, puis les assemble. Lorsque les lignes de visée sont toutes reconstruites, la formule FGPA est appliquée afin d’obtenir le champ de transmission F pour chaque ligne de visée. La dernière étape consiste à regrouper le résultat de chaque chunk. Le module `merge_qso.py` permet de lire tous les quasars tirés dans chaque tranche de chaque chunk et de créer un catalogue global appelé `master.fits`. Une fois le catalogue construit, le code `dla_saclay.py` tire les HCD le long de chaque ligne de visée. Le code est tourné sur les sept chunks en parallèle. Puis le module `merge_dla.py`, comme pour les quasars, permet de regrouper tous les HCD tirés et de construire le catalogue global `master_DLA.fits`. Enfin le code `make_transmissions.py` permet de mettre les fichiers contenant les forêts au bon format. Les forêts sont regroupées par HEALPix pixel (#prov expliquer ou mettre une ref?) dans des fichiers FITS, puis ces fichiers FITS sont regroupés par 100 selon leur HEALPix pixel :

`N/PIX/transmission-nside-PIX.fits.gz,`

où `nside = 16` est la résolution du schéma HEALPix utilisé, `PIX` donne le numéro du pixel HEALPix,

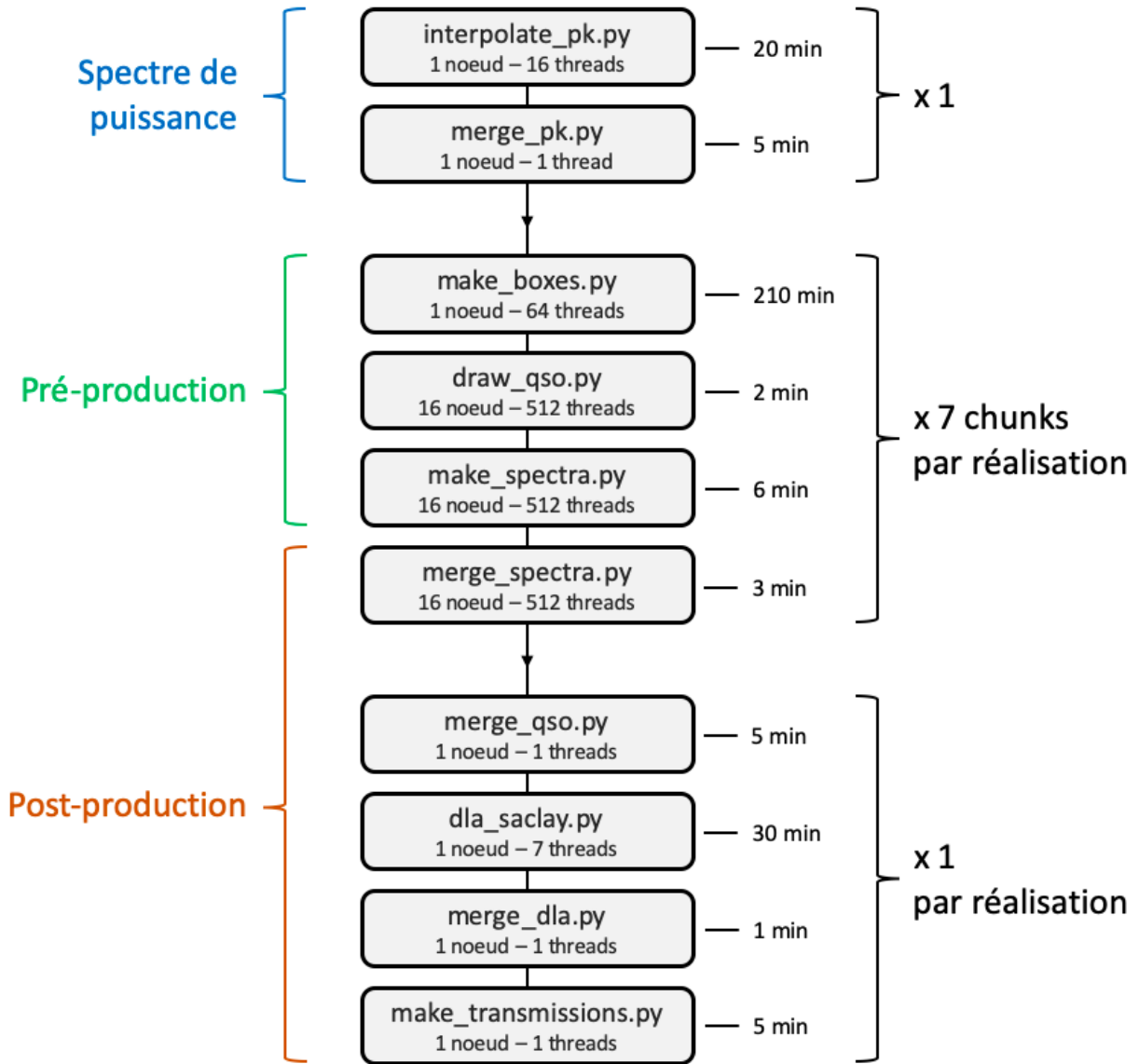


FIGURE 1.9 – Schéma illustratif du fonctionnement des mocks. Le premier bloc (bleu) génère le spectre de puissance nécessaire à la création des boîtes. Puis, pour chaque chunk de chaque réalisation, la pré-production (vert) génère les boîtes, tire les quasars, et interpole les lignes de visée. Enfin, la post-production (orange) rassemble les morceaux de lignes de visée et construit les forêts Ly α pour chaque chunk. Puis, elle regroupe le résultat de chaque chunk, construit les catalogues finaux de quasars et HCD et produit les fichiers transmissions pour chaque réalisation. Chaque case indique le nom du code, ainsi que les conditions dans lesquelles il est tourné. Les temps d'exécution indiqués sont approximatifs.

et \mathbf{n} est le résultat de la division euclidienne de \mathbf{pix} par 100. Cette dernière étape est effectuée sur un noeud, `make_transmissions.py` étant lancé en parallèle sur 64 sous-échantillons des HEALPix pixels. `#prov` : En fait je parle pas du tout des randoms QSO et randoms DLA. C'est sans doute intéressant que je le mentionne ?

Pour les analyses Ly α d'eBOSS et de DESI, nous avons décidé de produire 100 réalisations des mocks, afin d'avoir suffisamment de statistique pour étudier finement les potentielles systématiques. Nous avons organisé la production de ces 100 réalisations en deux étapes. Premièrement, nous avons effectué la *pré-production*. Cette étape consiste à créer les boîtes contenant les différents GRF, puis à tirer les quasars et enfin reconstruire la densité, les vitesses et gradients de vitesse, le long de chaque ligne de visée. Ceci correspond aux codes `make_boxes.py`, `draw_qso.py` et `make_spectra.py`. Cette étape est la plus coûteuse en temps de calcul : environ 43 heures CPU sur un noeud de Cori pour produire les sept chunks d'une réalisation. L'ensemble des fichiers temporaires propres à une réalisation pré-produite est estimé à $\sim 7 \times 550 \text{ Go}$ sur disque. Cependant, nous stockons uniquement les lignes de visée interpolées, et la boîte δ_k , ce qui représente $\sim 340 \text{ Go}$ sur disque, par réalisation.

Un certain nombre de problèmes informatiques, liés au centre de calcul NERSC, ont ralenti la phase de pré-production. Le principal problème venait du temps de lecture et d'écriture, qui par moment pouvait être multiplié par un facteur cent. Plus de trentes secondes étaient parfois nécessaires pour accéder à un simple fichier. Ce problème a été identifié comme venant du transfert des *meta data* sur le centre de calcul. Le code le plus affecté est `make_boxes.py`, car il écrit énormément de fichiers différents, correspondant aux différents slices des boîtes, et destinés à être lus par les codes `draw_qso.py` et `make_spectra.py`. Pour palier ce problème, nous avons décidé d'essayer de faire tourner les codes sur les noeuds de Cori appelés *Burst Buffer*. Ces noeuds possèdent des disques SSD (Solid-State drive), ce qui permet une lecture et une écriture très rapide. Une fois les codes exécutés, les données produites sont déplacées sur les disques durs habituels. Les noeuds Burst Buffer ont permis d'accélérer l'exécution du code `make_boxes.py` par un facteur ~ 2 , passant d'environ deux heures à une heure seulement. Une vingtaine de réalisation ont été produites en utilisant ces noeuds. Cependant, ils sont devenus instables au cours de la production. D'autre part, les problèmes liés aux transferts des meta data avaient été stabilisés entre temps. Nous sommes donc retournés à l'utilisation des noeuds classiques de Cori pour finir la production. Pour les 40 dernières réalisations, le temps d'exécution de `make_boxes.py` variait entre trois et quatre heures.

Le temps pris par les six mois nécessaires à la production nous a permis de choisir les paramètres Ly α souhaités, et d'ajuster les paramètres de la formule FGPA (équation 1.31) en conséquence. Ceci est expliqué dans la section suivante. Une fois ces paramètres ajustés et la phase de pré-production terminée, nous avons mené la phase de *post-production*. Cette phase consiste à créer les spectres d'absorption à partir des densités interpolées le long des lignes de visée, puis à regrouper tous les fichiers de sortie afin de les mettre au format décrit précédemment. Cette étape est beaucoup plus rapide, elle prend l'équivalent d'environ 6 heures CPU sur un noeud de Cori par réalisation. Une fois la production complète effectuée, la place sur disque d'une réalisation, sans compter les fichiers temporaires, correspond à environ 15 Go.

4 Ajustement des paramètres

Comme expliqué dans la section 2.4, le champ d'absorption Ly α est construit à partir des boîtes δ_l , δ_s et η_{\parallel} grâce à la formule FGPA. Contrairement aux quasars pour lesquels le biais est choisi, les paramètres physiques du champ d'absorption Ly α simulé dépendent des quatre paramètres $a(z)$, $b(z)$, $c(z)$ et $P_{miss}(z)$ utilisés dans l'équation 1.31. Nous décrivons dans cette section comment ajuster ces paramètres (paramètres FGPA dans la suite) afin d'obtenir les bons $b_{\text{Ly}\alpha}(z)$, $\beta_{\text{Ly}\alpha}(z)$, $\bar{F}(z)$ et $P^{1D}(k, z)$ (paramètres Ly α dans la suite). Nous nous servons des données décrites dans le chapitre ??

comme référence pour les paramètres $\text{Ly}\alpha$.

Afin d'ajuster ces paramètres FGPA, l'idée est de générer un mock avec un jeu de paramètres FGPA, calculer la fonction de corrélation du $\text{Ly}\alpha$, ajuster cette fonction de corrélation et mesurer les paramètres $\text{Ly}\alpha$. Puis une fois ces paramètres mesurés, itérer sur les paramètres FGPA afin de nous rapprocher des paramètres $\text{Ly}\alpha$ visés. Cependant cette méthode est très couteuse, car elle nécessite de générer des spectres et de produire la fonction de corrélation à chaque itération. Pour accélérer la procédure d'ajustement, nous tirons profit du fait que nos mocks possèdent une fonction de corrélation prédictible. Ainsi, au lieu de générer des spectres et calculer la fonction de corrélation sur ces derniers à chaque itération, nous générons la prédiction (section 2.4) puis nous l'ajustons directement afin de mesurer les paramètres $\text{Ly}\alpha$ correspondant au jeu de paramètres FGPA utilisé. Cet ajustement est fait dans les 5 bins en redshift $z_1 = 1,8$; $z_2 = 2,2$; $z_3 = 2,6$; $z_4 = 3,0$ et $z_5 = 3,6$.

Dans les lignes qui suivent, nous expliquons comment, dans chaque bin en redshift, nous choisissons le jeu de paramètres FGPA initial pour générer la prédiction à la première itération. Premièrement, nous choisissons le paramètre c . La somme des champs δ_l et η_{\parallel} permet de retrouver le champ δ^s dans l'espace des redshifts, via la formule de Kaiser (équation 1.11). Le champ δ^s possède un biais de 1, par construction. Son paramètre RSD est donc $\beta = f/b = f$. De plus, à grand redshift, nous pouvons considérer en bonne approximation que $f \sim 1$. Nous avons donc $\beta \sim 1$. Ainsi, lorsque nous considérons la somme des trois champs δ_l , δ_s et η_{\parallel} dans l'équation 1.31, le terme c devant η_{\parallel} donne, en première approximation, $\beta = c$. Dans chaque bin en redshift, nous choisissons donc $c(z) = \beta_{\text{Ly}\alpha}(z)$.

En ce qui concerne les paramètres a et b , ils sont ajustés afin de retrouver le bon $\bar{F}(z)$ et $P^{1D}(k, z)$ (#prov décrire).

Une fois les paramètres a , b et c choisis, nous déterminons le spectre de puissance P_{miss} à appliquer à $\delta_{k,s}$ afin d'obtenir le bon P_{mock}^{1D} . Nous avons besoin de P_{miss} car il nous faut connaître σ_g dans chaque bin en redshift afin de pouvoir calculer la prédiction. La variance du champ $g = bG(z)(\delta_l + \delta_s + c(z)\eta_{\parallel})$ est donné par

$$\sigma_g^2(z) = \langle \delta_l^2 \rangle + \langle \delta_s^2 \rangle + c(z)\langle \eta_{\parallel}^2 \rangle + c(z) \left(\langle \delta_l \eta_{\parallel} \rangle - \langle \delta_l^2 \rangle \langle \eta_{\parallel}^2 \rangle \right). \quad (1.45)$$

Nous commençons par construire le spectre de puissance $P_{\text{modèle}}^{1D}$ sur lequel le P_{mock}^{1D} des mocks sera ajusté. Tout d'abord, les données provenant de (#prov On fit les données DR12 sans les oscillations du Si, quel papier?) sont ajustées dans la gamme $0,2 < k < 2,0 h \text{ Mpc}^{-1}$. Les oscillations dans le spectre de puissance à une dimension causées par le silicium sont aussi ajustées afin de les retirer et de construire un modèle $P_{\text{modèle}}^{1D}$ sans ces oscillations. La fonction ajustée aux données est définie comme

$$f(k) = \exp\left(-ak + b + \frac{c}{k+d}\right) \times f_{Si}(k), \quad (1.46)$$

où a , b , c et d sont quatre paramètres à ajuster, et f_{Si} est la fonction qui prend en compte les oscillations dues au silicium. La forme choisie permet d'avoir $\log(P_{\text{modèle}}^{1D})$ linéaire à grand k , en accord avec ce qui est mesuré dans les simulations hydrodynamiques (?). L'ajustement est fait dans les bins en redshift 2,2; 2,4; 2,6; 2,8; 3,0; 3,2; 3,4 et 3,6. Du fait de la résolution $d_{\text{pix}} = 0,2 h^{-1} \text{ Mpc}$ des spectres simulés, nous avons besoin de calculer le P_{miss} jusqu'à $k_{Ny} = \pi/d_{\text{pix}} \sim 15,7 h^{-1} \text{ Mpc}$. Ainsi, le résultat de l'ajustement sur les données est extrapolé de $k = 2,0$ jusqu'à $k = 20 h \text{ Mpc}^{-1}$. Pour les k plus petits que $0,2 h \text{ Mpc}^{-1}$, nous utilisons le modèle décrit dans ? et ajusté sur des simulations hydrodynamiques. Il est défini comme

$$D_1(k, \mu) = \exp \left[\left(q_1 \Delta^2(k) + q_2 \Delta^4(k) \right) \left(1 - \left(\frac{k}{k_v} \right)^{a_v} \mu^{b_v} \right) - \left(\frac{k}{k_p} \right)^2 \right], \quad (1.47)$$

avec

$$\Delta^2(k) = \frac{1}{2\pi^2} k^3 P_L(k). \quad (1.48)$$

Les termes $q_1 \Delta^2(k)$ et $q_2 \Delta^4(k)$ représente l'augmentation de puissance due aux non-linéarités. L'ajustement que nous utilisons force $q_2 = 0$. Les autres paramètres ajustés sont donnés dans la section "Planck" de la table 7. Puis, dans chaque bin en redshift, la forme ajustée aux données et définie dans l'équation 1.46 est prolongée de $k = 0,2 h \text{ Mpc}^{-1}$ jusqu'à $k = 0$ par le modèle donné dans l'équation 1.47. Pour ce faire, nous calculons le modèle à $z = 2,4$, puis nous le multiplions par une constante, dépendant du bin en redshift, de façon à ce que le prolongement en $k = 0,2 h \text{ Mpc}^{-1}$ soit continu. La figure ?? (#prov faire la figure : modèle + données superposées. 1 seul bin en z ? Tous? Seulement 2.2, 2.6, 3.0 et 3.6?) montre le modèle ainsi construit, avec les données de #prov. Pour le bin $z = 1,8$, aucune donnée $\text{Ly}\alpha$ n'est disponible. Nous extrapolons donc le modèle à $z = 2,2$. Nous considérons que la forme du modèle. Nous considérons aussi que l'évolution en redshift de $z = 2,4$ à $z = 2,2$ est la même jusqu'à $z = 1,8$. Le modèle à $z = 1,8$ est donc donné par

$$P_{\text{modèle}}^{\text{1D}}(k, z = 1,8) = P_{\text{modèle}}^{\text{1D}}(k, z = 2,2) \left(\frac{P_{\text{modèle}}^{\text{1D}}(k, z = 2,2)}{P_{\text{modèle}}^{\text{1D}}(k, z = 2,4)} \right)^2. \quad (1.49)$$

Une fois le spectre de puissance modèle calculé, nous estimons le spectre de puissance à une dimension P_l^{1D} du champ interpolé $\delta_l + c\eta_{\parallel}$. Nous partons du spectre de puissance fournit par Camb. Nous multiplions ce spectre de puissance par le terme de kaiser $(1 + c\mu_k^2)^2$ puis par le terme W représentant l'effet du lissage gaussien (équation 1.40). Enfin, nous calculons P_l^{1D} grâce à l'équation 1.27, en nous restreignant aux $k < k_{\text{max}}$. Une première estimation du spectre de puissance à une dimension P_{miss}^0 à appliquer à $\delta_{k,s}$ afin d'obtenir le bon $P_{\text{mock}}^{\text{1D}}$ est donnée par

$$P_{\text{miss}}^0(k) = P_{\text{modèle}}^{\text{1D}}(k) - P_l^{\text{1D}}(k). \quad (1.50)$$

Le spectre de puissance P_{miss}^0 ainsi construit nous permet d'obtenir un $P_{\text{mock}}^{\text{1D}}$ convenable. Cependant, il reste des différences notables avec le spectre de puissance mesuré dans les données (#prov ça vient du fait qu'on applique FGPA? mettre un plot du pld des mocks sans shape tuning et des données?). Afin de corriger ces différences, nous ajustons itérativement la forme de P_{miss} . A chaque itération n , nous commençons par générer δ_s selon P_{miss}^n . Puis nous calculons le $P_{\text{mock}}^{\text{1D}}$, correspondant à ce P_{miss}^n . Le P_{miss}^{n+1} de l'itération suivante est alors donné par

$$P_{\text{miss}}^{n+1}(k) = P_{\text{miss}}^n(k) \left(1 + l \left(\frac{P_{\text{modèle}}^{\text{1D}}(k)}{P_{\text{mock}}^{\text{1D}}(k)} - 1 \right) \right). \quad (1.51)$$

Le terme l est un facteur de convergence. En pratique, nous fixons $l = 1$, et dix itérations sont suffisantes pour obtenir un P_{miss} qui donne un $P_{\text{mock}}^{\text{1D}}$ en accord avec les données.

A ce stade, nous disposons des informations nécessaires pour générer la prédiction. Nous générons donc la prédiction, puis nous l'ajustons avec le code `picca`. Nous comparons la mesure de $b_{\text{Ly}\alpha}$, $\beta_{\text{Ly}\alpha}$ et \bar{F} à ce qui est mesuré dans les données. Ceci constitue une itération complète de la procédure d'ajustement. Dans le cas où les paramètres $\text{Ly}\alpha$ mesurés dans les mocks ne sont pas en accord avec ce qui est mesuré dans les données, nous relançons une nouvelle itération : nous modifions légèrement les paramètres a , b et c , puis nous recalculons le nouveau P_{miss}^{10} afin de générer la nouvelle prédiction. Nous ajustons de nouveau la prédiction avec `picca` et comparons les valeurs de $b_{\text{Ly}\alpha}$, $\beta_{\text{Ly}\alpha}$ et \bar{F} mesurées aux données. Ces itérations sont répétées dans chaque bin en redshift jusqu'à obtenir des valeurs de $b_{\text{Ly}\alpha}$, $\beta_{\text{Ly}\alpha}$ et \bar{F} compatibles avec les données.

Une fois que cette procédure itérative a convergé dans les cinq bins en redshift, nous ajustons les paramètres $a(z)$, $b(z)$ et $c(z)$ afin d'obtenir une fonction continue de ces paramètres sur $z \in [1,8; 3,6]$. Les paramètres $\log a(z)$ et $\log b(z)$ sont ajustés par des polynômes de degré 4 de $\log z$. Ceci est fait dans le but d'éviter les valeurs négatives pour $a(z)$, ainsi que les points d'inflexion. Le paramètre $c(z)$ est ajusté par un polynôme de degré 4 de z . La figure 1.10 présente ces trois paramètres ajustés. En ce qui concerne les cinq spectres de puissance $P_{\text{miss}}(k)$ construits précédemment dans chaque bin en

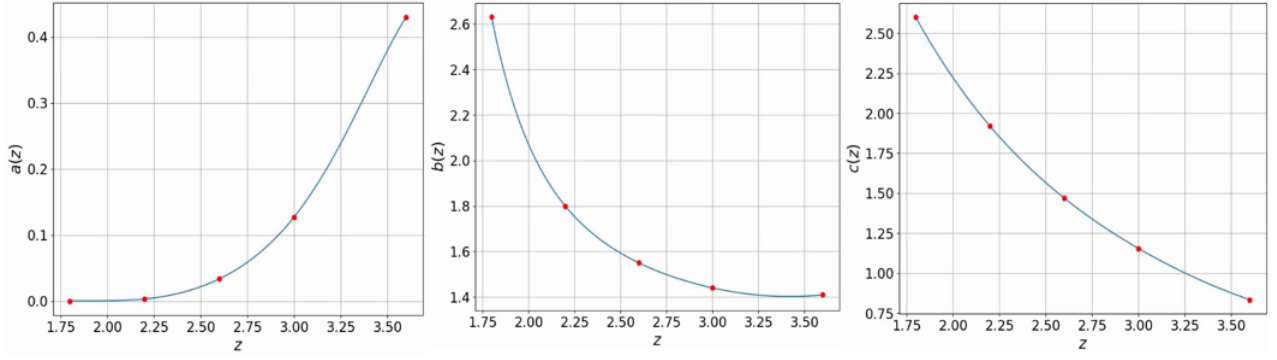
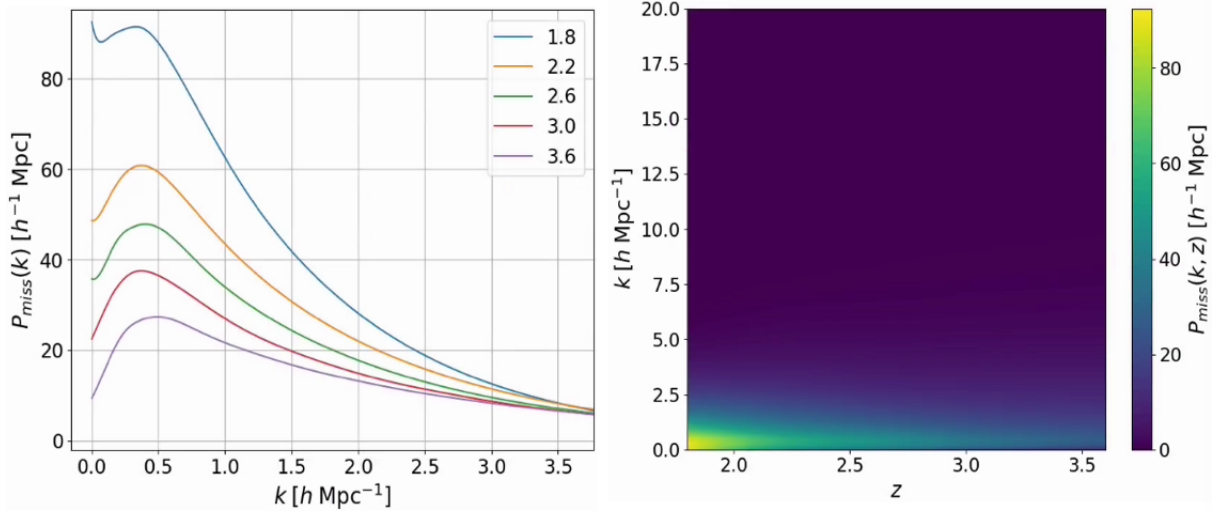


FIGURE 1.10 – Résultat de la procédure d’ajustement des paramètres FGPA.


 FIGURE 1.11 – Gauche : les cinq spectres de puissance $P_{miss}(k)$ ajustés dans chaque bin en redshift. Droite : interpolation cubique de ces cinq spectres de puissance sur une grille (k, z) .

redshift, nous créons une grille (k, z) couvrant $z \in [1.8; 3.6]$ et $k \in [0; 20] h \text{Mpc}^{-1}$. Puis, $P_{miss}(k, z)$ est obtenu via une interpolation cubique sur cette grille. Du fait des très faibles valeurs de P_{miss} à grand k et de l’interpolation cubique utilisée, certaines valeurs de l’interpolation sont négatives. Pour chaque redshift de la grille, nous forçons tous les pixels de l’interpolation de P_{miss} à zéro pour tous les $k > k_0$, où k_0 est le premier k pour lequel l’interpolation est nulle. La figure 1.11 montre l’interpolation de $P_{miss}(k)$ pour différentes valeurs de z .

Le résultat de la procédure d’ajustement décrite précédemment a été utilisé pour produire 100 réalisations indépendantes des mocks. Les figures ?? présentent les paramètres $\text{Ly}\alpha$ obtenus avec la prédiction (couleur), sur le stack des 100 rea (couleur) et mesuré dans les données (couleur). #prov montrer biais(z), beta(z), $\langle F \rangle(z)$ et P1D(z).

#prov : expliquer sigma_g : dépend de Pmiss

Bibliographie

- Katrin Heitmann, Hal Finkel, Adrian Pope, Vitali Morozov, Nicholas Frontiere, Salman Habib, Esteban Rangel, Thomas Uram, Danila Korytov, Hillary Child, Samuel Flender, Joe Insley, and Silvio Rizzi. The Outer Rim Simulation : A Path to Many-Core Supercomputers. apr 2019. doi : 10.3847/1538-4365/ab4da1. URL <http://arxiv.org/abs/1904.11970><http://dx.doi.org/10.3847/1538-4365/ab4da1>.
- Aniket Agrawal, Ryu Makiya, Chi-Ting Chiang, Donghui Jeong, Shun Saito, and Eiichiro Komatsu. Generating {Log}-normal {Mock} {Catalog} of {Galaxies} in {Redshift} {Space}. *arXiv :1706.09195 [astro-ph]*, jun 2017. URL <http://arxiv.org/abs/1706.09195>.
- L. Clerkin, D. Kirk, M. Manera, O. Lahav, F. Abdalla, A. Amara, D. Bacon, C. Chang, E. Gaztañaga, A. Hawken, B. Jain, B. Joachimi, V. Vikram, T. Abbott, S. Allam, R. Armstrong, A. Benoit-Lévy, G. M. Bernstein, E. Bertin, D. Brooks, D. L. Burk, A. Carnero Rosell, M. Carrasco Kind, M. Crocce, C. E. Cunha, C. B. D’Andrea, L. N. da Costa, S. Desai, H. T. Diehl, J. P. Dietrich, T. F. Eifler, A. E. Evrard, B. Flaugher, P. Fosalba, J. Frieman, D. W. Gerdes, D. Gruen, R. A. Gruendl, G. Gutierrez, K. Honscheid, D. J. James, S. Kent, K. Kuehn, N. Kuropatkin, M. Lima, P. Melchior, R. Miquel, B. Nord, A. A. Plazas, A. K. Romer, E. Sanchez, M. Schubnell, I. Sevilla-Noarbe, R. C. Smith, M. Soares Santos, F. Sobreira, E. Suchyta, M. E. C. Swanson, G. Tarle, and A. R. Walker. Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps. may 2016. doi : 10.1093/mnras/stw2106. URL <http://arxiv.org/abs/1605.02036><http://dx.doi.org/10.1093/mnras/stw2106>.
- Peter Coles and Bernard Jones. A lognormal model for the cosmological mass distribution. *Monthly Notices of the Royal Astronomical Society*, 248 :1–13, jan 1991. ISSN 0035-8711. URL <http://adsabs.harvard.edu/abs/1991MNRAS.248....1C>.
- Antony Lewis, Anthony Challinor, and Anthony Lasenby. Efficient Computation of CMB anisotropies in closed FRW models. nov 1999. doi : 10.1086/309179. URL <http://arxiv.org/abs/astro-ph/9911177><http://dx.doi.org/10.1086/309179>.
- M. M. Pieri, S. Bonoli, J. Chaves-Montero, I. Paris, M. Fumagalli, J. S. Bolton, M. Viel, P. Noterdaeme, J. Miralda-Escudé, N. G. Busca, H. Rahmani, C. Peroux, A. Font-Ribera, S. C. Trager, and The WEAVE Collaboration. WEAVE-QSO : A Massive Intergalactic Medium Survey for the William Herschel Telescope. nov 2016. URL <https://arxiv.org/abs/1611.09388>.
- James Farr, Andreu Font-Ribera, Hélión du Mas des Bourboux, Andrea Muñoz-Gutiérrez, Francisco Javier Sanchez Lopez, Andrew Pontzen, Alma Xochitl González-Morales, David Alonso, David Brooks, Peter Doel, Thomas Etourneau, Julien Guy, Jean-Marc Le Goff, Axel de la Macorra, Nathalie Palanque-Delabrouille, Ignasi Pérez-Ràfols, James Rich, Anže Slosar, Gregory Tarle, Duan Yutong, and Kai Zhang. LyCoLoRe : Synthetic Datasets for Current and Future Ly-alpha Forest BAO Surveys. dec 2019. doi : 10.1088/1475-7516/2020/03/068. URL <https://arxiv.org/abs/1912.02763>.

- Andreu Font-Ribera, Patrick Mcdonald, and Jordi Miralda-Escudé. Generating mock data sets for large-scale Lyman- α forest correlation measurements Generating mock data sets for large-scale Lyman- α forest correlation measurements. *Journal of Cosmology and Astroparticle Physics*, 2012. doi : 10.1088/1475-7516/2012/01/001. URL <http://iopscience.iop.org/article/10.1088/1475-7516/2012/01/001/pdf>.
- A J S Hamilton. Measuring $\{\Omega\}$ and the real correlation function from the redshift correlation function. *The Astrophysical Journal Letters*, 385 :L5—L8, jan 1992. ISSN 0004-637X. doi : 10.1086/186264. URL <http://adsabs.harvard.edu/abs/1992ApJ...385L...5H>.