

# 1

## Réalisation des mocks

---

Dans ce chapitre, nous présentons la construction des *mocks* : des simulations qui visent à reproduire les données d’eBOSS et de DESI. Ces mocks, nommés **SaclayMocks** et présentés dans ?, sont le cœur de ce manuscrit. Le code est écrit en Python<sup>1</sup> et se trouve en accès libre sur GitHub<sup>2</sup>. 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 $\alpha$ . Mais ces simulations sont très couteuses car elles nécessitent de modéliser les effets physiques qui affectent les paramètres mesurés. Les mocks, quant à eux, sont conçus afin de répliquer fidèlement et 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 due à l’ajustement du continuum du quasar est-elle correctement prise en compte par la matrice de distorsion ?
- de vérifier le calcul 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 $\alpha$  des données complète d’eBOSS (?), et seront utilisés dans l’analyse Ly $\alpha$  de DESI. L’objectif de ces mocks est donc de répliquer fidèlement les données Ly $\alpha$  d’eBOSS et de DESI. Ces relevés couvrent un volume de plusieurs dizaines de  $h^{-1}$  Gpc<sup>3</sup>, et les échelles sondées grâce au Ly $\alpha$  descendent jusqu’à la centaine de  $h^{-1}$  kpc. Les mocks nécessitent donc de reproduire un volume immense, avec une bonne résolution. Les simulations dites N-body sont des simulations qui traitent le problème à N corps. Elles sont initialisées avec une distribution de matière noire, représentée par des macro-particules de masse  $\sim 10^9 M_{\odot}$ , à un redshift élevé ( $z > 100$ ). 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-body. 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 des simulations N-body. En contre partie, 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. Cependant, ces simulations ne sont pas non plus adaptées à notre utilisation car le volume d’univers simulé est bien

---

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

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

trop petit : quelques dizaines de  $h^{-1}$  Mpc<sup>3</sup>. Ainsi, avoir un grand volume et une grande résolution requiert l'utilisation des *champs aléatoires gaussiens* (GRF pour Gaussian Random Field). 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. Les GRF sont donc idéaux pour simuler le champ de matière à grande échelle. Cependant, l'utilisation des GRF ne donne pas accès aux non linéarités qui peuvent émerger dans l'évolution des simulations N-body 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.

## 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 de ces champs de densité, comment sont tirés les quasars. Nous expliquons ensuite comment, de la position de chaque quasar, nous créons sa ligne de visée. Enfin, nous présentons comment de la densité le long de la ligne de visée nous calculons la 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. D'abord, un GRF est généré dans une boîte  $2560 \times 2560 \times 1536$ , chaque voxel faisant  $d_{cell}^3 = (2,19 h^{-1} \text{ Mpc})^3$ . Afin que le champ  $\delta$  possède la bonne fonction de corrélation, une transformation de Fourier 3D<sup>1</sup> est appliquée sur la boîte, puis le champ  $\delta_k$  ainsi obtenu est multiplié par

$$\sqrt{\frac{P(k)}{V_{cell}}}, \quad (1.1)$$

où  $P(k)$  est le spectre de puissance désiré, et  $V_{cell} = (2,19 h^{-1} \text{ Mpc})^3$  est le volume d'un voxel. Le GRF pourrait être tiré directement dans l'espace  $k$ , mais nous ne perdons pas beaucoup de temps CPU<sup>2</sup> à procéder comme cela. Nous distinguons ici deux champs : le champ utilisé pour tirer les quasars, et le champ utilisé pour créer l'absorption Ly $\alpha$ . Ces deux champs requièrent deux spectres de puissance différents, et donc deux boîtes de densité différentes.

#### Les quasars

Comme expliqué précédemment, l'utilisation des GRF ne permet pas de capturer l'évolution non linéaire du champ de matière. Plutôt que de modéliser ces non linéarités, nous considérons que les quasars suivent une distribution log-normale. 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, au lieu de placer les quasars dans les voxels qui possèdent une densité plus élevée qu'un certain seuil, nous tirons les quasars dans chaque voxel avec une probabilité

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

où  $\delta_q$  est le champ de densité dans le voxel considéré. Afin que les quasars suivent la fonction de corrélation  $\xi(r)$ , le champ  $\delta_q$  doit suivre la fonction de corrélation

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

---

1. <https://github.com/pyFFTW/pyFFTW> # prov ou citer directement FFTW ?

2. Le temps CPU (Central Processing Unit) désigne le temps utilisé par les processeurs d'une machine pour exécuter un code.

De manière à obtenir un relevé synthétique de quasar dont le biais dépend de  $z$ , nous utilisons trois champs qui suivent des distributions log-normales, à des redshifts différents. La probabilité pour tirer les quasars dépend de l'interpolation de ces 3 champs. Ces champs sont construits aux redshifts  $z_1 = 1,9$ ,  $z_2 = 2,75$ , et  $z_3 = 3,6$ . Pour chacun des champs, nous partons du spectre de puissance de la matière  $P_{matière}(k)$  à  $z = 0$ , fournit par Camb (Lewis et al., 1999). Nous multiplions ensuite ce spectre de puissance par  $(b_{QSO}(z_i)G(z_i))^2$ , 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  $P_{QSO,i}(k)$ , à appliquer au champ  $\delta_k$ , comme la transformée de Fourier de  $\xi_{QSO,i}(r) = \ln(1 + \xi_i(r))$  (équation ??). Une fois les trois spectres de puissances  $P_{QSO,i}(k)$  obtenus, nous construisons 3 champs

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

où  $\delta_k$  est le GRF dans l'espace de Fourier. Une fois ces 3 champs construits, nous appliquons à chacun d'entre eux une transformation de Fourier inverse afin d'obtenir les champs  $\delta_{QSO,i}$ . Ces champs seront interpolés en  $z$ , puis les quasars seront ensuite tirés avec une probabilité  $\propto \exp(\delta_{QSO}(z))$ , où  $\delta_{QSO}$  est le champ interpolé. Nous expliquons cette étape dans la section 2.2.

### Le champ Ly $\alpha$

Afin de construire le champ d'absorption Ly $\alpha$ , 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  $\tau$  par

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

De plus, la formule FGPA (Fluctuating Gunn Peterson Approximation) permet de relier la profondeur optique  $\tau$  au contraste de densité  $\delta$  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 $\alpha$  et la bonne transmission moyenne  $\bar{F}$ . Leur détermination est décrite dans la section ?. Le facteur de croissance  $G$  prend en compte l'évolution avec le redshift du champ de densité  $\delta$ . 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 $\alpha$ . Pour ce faire, nous partons du même champ  $\delta_k$  utilisé pour construire les 3 champs log-normaux des quasars. Ceci garanti la corrélation croisée entre les quasars et le champ d'absorption Ly $\alpha$ . Le spectre de puissance de la matière  $P_{matière}(k)$  à  $z = 0$  est ensuite appliqué au champ  $\delta_k$ . Enfin, nous obtenons le champ de densité  $\delta_{matière}$  à  $z = 0$  qui servira au calcul du champ d'absorption du Ly $\alpha$  en effectuant la transformée de Fourier du champ

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

### Les champs des vitesses

Afin d'inclure les RSD dans nos mocks, nous simulons aussi le champ des vitesses. En première approximation, le champs 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  $\delta_k$  par la relation

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

Il est fréquent 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 les déplaçant proportionnellement à leur vitesse le long de la ligne de visée (équation ??). Dans ce but, nous calculons les trois champs 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, le champ  $\delta_{k,matière}$  est le même que celui utilisé pour construire les champs  $\delta_{QSO,i}$  des quasars et le champ  $\delta_{matière}$  utilisé pour le Ly $\alpha$ , ceci afin de garantir la corrélation entre la densité des traceurs et leurs vitesses particulières. A l'aide d'une transformation de Fourier, nous obtenons les trois champs de vitesse à  $z = 0$  dans l'espace réel  $v_x$ ,  $v_y$  et  $v_z$ .

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

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

Cette équation nous permet de retrouver la formule de kaiser :

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

Nous construisons donc 6 champs de gradients de vitesse, à  $z = 0$ , comme

$$\eta_{nm}(\vec{k}) = \frac{k_n k_m}{k^2} \delta_k(\vec{k}) ; \quad (n, m) \in [x, y, z]^2 . \quad (1.12)$$

Le champ  $\delta_k$  utilisé est le GRF initial, afin de garantir les corrélations entre les différents champs. Nous omettons volontairement le facteur  $f(z = 0)$  à ce stade. Il sera pris en compte lors de l'ajout de la dépendance en redshift (voir section 2.4). A l'aide d'une transformation de Fourier, nous obtenons les 6 champs de gradients de vitesses à  $z = 0$  dans l'espace réel  $\eta_{xx}$ ,  $\eta_{yy}$ ,  $\eta_{zz}$ ,  $\eta_{xy}$ ,  $\eta_{yz}$  et  $\eta_{xz}$ .

## 2.2 Le relevé de quasars

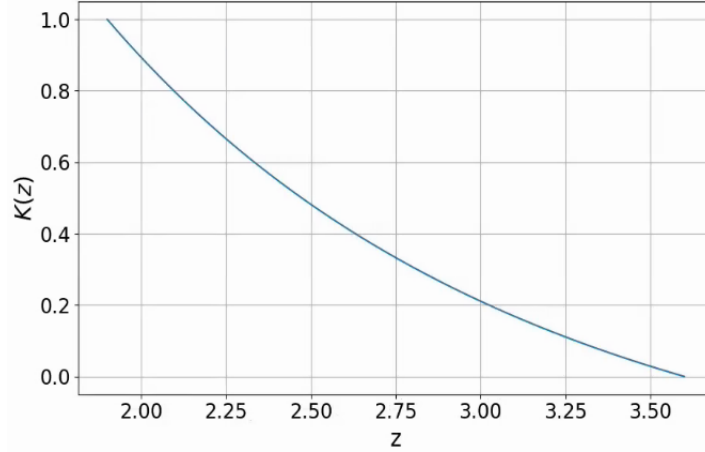
Une fois tous ces champs construits, nous définissons la géométrie du relevé. Les boîtes, d'une taille  $2560 \times 2560 \times 1536$  selon les axes  $x$ ,  $y$  et  $z$  respectivement, sont placées à un redshift central  $z_0 = 1.70975268202$ , et à une ascension droite  $\alpha_0$  et une déclinaison  $\delta_0$  (voir équation 1.17). 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 de construire le relevé de quasars, nous utilisons les trois champs  $\delta_{QSO,i}$  construits précédemment, aux redshifts  $z_1 = 1.9$ ,  $z_2 = 2.75$ , et  $z_3 = 3.6$ . Dans chacun des cas, nous calculons

$$\hat{\delta}_{QSO,i}(z) = \exp \left( \delta_{QSO,i} \frac{b_{QSO}(z)(1 + z_i)}{b_{QSO}(z_i)(1 + z)} \right) , \quad (1.13)$$

où  $b_{QSO}$  est le biais des quasars. Le redshift dans chaque voxel est calculé en utilisant l'équation ??. Les paramètres cosmologiques utilisés sont donnés dans l'équation ??. Toutes les distances sont comobiles. Une fois les 3 champs  $\hat{\delta}_{QSO,i}$  construits, nous construisons les deux champs

$$\hat{\delta}_{QSO,12}(z) = \hat{\delta}_{QSO,1}(z) \frac{z_2 - z}{z_2 - z_1} + \hat{\delta}_{QSO,2}(z) \frac{z - z_1}{z_2 - z_1} , \quad (1.14)$$

$$\hat{\delta}_{QSO,23}(z) = \hat{\delta}_{QSO,2}(z) \frac{z_3 - z}{z_3 - z_2} + \hat{\delta}_{QSO,3}(z) \frac{z - z_2}{z_3 - z_2} , \quad (1.15)$$


 FIGURE 1.1 – Coefficient  $K(z)$  défini dans l'équation 1.16.

puis, nous construisons le champ interpolé

$$\hat{\delta}_{\text{QSO}}(z) = K(z) \left( \hat{\delta}_{\text{QSO},12}(z) - \hat{\delta}_{\text{QSO},23}(z) \right) + \hat{\delta}_{\text{QSO},23}(z), \quad (1.16)$$

où  $K(z)$  est un coefficient qui varie entre 0 et 1. Il est représenté sur la figure ???. Les quasars sont ensuite tirés dans chaque voxel, avec une probabilité  $P \propto \hat{\delta}_{\text{QSO}}$ . Pour ce faire, nous générons un champ  $\phi$  aléatoire uniforme entre 0 et 1, et les voxels pour lesquelles  $\phi < N(z)\hat{\delta}_{\text{QSO}}$  hébergent potentiellement un quasar. Le facteur  $N(z)$  est 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  $\alpha$  et la déclinaison  $\delta$  du point  $(x, y, z)$  sont définies comme

$$\alpha = \arctan\left(\frac{\cos(\alpha_0)x - \sin(\delta_0)\sin(\alpha_0)y + \cos(\delta_0)\sin(\alpha_0)z}{-\sin(\alpha_0)x - \sin(\delta_0)\cos(\alpha_0)y + \cos(\delta_0)\cos(\alpha_0)z}\right), \quad (1.17)$$

$$\delta = \arcsin\left(\frac{\cos(\delta_0)y + \sin(\delta_0)z}{\sqrt{x^2 + y^2 + z^2}}\right). \quad (1.18)$$

Enfin, grâce au facteur de normalisation  $N(z)$ , les quasars sont tirés selon la distribution en  $z$  prédite pour DESI. Cette distribution est présentée sur la figure ??. A la fin, nous obtenons environ 100 quasars à  $z > 2,1$  par degré carré.

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

$$v_{\parallel} = \frac{v_x x + v_y y + v_z z}{\sqrt{x^2 + y^2 + z^2}}. \quad (1.19)$$

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

$$R \rightarrow R + (1+z) \frac{dG}{dz} \frac{1}{H_0 \frac{dG}{dz}(z=0)} v_{\parallel}. \quad (1.20)$$

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.

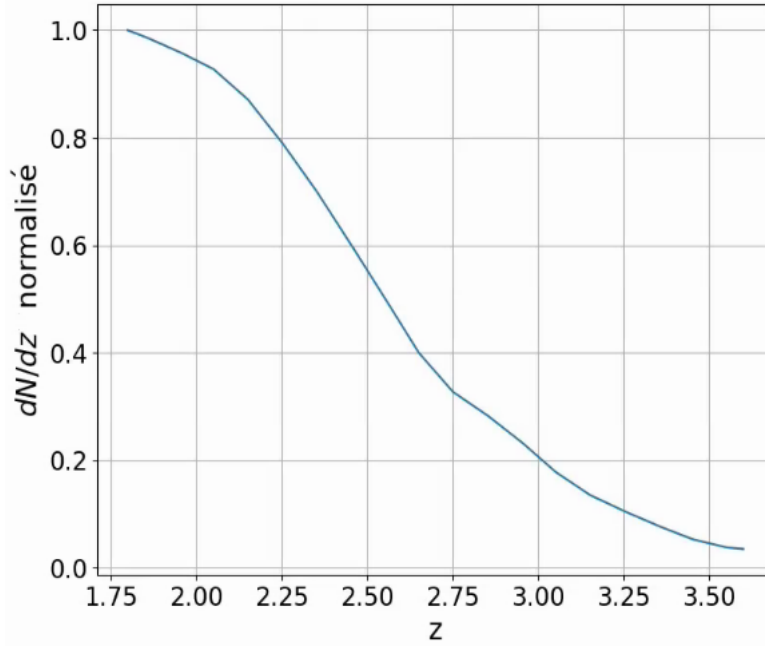


FIGURE 1.2 – Distribution normalisée en redshift des quasars tirés dans les mocks.

### 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 $\alpha$ . Nous pouvons donc créer les lignes de visées à partir de chaque quasar, et interpoler le champ de densité le long de celles-ci. Dans un premier temps, nous commençons par créer le vecteur en longueurs d'onde observées, sur lequel sera interpolé le champ 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. 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 (voir ??). Les longueurs d'onde observées couvrent donc  $3476,1877 < \lambda < 5591,566 \text{ \AA}$  à l'aide de 6524 pixels.

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  $\lambda_{QSO} = (1 + z_{QSO})\lambda_{Ly\alpha}$  est placé à la position  $(x_{QSO}, y_{QSO}, z_{QSO})$  du quasar, et le vecteur est dirigé vers l'observateur. Puis, pour chaque pixel  $i$  entre  $\lambda_{min} = 3476,1877$  et  $\lambda_{QSO}$ , la position  $(x_i, y_i, z_i)$  du pixel est déterminée. Le champ est alors interpolé puis lissé. Pour chaque pixel, nous considérons les voxcells appartenant au cube de 7 voxcells de côté, centré sur le voxcell dans lequel se trouve le pixel. Ceci représente donc 343 pixel à interpoler puis lisser. Le pixel  $i$  est alors donné par

$$\delta_i = \sum_{j=0}^{342} \frac{\delta(\vec{r}_j) e^{-\frac{(\vec{r}_j - \vec{r}_i)^2}{\sigma_j^2}}}{\sigma_j^2}, \quad (1.21)$$

#prov c'est pas correcte (voir le code)

où  $\vec{r}_j$  est la position du voxcell  $j$ ,  $\vec{r}_i$  celle du pixel  $i$ ,  $\delta$  est le champ à interpoler, et enfin  $\sigma^2 = 2d_{cell}^2$  est la largeur du lissage gaussien appliqué. Ce lissage est nécessaire afin d'éviter le crénelage aux petites échelles. Ce calcul est effectué pour tous les pixels qui vérifient  $\lambda_i < \lambda_{QSO}$ , pour chaque quasar. Les champs interpolés sont le champ  $\delta_{matière}$  utilisé pour construire l'absorption Ly $\alpha$ , les trois champs de vitesse utilisés pour ajouter les RSD aux HCD tirés dans chaque ligne de visée (voir section 2.5), et les six champs de gradient de vitesse afin d'ajouter les RSD au champ Ly $\alpha$ .

## 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 $\alpha$ . Ceci est fait via la formule FGPA :

$$F = \exp\left(-a(z)e^{b(z)G(z)\delta_{matière}}\right). \quad (1.22)$$

### Les petites échelles

Le champ  $\delta_{matière}$ , utilisé dans l'équation 1.22, est le champ de matière à grand échelle, construit grâce à la transformation de Fourier du champ  $\delta_k$ . Cependant, ce champ est construit sur une grille de taille  $d_{cell} = 2,19 h^{-1} \text{ Mpc}$ . Par conséquent, la plus petite échelle accessible est

$$k_N = \frac{2\pi}{d_{cell}} \sim 2,87 h \text{ Mpc}^{-1}. \quad (1.23)$$

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

$$P^{1D}(k_{\parallel}) = \frac{1}{2\pi} \int_{k_{\parallel}}^{\infty} k P(k) dk, \quad (1.24)$$

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

$$F = \exp\left(-a(z)e^{b(z)G(z)(\delta_l + \delta_s)}\right);. \quad (1.25)$$

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 dimensions<sup>1</sup>. 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

$$\sqrt{\frac{P_{miss}(k)}{d_{pix}}}, \quad (1.26)$$

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 ???. Enfin, nous obtenons  $\delta_s$  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

$$k_{max} = \frac{2\pi}{1536 d_{cell}} \sim 1,9 \times 10^{-3} h \text{ Mpc}^{-1}, \quad (1.27)$$

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

---

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 ??)



isotrope. Initialement, nous pensions ajouter les RSD au niveau du spectre de puissance : multiplier le GRF initial 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 parallèles (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'utiliser 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  $\delta_s$ , au champ  $\delta_l$ . 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

$$F = \exp\left(-a(z)e^{b(z)G(z)(\delta_l + \delta_s + c(z)\eta_{||})}\right); \quad (1.28)$$

Les champs  $\delta_l$ ,  $\delta_s$  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 côté 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  $\beta_{Ly\alpha}(z)$ ,  $\beta_{Ly\alpha}(z)$ ,  $\bar{F}(z)$  et  $P_{Ly\alpha}^{1D}(z)$ . L'ajustement est décrit dans la section ??.

## 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 considérée, 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  $Ly\alpha$  est la méthode choisie par ?. 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 ?, il est possible de relier la fonction de corrélation  $\xi_F$  du champ  $F$  à la fonction de corrélation  $\xi_g$  du champ  $\delta_g$ . Ces deux fonctions de corrélations sont reliées par

$$\xi_F(r_{12}) = \int_{-\infty}^{\infty} d\delta_{g1} \int_{-\infty}^{\infty} d\delta_{g2} \frac{\exp\left[-\frac{\delta_{g1}^2 + \delta_{g2}^2 - 2\delta_{g1}\delta_{g2}\xi_g(r_{12})}{2(1 - \xi_g^2(r_{12}))}\right]}{2\pi\sqrt{1 - \xi_g^2(r_{12})}} \delta_F(\delta_{g1})\delta_F(\delta_{g2}), \quad (1.29)$$

où  $\delta_g$  est un GRF de variance 1,  $\delta_F$  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  $\delta_g$  et  $\delta_F$ . Dans notre cas, le champ  $\delta_g$  représente le champ  $G(z)(\delta_l + \delta_s + c(z)\eta_{||})$ . Ce champ est un champ gaussien, de valeur moyenne nulle et de variance  $\sigma_g^2$ . Nous compensons le fait que  $\sigma_g^2 \neq 1$  en remplaçant  $\xi_g$  par  $\xi_g/\sigma_g^2$  dans l'équation précédente. L'équation 1.29 ne dépendant que de la valeur de  $\xi_g$ , nous construisons une table qui permet de relier chaque valeur de  $\xi_g \in [-1; 1]$  à la valeur  $\xi_F$  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  $\delta_l$  :

$$\xi(r, \mu) = \xi_0(r) + \frac{1}{2}(3\mu^2 - 1)\xi_2(r) + \frac{1}{8}(35\mu^4 - 30\mu^2 + 3)\xi_4(r), \quad (1.30)$$

avec

$$\xi_0(r) = \left(1 + \frac{2}{3}f + \frac{1}{5}f^2\right) \xi(r), \quad (1.31)$$

$$\xi_2(r) = \left(\frac{4}{3}f + \frac{4}{7}f^2\right) [\xi(r) - \bar{\xi}(r)], \quad (1.32)$$

$$\xi_4(r) = \frac{8}{35}f^2 \left[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)\right], \quad (1.33)$$

et

$$\bar{\xi}(r) = 3r^{-3} \int_0^r \xi(s) s^2 ds, \quad (1.34)$$

$$\bar{\bar{\xi}}(r) = 5r^{-5} \int_0^r \xi(s) s^4 ds. \quad (1.35)$$

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

$$P(k) = W^2(k) P_{matière}(z=0), \quad (1.36)$$

où  $W$  est le terme représentant le lissage gaussien appliqué au champ  $\delta_l$  (voir section 2.3) :

$$W(k) = e^{-\frac{k^2 d_{cell}^2}{2}}. \quad (1.37)$$

(#prov il y a pas un facteur  $\frac{1}{2}$  en trop ? on a  $\sigma_{smooth}^2 = 2d_{cell}^2$ )

A l'aide d'une transformation de Fourier, nous obtenons la fonction de corrélation  $\xi(r)$  que suit le champ interpolé  $\delta_l$ . Puis, nous calculons la fonction de corrélation dans l'espace des redshifts  $\xi_g(r, \mu)$  que suit le champ  $\delta_g$  grâce à l'équation 1.30. Enfin, pour tous les couples  $(r, \mu)$  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  $\xi_g$ . (#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 le champ de densité  $\delta_l$  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  $\delta_s$  bruyerait la corrélation entre les HCD et les autres traceurs.

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

$$b_\nu = \frac{pdf(\nu)}{cdf(-\nu)}, \quad (1.38)$$

où  $pdf(\nu)$  donne la densité de probabilité de  $\nu$ , et  $cdf(-\nu)$  est la fonction de répartition : la probabilité d'être au dessus du seuil  $\nu$ . Ainsi, pour avoir un biais de 2, il faut que la probabilité que le champ prenne la valeur  $\nu$  soit 2 fois plus grande que la probabilité que le champ soit au dessus du seuil  $\nu$ . Afin d'obtenir le seuil pour un biais donné, nous calculons  $b_\nu$  pour une large gamme de seuils  $\nu$  puis nous interpolons  $b_\nu$  sur  $\nu$ . Ainsi, pour un biais  $b$ , nous connaissons le seuil  $\nu(b)$  à choisir. Dans notre cas, le champ  $\delta_l$  suit une densité de probabilité gaussienne. Cependant, sa variance n'est pas égale à 1. De plus, le champ  $\delta_l$  interpolé le long des lignes de visée correspond au champ de matière à  $z=0$ .

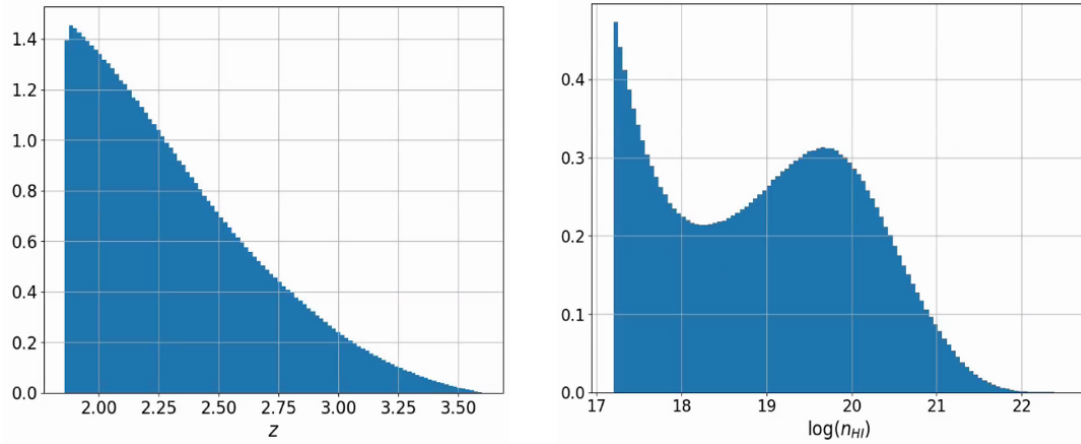


FIGURE 1.3 – Gauche : distribution normalisée en redshift des HCD. Droite : distribution normalisée de  $\log(n_{HI})$  des HCD. Ces distributions proviennent de la librairie `pyigm`.

Ainsi, pour obtenir un biais  $b_{HCD}$ , pour chaque redshift nous calculons le seuil  $\nu$  comme si nous visions un biais  $b = b_{HCD}\sigma_l G(z)$ . Le terme  $\sigma_l$  prend en compte la variance du champ  $\delta_l$ , et  $G(z)$  le fait que  $\delta_l$  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

$$\lambda(z) = \frac{N(z)}{cdf(-\nu(z))}, \quad (1.39)$$

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`<sup>1</sup>. La distribution en redshift des HCD est présentée sur le graphique de gauche de la figure 1.3. Une fois tous les HCD tirés, nous leur assignons une densité de colonne dans la gamme  $17,2 < \log(n_{HI}) < 22,5$ , selon la distribution donnée par `pyigm`. Cette distribution est présentée sur le graphique de droite de la figure 1.3. Enfin, nous ajoutons les RSD aux HCD tirés. Chaque HCD tiré est déplacé le long de la ligne de visée proportionnellement à la vitesse

$$v_{\parallel} = \frac{v_x X + v_y Y + v_z Z}{\sqrt{X^2 + Y^2 + Z^2}}, \quad (1.40)$$

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

$$z \rightarrow z + (1+z)H(z) \frac{dG}{dz} \frac{1}{H_0 \frac{dG}{dz}(z=0)} v_{\parallel}. \quad (1.41)$$

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` (présenté dans la section ??), qui utilise le catalogue de HCD que nous produisons.

### 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 ( $2560 \times 2560 \times 1536$ )

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

et leur résolution ( $2,19 h^{-1}$  Mpc) ne suffisent pas à couvrir les 14 000 degrés carrés de DESI. Pour pallier ce problème, nous construisons sept *chunks* indépendants, que nous assemblons pour former le relevé de DESI. Le découpage du relevé en 7 chunks est montré sur la figure ???. 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 NERSC<sup>1</sup>, avec la machine CORI. Sur cette machine, nous avons utilisé les noeuds “Haswell”, au nombre de 2388. Chaque noeud possède 32 cœurs, chacun possédant 2 hyper-threads. Chaque noeud peut donc gérer 64 tâches simultanément. De plus, ces noeuds possèdent 128 Go 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 avons choisi de ne pas utiliser cette librairie et d’effectuer les transformations de Fourier sur un seul noeud, ce qui explique le découpage du relevé en sept chunks indépendants. Nous profitons néanmoins des 64 threads de chaque noeud pour paralléliser notre code.

Dans les lignes qui suivent, nous détaillons les différents éléments du code. Le premier module, `interpolate_pk.py` permet d’interpoler le spectre de puissance

---

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.

# Bibliographie

---

- 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>.
- 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>.