

# 1

## Validation des mocks

---

Ce chapitre a pour vocation de présenter l’analyse menée sur les mocks, afin de valider leur construction et le choix des différents paramètres. Nous présentons d’abord le calcul des diverses fonctions de corrélations, puis l’analyse des 100 réalisations produites. Pour faire cette analyse, nous utilisons le code `picca` qui permet de calculer les champs  $\delta$ , les fonctions de corrélations, les matrices de distorsion, les matrices de covariance ainsi que de réaliser l’ajustement du modèle.

## 1 Les estimateurs

Nous présentons ici les estimateurs utilisés pour calculer la fonction d’auto-corrélation du  $\text{Ly}\alpha$ , la fonction de corrélation croisée  $\text{Ly}\alpha$ -QSO, ainsi que la fonction d’auto-corrélation d’objets ponctuels tels les quasars. Dans le cas des objets ponctuels, la seule information nécessaire au calcul de la fonction de corrélation est le catalogue fournissant pour chaque objet l’ascension droite, la déclinaison et le redshift. Pour le calcul des fonctions de corrélation impliquant le  $\text{Ly}\alpha$ , il est nécessaire de calculer au préalable le champ  $\delta_F$  (équation ??). Nous distinguons deux cas : le cas où nous analysons les *raw mocks* et le cas où nous analysons les *cooked mock* ou les données. Les raw mocks (les mocks bruts, non préparés) désignent les mocks avant d’avoir tourné le code `quickquasars`. Les fonctions de corrélation sont alors calculées en utilisant directement la fraction de flux transmise  $F$  donnée dans les fichiers de transmissions. Etant donné que pour chaque forêt, nous avons directement accès à  $F$ , le champ  $\delta_F$  est donné par

$$\delta_F = \frac{F}{\overline{F}} - 1. \quad (1.1)$$

Les cooked mocks quant à eux désignent les mocks après avoir appliqué `quickquasars` à chaque forêt. Dans ce cas, comme pour les données, la fraction de flux transmise  $F$  n’est pas accessible. Il faut alors utiliser la procédure d’ajustement du continuum décrite dans la section ?. Une fois le champ  $\delta_F$  calculé, nous pouvons calculer les différentes fonctions de corrélation.

### 1.1 L’auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$

Comme expliqué dans le chapitre d’introduction, la fonction d’auto-corrélation du champ d’absorption  $F$  du  $\text{Ly}\alpha$  est définie comme

$$\xi_{\text{Ly}\alpha}(\vec{r}) = \langle \delta_F(\vec{r}') \delta_F(\vec{r} + \vec{r}') \rangle. \quad (1.2)$$

Elle peut-être mise sous la forme

$$\xi(\vec{r}) = \langle \delta_i \delta_j \rangle, \quad (1.3)$$

où  $\langle . \rangle$  désigne la moyenne sur tous les pixels  $i$  et  $j$  qui vérifient  $\vec{r}_{ij} = \vec{r}$ . Afin de calculer  $\xi_{\text{Ly}\alpha}$ , nous créons une grille en  $(r_{\parallel}, r_{\perp})$ . Les bins mesurent  $4 h^{-1}$  Mpc dans chaque direction. Cette taille de bin permet à la fois de regrouper suffisamment de pixels pour avoir une bonne statistique dans chaque bin, mais aussi d’avoir suffisamment de bins pour résoudre correctement la région du pic BAO (#prov est-ce que c’est vrai ? toujours la même quantité d’info, donc ça améliore pas le fit ?). Une fois cette grille construite, nous transformons les coordonnées  $(\alpha, \delta, z)$  des pixels  $\delta_F$  en distances (#prov donner les transformations géométriques ?). Puis, pour chaque bin  $A$  de la grille en  $(r_{\parallel}, r_{\perp})$ , nous considérons toutes les paires de pixels dont la distance de séparation se trouve dans ce bin  $A$ . La fonction de corrélation dans ce bin est alors donnée par

$$\xi_A = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i \delta_j}{\sum_{(i,j) \in A} w_i w_j}, \quad (1.4)$$

où  $w_i$  est le poids associé au pixel  $i$ . Ces poids sont définis dans l’équation ?. La correction de la dépendance en redshift du biais du  $\text{Ly}\alpha$  permet de donner plus de poids aux pixels à grand redshift,

où l'amplitude de la fonction de corrélation est plus importante. Le calcul de la fonction de corrélation est donc effectué à l'aide d'une double boucle sur les pixels. Afin de réduire le temps de calcul, nous calculons la fonction de corrélation uniquement pour les paires de pixels pour lesquelles  $r_{\parallel}$  et  $r_{\perp}$  sont inclus dans  $[0; 200]h^{-1}$  Mpc. La fonction de corrélation est donc calculée dans  $50 \times 50 = 2500$  bins. Les paires formées par des pixels provenant de la même forêt sont exclues du calcul afin d'éviter que les erreurs sur l'ajustement du continuum biaisent la mesure de la fonction de corrélation.

L'analyse Ly $\alpha$  des données complète d'eBOSS (?) utilise deux fonctions de corrélation distinctes : les fonctions de corrélation Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\alpha$ ) et Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\beta$ ). La région Ly $\beta$  ne fournit pas suffisamment de pixels pour pouvoir calculer la fonction de corrélation Ly $\alpha$ (Ly $\beta$ ) $\times$ Ly $\alpha$ (Ly $\beta$ ) et y détecter le pic BAO. L'utilisation de ces deux fonctions de corrélation permet de mesurer la position du pic BAO avec une plus grande précision. Dans ce manuscrit, nous nous intéressons à la mesure de  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  afin de construire correctement nos mocks. Pour limiter les potentielles systématiques, nous considérons donc uniquement la fonction de corrélation Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\alpha$ ). Le graphique de droite de la figure ?? présente les distributions pondérées en redshift des paires Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\alpha$ ) et Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\beta$ ). L'analyse des données DR16 que nous présentons ici considère donc uniquement la distribution indiquée en orange.

## 1.2 La corrélation croisée Ly $\alpha$ $\times$ QSO

Nous donnons dans cette section l'estimateur de la fonction de corrélation croisée Ly $\alpha$  $\times$ QSO. De la même manière que précédemment, le calcul s'effectue dans des bins en  $(r_{\parallel}, r_{\perp})$ . La fonction de corrélation dans le bin  $A$  est donnée par

$$\xi_A = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i}{\sum_{(i,j) \in A} w_i w_j}, \quad (1.5)$$

où l'indice  $i$  court sur les pixels des forêts et  $j$  sur les quasars pour lesquels la distance de séparation est comprise dans le bin  $A$ . Comme précédemment, les paires  $ij$  où le pixel  $i$  appartient à la forêt du quasar  $j$  sont rejetées. Les poids  $w_j$  associés aux quasars, similairement au Ly $\alpha$ , favorisent les quasars à plus grand redshift. Ces poids sont définis comme

$$w_j = \left( \frac{1 + z_j}{1 + 2.25} \right)^{\gamma_{\text{QSO}}}, \quad (1.6)$$

où  $\gamma_{\text{QSO}} = 1.44 \pm 0.08$  (des Bourboux et al., 2019). Dans le calcul de la fonction de corrélation croisée Ly $\alpha$  $\times$ QSO, nous nous restreignons aux paires pour lesquelles  $r_{\perp} \in [0; 200]h^{-1}$  Mpc. Contrairement à l'auto-corrélation du Ly $\alpha$ , la fonction de corrélation croisée n'est pas symétrique en  $r_{\parallel}$ . Nous la calculons donc dans les bins pour lesquels  $r_{\parallel} \in [-200; 200]h^{-1}$  Mpc. La fonction de corrélation croisée est donc calculée dans  $50 \times 50 = 2500$  bins.

## 1.3 Le spectre de puissance à une dimension

Comme expliqué dans le chapitre précédent, nous ajustons le spectre de puissance  $P_{\text{miss}}$  appliqué à  $\delta_s$  de façon à obtenir un spectre de puissance à une dimension  $P^{1D}$  en accord avec les données. Nous présentons donc maintenant la mesure de ce spectre de puissance. Pour chaque forêt, nous appliquons une transformation de Fourier au champ  $\delta_F$  afin d'obtenir le champ  $\delta_k$ . Puis, le spectre de puissance de cette forêt est obtenu comme

$$P^{1D}(k) = |\delta_k^2|. \quad (1.7)$$

Nous répétons cette procédure pour toutes les forêts, puis le spectre de puissance total est obtenu comme la moyenne du spectre de puissance de chaque forêt.

En ce qui concerne les données, la mesure du spectre de puissance à une dimension est plus complexe. Un certain nombre d'effets liés à la mesure doivent être pris en compte. Le bruit, par exemple, doit être estimé afin d'être pris en compte pour ne pas fausser l'estimation du spectre de puissance. Cette analyse est détaillée dans Chabanier et al. (2018).

## 1.4 La fonction de corrélation à une dimension

Important / nécessaire ?

## 2 Les matrices de distorsion

L'ajustement du continuum nécessaire au calcul du champ  $\delta_F$  dans les données et les coocked mocks biaise le champ mesuré. Cependant, grâce à la transformation (équation ??) décrite dans la section ??), l'effet sur la fonction de corrélation peut-être pris en compte. L'idée est la suivante : modéliser la distorsion induite par l'ajustement du continuum et par la transformation ?? sur la fonction de corrélation, appliquer cette distorsion au modèle, puis ajuster le modèle "distordu" aux données. L'ajustement du continuum et la transformation ?? induisent des corrélation le long de la ligne de visée. Au premier ordre, nous pouvons considérer que chaque  $\delta_F$  d'une forêt est une combinaison linéaire de tous les  $\delta_F$  de cette forêt. La fonction de corrélation distordue dans le bin  $A$  peut alors être reliée à la vraie fonction de corrélation comme

$$\xi_{distortion}(A) = \sum_B D_{AB} \xi_{vraie}(B), \quad (1.8)$$

où  $D_{AB}$  est appelée la *matrice de distorsion*. Celle-ci s'exprime en fonction du projecteur  $\eta_{ij}^q$  défini dans l'équation ?. Pour l'auto-corrélation, elle est défini comme

$$D_{AB} = \frac{1}{W_A} \sum_{(i,j) \in A} w_i w_j \left( \sum_{(i',j') \in B} \eta_{ii'} \eta_{jj'} \right) \quad (1.9)$$

où  $W_A = \sum_{(i,j) \in A} w_i w_j$  est le poids du bin  $A$ . Pour la corrélation croisée, la matrice de distorsion est donnée par

$$D_{AB} = \frac{1}{W_A} \sum_{(i,j) \in A} w_i w_j \left( \sum_{(i',j') \in B} \eta_{ii'} \right). \quad (1.10)$$

Comme précédemment, les indices  $i$  correspondent aux pixels des forêts, et  $j$  aux quasars. A cause de la double somme, le calcul de la matrice de distorsion est très long. Afin de rendre possible l'estimation de cette dernière, le calcul est fait sur 1 % des paires, tirées au hasard (#prov Bautista 2017 montre que c'est OK avec 5%, est-ce que y a une etude qui montre que c'est ok avec 1% ? Dans DR14, ils utilisent 5% je pense).

La figure ?? (#prov faire la figure) montre l'effet de la distorsion sur l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  dans les mocks. La fonction de corrélation est présentée dans quatres gammes en  $\mu$ . La courbe noire indique la fonction de corrélation calculée sur les raw mocks, et la courbe bleu la fonction de .... #prov Refaire la figure 11 de Bautista : CF des raw mocks (stack) + CF sur les coocked mocks (stack) avec le fit sur les raw + fit \* DM

La matrice de distorsion est un objet uniquement géométrique. Son calcul est indépendant du champ  $\delta_F$ . Elle ne dépend uniquement de la géométrie du relevé (#prov comment ?) et de la distribution des poids. La figure ?? montre la différence entre #prov montre la difference du fit d'une réa avec sa DM et avec une autre DM et/ou le stack de 10 avec les 10 DM ou dix fois la meme. Est-ce que les DM eboss-0.0 et eboss-0.2 sont différentes ? Ainsi, il n'est pas nécessaire de calculer la matrice

de distorsion pour les 100 réalisations des mocks. Dans l'analyse présentée dans la suite de ce chapitre, nous calculons la matrice de distorsion pour l'auto-corrélation et pour la corrélation croisée une seule fois (#prov ou une fois pour chaque run de quickquasars?). L'ajustement de chaque fonction de corrélation utilise une de ces deux matrices de distorsion.

### 3 Les matrices de covariance

Afin de réaliser l'ajustement de chaque fonction de corrélation, nous avons besoin de calculer les matrices de covariance associées à ces fonctions de corrélation. La covariance de la fonction de corrélation  $\xi$  dans le bin  $A$  et de  $\xi$  dans le bin  $B$  est définie comme

$$C_{AB} = \langle \xi_A \xi_B \rangle - \langle \xi_A \rangle \langle \xi_B \rangle . \quad (1.11)$$

De cette matrice de covariance, nous définissons la matrice de corrélation comme

$$Corr_{AB} = \frac{C_{AB}}{\sqrt{C_{AA}C_{BB}}} . \quad (1.12)$$

La matrice de corrélation donne la corrélation, comprise dans  $[-1;1]$ , d'un bin  $A$  avec un bin  $B$ . Afin d'estimer la matrice de covariance, le relevé est divisé en HEALPix pixels, en utilisant `nside` = 16. Cette résolution produit des pixels d'une taille sur le ciel de  $3,7 \times 3,7 = 13,4 \text{ deg}^2$ , correspondant à  $250 \times 250 (h^{-1} \text{ Mpc})^2$  à un redshift  $z = 2,33$ . Ces sous-échantillons sont suffisamment grands pour pouvoir négliger les corrélations entre différents HEALPix pixels et ainsi estimer la matrice de covariance comme la variance d'un sous-échantillon à un autre. La matrice de covariance est donc calculée comme

$$C_{AB} = \frac{1}{W_A W_B} \sum_s W_A^s W_B^s (\xi_A^s \xi_B^s - \xi_A \xi_B) , \quad (1.13)$$

où  $s$  est un sous-échantillon, et  $W_A^s$  les poids du bin  $A$  de ce sous-échantillon. Les principaux éléments de cette matrice sont les éléments diagonaux : la variance dans chaque bin. Les éléments non-diagonaux, les covariances entre deux bins distincts, sont faibles (#prov donner comme `Var_A` est modélisé? Donner comment on modélise les éléments non diag?). Leur estimation est bruitée. La matrice de covariance est donc lissée après avoir été estimée. En ce qui concerne l'auto-corrélation, la matrice de covariance possède  $2500 \times 2500$  bins. Pour la corrélation croisée, elle en possède  $5000 \times 5000$ .

## 4 Modélisation des fonctions de corrélation

Dans cette section, nous présentons les modèles utilisés pour ajuster les fonctions de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  et  $\text{Ly}\alpha \times \text{QSO}$ . Nous présentons d'abord les modèles utilisés dans l'analyse des données DR16, puis nous donnons les modèles utilisés pour analyser les mocks.

### 4.1 Modélisation des données

Pour l'analyse des données DR16, nous utilisons le modèle décrit dans ?. L'analyse décrite dans cette étude est une analyse BAO : l'auto-corrélation et la corrélation croisée sont modélisées de façon à mesurer au mieux les paramètres BAO  $\alpha_{\parallel}$  et  $\alpha_{\perp}$ . Pour ce faire, le modèle est séparé en deux composantes. La première,  $\xi_{\text{smooth}}$ , correspond à la forme globale de la fonction de corrélation. La seconde,  $\xi_{\text{peak}}$ , correspond au pic BAO. C'est cette seconde composante qui dépend des paramètres BAO :

$$\xi(r_{\parallel}, r_{\perp}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\text{smooth}}(r_{\parallel}, r_{\perp}) + \xi_{\text{peak}}(\alpha_{\parallel} r_{\parallel}, \alpha_{\perp} r_{\perp}) . \quad (1.14)$$

Cette séparation est opérée au niveau du spectre de puissance. Le modèle de la fonction de corrélation est ensuite obtenue à l'aide d'une transformation de Fourier.

Le spectre de puissance utilisé dans la modélisation de la fonction de corrélation des traceurs  $i$  et  $j$  s'exprime comme

$$P(\vec{k}) = b_i b_j (1 + \beta_i \mu_k^2) (1 + \beta_j \mu_k^2) P_{\text{QL}}(\vec{k}) F_{\text{NL}}(\vec{k}) G(\vec{k}). \quad (1.15)$$

Les termes  $b_i(1 + \beta_i \mu_k^2)$  et  $b_j(1 + \beta_j \mu_k^2)$  sont les facteurs de Kaiser (équation ??) relatifs aux traceurs  $i$  et  $j$ .  $P_{\text{QL}}$  est le spectre de puissance *quasi-linéaire*. Il est découpé en deux composantes  $P_{\text{smooth}}$  et  $P_{\text{peak}}$  comme

$$P_{\text{QL}}(\vec{k}, z) = P_{\text{smooth}}(k, z) + \exp\left(-\frac{k_{\parallel}^2 \Sigma_{\parallel}^2 + k_{\perp}^2 \Sigma_{\perp}^2}{2}\right) P_{\text{peak}}(k, z). \quad (1.16)$$

$P_{\text{smooth}}$  est le spectre de puissance linéaire, sans les BAO. Il est construit à partir du spectre de puissance linéaire  $P_{\text{L}}$  donné par Camb, puis les BAO sont retirées en utilisant la technique *side-band* décrite dans Kirkby et al. (2013). Le spectre de puissance  $P_{\text{peak}}$  est alors obtenu comme la différence  $P_{\text{L}} - P_{\text{smooth}}$  : il contient uniquement les oscillations dues aux BAO présentes dans le spectre de puissance linéaire. Le terme exponentiel devant  $P_{\text{peak}}$ , paramétré par  $\Sigma_{\parallel}$  et  $\Sigma_{\perp}$ , prend en compte l'élargissement non linéaire du pic BAO (Eisenstein et al., 2007). Nous utilisons  $\Sigma_{\parallel} = 6,42 h^{-1} \text{ Mpc}$  et  $\Sigma_{\perp} = 3,26 h^{-1} \text{ Mpc}$ . Le terme  $F_{\text{NL}}$  prend en compte les non linéarités aux petites échelles. Nous distinguons  $F_{\text{NL}}^{\text{auto}}$  et  $F_{\text{NL}}^{\text{cross}}$ . Pour l'auto-corrélation, les effets non linéaires proviennent de l'élargissement thermique, des vitesses particulières et de la croissance des structures non linéaire. Comme lors de la modélisation du  $P^{1D}$ , nous utilisons le modèle décrit dans Arinyo-i Prats et al. (2015). Nous avons donc  $F_{\text{NL}}^{\text{auto}}(k, \mu) = D(k, \mu)$ , où  $D$  est défini dans l'équation ?. Les paramètres utilisés sont une interpolation à  $z = 2,334$  de ceux donnés dans la section "Planck" de la table 7 de Arinyo-i Prats et al. (2015). Pour la corrélation croisée, l'effet dominant est dû aux vitesses non linéaires des quasars. Cet effet est modélisé par une lorentzienne :

$$F_{\text{NL}}^{\text{cross}}(k_{\parallel}) = \frac{1}{1 + (k_{\parallel} \sigma_v)^2}, \quad (1.17)$$

où l'inverse de la demi-largeur à mi-hauteur  $\sigma_v$  est un paramètre libre. L'effet dû aux erreurs statistiques sur la mesure du redshift des quasars étant confondu avec l'effet des vitesses non linéaires des quasars, il est aussi pris en compte par le terme  $F_{\text{NL}}^{\text{cross}}$ . Enfin, le terme  $G(\vec{k})$  prend en compte l'effet du binning utilisé lors du calcul de la fonction de corrélation. Il est défini comme le produit des transformés de Fourier de la fonction porte :

$$G(\vec{k}) = \text{sinc}\left(\frac{k_{\parallel} R_{\parallel}}{2}\right) \text{sinc}\left(\frac{k_{\perp} R_{\perp}}{2}\right), \quad (1.18)$$

avec  $R_{\parallel}$  et  $R_{\perp}$  la largeur des bins, soit  $4 h^{-1} \text{ Mpc}$ .

Afin de modéliser la corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ , nous ajoutons le paramètre  $\Delta_{r_{\parallel}, \text{QSO}}$ , inclus comme

$$r_{\parallel} = r_{\parallel, \text{measure}} + \Delta_{r_{\parallel}, \text{QSO}}, \quad (1.19)$$

où  $r_{\parallel, \text{measure}}$  est la séparation mesurée des paires  $(i, j)$ , et  $r_{\parallel}$  est la séparation utilisée dans le modèle de la corrélation croisée. L'ajout de ce paramètre permet de prendre en compte les erreurs systématiques sur la mesure du redshift des quasars, qui rendent asymétrique la fonction de corrélation  $\text{Ly}\alpha \times \text{QSO}$ .

La présence des facteurs de Kaiser dans l'équation 1.15 permet de mesurer le biais et le paramètre RSD de nos traceurs. En ce qui concerne l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ , la fonction de corrélation est proportionnelle à  $b_{\text{Ly}\alpha}^2 (1 + \beta_{\text{Ly}\alpha} \mu^2)^2$ . Cependant, la présence de HCD dans les données modifie le biais effectif du  $\text{Ly}\alpha$ . En effet, l'efficacité de l'algorithme de détection n'étant pas de 100 %, il subsiste des DLA non identifiés dans les forêts. De plus, les HCD avec  $\log n_{\text{HI}} < 20.3$  ne sont pas identifiés.

Ces deux effets participent à augmenter le biais mesuré du Ly $\alpha$  significativement. Nous utilisons les paramètres effectifs

$$b'_{\text{Ly}\alpha} = b_{\text{Ly}\alpha} + b_{\text{HCD}} F_{\text{HCD}}(k_{\parallel}) , \quad (1.20)$$

$$b'_{\text{Ly}\alpha} \beta'_{\text{Ly}\alpha} = b_{\text{Ly}\alpha} \beta_{\text{Ly}\alpha} + b_{\text{HCD}} \beta_{\text{HCD}} F_{\text{HCD}}(k_{\parallel}) , \quad (1.21)$$

où  $F_{\text{HCD}}$  est une fonction qui dépend de la distribution en  $z$  et en  $\log n_{\text{HI}}$  des HCD (Font-Ribera and Miralda-Escudé, 2012). Cette fonction est estimée sur des simulations hydrodynamiques (Rogers et al., 2017). Nous la modélisons comme

$$F_{\text{HCD}}(k_{\parallel}) = \exp(-L_{\text{HCD}} k_{\parallel}) , \quad (1.22)$$

où  $L_{\text{HCD}}$  est la taille typique des HCD non masqués. La résolution du spectrographe d'eBOSS rend possible l'identification des DLAs dont la largeur est supérieure à 2 nm, correspondant à une taille d'environ  $14 h^{-1}$  Mpc à redshift effectif de la mesure. De plus,  $L_{\text{HCD}}$  est très dégénéré avec les autres paramètres du modèle, comme le biais des HCD ou les paramètres du Ly $\alpha$ . Nous fixons sa valeur à  $10 h^{-1}$  Mpc dans l'ajustement du modèle.

Afin de pouvoir ajuster le même modèle sur tous les bins  $(r_{\parallel}, r_{\perp})$ , nous tenons compte de la dépendance en redshift de  $\delta_F$ . Nous avons  $\delta_F(z) \propto G(z) b_{\text{Ly}\alpha}(z)$ , avec  $G(z) \propto (1+z)^{-1}$ . Concernant le biais du Ly $\alpha$ , nous utilisons la même dépendance que celle choisie lors du calcul des poids (équation ??), c'est à dire  $b_{\text{Ly}\alpha} \propto (1+z)^{\gamma_{\text{Ly}\alpha}}$ , avec  $\gamma_{\text{Ly}\alpha} = 2.9$  (McDonald et al., 2006). En ce qui concerne  $\beta_{\text{Ly}\alpha}$ , nous considérons lors de l'ajustement du modèle qu'il est indépendant du redshift. Comme montré dans l'analyse présentée dans le chapitre ??,  $\beta_{\text{Ly}\alpha}$  n'est pas indépendant du redshift. Cependant, le redshift moyen dans chaque bin  $(r_{\parallel}, r_{\perp})$  varie peu. Lors de l'analyse de l'ensemble des données DR16, il varie dans la gamme  $2,31 < z < 2,39$ . Cette variation correspond à une variation de  $\beta_{\text{Ly}\alpha}$  de moins de 5 %. De plus, elle est d'autant plus faible lorsque l'analyse est faite dans différents bins en redshift.

La fonction de corrélation croisée Ly $\alpha$   $\times$  QSO est sensible au produit  $b_{\text{Ly}\alpha}(1 + \beta_{\text{Ly}\alpha} \mu^2) b_{\text{QSO}}(1 + \beta_{\text{QSO}} \mu^2)$ . L'ajustement seul de cette fonction de corrélation ne permet donc pas de distinguer les paramètres Ly $\alpha$  des paramètres des QSO. Nous fixons donc  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$ . Concernant le biais, comme pour le Ly $\alpha$  nous prenant en compte l'évolution avec le redshift. Il est paramétrisé comme

$$b_{\text{QSO}}(z) = 3,77 \left( \frac{1+z}{1+2,334} \right)^{1,44} . \quad (1.23)$$

Cette paramétrisation est sensiblement équivalente à celle utilisée dans les mocks (équation ??) dans la gamme  $2,31 < z < 2,39$ .  $\beta_{\text{QSO}}$  est choisi constant et vaut  $\beta_{\text{QSO}} = f/b_{\text{QSO}}(2,334) = 0,257$ .

Une fois les composantes multiplicatives incluses au modèle, nous pouvons transformer le  $P(\vec{k})$  modèle défini dans l'équation 1.15 en fonction de corrélation  $\xi(r, \mu)$ . Cette transformation est faite en utilisant FFTLog (Hamilton, 1999) : la fonction de corrélation est décomposée en polynômes de Legendre  $P_l$  jusqu'à  $l_{\text{max}} = 6$ . Pour chaque  $l \in [0; 2; 4; 6]$ , une transformation de Fourier est appliquée au spectre de puissance. Etant donné que la transformation  $P(\vec{k}) \rightarrow \xi(r, \mu)$  est faite à chaque étape de la minimisation lors de l'ajustement du modèle, il est important que cette transformation se fasse très rapidement. Nous utilisons donc l'algorithme FFTLog, qui apporte une rapidité d'exécution pour la précision souhaitée. Ainsi, en suivant cette procédure, nous obtenons les fonctions de corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  en choisissant  $i = j = \text{Ly}\alpha$ , et  $\xi_{\text{Ly}\alpha \times \text{QSO}}$  en choisissant  $i = \text{Ly}\alpha$  et  $j = \text{QSO}$ . Afin de pouvoir ajuster les fonctions de corrélation calculées avec les données, nous devons prendre en compte dans nos modèles les corrélations parasites. Ces corrélations s'ajoutent au  $\xi$  modèle calculé précédemment. A ce stade, nous distinguons le modèle utilisé pour l'auto corrélation et la corrélation croisée. Le modèle de l'auto corrélation Ly $\alpha$   $\times$  Ly $\alpha$  est défini comme

$$\xi = \xi_{\text{Ly}\alpha \times \text{Ly}\alpha} + \sum_{m,n} \tilde{\xi}_{m \times n} + \xi_{\text{ciel}} , \quad (1.24)$$

où  $\tilde{\xi}_{m \times n}$  est la corrélation de l'absorbeur  $m$  avec l'absorbeur  $n$ , interprétés comme des absorptions Ly $\alpha$ . Ces absorbeurs peuvent être du Ly $\alpha$  ou des métaux ( $m$  et  $n$  ne peuvent pas être tous les deux du Ly $\alpha$ ). Le tableau 1.1 liste les métaux ajustés dans les données.  $\xi_{ciel}$  est la corrélation induite par le masquage des lignes de ciel. Le modèle de la corrélation croisée Ly $\alpha$   $\times$  QSO est défini comme

$$\xi = \xi_{\text{Ly}\alpha \times \text{QSO}} + \sum_m \tilde{\xi}_{m \times \text{QSO}} + \xi_{prox} , \quad (1.25)$$

où  $\tilde{\xi}_{m \times \text{QSO}}$  est la corrélation du métal  $m$  avec les quasars, et  $\xi_{prox}$  donne la corrélation induite par l'effet de proximité des quasars.

Afin de modéliser la corrélation des métaux, nous utilisons le modèle défini précédemment, utilisé pour décrire la corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  et  $\xi_{\text{Ly}\alpha \times \text{QSO}}$ . Comme expliqué dans la section ??, toutes les absorptions sont supposées être des absorptions Ly $\alpha$ . Les absorptions causées par les métaux sont donc reconstruites à un mauvais redshift. Ceci résulte dans un décalage de la fonction de corrélation. De manière générale, pour la corrélation d'un absorbeur  $m$  avec un autre absorbeur  $n$  ( $m \neq n$ ), le décalage se fait le long de la ligne de visée. Lorsque la séparation physique de ces deux absorbeurs est nulle, la séparation, en supposant que ces deux absorptions sont causées par le Ly $\alpha$ , est reconstruite à  $r_{\perp} = 0$  et  $r_{\parallel} \sim (1+z)D_H(z)(\lambda_m - \lambda_n)/\lambda_{\text{Ly}\alpha}$ , où  $z$  est le redshift moyen des deux absorbeurs. La fonction de corrélation étant maximale pour  $r = 0$ , nous attendons donc un excès de corrélation pour ces séparations dans les fonctions de corrélation estimées à partir des données. Les métaux étant beaucoup moins présents dans le milieu intergalactique, les corrélations mettant en jeu deux métaux sont beaucoup moins importantes que les corrélations Ly $\alpha$   $\times$   $m$  ( $m \neq \text{Ly}\alpha$ ). L'effet principal vient donc des corrélations Ly $\alpha$   $\times$   $m$ . Le tableau 1.1 donne les séparations associés aux corrélations entre le Ly $\alpha$  et les métaux ajustés sur les données. En ce qui concerne les corrélations  $m \times m$ , le décalage est d'origine différente. Une séparation physique  $r = 0$  correspond bien à une reconstruction  $r_{\parallel} = r_{\perp} = 0$ . Cependant, le redshift de la paire est mal estimé. Chaque séparation physique  $(r_{\parallel}, r_{\perp})$  est donc reconstruite à  $(D_H(z_m)/D_H(z))r_{\parallel}$  et  $(D_M(z_m)/D_M(z))r_{\perp}$ . Pour la corrélation  $m \times \text{QSO}$ , le décalage est le même que dans le cas  $m \times n$ , en prenant  $z_n = z_{\text{QSO}}$ .

Pour chaque couple  $(m, n)$ , le décalage de la fonction de corrélation  $\xi_{m \times n}$  est pris en compte par la matrice des métaux  $M_{AB}^{m \times n}$ . Cette matrice permet donc de prendre en compte le décalage des séparations décrit précédemment. Nous ne détaillons pas son calcul ici, mais il est donné dans la thèse ?. La matrice des métaux permet donc de relier la fonction de corrélation  $\xi_{m \times n}$  à la fonction de corrélation  $\tilde{\xi}_{m \times n}$ , où les absorbeurs  $m$  et  $n$  sont interprétés comme des absorptions Ly $\alpha$ , utilisée comme modèle dans l'équation 1.24 et 1.25 :

$$\tilde{\xi}_{m \times n}(A) = \sum_B M_{AB}^{m \times n} \xi_{m \times n}(B) , \quad (1.26)$$

Les bins  $A$  correspondent aux séparations calculées en supposant une absorption Ly $\alpha$ . Les bins  $B$  correspondent aux séparations physiques, calculées en utilisant les redshifts  $z_m$  et  $z_n$  des absorbeurs. Ainsi, pour chaque couple  $(m, n)$ , la matrice des métaux  $M_{AB}^{m \times n}$  est calculée, puis la fonction de corrélation  $\tilde{\xi}_{m \times n}$  est estimée et ajoutée à la fonction de corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  (équation 1.24). Dans le cas de la corrélation croisée, la matrice est calculée pour tous les couples  $(m, \text{QSO})$ , puis la fonction de corrélation  $\tilde{\xi}_{m \times \text{QSO}}$  est estimée et ajoutée à la fonction de corrélation  $\xi_{\text{Ly}\alpha \times \text{QSO}}$  (équation 1.25).

Le modèle utilisé pour construire les fonctions de corrélation  $\xi_{m \times n}$  et  $\xi_{m \times \text{QSO}}$  et le même que celui utilisé pour construire les fonctions de corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  et  $\xi_{\text{Ly}\alpha \times \text{QSO}}$  (équation 1.15). Les paramètres  $b_i$ ,  $b_j$ ,  $\beta_i$  et  $\beta_j$  sont choisis comme étant ceux des métaux. Du fait que les métaux sont mesurables principalement le long de la ligne de visée, il est difficile d'ajuster à la fois le biais et le paramètre RSD de chaque métal. Nous ajustons donc uniquement le biais de chaque métal, le paramètre RSD étant fixé à  $\beta_m = 0,5$ . Aussi, nous négligeons l'impact des HCD sur les fonctions de corrélations  $\xi_{m \times n}$  et  $\xi_{m \times \text{QSO}}$ .



TABLE 1.1 – Liste des métaux inclus dans le modèle ajusté aux données. La 3<sup>e</sup> colonne donne la séparation reconstruite pour une séparation réelle  $r = 0$ . CIV(eff) indique la raie effective du carbon IV : la résolution du spectrographe d’eBOSS étant trop faible pour distinguer le doublet du CIV, nous ajustons la combinaison des deux raies. La séparation liée au CIV est bien supérieure à  $200 h^{-1}$  Mpc, la corrélation  $\text{Ly}\alpha \times \text{CIV}$  n’a donc pas d’effet sur nos mesures. Nous modélisons cependant l’effet lié à l’auto-corrélation  $\text{CIV} \times \text{CIV}$ .

Raie	$\lambda_m [\text{\AA}]$	$r_{\parallel} [h^{-1} \text{ Mpc}]$
SiIIa	1190,4158	−64
SiIIb	1193,2897	−56
SiIII	1206,500	−21
SiIIc	1260,4221	+111
CIV(eff)	1549,06	> 200

Le terme additionnel suivant est le terme  $\xi_{ciel}$ . Ce terme prend en compte les corrélations induites par la soustraction du fond de ciel. Lors de la réduction des données, décrite dans la section ??, le fond de ciel est soustrait à tous les spectres d’une même demi-plaque. Ceci induit alors des corrélations dans tous ces spectres pour  $r_{\parallel} = 0$ . A cause de la distorsion induite par l’ajustement du continuum, cet effet ne se limite pas à  $r_{\parallel} = 0$ . L’effet est modélisé par une fonction gaussienne de  $r_{\perp}$  :

$$\xi_{ciel}(r_{\parallel}, r_{\perp}) = \begin{cases} \frac{A_{sky}}{\sqrt{2\pi\sigma_{sky}^2}} \exp\left(-\frac{r_{\perp}^2}{2\sigma_{sky}^2}\right) & , \text{ si } r_{\parallel} = 0 \\ 0 & , \text{ si } r_{\parallel} \neq 0 \end{cases} . \quad (1.27)$$

Les paramètres  $A_{sky}$  et  $\sigma_{sky}$  sont laissés libres lors de l’ajustement des données. Ils donnent l’amplitude et la largeur de la gaussienne. Le terme  $\xi_{ciel}$  n’est présent que dans l’ajustement de l’auto corrélation, car ces corrélations parasites ne sont induites que lorsqu’on corréle des pixels d’absorption issus de deux spectres présents sur la même demi-plaque. Cet effet n’a donc pas lieu d’être pour la corrélation croisée.

Enfin, le dernier terme additionnel est le terme  $\xi_{prox}$ . Ce terme n’est présent que dans la fonction de corrélation croisée  $\xi_{\text{Ly}\alpha \times \text{QSO}}$ . Il prend en compte l’effet du rayonnement produit par le quasar sur l’hydrogène environnant. En effet, à cause de leur grande luminosité, et en particulier selon la direction de leur jet, les quasars ionise le gas qui les entoure. Ceci réduit donc la fraction d’hydrogène neutre au voisinage de chaque quasar, ce qui induit des corrélations supplémentaire entre le champ d’absorption  $\text{Ly}\alpha$  et la position des quasars. Cet effet est modélisé comme (?) :

$$\xi_{prox} = \xi_{0,prox} \left( \frac{1 h^{-1} \text{ Mpc}}{r} \right)^2 \exp\left(-\frac{r}{\lambda_{UV}}\right) , \quad (1.28)$$

où  $\xi_{0,prox}$  donne l’amplitude de l’effet. L’émission est supposée isotrope, et le paramètre  $\lambda_{UV}$  est fixé à  $300 h^{-1}$  Mpc.

#prov Matrice de distorsion

## 4.2 Modélisation des mocks

Afin d’analyser les fonctions de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  et  $\text{Ly}\alpha \times \text{QSO}$  des mocks, nous utilisons les modèles ajustés sur les données et décrits dans la section précédente. Cependant, un certain nombre d’effets modélisés dans les données ne sont pas présents dans les mocks. Nous modifions donc légèrement les modèles décrits précédemment.



# Bibliographie

---

- Hélien du Mas des Bourboux, Kyle S. Dawson, Nicolás G. Busca, Michael Blomqvist, Victoria de Sainte Agathe, Christophe Balland, Julian E. Bautista, Julien Guy, Vikrant Kamble, Adam D. Myers, Ignasi Pérez-Ràfols, Matthew M. Pieri, James Rich, Donald P. Schneider, and Anže Slosar. The extended Baryon Oscillation Spectroscopic Survey : measuring the cross-correlation between the MgII flux transmission field and quasars and galaxies at  $z=0.59$ . jan 2019.
- Solène Chabanier, Nathalie Palanque-Delabrouille, Christophe Yèche, Jean-Marc Le Goff, Eric Armengaud, Julian Bautista, Michael Blomqvist, Nicolas Busca, Kyle Dawson, Thomas Etourneau, Andreu Font-Ribera, Youngbae Lee, Hélien du Mas des Bourboux, Matthew Pieri, James Rich, Graziano Rossi, Donald Schneider, and Anže Slosar. The one-dimensional power spectrum from the SDSS DR14 Ly-alpha forests. dec 2018. doi : 10.1088/1475-7516/2019/07/017. URL <http://arxiv.org/abs/1812.03554><http://dx.doi.org/10.1088/1475-7516/2019/07/017>.
- David Kirkby, Daniel Margala, Anže Slosar, Stephen Bailey, Nicolás G. Busca, Timothée Delubac, James Rich, Michael Blomqvist, Joel R. Brownstein, Bill Carithers, Rupert A. C. Croft, Kyle S. Dawson, Andreu Font-Ribera, Jordi Miralda-Escudé, Adam D. Myers, Robert C. Nichol, Nathalie Palanque-Delabrouille, Isabelle Pâris, Patrick Petitjean, Graziano Rossi, David J. Schlegel, Donald P. Schneider, Matteo Viel, David H. Weinberg, and Christophe Yèche. Fitting Methods for Baryon Acoustic Oscillations in the Lyman- $\alpha$  Forest Fluctuations in BOSS Data Release 9. jan 2013. doi : 10.1088/1475-7516/2013/03/024. URL <https://arxiv.org/abs/1301.3456>.
- D J Eisenstein, H.-J. Seo, and M White. On the {Robustness} of the {Acoustic} {Scale} in the {Low}-{Redshift} {Clustering} of {Matter}. *\textbackslash textbackslash apj*, 664 :660–674, 2007. doi : 10.1086/518755.
- Andreu Arinyo-i Prats, Jordi Miralda-Escudé, Matteo Viel, and Renyue Cen. The Non-Linear Power Spectrum of the Lyman Alpha Forest. jun 2015. doi : 10.1088/1475-7516/2015/12/017. URL <http://arxiv.org/abs/1506.04519><http://dx.doi.org/10.1088/1475-7516/2015/12/017>.
- Andreu Font-Ribera and Jordi Miralda-Escudé. The {Effect} of {High} {Column} {Density} {Systems} on the {Measurement} of the {Lyman}  $\alpha$  {Forest} {Correlation} {Function}. *Journal of Cosmology and Astroparticle Physics*, 2012(07) :28, jul 2012. ISSN 1475-7516. doi : 10.1088/1475-7516/2012/07/028. URL <http://arxiv.org/abs/1205.2018>.
- Keir K. Rogers, Simeon Bird, Hiranya V. Peiris, Andrew Pontzen, Andreu Font-Ribera, and Boris Leistedt. Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers. nov 2017. URL <http://arxiv.org/abs/1711.06275>.
- P McDonald, U Seljak, and et al. Burles. The  $\text{Ly}\alpha$  {Forest} {Power} {Spectrum} from the {Sloan} {Digital} {Sky} {Survey}. *\textbackslash textbackslash apjs*, 163 :80–109, mar 2006. doi : 10.1086/444361.
- A. J. S. Hamilton. Uncorrelated Modes of the Nonlinear Power Spectrum. may 1999. doi : 10.1046/j.1365-8711.2000.03071.x. URL <http://arxiv.org/abs/astro-ph/9905191><http://dx.doi.org/10.1046/j.1365-8711.2000.03071.x>.