

# 1

## Présentation des données

---

Dans le chapitre précédent, nous présentions l'instrument SDSS et sa prise de données, qui constitue le relevé eBOSS. Nous présentons ici, comment des photons acquis avec les CCD, les spectres sont déduits, puis comment du spectre de chaque quasar, le champ d'absorption est reconstruit.

phrase à l'envers, lourd

## 1 Réduction des données

image sur le CCD, mettre une image ?

L'observation de chaque plaque fournit des données en deux dimensions : la première correspond aux différentes longueurs d'onde, la seconde aux différentes fibres. La chaîne de réduction des données transforme ces informations en une liste de spectres. Premièrement, chaque spectre est extrait de l'image acquise par le CCD. Le spectre est étalonné en longueur d'onde à l'aide de lampes à arc. Le flux est calibré en utilisant les spectres des fibres dédiées aux étoiles standards. Puis, le fond du ciel est estimé dans chaque demi plaque grâce aux fibres dédiées, et soustrait à chaque spectre. La variance dans chaque pixel est ensuite estimée. Elle prend en compte le bruit de photon, le bruit de l'électronique, et aussi les éventuels défauts que présentent les CCD. Les pixels affectés par des rayons cosmiques sont rejetés. Enfin, pour chaque objet, toutes les expositions sont ajoutées pour former un seul et même spectre, avec un plus grand rapport signal sur bruit. Le nombre typique d'expositions par objet varie entre 4 et 6.

ont été

Les données que nous utilisons dans ce manuscrit sont rendus publiques lors de la sixième publication de données SDSS (DR16<sup>1</sup>). Elles sont décrites dans Ahumada et al. (2019).

par

## 2 Le catalogue de quasar

ne pas inverser

Une fois les spectres extraits, il est important de les classifier afin de construire des catalogues utilisables par les différentes analyses. Pour ce faire, chaque spectre est traité par la chaîne d'identification de SDSS, décrite dans Bolton et al. (2012). A chaque spectre est ajusté plusieurs modèles d'étoiles, de galaxies et de quasars. Les modèles sont construits en utilisant les PCA templates : des spectres typiques qui sont combinés afin d'ajuster au mieux le spectre à classifier. Une fois tous les modèles ajustés, ils sont triés selon leur  $\chi^2$  réduit. Le spectre est alors classifié selon le modèle possédant le plus faible  $\chi^2$  réduit. Tous les spectres mesurés par SDSS et classifiés comme quasar constituent l'échantillon superset. Il contient 1 440 627 objets.

inversion

certains

Les quasars, à cause des absorptions intenses comme les DLA ou les BAL<sup>2</sup>, peuvent être difficiles à classifier. Lors de BOSS, tous les quasars ont été inspectés visuellement (Pâris et al., 2016) afin de confirmer la classification, et d'estimer leur redshifts. Cet échantillon représente 297 301 objets. Cependant, à cause du grand nombre de quasars observés par eBOSS (voir figure 1.1), l'observation visuelle n'a pas pu être effectuée pour tous ces objets. Une seconde classification, fondée sur la chaîne pipeline d'identification de SDSS, est alors appliquée (?). Elle est complétée par l'algorithme QuasarNET (Busca and Balland, 2018) afin de réduire le nombre de mauvaises et non-identifications, et donc le nombre d'inspections visuelles nécessaires. Grâce à QuasarNET, ce chiffre est réduit à 0,6 % des spectres, soit 8581 spectres. Une sous-échantillon du superset est alors construit. Il contient 750 426 objets, confirmés comme quasar par la chaîne de traitement précédente.

Les spectres BAL sont exclus

Pour chaque objet, le catalogue fournit plusieurs estimations de redshift. La chaîne de classification SDSS produit une première estimation. L'algorithme QuasarNET en fournit une seconde. Les spectres inspectés visuellement possèdent une autre estimation. Enfin, l'algorithme redvsblue<sup>3</sup> produit plusieurs estimations de redshift, parmi lesquelles figure Z\_PCA et Z\_LYAWG.

figurent tu vas en parler après ?

1. <https://www.sdss.org/dr16/>

une figure<sup>2</sup> 2. Les BAL (Broad Absorption Line) sont des quasars qui présentent des absorptions intenses au voisinage de leurs raies d'émission. Cette absorption est interprétée comme étant due à un absorbeur dense situé juste devant le quasar.

3. <https://github.com/londumas/redvsblue>

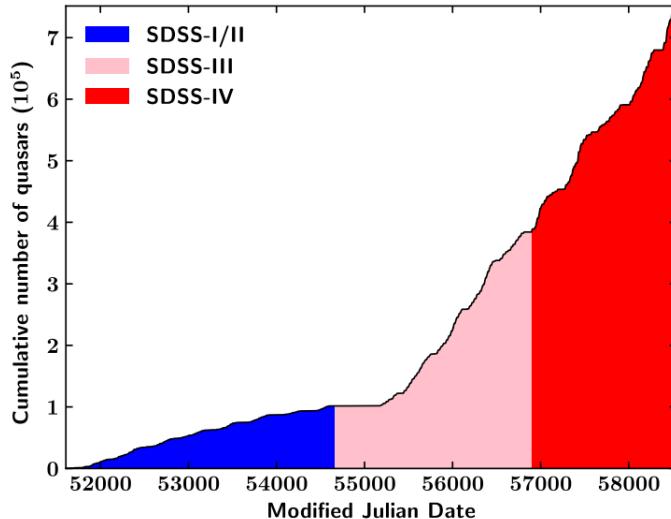


FIGURE 1.1 – Evolution du nombre de quasars observés par les différentes générations de SDSS en fonction du temps. Les générations SDSS I et II correspondent à  $MJD < 54663$  (bleu), SDSS III à  $54663 \leq MJD < 56898$  (rose) et SDSS IV à  $56898 < MJD < 58543$  (rouge). Crédits : #prov Lyke in prep

Différents algorithmes sont alors appliqués au catalogue, afin d’identifier les DLA et les BAL présents. L’identification, qui prend place sur les spectres, utilise les différentes expositions additionnées pour chaque objet. Concernant les DLA, l’algorithme de détection est décrit dans Parks et al. (2017). Il est appliqué sur les quasars à un redshift  $2 \leq Z_{PCA} \leq 6$ , afin d’avoir suffisamment de pixel dans la zone  $900 < \lambda_{RF} < 1346 \text{ \AA}$ . Parmi les 270 315 spectres inspectés, 39 514 DLA ont été identifiés, distribués dans 35 686 spectres. Concernant les BAL, l’algorithme utilisé est très similaire à celui décrit dans Guo and Martini (2019). Les BAL sont recherchés dans les spectres ayant un redshift entre 1,57 et 5,6. L’algorithme fournit la probabilité qu’un spectre possède un BAL. Le champ BAL\_PROB du catalogue indique cette probabilité.

A la fin, le catalogue ainsi construit (DR16Q dans la suite de ce manuscrit) contient 750 426 quasars confirmés. Nous référons le lecteur à l’article ? pour davantage d’informations.

L’analyse Ly $\alpha$  finale d’eBOSS (?) utilise le catalogue DR16Q. Le redshift des quasars est choisi comme étant  $Z_{LYAWG}$ . Les quasars sont sélectionnés avec un redshift  $1.77 < z \leq 4$ . L’échantillon correspondant représente alors 341 468 objets. La distribution en redshift de ces quasars traceurs est présentée dans le graphique de gauche de la figure 1.2.

expliqué page suivante

### 3 La sélection des forêts

#### analyse du set de données complet

L’analyse Ly $\alpha$  finale d’eBOSS, dont nous nous servons dans ce manuscrit, est présentée dans l’article ?. Les informations que nous donnons dans la suite de ce chapitre sont tirés de cette article. Nous y référons le lecteur pour davantage d’informations.

Les spectres produits par la chaîne de réduction des données SDSS sont rebinnés : 3 pixels du spectre original, d’une taille  $\Delta \log_{10}(\lambda) \sim 10^{-4}$ , sont combinés en 1 seul pixel d’analyse, d’une taille  $\Delta \log_{10}(\lambda) \sim 3 \times 10^{-4}$ . Ceci est fait afin d’augmenter le rapport signal sur bruit des absorptions Ly $\alpha$ . Dans la suite, l’utilisation de “pixel” réfère à ces pixels d’analyse.

non, c’est pour gagner en CPU sur picca

L’absorption Ly $\alpha$  est mesurée dans deux régions distinctes du spectre. La première, dénommée région Ly $\alpha$ , correspond aux longueurs d’onde  $1040 \leq \lambda_{RF} \leq 1200 \text{ \AA}$ , c’est à dire entre les raies d’émission

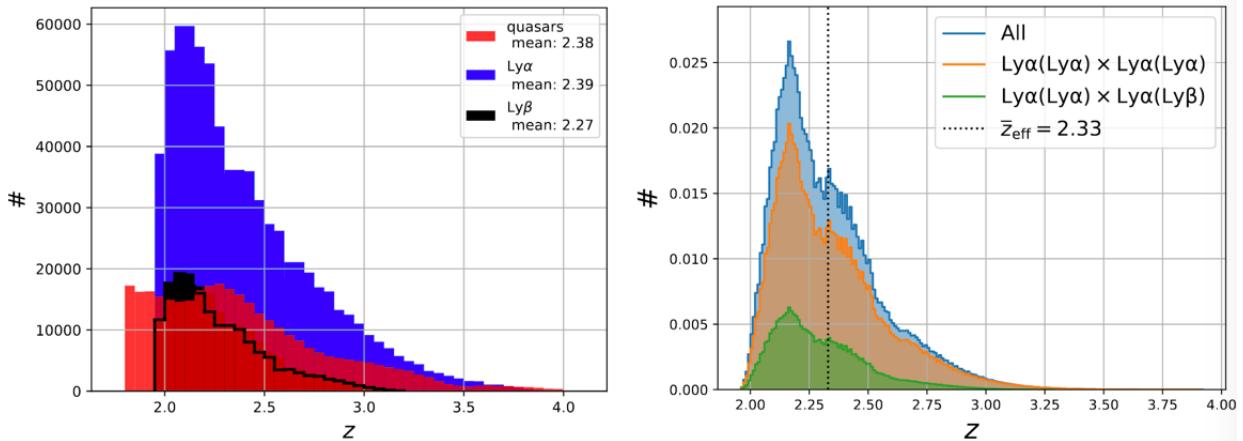


FIGURE 1.2 – Gauche : distribution en redshift des pixels Ly $\alpha$ (Ly $\alpha$ ) (bleu) et des pixels Ly $\alpha$ (Ly $\beta$ ) (noir) (nombres divisés par 50), ainsi que celle des quasars traceurs (rouge). Droite : distribution pondérée du redshift efficace de chaque paire de pixels utilisée pour le calcul de la fonction corrélation du Ly $\alpha$ . La distribution orange présente les paires de pixels Ly $\alpha$ (Ly $\alpha$ )-Ly $\alpha$ (Ly $\alpha$ ), la verte les paires de pixels Ly $\alpha$ (Ly $\alpha$ )-Ly $\alpha$ (Ly $\beta$ ), et la bleue la somme des deux. Les paires représentées sont celles nécessaires pour la mesure de la position du pic BAO, c'est à dire avec une séparation  $r \in [80; 120] h^{-1}$  Mpc. La ligne en pointillés indique le redshift effectif de la mesure, c'est à dire la moyenne de la distribution bleue.

Ly $\beta$  et Ly $\alpha$ . La seconde, dénommée région Ly $\beta$ , correspond aux longueurs d'onde  $920 \leq \lambda_{RF} \leq 1020 \text{ \AA}$ , c'est à dire entre la limite de la série de Lyman et la raie d'émission Ly $\beta$ . Les pixels d'absorption Ly $\alpha$  dans la région Ly $\alpha$  sont dénommés pixels Ly $\alpha$ (Ly $\alpha$ ), et ceux dans la région Ly $\beta$  sont dénommés pixels Ly $\alpha$ (Ly $\beta$ ). De plus, l'analyse se limite aux pixels dont la longueur d'onde observée est comprise entre  $3000 \leq \lambda_{obs} \leq 6000 \text{ \AA}$ , à cause notamment des absorptions atmosphériques intenses dans l'UV, et des raies d'émissions du ciel dans le proche infrarouge. Ces limites sur  $\lambda_{RF}$  correspondent à un redshift minimal  $z_{QSO} = 2$  pour les quasars Ly $\alpha$ , et  $z_{QSO} = 2.53$  pour les quasars Ly $\beta$ . Parmi les 341 468 quasars utilisés comme traceur (dénommés quasars traceurs), 256 328 sont utilisés pour leurs pixels Ly $\alpha$ (Ly $\alpha$ ) (dénommés quasar Ly $\alpha$ ), et 103 080 sont utilisés pour leurs pixels Ly $\alpha$ (Ly $\beta$ ) (dénommés quasar Ly $\beta$ ). jargon eBOSS ? mettre la définition au début si c'est nécessaire

sensibilité  
des CCD

D'autres sélections sont aussi appliquées. Les quasars pour lesquels la probabilité d'avoir un BAL est supérieure à 0.9 sont écartés. Les mauvaises observations durant BOSS ou eBOSS sont mises de côté. De plus, chaque région nécessite au moins 50 pixels. Cette sélection décale le redshift minimal à  $z_{QSO} = 2.10$  pour les quasars Ly $\alpha$ , et à  $z = 2.65$  pour les quasars Ly $\beta$ . Enfin, l'ajustement du continuum (voir section suivante) échoue pour certains spectres. Ces spectres sont aussi écartés. Après toutes ces sélections, l'échantillon final contient 210 005 quasars Ly $\alpha$  et 69 656 quasars Ly $\beta$ .

rare ?

Enfin, des corrections sur le flux et la variance du flux sont appliquées afin de corriger les imperfections de la chaîne de réduction. De plus, certaines régions spectrales en longueur d'onde observée sont masquées, à cause de l'augmentation de la variance du flux causée par raies du ciel. Tout ceci est décrit en détail dans ?.

## 4 Définition du champ d'absorption

Dans cette section, nous décrivons comment, du flux mesuré  $f_q(\lambda)$  du quasar  $q$  à la longueur d'onde observée  $\lambda$ , nous pouvons déduire les fluctuations du champ d'absorption Ly $\alpha$ . La démarche que nous décrivons ici est très bien expliquée dans la thèse ?. Nous y référons le lecteur pour plus de détails.

## 4.1 Calcul des $\delta$

Comme présenté dans la section ??, le contraste de l'absorption Ly $\alpha$   $\delta_q(\lambda)$  est défini comme

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{\bar{F}(\lambda)C_q(\lambda)} - 1, \quad (1.1)$$

où  $\bar{F}(\lambda)$  est la transmission moyenne au redshift  $z = \lambda/\lambda_{\text{Ly}\alpha} - 1$ , et  $C_q(\lambda)$  donne le flux du spectre sans absorption. Il est dénommé *continuum* et est différent pour chaque quasar. Le produit  $\bar{F}(\lambda)C_q(\lambda)$  représente donc le flux moyen attendu du quasar  $q$ . Du fait de la faible résolution et du S/N des données,  $\bar{F}(\lambda)$

pas clair, référence aux fig n'est pas mesurable. Pour ne pas dépendre d'analyses externes, on détermine directement le produit  $\bar{F}(\lambda)C_q(\lambda)$  pour chaque spectre. Afin de prendre en compte la variabilité d'un spectre à un autre, le produit  $\bar{F}(\lambda)C_q(\lambda)$  est modélisé par une relation linéaire en  $\log_{10}\lambda$  :

$$\bar{F}(\lambda)C_q(\lambda) = \bar{C}(\lambda_{\text{RF}})(a_q + b_q \log_{10}(\lambda)), \quad (1.2)$$

où  $\bar{C}(\lambda_{\text{RF}})$  donne le continuum moyen sur tous les spectres, en fonction de la longueur d'onde dans le référentiel du quasar.  $a_q$  et  $b_q$  sont propres à chaque quasar, ils modélisent la diversité des quasars. Les termes  $a_q$ ,  $b_q$  et  $\bar{C}(\lambda_{\text{RF}})$  sont déterminés en minimisant la fonction de vraisemblance  $L$ , qui s'exprime comme

$$-2 \ln L = -\sum_i \frac{[f_i - \bar{F}C_q(\lambda_i, a_q, b_q)]^2}{\sigma_q^2(\lambda_i)} + \ln[\sigma_q^2(\lambda_i)]. \quad (1.3)$$

commenter ou retirer

La variance  $\sigma_q^2(\lambda)$  dans chaque pixel est donnée par

$$\sigma_q^2(\lambda) = \eta(\lambda)\sigma_{\text{instru},q}^2(\lambda) + \sigma_{\text{cosmo}}^2(\lambda)(\bar{F}C_q(\lambda))^2 + \frac{\epsilon(\lambda)(\bar{F}C(\lambda))^4}{\sigma_{\text{instru},q}^2(\lambda)}. \quad (1.4)$$

a voir

Le terme  $\eta(\lambda)\sigma_{\text{instru},q}^2(\lambda)$  rend compte de la variance provoquée par l'instrument, et liée à l'incertitude sur la mesure des flux. Le terme  $\sigma_{\text{cosmo}}(\lambda)$  est purement cosmologique et donne la variance propre au champ d'absorption du Ly $\alpha$ . Il traduit le fait que certaines régions de l'univers peuvent absorber plus ou moins que la moyenne. Du fait de l'isotropie,  $\sigma_{\text{cosmo}}$  ne dépend que de  $z$ . Enfin, le terme  $\epsilon(\lambda)(\bar{F}C(\lambda))^4/\sigma_{\text{instru},q}^2(\lambda)$  rend compte du fait que la variance augmente avec le rapport signal sur bruit, probablement à cause de la diversité des quasars et de l'erreur sur la mesure de leur redshift.

ad hoc

Afin d'obtenir le produit  $\bar{F}C_q$  pour chaque quasar, il faut minimiser la fonction de vraisemblance  $L$  (équation 1.3), et donc ajuster les paramètres  $a_q$ ,  $b_q$ ,  $\eta(\lambda)$ ,  $\sigma_{\text{cosmo}}(\lambda)$  et  $\epsilon(\lambda)$ . Cependant, celle fonction dépend elle-même du produit  $\bar{F}C_q$ . Pour résoudre ce problème, le calcul l'ajustement est fait de manière itérative. Premièrement, les 5 paramètres à ajuster, ainsi que  $\bar{C}(\lambda_{\text{RF}})$ , sont initialisés. Ceci permet d'ajuster  $\bar{F}C_q$  sur  $f_q(\lambda)$  en minimisant la fonction de vraisemblance  $L$ . Dans cet ajustement, chaque pixel observé à une longueur d'onde  $\lambda$  est pondéré par un poid

$$w_q(\lambda) = \frac{1}{\sigma_q^2(\lambda)}. \quad (1.5)$$

Une fois cette ajustement fait,  $\bar{C}(\lambda_{\text{RF}})$  est de nouveau calculé. Puis, les paramètres  $\eta(\lambda)$ ,  $\sigma_{\text{cosmo}}(\lambda)$  et  $\epsilon(\lambda)$  sont ajustés de manière à ce que, à lambda fixé,  $\sigma_q(\lambda)$  reproduise la variance estimée à partir des flux  $f_q(\lambda)$  de tous les quasars. Ce nouvel ajustement permet une nouvelle fois d'ajuster  $\bar{F}C_q$ , etc. Cette procédure est répétée jusqu'à obtenir des valeurs stables pour les paramètres ajustés. En pratique, 5 itérations sont nécessaires. Cette procédure permet donc d'estimer, pour chaque quasar, le produit  $\bar{F}(\lambda)C_q(\lambda)$ , et ainsi d'estimer les  $\delta_q(\lambda)$  (équation 1.1) indépendamment dans chaque région.

## 4.2 Prise en compte de l'ajustement du continuum

La procédure décrite dans la section précédente permet d'ajuster le continuum de chaque quasar afin d'estimer le champ  $\delta$  dans chaque région d'absorption Ly $\alpha$ . Cependant, cet ajustement biaise le champ  $\delta$  mesuré, et introduit de fausses corrélations lors du calcul des fonctions de corrélation. Il est donc important de prendre en compte cet effet.

En redéfinissant les paramètres  $a_q$  et  $b_q$ , l'équation 1.2 peut être mise sous la forme

$$\bar{F}(\lambda)C_q(\lambda) = \bar{C}(\lambda_{\text{RF}})(a_q + b_q(\Lambda - \bar{\Lambda}_q)) , \quad (1.6)$$

où  $\Lambda = \log_{10} \lambda$ , et  $\bar{\Lambda}_q$  est la moyenne de  $\Lambda$  dans chaque forêt :

$$\bar{\Lambda}_q = \frac{\sum_{\lambda} w_q(\lambda)\Lambda}{\sum_{\lambda} w_q(\lambda)} . \quad (1.7)$$

Le champ  $\delta_q$  mesuré est donc donné par

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{\bar{C}(\lambda_{\text{RF}})(a_q + b_q(\Lambda - \bar{\Lambda}_q))} - 1 . \quad (1.8)$$

Du fait que  $a_q$  et  $b_q$  sont ajustés sur les  $f_q(\lambda)$ , l'utilisation de l'équation 1.6 pour déterminer le continuum force la moyenne et la pente de chaque région spectrale à être nulle. Ainsi le champ  $\delta$  mesuré est biaisé, et de fausses corrélations sont induites dans le calcul de la fonction de corrélation. Afin de prendre en compte cet effet, le champ  $\delta$  est transformé selon la relation

$$\tilde{\delta}_q(\lambda_i) = \sum_j \eta_{ij}^q \delta_q(\lambda_j) , \quad (1.9)$$

avec

$$\eta_{ij}^q = \delta_{ij}^K - \frac{w_j}{\sum_k w_k} - \frac{w_j (\Lambda_i - \bar{\Lambda}_q) (\Lambda_j - \bar{\Lambda}_q)}{\sum_k w_k (\Lambda_k - \bar{\Lambda}_q)^2} , \quad (1.10)$$

où  $\delta_{ij}^K$  est le symbole de Kronecker. Dans ce calcul, les poids  $w_q(\lambda)$  sont corrigés de la dépendance en redshift du biais du Ly $\alpha$  :

$$w_i = \sigma_q^{-2}(\lambda_i) \left( \frac{1+z_i}{1+2.25} \right)^{\gamma_{\text{Ly}\alpha}-1} , \quad (1.11)$$

avec  $\gamma_{\text{Ly}\alpha} = 2.9$  (McDonald et al., 2006). Grâce à cette transformation, la distorsion due à l'ajustement du continuum est exacte et connue. L'effet sur la fonction de corrélation peut ainsi être modélisé et pris en compte. Sa modélisation passe par le calcul de la *matrice de distorsion*, détaillée dans la section ??.

Enfin, l'utilisation de  $\bar{C}(\lambda_{\text{RF}})$  dans l'équation 1.2 force la moyenne des  $\delta_q$  dans chaque bin de longueur d'onde observée à nulle. Cependant, la transformation 1.9 modifie légèrement la moyenne des  $\delta_q$  dans chaque bin. Afin que la fonction de corrélation croisée Ly $\alpha$ -QSO tende vers zéro à grand séparation, l'effet est corrigé en soustrayant explicitement la moyenne dans chaque bin :

$$\hat{\delta}_q(\lambda_i) = \tilde{\delta}_q(\lambda_i) - \bar{\delta}(\lambda) . \quad (1.12)$$

Dans la suite de ce manuscript, lorsque nous parlons du champ  $\delta$ , nous référons à ces  $\hat{\delta}_q(\lambda_i)$ . Le graphique de gauche de la figure 1.2 montre la distribution en redshift des  $\delta$  dans chaque région d'absorption. Le graphique de droite présente la distribution pondérée par les poids  $w_i$  (équation 1.11) du redshift effectif de chaque paire, utilisé dans le calcul de la fonction de corrélation.

# Bibliographie

---

Romina Ahumada, Carlos Allende Prieto, Andres Almeida, Friedrich Anders, Scott F. Anderson, Brett H. Andrews, Borja Anguiano, Riccardo Arcodia, Eric Armengaud, Marie Aubert, Santiago Avila, Vladimir Avila-Reese, Carles Badenes, Christophe Balland, Kat Barger, Jorge K. Barrera-Ballesteros, Sarbani Basu, Julian Bautista, Rachael L. Beaton, Timothy C. Beers, B. Izamar T. Benavides, Chad F. Bender, Mariangela Bernardi, Matthew Bershad, Florian Beutler, Christian Moni Bidin, Jonathan Bird, Dmitry Bizyaev, Guillermo A. Blanc, Michael R. Blanton, Mederic Boquien, Jura Borissova, Jo Bovy, W. N. Brandt, Jonathan Brinkmann, Joel R. Brownstein, Kevin Bundy, Martin Bureau, Adam Burgasser, Etienne Burtin, Mariana Cano-Diaz, Raffaella Capasso, Michele Cappellari, Ricardo Carrera, Solene Chabanier, William Chaplin, Michael Chapman, Brian Cherinka, Cristina Chiappini, Peter Doohyun Choi, S. Drew Chojnowski, Haeun Chung, Nicolas Clerc, Damien Coffey, Julia M. Comerford, Johan Comparat, Luiz da Costa, Marie-Claude Cousinou, Kevin Covey, Jeffrey D. Crane, Katia Cunha, Gabriele da Silva Ilha, Yu Sophia Dai, Sanna B. Damsted, Jeremy Darling, Danny Horta Darrington, James W. Davidson, Roger Davies, Kyle Dawson, Nikhil De, Axel de la Macorra, Nathan De Lee, Anna Barbara de Andrade Queiroz, Alice Deconto Machado, Sylvain de la Torre, Flavia Dell'Agli, Helion du Mas des Bourboux, Aleksandar M. Diamond-Stanic, Sean Dillon, John Donor, Niv Drory, Chris Duckworth, Tom Dwelly, Garrett Ebelke, Sarah Eftekharzadeh, Arthur Davis Eigenbrot, Yvonne P. Elsworth, Mike Eracleous, Ghazaleh Erfanianfar, Stephanie Escoffier, Xiaohui Fan, Emily Farr, Jose G. Fernandez-Trincado, Diane Feuillet, Alexis Finoguenov, Patricia Fofie, Amelia Fraser-McKelvie, Peter M. Frinchaboy, Sebastien Fromenteau, Hai Fu, Lluis Galbany, Rafael A. Garcia, D. A. Garcia-Hernandez, Luis Alberto Garma Oehmichen, Junqiang Ge, Marcio Antonio Geimba Maia, Doug Geisler, Joseph Gelfand, Julian Goddy, Jean-Marc Le Goff, Violeta Gonzalez-Perez, Kathleen Grabowski, Paul Green, Catherine J. Grier, Hong Guo, Julien Guy, Paul Harding, Sten Hasselquist, Adam James Hawken, Christian R. Hayes, Fred Hearty, S. Hekker, David W. Hogg, Jon Holtzman, Jiamin Hou, Bau-Ching Hsieh, Daniel Huber, Jason A. S. Hunt, J. Ider Chitham, Julie Imig, Mariana Jaber, Camilo Eduardo Jimenez Angel, Jennifer A. Johnson, Amy M. Jones, Henrik Jonsson, Eric Jullo, Yerim Kim, Karen Kinemuchi, Charles C. Kirkpatrick, George W. Kite, Mark Klaene, Jean-Paul Kneib, Juna A. Kollmeier, Hui Kong, Marina Kounkel, Dhanesh Krishnarao, Ivan Lacerna, Ting-Wen Lan, Richard R. Lane, David R. Law, Henry W. Leung, Hannah Lewis, Cheng Li, Jianhui Lian, Lihwai Lin, Dan Long, Penelope Longa-Pena, Britt Lundgren, Brad W. Lyke, J. Ted Mackereth, Chelsea L. MacLeod, Steven R. Majewski, Arturo Manchado, Claudia Maraston, Paul Martini, Thomas Masseron, Karen L. Masters, Savita Mathur, Richard M. McDermid, Andrea Merloni, Michael Merrifield, Szabolcs Meszaros, Andrea Miglio, Dante Minniti, Rebecca Minsley, Takamitsu Miyaji, Faizan Gohar Mohammad, Benoit Mosser, Eva-Maria Mueller, Demitri Muna, Andrea Munoz-Gutierrez, Adam D. Myers, Seshadri Nadathur, Preethi Nair, Janaina Correa do Nascimento, Rebecca Jean Nevin, Jeffrey A. Newman, David L. Nidever, Christian Nitschelm, Pasquier Noterdaeme, Julia E. O'Connell, Matthew D Olmstead, Daniel Oravetz, Audrey Oravetz, Yeisson Osorio, Zachary J. Pace, Nelson Padilla, Nathalie Palanque-Delabrouille, Pedro A. Palicio, Hsi-An Pan, Kaike Pan, James Parker, Romain Paviot, Sebastien Peirani, Karla Pena Ramrez, Samantha Penny, Will J. Percival, Ismael Perez-Fournon,

Ignasi Perez-Rafols, Patrick Petitjean, Matthew M. Pieri, Marc Pinsonneault, Vijith Jacob Poovelil, Joshua Tyler Povick, Abhishek Prakash, Adrian M. Price-Whelan, M. Jordan Raddick, Anand Raichoor, Amy Ray, Sandro Barboza Rembold, Mehdi Rezaie, Rogemar A. Riffel, Rogerio Riffel, Hans-Walter Rix, Annie C. Robin, A. Roman-Lopes, Carlos Roman-Zuniga, Benjamin Rose, Ashley J. Ross, Graziano Rossi, Kate Rowlands, Kate H. R. Rubin, Mara Salvato, Ariel G. Sanchez, Laura Sanchez-Menguiano, Jose R. Sanchez-Gallego, Conor Sayres, Adam Schaefer, Ricardo P. Schiavon, Jaderson S. Schimoia, Edward Schlafly, David Schlegel, Donald P. Schneider, Mathias Schultheis, Axel Schwope, Hee-Jong Seo, Aldo Serenelli, Arman Shafieloo, Shoaib Jamal Shamsi, Zhengyi Shao, Shiyin Shen, Matthew Shetrone, Raphael Shirley, Victor Silva Aguirre, Joshua D. Simon, M. F. Skrutskie, Anze Slosar, Rebecca Smethurst, Jennifer Sobeck, Bernardo Cervantes Sodi, Diogo Souto, David V. Stark, Keivan G. Stassun, Matthias Steinmetz, Dennis Stello, Julianna Stermer, Thaisa Storchi-Bergmann, Alina Streblyanska, Guy S. Stringfellow, Amelia Stutz, Genaro Suarez, Jing Sun, Manuchehr Taghizadeh-Popp, Michael S. Talbot, Jamie Tayar, Aniruddha R. Thakar, Riley Theriault, Daniel Thomas, Zak C. Thomas, Jeremy Tinker, Rita Tojeiro, Hector Hernandez Toledo, Christy A. Tremonti, Nicholas W. Troup, Sarah Tuttle, Eduardo Unda-Sanzana, Marica Valentini, Jaime Vargas-Gonzalez, Mariana Vargas-Magana, Jose Antonio Vazquez-Mata, M. Vivek, David Wake, Yuting Wang, Benjamin Alan Weaver, Anne-Marie Weijmans, Vivienne Wild, John C. Wilson, Robert F. Wilson, Nathan Wolthuis, W. M. Wood-Vasey, Renbin Yan, Meng Yang, Christophe Yéche, Olga Zamora, Pauline Zarrouk, Gail Zasowski, Kai Zhang, Cheng Zhao, Gongbo Zhao, Zheng Zheng, Zheng Zheng, Guangtun Zhu, and Hu Zou. The Sixteenth Data Release of the Sloan Digital Sky Surveys : First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra. dec 2019. URL <http://arxiv.org/abs/1912.02905>.

Adam S. Bolton, David J. Schlegel, Eric Aubourg, Stephen Bailey, Vaishali Bhardwaj, Joel R. Brownstein, Scott Burles, Yan-Mei Chen, Kyle Dawson, Daniel J. Eisenstein, James E. Gunn, G. R. Knapp, Craig P. Loomis, Robert H. Lupton, Claudia Maraston, Demitri Muna, Adam D. Myers, Matthew D. Olmstead, Nikhil Padmanabhan, Isabelle Paris, Will J. Percival, Patrick Petitjean, Constance M. Rockosi, Nicholas P. Ross, Donald P. Schneider, Yiping Shu, Michael A. Strauss, Daniel Thomas, Christy A. Tremonti, David A. Wake, Benjamin A. Weaver, and W. Michael Wood-Vasey. Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey. jul 2012. doi : 10.1088/0004-6256/144/5/144. URL <http://arxiv.org/abs/1207.7326><http://dx.doi.org/10.1088/0004-6256/144/5/144>.

Isabelle Pâris, Patrick Petitjean, Nicholas P. Ross, Adam D. Myers, Éric Aubourg, Alina Streblyanska, Stephen Bailey, Éric Armengaud, Nathalie Palanque-Delabrouille, Christophe Yèche, Fred Hamann, Michael A. Strauss, Franco D. Albareti, Jo Bovy, Dmitry Bizyaev, W. Niel Brandt, Marcella Brusa, Johannes Buchner, Johan Comparat, Rupert A. C. Croft, Tom Dwelly, Xiaohui Fan, Andreu Font-Ribera, Jian Ge, Antonis Georgakakis, Patrick B. Hall, Linhua Jian, Karen Kinemuchi, Elena Malanushenko, Viktor Malanushenko, Richard G. McMahon, Marie-Luise Menzel, Andrea Merloni, Kirpal Nandra, Pasquier Noterdaeme, Daniel Oravetz, Kaise Pan, Matthew M. Pieri, Francisco Prada, Mara Salvato, David J. Schlegel, Donald P. Schneider, Audrey Simmons, Matteo Viel, David H. Weinberg, and Liu Zhu. The Sloan Digital Sky Survey Quasar Catalog : twelfth data release. aug 2016. doi : 10.1051/0004-6361/201527999. URL <https://arxiv.org/abs/1608.06483>.

Nicolas Busca and Christophe Balland. QuasarNET : Human-level spectral classification and redshifting with Deep Neural Networks. aug 2018. URL <http://arxiv.org/abs/1808.09955>.

David Parks, J. Xavier Prochaska, Shawfeng Dong, and Zheng Cai. Deep Learning of Quasar Spectra to Discover and Characterize Damped Ly $\alpha$  Systems. sep 2017. doi : 10.1093/mnras/sty196. URL <http://arxiv.org/abs/1709.04962><http://dx.doi.org/10.1093/mnras/sty196>.

Zhiyuan Guo and Paul Martini. Classification of Broad Absorption Line Quasars with a Convolutional Neural Network. jan 2019. doi : 10.3847/1538-4357/ab2590. URL <http://arxiv.org/abs/1901.04506> <http://dx.doi.org/10.3847/1538-4357/ab2590>.

P McDonald, U Seljak, and et al. Burles. The {Ly $\alpha$ } {Forest} {Power} {Spectrum} from the {Sloan} {Digital} {Sky} {Survey}. *\textbackslash apjs*, 163 :80–109, mar 2006. doi : 10.1086/444361.