1 Introduction

2 Earlier mocks

For Ly-α analysis we need mock data sets that cover tenths of Gpc^3, while the absorption in a quasar spectrum is probing the intergalactic medium at the Jean's scale, i.e. around $100h^{-1}$ kpc. It is not possible to perform N-body simulations with such volume and resolution. So we are limited to generating Gaussian random fields with a correlation close to that of the data, in order to test the analysis pipeline and study possible astrophysical effects.

The BOSS collaboration developed Ly-α mocks using a Choleski decomposition of the correlation matrix between pixel pairs and generating a correlated field only along the quasar lines-of-sight [1]. However, this approach results in no cross correlation between the transmission in the Ly-α pixels and the quasar positions, while it was latter realized that this cross correlation provides similar statistics for BAO than the Ly-α auto-correlation [2]. Dedicated mocks were produced for the final eBOSS analysis of the cross-correlation [3]. They are based on a simpler approach, originally developed to study BOSS Ly-α analysis feasibility [4]. The Gaussian random field boxes of density have 3.2 h^{-1}Mpc voxels, RSD are implemented using the corresponding velocity boxes and the approximation of parallel lines of sight is made. In theses mocks the positions of the quasars are drawn randomly over the box, so for cross correlation studies they were modified such that the quasar positions are selected in pixels above a given threshold.

The eBOSS DR14 autocorrelation [5] and cross-correlation [6] analyses made use of BOSS mocks. Two mock projects were developed for the final eBOSS DR16 analysis and for DESI. They are described in [7] and in this paper.

3 Mock generation

We define a box of volume V in real space with cubic voxels of typically a few Mpc side. An uncorrelated complex gaussian random field with $\delta(-\boldsymbol{k}) = \delta^*(\boldsymbol{k})$ is generated in the corresponding box in Fourrier space. Each mode k is multiplied by $\sqrt{P_m(k)/V}$, where P_m is the matter power spectrum at $z = 0$ obtained with CAMB with parameters $\Omega_M = 0.31457$, $\Omega_\Lambda = 0.68543$ and $\Omega_k = 0$. An inverse 3D Fourier transform is then performed using fast Fourier transform algorithm (FFT) to produce a correlated box, δ, in real space with power spectrum P_m.

In Refs. [4] and [7] the quasars are randomly drawn in cells where the field is above a given threshold, which is selected to produce the required bias. This procedure results in a quasar-quasar correlation function that is significantly different from both the linear prediction and the measured one on small scales. Here, we produce a quasar-density box, filled with field δ_q, and draw a quasar in each cell with a probability proportional to $\exp \delta_q$. This box is obtained from the same box in Fourier space as the δ box but using a different power spectrum. This power spectrum is such that the corresponding correlation function is $c(r) = \log(1 + b_q^2 G^2(z_0)\xi_m(r))$, where b_q is the quasar bias, G the linear growth factor (with the convention $G(z = 0) = 1$) and ξ_m is the matter correlation function at $z = 0$. This ensures that the power spectrum of $\exp \delta_q$ has a bias b_q relative to matter at z_0 [**?**].

The z dependance of the quasar correlation function can be introduced by drawing quasar with a probability proportional to $\exp[\alpha(z)\delta_q]$ with $\alpha(z) = b(z)(1 + z_0)/[b(z_0)(1 + z)]$. To first order this is producing quasars with a correlation function $\alpha^2(z)\xi(z_0)$, which is what we need. The deviations are, however, significant at low values of r and a better result is obtained by producing two quasar-density boxes at z_1 and z_2, and linearly interpolating between the probabilities obtained from the two boxes. This interpolation results in a value of the probability that is slightly too low. On the other hand extrapolating outside the $[z_1, z_2]$ range results in a slightly too high value. The shape of these under and over-estimations as a function of r happen to be very close. So finally we produce three quasar-density boxes at $z_1 = 1.9$, $z_2 = 2.75$ and $z_3 = 3.6$. To produce the probability field at e.g. $z = 2.3$, we combine the probabilities from the interpolation between z_1 and z_2 and from the extrapolation from z_2 and z_3, with coefficients that accurately compensate the under and the over-estimates. As a result, we get the expected correlation function at better than 4×10^{-4} down to $r = 5\,h^{-1}$ Mpc. *(should we put this rather technical paragraph in an appendix with some more details and may be a plot ?)*

Using the same box in Fourier space again, we also produce three other boxes ($j = 1, 2, 3$) that contains the three coordinates of the velocity field at $z = 0$. These boxes are used to compute the radial velocity of each quasar, which is corrected to get its value at the considered quasar redshift and then added to the "real" redshift to give the "measured" redshift, including redshift space distortions (RSD).

The value of the δ field is computed along slanted lines-of-sight towards the quasars. Using the closest grid point would result in discontinuities and then aliasing to small scales, i.e. spurious power on k larger than $k_N = \pi/a$ (Nyquist k), where a is the side of the box voxels. To avoid that, we define the value of the field at a given point as a Gaussian-kernel average over neighboring grid points. We use the voxel side as the sigma of this Gaussian smoothing, which cuts the power spectrum at large k, making the aliasing terms at $k > k_N$ completely negligible. The effect on the correlation function can be computed and, for the nominal voxel side used, $a = 2.19\,h^{-1}$ Mpc, it is hardly visible on a plot[1].

This provides a field with the proper 3D correlation. However, the field is averaged over the size of a voxel in the transverse directions, while in the real data the average is only over the Jean's scale. This averaging is removing a lot of low scale fluctuations and results in a 1D correlation or power spectrum of the field that is significantly lower than in real data. More precisely, the 1D power spectrum is an integral of the 3D power spectrum over $k_\perp$ [**?**],

$$P^{1D}(k_\parallel) = \frac{1}{2\pi}\int_0^\infty P(k_\parallel, k_\perp)k_\perp dk_\perp \;, \tag{3.1}$$

which in our mocks is truncated at k_N, so that a constant power, independent of $k_\parallel$ and corresponding to the integral from k_N to infinity, is missing. Solving this issue by reducing the voxel size to the Jean's scale would require 3D FFT with more than 10^{13} voxels, which is not reasonable. The idea is then to write the field as the sum of a large scale and a small scale fields, $\delta = \delta_l + \delta_s$, where δ_l is the result of the 3D FFT, while δ_s is a random field that is independent between different lines of sight. The 3D correlation function of δ is then that of δ_l, which is very close to CAMB correlation function, as discussed above. The power spectrum of δ_s is chosen to get the measured 1D power spectrum of the flux.

Refs. [4] and [7] introduced the redshift space distortions (RSD) using the velocity field. Here we use the velocity-gradient tensor field, which reads in Fourier space

$$\eta^k_{mn} = f\frac{k_m k_n}{k^2}\delta^k \;, \tag{3.2}$$

where $f = \mathrm{d}\ln G/\mathrm{d}\ln a$ is the linear growth rate. We use the lognormal approach [**?**] to get the baryon density from the δ field and the baryon density is transformed into transmitted flux fractions F using the fluctuating Gunn-Peterson approximation [**? ?**], which is modified with

[1] On a plot of $r^2\xi(r)$ with a maximum of 120 and a peak height of 50, the maximum effect is 3 for $r > 7h^{-1}$Mpc and reaches 13 at small r.

an additional $\eta_{\parallel}$ term to include RSD, namely

$$F = \exp[-a(z) \exp g] \quad \text{with} \quad g = b\, G(z)(\delta + c(z)\eta_{\parallel}) \; . \tag{3.3}$$

Here b was fixed to $b = 2 - 0.7(\gamma - 1) = 1.58$ for an intergalactic-medium equation-of-state parameter $\gamma = 1.6$'[**?**] and $\eta_{\parallel} = u_m u_n \eta^{mn}$ is the velocity gradient along the line-of-sight with unit vector u. The parameters $a(z)$ and $c(z)$ are adjusted to match the observed Ly-α bias and β parameters. The parameter $c(z)$ controls redshift space distortions, and its value is actually close to β while $a(z)$ controls the bias.

The power spectrum of the g field is given by the Kaiser formula, $P_g(\boldsymbol{k}) = b^2 G^2 \left(1 + \beta\mu_k^2\right)^2 \times P_m(k, z = 0)W^2(k)$, where $W(k)$ is the Gaussian smearing in k space. The correlation function of g can then be obtained using Hamilton's formula [**?**] and the correlation function of F is deduced from that of g using Eqn (2.6) of ref. [2]. We therefore have a prediction for the correlation function of our mocks. This is very convenient when tuning the mocks parameters.

References

[1] Font-Ribera, A., P. McDonald, and J. Miralda-Escudé. Generating mock data sets for large-scale Lyman-α forest correlation measurements. *JCAP*, 1:1, 2012. arXiv:1108.5606.

[2] Font-Ribera, A., E. Arnau, J. Miralda-Escudé, et al. The large-scale quasar-Lyman α forest cross-correlation from BOSS. *JCAP*, 5:018, 2013. arXiv:1303.1937.

[3] du Mas des Bourboux, H., J.-M. Le Goff, M. Blomqvist, et al. Baryon acoustic oscillations from the complete SDSS-III Lyα-quasar cross-correlation function at z = 2.4. *Astron. and Astrophys.*, 608: A130, 2017. arXiv:1708.02225.

[4] Le Goff, J. M., C. Magneville, E. Rollinde, et al. Simulations of BAO reconstruction with a quasar Ly-α survey. *Astron. and Astrophys.*, 534:A135, 2011. arXiv:astro-ph.CO/1107.4233.

[5] de Sainte Agathe, V., C. Balland, H. du Mas des Bourboux, et al. Baryon acoustic oscillations at z = 2.34 from the correlations of Lyα absorption in eBOSS DR14. *Astron. and Astrophys.*, 629:A85, 2019. arXiv:astro-ph.CO/1904.03400.

[6] Blomqvist, M., H. du Mas des Bourboux, N. G. Busca, et al. Baryon acoustic oscillations from the cross-correlation of Lyα absorption and quasars in eBOSS DR14. *Astron. and Astrophys.*, 629:A86, 2019. arXiv:astro-ph.CO/1904.03430.

[7] Farr, J., A. Font-Ribera, H. du Mas des Bourboux, et al. LyaCoLoRe: Synthetic Datasets for Current and Future Lyman-α Forest BAO Surveys. *arXiv e-prints*, page arXiv:1912.02763, 2019. arXiv:astro-ph.CO/1912.02763.