

# 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 coûteuses car elles nécessitent de modéliser les effets physiques qui affectent les paramètres mesurés. Les mocks, quant à eux, sont conçus afin de répliquer fidèlement 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 $\alpha$  finale des données 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. De plus, la densité, 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 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*

---

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

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

*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 DLA.

### 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 cellule faisant  $(2,19h^{-1} \text{ Mpc})^3$ . 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  $V_{cell} = (2,19h^{-1} \text{ Mpc})^3$  est le volume d'une cellule. 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 (?), et est en accord avec ce qui est observé dans les données (?). Ainsi, au lieu de placer les quasars dans les cellules qui possèdent une densité plus élevée qu'un certain seuil, nous tirons les quasars dans chaque cellule avec une probabilité

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

où  $\delta_q$  est le champ de densité dans la cellule considérée. 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 (?). 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 ??.

### 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

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

### Le champ 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  $\delta_{matière}$  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 :

$$\delta_k^s(\vec{k}) = \delta_k(\vec{k}) + \eta_{||}(\vec{k}) , \quad (1.11)$$

$$= (1 + f \mu_k^2) \delta_k(\vec{k}) . \quad (1.12)$$

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

$$\eta_{nm}(\vec{k}) = \frac{k_n k_m}{k^2} \delta_k(\vec{k}) ; \quad (n, m) \in [x, y, z]^2 . \quad (1.13)$$

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 ??). 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

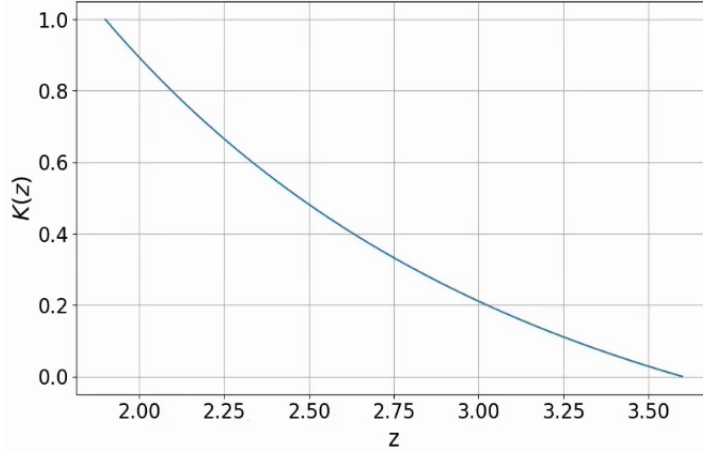
Une fois tous ces champs construits, nous définissons la géométrie du relevé. Les boîtes, d'une taille  $2560 \times 2560 \times 1536$  selon les axes  $X$ ,  $Y$  et  $Z$  respectivement, sont placées à un redshift central  $z_0 = 1.70975268202$ , et à une ascension droite  $\alpha_0$  et une déclinaison  $\delta_0$  (voir équation 1.18). 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 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

$$\hat{\delta}_{QSO,i}(z) = \exp \left( \delta_{QSO,i} \frac{b_{QSO}(z)(1+z_i)}{b_{QSO}(z_i)(1+z)} \right) , \quad (1.14)$$

où  $b_{QSO}$  est le biais des quasars. Le redshift dans chaque cellule 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

$$\hat{\delta}_{QSO,12}(z) = \hat{\delta}_{QSO,1}(z) \frac{z_2 - z}{z_2 - z_1} + \hat{\delta}_{QSO,2}(z) \frac{z - z_1}{z_2 - z_1} , \quad (1.15)$$

$$\hat{\delta}_{QSO,23}(z) = \hat{\delta}_{QSO,2}(z) \frac{z_3 - z}{z_3 - z_2} + \hat{\delta}_{QSO,3}(z) \frac{z - z_2}{z_3 - z_2} , \quad (1.16)$$


 FIGURE 1.1 – Coefficient  $K(z)$  défini dans l'équation 1.17.

puis, nous construisons le champ interpolé

$$\hat{\delta}_{QSO}(z) = K(z) \left( \hat{\delta}_{QSO,12}(z) - \hat{\delta}_{QSO,23}(z) \right) + \hat{\delta}_{QSO,23}(z), \quad (1.17)$$

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 cellule, avec une probabilité  $P \propto \hat{\delta}_{QSO}$ . Pour ce faire, nous générons un champ  $\phi$  aléatoire uniforme entre 0 et 1, et les cellules pour lesquelles  $\phi < N(z)\hat{\delta}_{QSO}$  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

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

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

Enfin, grâce au facteur 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

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

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

$$R \rightarrow R + (1+z) \frac{dG}{dz} \frac{1}{H_0 \frac{dG}{dz}(z=0)} v_{\parallel}. \quad (1.21)$$

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.

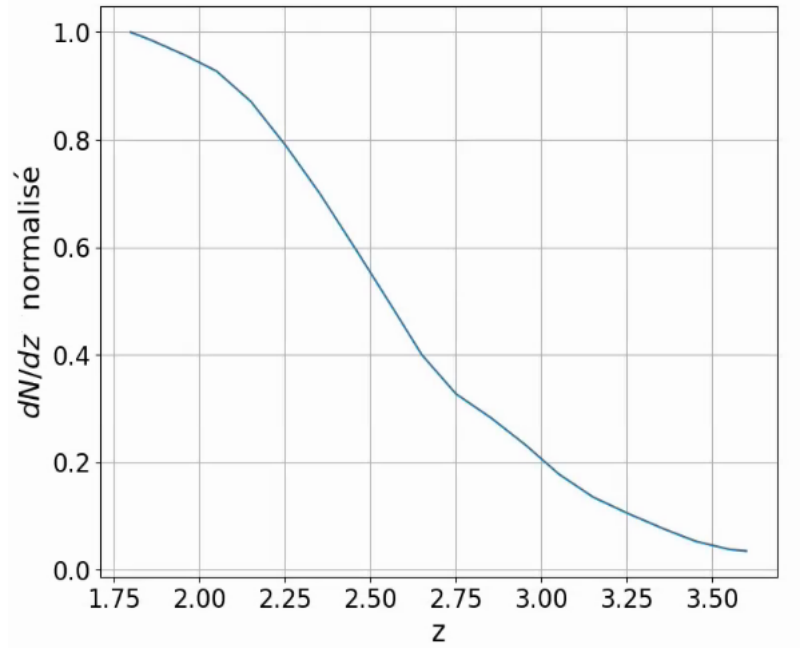


FIGURE 1.2 – Distribution normalisée en redshift des quasars tirés dans les mocks.

### 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  $\text{Ly}\alpha$ . Nous pouvons donc créer les lignes de visées à partir de chaque quasar, et interpoler le champ de densité le long de celles-ci.