Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **535,** 826–852 (2024)
Advance Access publication 2024 October 14

https://doi.org/10.1093/mnras/stae2358

Probing the cosmic web in Lyα emission over large scales: an intensity mapping forecast for DECaLS/BASS and DESI

Pablo Renard [1]★ Daniele Spinoso [1] Paulo Montero-Camacho [2] Zechang Sun,[1] Hu Zou[3] and Zheng Cai[1]

[1]*Department of Astronomy, Tsinghua University, Beijing 100084, China*
[2]*Department of Mathematics and Theory, Peng Cheng Laboratory, Shenzhen, Guangdong 518066, China*
[3]*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China*

Accepted 2024 October 9. Received 2024 October 9; in original form 2024 June 26

ABSTRACT

Being the most prominent H I line, Lyα permeates the cosmic web in emission. Despite its potential as a cosmological probe, its detection on large scales remains elusive. We present a new methodology to perform Lyα intensity mapping with broad-band optical images, by cross-correlating them with Lyα forest data using a custom one-parameter estimator. We also develop an analytical large-scale Lyα emission model with two parameters (average luminosity $\langle L_{\mathrm{Ly}\alpha}\rangle$ and bias b_{e}) that respects observational constraints from quasar (QSO) luminosity functions. We compute a forecast for Dark Energy Camera Legacy Survey (DECaLS)/Beijing–Arizona Sky Survey (BASS) g-band images cross-correlated with Dark Energy Spectroscopic Instrument (DESI) Lyα forest data, setting guidelines for reducing images into Lyα intensity maps. Given the transversal scales of our cross-correlation (26.4 arcmin, $\sim$33 cMpc h^{-1}), our study effectively integrates Lyα emission over all the cosmic volume inside the DESI footprint at $2.2 < z < 3.4$ (the g-band Lyα redshift range). Over the parameter space ($\langle L_{\mathrm{Ly}\alpha}\rangle$, b_{e}) sampled by our forecast, we find a 3σ of large-scale structure in Lyα likely, with a probability of detection of 23.95 per cent for DESI–DECaLS/BASS, and 54.93 per cent for a hypothetical DESI phase-II with twice as much Lyα QSOs. Without a detection, we derive upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ competitive with optimistic literature estimates [$2.3 \pm 1 \times 10^{41}$ erg s^{-1} (cMpc^3)$^{-1}$ for DESI, and $\sim$35 per cent lower for its hypothetical phase-II]. Extrapolation to the DESI-Rubin overlap shows that a detection of large-scale structure with Lyα intensity mapping using next-generation imaging surveys is certain. Such detection would allow constraining $\langle L_{\mathrm{Ly}\alpha}\rangle$, and explore the constraining power of Lyα intensity mapping as a cosmological probe.

Key words: intergalactic medium – large-scale structure of Universe – cosmology: observations.

1 INTRODUCTION

The observational study of the Universe consists mostly in the detection and analysis of radiation from cosmic origin, emitted by baryonic matter, and bent by the gravitational potential of matter itself. In the visible spectrum, the vast majority of this radiation is due to thermal and line emission of resolved objects, such as galaxies and quasars (QSOs). These resolved objects cluster in filaments and nodes, forming a cosmic web that traces the underlying dark matter large-scale structure (LSS). The study of the LSS statistical properties is a cornerstone of observational cosmology, constraining vastly different aspects of cosmological models (Bernardeau et al. 2002), such as inflationary fields (Planck Collaboration 2020), dark matter properties (Kunz, Nesseris & Sawicki 2016), or neutrino masses (Palanque-Delabrouille et al. 2015).

Most of the baryon budget of the Universe lies in the intergalactic medium (IGM), in the form of diffuse gas (Shull, Smith & Danforth 2012). One observational technique that traces diffuse baryons in the IGM, as well as objects too faint to be directly resolved, is intensity mapping (IM). IM consists on the analysis of a single emission line across large sky areas, without resolving any objects (Peterson et al. 2009). Since the rest-frame wavelength of the emission line is well known, observing in a given wavelength range is equivalent to observing in a certain redshift interval. This technique was originally devised for 21-cm line emission (Madau, Meiksin & Rees 1997; Loeb & Zaldarriaga 2004; Hu et al. 2020), but its application has also been studied for other emission lines (Bernal & Kovetz 2022), such as CO (Breysse, Kovetz & Kamionkowski 2014; Li et al. 2016; Keating et al. 2020) and C II (Gong et al. 2012; Yue et al. 2015; Yang, Pullen & Switzer 2019) in the infrared, Hα in the optical (Parsons et al. 2022), or Lyα in the ultraviolet (UV, Silva et al. 2013; Pullen, Doré & Bock 2014). The case for Lyα IM is of particular interest, given that it is the strongest emission line from neutral hydrogen (H I), hence its presence should be ubiquitous in the IGM.

In fact, IGM Lyα absorption has been widely observed in the spectra of high-redshift QSOs, where it produces the so-called Lyα forest (Rauch 1998), i.e., a series of absorption lines bluewards of the prominent QSO Lyα emission. Compiling large Lyα forest catalogues has been a key goal of many spectroscopic surveys (e.g. the Baryon Oscillation Spectroscopic Survey, BOSS, Dawson et al. 2013; the extended Baryon Oscillation Spectroscopic Survey,

★ E-mail: p.renard.guiral@gmail.com

eBOSS, Ahumada et al. 2020; the Dark Energy Spectroscopic Instrument, DESI, Ramírez-Pérez et al. 2024; or the WHT Enhanced Area Velocity Explorer QSO survey, WEAVE-QSO, Pieri et al. 2016), from which LSS features have been observed, such as baryon acoustic oscillations (BAO, e.g. Bautista et al. 2017; DuMasDesBourboux et al. 2020; DESI Collaboration 2024), and smaller scale features such as the one-dimensional power spectrum (Palanque-Delabrouille et al. 2013; Iršič et al. 2017; Chabanier et al. 2019). The absorption features of the QSOs Lyα forest are due to absorption by neutral gas within relatively low-density regions. This means that the Lyα forest traces a density field more linear than galaxy clustering (i.e. its evolution can be described with reasonable accuracy with linearized equations, Bernardeau et al. 2002). However, if the H I density is too high, Lyα absorption saturates, resulting in Lyman limit systems (Songaila & Cowie 2010) and damped Lyα systems (DLAs, Wolfe, Gawiser & Prochaska 2005; Wang et al. 2022a). Since the Lyα forest traces the underlying dark matter distribution in a low-density regime that can primarily be described by linearized equations, it is a particularly useful tracer to place valuable constraints on dark matter models (Viel et al. 2013; Palanque-Delabrouille et al. 2020; Iršič et al. 2024) neutrino masses (Palanque-Delabrouille et al. 2015; Yèche et al. 2017), or even reionization physics (Montero-Camacho & Mao 2021).

On the other hand, Lyα emission traces LSS in a very different manner. UV emission from galaxies, QSOs or the UV background (Chiang, Ménard & Schiminovich 2019; Gallego et al. 2021) excites H I, which in turn emits in Lyα. Being a resonant line, the diffuse Lyα emission scatters into the IGM before reaching the observer (Cantalupo et al. 2005; Byrohl & Nelson 2023). This Lyα emission generates primarily in overdense regions (Byrohl & Nelson 2023), which makes it complementary to the Lyα forest absorption tracing the underdense regime. Since it scatters before reaching us, the observed Lyα emission is directly associated to baryons enclosed within the cosmic web. Hence, IGM Lyα emission highly depends on IGM physical properties, such as temperature or ionization state (Ouchi, Ono & Shibuya 2020). Therefore, combining Lyα forest measurements with Lyα emission LSS may help disentangling how both astrophysical and cosmological processes shape the baryon distribution across the cosmic web.

This diffuse Lyα emission has already been well documented up to cMpc scales. At halo scales (∼100 pkpc) diffuse Lyα emission around QSOs and star-forming galaxies has been extensively studied (Lyα blobs, Taniguchi, Shioya & Kakazu 2001; Matsuda et al. 2004; Wisotzki et al. 2018), and more recently, larger Lyα nebulae of scales $<$ 1p.m.pc sourced by QSO UV emission have also been reported (enourmous Lyα nebulae, Cai et al. 2017, 2019; Battaia et al. 2022). On even larger scales, filamentary emission between galaxies at scales of few cMpc has been observed (Bacon et al. 2021). Generally, these studies are based on observations on small, ultradeep fields, since great image depth is usually needed to securely measure faint, extended Lyα emission.

Another possible approach to statistically measure LSS in Lyα emission is to perform IM, i.e., to average observed cosmic emission with a given correlation estimator over large cosmological volumes (i.e. cosmological surveys with much wider but shallower fields than deep, targeted observations). Such studies would allow to detect LSS Lyα emission on scales much larger than those probed by ultradeep fields (e.g. $\sim$ 5 cMpc field size for Bacon et al. 2021); however, all studies so far have only partially succeeded at this endeavour. For example, Croft et al. (2016) and Lin, Zheng & Cai (2022) detect Lyα emission at scales $<$30 cMpc h^{-1} at $z \sim 2.5$ by cross-correlating the residuals in galaxy spectra with QSOs in BOSS/eBOSS. However, these cross-correlations only sample the overdense regions surrounding QSOs; a small fraction of the total survey volume that might not be fully representative of the cosmic average. When repeating this cross-correlation with the Lyα forest to sample a much larger volume (Croft et al. 2018), no significant LSS in Lyα emission is found. The detection of a large-scale cross-correlation between Lyα emission in a narrow-band photometric survey (the Physics of the Accelerating Universe Survey, PAUS, Padilla et al. 2019) and Lyα forest data are also simulated and evaluated in Renard et al. (2021). The results of the Renard et al. (2021) forecast were that a detection of large-scale Lyα emission with Lyα forest cross-correlation was seemingly impossible for a projected survey footprint of 100 deg^2, but for similar observation times expanding the survey footprint was far more efficient in increasing the signal-to-noise ratio (SNR) than going deeper. It is worth noting that future IM experiments, such as *SPHEREx* (Doré et al. 2014), also aim to perform IM with Lyα among other lines (Visbal & McQuinn 2023).

Here, we present a new methodology to perform Lyα IM by cross-correlating broad-band images and Lyα forest data: being the standard for photometric surveys, most of the night sky has already been observed in broad bands, which potentially allows IM to be applied on very large cosmological volumes. We develop a specific one-parameter correlation estimator to evaluate the fluctuation of Lyα luminosity versus Lyα forest absorption, as well as a simple analytical Lyα emission model for cosmological scales. We perform a forecast for DESI and the g band of its Legacy Imaging Surveys (Dark Energy Camera Legacy Survey, DECaLS/Beijing–Arizona Sky Survey, BASS; Dey et al. 2019), providing realistic expectations for a detection of Lyα LSS, as well as the upper bounds derived from a non-detection.

This work is divided as follows. In Section 2, we describe the relevant specifications of the cosmological surveys used in our forecast. In Section 3, we provide an in-depth explanation of the proposed Lyα IM methodology. In Section 4, the simulation for our forecast is described, with special emphasis on image reduction and computation of intensity maps; the forecast results are displayed in Section 5. Finally, in Section 6, we discuss the implications of our results for upcoming surveys, as well as the lines of research that our analysis may open up. We summarize our conclusions in Section 7. Throughout this paper, we assume a flat Lamda cold dark matter cosmology with $h = 0.702$, $\Omega_m = 0.275$, $\Omega_\Lambda = 0.725$, $\Omega_b = 0.046$, $n_s = 0.968$, and $\sigma_8 = 0.82$, as this is the parameter set of the hydrodynamic simulation used in Section 4.

2 COSMOLOGICAL SURVEYS

Optical cosmological surveys are broadly divided in two categories: imaging/photometric. The former take images of the sky with CCDs covering the field of view (FOV) of a camera (e.g. Flaugher et al. 2015; Padilla et al. 2019); their wavelength sensitivity is determined by different sets of filters. The latter generally observe spectra of pre-designated targets (e.g. Eisenstein et al. 2011; DESI Collaboration 2022), trading off speed and full sky coverage for much higher spectral resolution. Our work consists in detecting the diffuse Lyα emission tracing LSS that must be contained in photometric images, by cross-correlating said images with the Lyα forest contained in QSO spectra. Hence, both spectroscopic and imaging data are required.

2.1 DESI

The DESI (DESI Collaboration 2022) is a spectroscopic cosmological survey, currently ongoing at the Mayall telescope in Kitt Peak, Arizona. The survey is expected to cover a total of 14 000 deg^2 in its phase-I, thus becoming the largest spectroscopic survey yet.

DESI targets four different galaxy samples, in ascending redshift order: the bright galaxy sample, luminous red galaxies, emission-line galaxies, and QSO. The latter are split into two subsamples, according to their redshift. The sample at $z < 2.1$ is only detected via strong QSO emission lines, such as Lyα, C IV, C III, and S IV typically occurring in QSO spectra (e.g. Vanden Berk et al. 2001; Selsing et al. 2016), and thus only QSO positions are used as tracers. On the other hand, the QSOs at $z > 2.1$ (i.e. the Lyα QSOs) show Lyα forest absorption lines in their spectra, in addition to the aforementioned QSO emission lines. This is because at $z > 2.1$, the spectral range affected by the Lyα forest (bluewards of the Lyα line at the QSO redshift) becomes observable by the DESI spectrograph (3600–9800 Å, DESI Collaboration 2022). Our work is focused on the Lyα QSO subsample at $z > 2.1$.

The DESI QSO target selection is based on a colour cut of point-like sources using the optical grz magnitudes from the DESI Legacy Imaging Surveys (Dey et al. 2019), together with the $W1$ and $W2$ infrared magnitudes of the Wide-Field Infrared Survey Explorer satellite (*WISE*, Wright et al. 2010). This QSO colour selection is expected to yield an approximate completeness of 60 per cent for a limiting magnitude of $r < 23$ (DESI Collaboration 2016). This selection results in a preliminary Lyα QSO density of 40 deg^{-2}, and the QSO redshift distribution shown in Fig. 1 (only for the redshift range of interest in this work). We will work with this preliminary estimate, but it is worth noting that current DESI observations reach significantly higher Lyα QSO completeness, yielding a Lyα QSO density of 60 deg^{-2} (Chaussidon et al. 2023), well above the DESI science requirement of 50 deg^{-2}. Higher Lyα QSO density on an eventual final DESI phase-I data release means a larger total number of Lyα forests, and thus higher statistical power and SNR for our Lyα IM methodology.

2.2 DECaLS–BASS

Spectroscopic surveys require preliminary target selection, usually based on already existing photometric data. For DESI, three imaging surveys where carried out for this purpose (Dey et al. 2019): the DECaLS (grz bands), the BASS (gr bands, Zou et al. 2017a), and the Mayall z-band Legacy Survey (z band). Since this work uses the g band to perform an IM forecast, we focus only on the first two. The combination of DECaLS and BASS covers the entirety of the DESI footprint on the gr bands.

DECaLS has been carried out at the Blanco Telescope, with the Dark Energy Camera (DECam, Flaugher et al. 2015), a large camera with an hexagonal FOV of approximately 3.2 deg^2. BASS was observed at the Bok Telescope with the 90Prime camera (Williams et al. 2004), which provides a square FOV of 1.08×1.03 deg^2. Both surveys reach similar magnitudes: for a point source, the detection limit at 5σ is $g = 23.95$ for DECaLS and $g = 23.65$ for BASS (Dey et al. 2019, see table 4).

For the purposes of our forecast, the survey features we are interested in are the survey depth and the FOV of the camera. A deeper survey means a fainter magnitude detection limit, and lower photometric noise in the images. The FOV will determine the maximum angular area in which sky subtraction can be performed without removing clustered diffuse Lyα emission (see Section 4.2.1), and thus the largest scales we can sample. DECaLS both has a larger FOV and slightly higher survey depth than BASS (Dey et al. 2019), therefore, we will be conservative and model our forecast only after BASS, even though DECaLS covers most of the DESI footprint (9850 deg^2 versus 5500 deg^2).

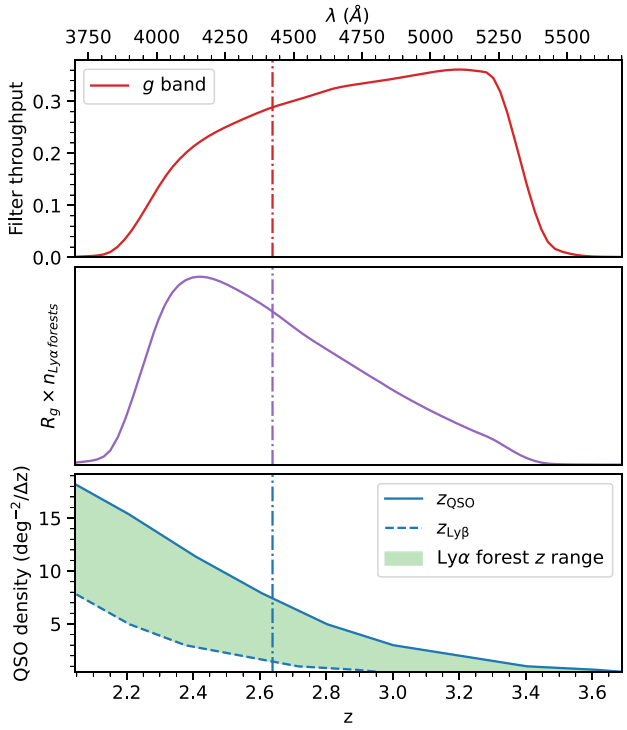

Figure 1. Upper panel: standard SDSS g-band response function. The g-band wavelength in the upper x-axis matches the redshift for the Lyα line in the x-axis of the lower panel. Middle panel: product of the g-band response function times the number density of Lyα forest sightlines per redshift, normalized. This is the redshift kernel of our cross-correlation study; the dash–dotted line (in all three panels) represents the average redshift of our convolution kernel. Lower panel: DESI projected QSO angular density versus redshift (solid blue line). The accompanying dashed blue line represents the lower redshift limit of the observable Lyα forest, marked by the start of the Lyβ forest. Therefore, the green-shaded region is the redshift range covered by the Lyα forest for a given QSO redshift.

For the sake of homogeneity, we use in this work the original standard g band from the Sloan Digital Sky Survey (SDSS, York et al. 2000); hypothetical deviations in the response function of the g-band filters used in DECaLS/BASS are expected to be negligible. We show in Fig. 1 the response function of the SDSS g band, with the corresponding redshift range for the Lyα line. Together with the DESI QSO redshift distribution, this figure gives an idea of the redshift range sampled by our study: the DECalS/BASS images contain an intensity map of Lyα emission convolved in redshift with the g band, and the Lyα forest of DESI samples Lyα-sightlines with a spatial density given by the QSO distribution. Both are convolved in the middle panel of Fig. 1 to show the redshift kernel of our cross-correlation study, whose average redshift is $\langle z \rangle = 2.64$.

3 METHODOLOGY

The optical depth of the Lyα forest is a tracer of the H I density along the line of sight; hence regions with higher optical depths corresponding to higher H I densities. This relationship between Lyα

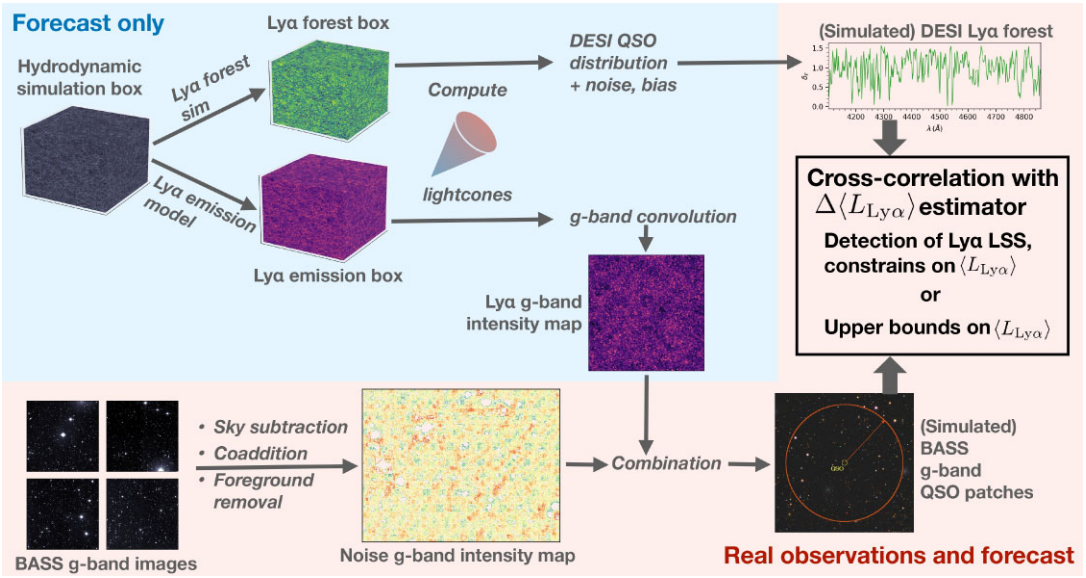

Figure 2. Diagram of the Lyα IM forecast presented in this paper. Blue region corresponds to forecast-only steps (i.e. simulation of the Lyα absorption/emission signals). Red region corresponds to steps carried out in this forecast, but also expected to be performed on an eventual observational study (e.g. image reduction and cross-correlation).

absorption and H I column density depends on both physical parameters of the IGM (e.g. temperature, photoionization rate, and line-of-sight velocity) and cosmological parameters (*Hubble* constant and baryon density). For an analytical model valid on linear scales, see the fluctuating Gunn–Peterson approximation (e.g. Weinberg et al. 1997; Kooistra, Lee & Horowitz 2022). The relationship between H I density and Lyα emission is also convoluted, given that it is a resonant line (Dijkstra 2017; Ouchi et al. 2020), and the bulk of Lyα emission seems to be sourced from faint Lyα emitters (LAEs) and their surrounding circumgalactic medium (Bacon et al. 2021; Byrohl & Nelson 2023).

In order to evaluate the correlation between the Lyα absorption in the DESI QSO spectra and the Lyα emission contained in *g*-band images, we will not use the conventional two-point correlation function (2PCF) and its Fourier space analogue, the power spectrum monopole (Bernardeau et al. 2002), but a custom estimator based on stacking with Bayesian inference the fluctuations of Lyα emission binned in bins of Lyα forest absorption.

The 2PCF/power spectrum and its higher order counterparts (e.g. bispectrum), are the backbone of the study of LSS (e.g. Bernardeau et al. 2002), including IM (Bernal & Kovetz 2022). However, alternative summary statistics have also been described in the literature and applied to constrain the properties of different LSS tracers. For example, the wavelet transform has seen use to extract non-Gaussian information (Maksora Tohfa et al. 2024) or capture environmental information (Wang, Yang & He 2022b), among other works. Another example of alternative statistic is the *k*-estimator (Adelberger et al. 2005), which evaluates clustering strength and constrains linear tracer bias in small fields, using only fine redshift bins (e.g. Diener et al. 2017; Herrero Alonso et al. 2021). In the context of IM, the voxel intensity distribution, i.e., the histogram of voxel fluxes, has also been studied as a complementary statistic to the power spectrum (e.g. Ihle et al. 2019; Sato-Polito & Bernal 2022).

Following these examples, in this work we define a custom estimator instead of using the 2PCF/power spectrum. Our main reason is the degeneracy between the absolute value of the 2PCF and the foreground value in the Lyα intensity map (Renard et al. 2021). Due to this degeneracy, we would need a significant detection of the 2PCF in at least two distance bins to place any constraints on the combined bias of diffuse Lyα emission as an LSS tracer, by constraining the shape of the 2PCF. Unfortunately, the SNR we expect with DESI–DECaLS/BASS is very likely to not be enough (Section 5). Moreover, the bias of the Lyα emission field is not properly constrained (see Section 4.1 for a short discussion), which adds an extra degeneracy to potential constraints obtained through the 2PCF.

Therefore, to fully exploit the low SNR of the Lyα emission intensity map, we use a single-parameter estimator with the fewest assumptions possible, namely: the cosmological principle. We assume both the Lyα absorption and emission fields to be isotropic, which should be sufficient as a first approximation (especially given the very large redshift convolution kernel, see Fig. 1). We assume a correlation monotonically decreasing with comoving distance between two arbitrary points, and both Lyα emission and absorption fields being virtually homogeneous on large enough scales (the homogeneity scale χ_{h}, e.g. see Sarkar et al. 2009; Pandey & Sarkar 2015; Gonçalves et al. 2018). Despite the use of a custom estimator, we consider our work an IM study, since we constrain the properties of an emission-line tracing LSS by integrating all observed emission, without resolving any objects (as in the general definitions given in e.g. Peterson et al. 2009; Kovetz et al. 2017; Bernal & Kovetz 2022). Moreover, the transversal scales we sample are large enough to be considered LSS (Section 5.1.2), and our Lyα emission model is defined in the linear regime, where the effects of gravitational collapse become negligible (Section 4.1).

In Fig. 2, we provide a diagram which illustrates the overall structure of our forecast: how different simulated and observed data are combined to simulate DESI and DECaLS/BASS, and perform Lyα IM by cross-correlating them. This is intended just as a visual reference; the cross-correlation estimator is fully described in this section, and in Section 4, the simulation of our forecast is thoroughly laid out.

3.1 Imaging data: QSO patches

The g-band images are a two-dimensional projection of the Lyα emission field, convolved over a large redshift range. Given our assumption that the correlation between Lyα emission and absorption decreases with distance and that isotropy holds, our analysis must concentrate on circular regions centred on the Lyα DESI QSOs. This approach will maximize the correlation and consequently yield the highest SNR. Throughout the paper, by distance r, we will refer to transversal angular distance, projected onto the sky. We will assume that the flat-sky approximation holds true on the scales we sample (of the order of arcminutes), and thus our assumptions derived from the cosmological principle also hold true.

Hence, we define a patch of a given QSO i as a circular aperture of radius r centred on such QSO i, on the image of a given broad-band filter (the g band in our case). The patch flux in the g band, $f_{\rm i}^{g}$, will be the average integrated flux in such patch, minus the flux of all resolved sources inside the patch (below a given magnitude threshold):

$$f_{\rm i}^{g} \equiv \frac{1}{\pi r_{\rm patch}^2}\left[\int_0^{2\pi}\int_0^{r_{\rm patch}} f^{g}(r,\theta) r^2\,{\rm d}r\,{\rm d}\theta - \sum_{\substack{g<g_{\rm lim}\\ r<r_{\rm patch}}} f_{\rm src}^{g}\right], \tag{1}$$

where $f^{g}(r,\theta)$ is the observed flux in the g-band image, in polar coordinates centred on the respective QSO, $r_{\rm patch}$ the patch radius, and $f_{\rm src}^{\rm g}$ the flux of resolved sources in the image, brighter than the foreground threshold $g_{\rm lim}$. This source flux is summed over all sources inside the patch ($r < r_{\rm patch}$) and with magnitude brighter than the foreground cut magnitude limit $g < g_{\rm lim}$. Even if the QSO at the centre of the patch itself has $g < g_{\rm lim}$, it is not included in this sum, since we do not consider it foregrounds (a significant part of its integrated emission will be Lyα).

The only criterion we have adopted to choose to remove a given resolve source is $g < g_{\rm lim}$. Therefore, if any QSO with available Lyα forest data is inside another Lyα QSO patch, it will only be removed if its g magnitude is brighter than $g_{\rm lim}$. However, we will choose a conservative foreground cut ($g_{\rm lim} = 19$) that preserves the vast majority of DESI Lyα QSOs, as discussed in Section 4.2.1.

In a more general sense, we could define a patch as

$$f_{\rm i}^{g} \equiv \frac{\int_0^{2\pi}\int_0^{\infty} W(r) f^{g}(r,\theta) r^2\,{\rm d}r\,{\rm d}\theta - \sum_{\substack{g<g_{\rm lim}\\ r<r_{\rm patch}}} f_{\rm src}^{g}}{2\pi\int_0^{\infty} W(r)\,{\rm d}r}, \tag{2}$$

with $W(r)$ being a radial kernel such that $\lim_{r\to\infty} W(r) = 0$, normalized by its integral. Under this definition, equation (1) represents the particular case where $W(r)$ is a step function with $W(r) = 1$ for $r < r_{\rm patch}$ and $W(r) = 0$ elsewhere. A different radial kernel $W(r)$ that weights more the more correlated (closer) regions might result in a higher SNR of our estimator (i.e. a Gaussian kernel, or any monotonically decreasing function). However, since we will perform our forecast for the $r_{\rm patch}$ that yields the maximum SNR for our estimator (Section 5.1.2), we argue that the gains in SNR we may derive from using a different $W(r)$ must be fairly limited.

Moreover, we also choose a step kernel and follow equation (1) to ensure that we reach radial scales large enough to sample the whole cosmological volume defined by the DESI footprint and the g-band Lyα redshift range. A $W(r)$ preferentially weighting closer regions could lead to an overestimation the cross-correlation radial length, since larger radial scales would be unfavourably weighted, regardless of the correlation length we define (e.g. the full width at half-maximum, FWHM of $W(r)$ or its second moment). However, with a step function, $r_{\rm patch}$ is unequivocally the maximum radial scale that contributes to our cross-correlation.

Being able to sample the entire survey cosmological volume is an important feature of our methodology. Previous Lyα IM works cross-correlating QSO positions with residuals in galaxy spectra (Croft et al. 2016; Lin et al. 2022) only sample ∼20–30 cMpc around QSOs in both transversal and line-of-sight directions, which is a small fraction of the total survey volume: overdense regions affected by the QSO proximity effect that may not be representative of the whole survey volume.

Throughout the paper, we will adopt the convention of applying a superscript $^{\rm g}$ to cosmic emission/absorption fields that have been observed and convolved in redshift/wavelength with the g band. Moreover, we will assume flux units of erg s^{-1} cm^{-2} Å^{-1} through this work, and erg s^{-1} cm^{-2} Å^{-1} arcmin^{-2} units for integrated patch fluxes.

Most of the flux f^{g} observed inside a given QSO patch will not come from Lyα itself ($f_{{\rm Ly}\alpha}^{g}$), but contamination of various origins ($f_{\rm noise}^{g}$): instrumental (CCD thermal noise), atmospheric (telluric lines and airglow), solar/lunar (moonlight and zodiacal light), and cosmic (unresolved/non-removed emission from other rest-frame wavelengths than Lyα). For this reason, it is crucial to understand how the emitted Lyα luminosity of cosmic origin, $L_{{\rm Ly}\alpha}$, is transformed into the signal we observe, $f_{{\rm Ly}\alpha}^{g}$. If we consider an infinitesimal element of volume in redshift and celestial equatorial coordinates (dz dα dδ), the Lyα luminosity emitted in such element of volume will be $L_{{\rm Ly}\alpha}(z,\alpha,\theta)\,{\rm d}z\,{\rm d}\alpha\,{\rm d}\delta$. Therefore, the contribution of each volume element to the total observed Lyα flux density would be:

$$f_{{\rm Ly}\alpha}(z,\alpha,\delta)\,{\rm d}z\,{\rm d}\alpha\,{\rm d}\delta = \frac{L_{{\rm Ly}\alpha}(z,\alpha,\delta)\,{\rm d}z\,{\rm d}\alpha\,{\rm d}\delta}{4\pi\,D_L(z)^2\,\Delta\lambda^{\rm obs}}, \tag{3}$$

where $D_L(z)$ is the luminosity distance to said volume element, and $\Delta\lambda^{\rm obs}$ the wavelength width of the Lyα line emitted by this element. Neglecting line broadening, this wavelength width is:

$$\Delta\lambda^{\rm obs} = \lambda_{{\rm Ly}\alpha}\,{\rm d}z. \tag{4}$$

Given that any line broadening due to thermal motion, turbulence or line resonance (Ouchi et al. 2020) is at least an order of magnitude smaller than the FWHM of any broad-band filter (∼1000 Å), this assumption fully holds for g-band images. Finally, to determine the total Lyα flux of cosmic origin observed with the g band, we just need to convolve equation (3) with the g-band transmission curve/response function:

$$f_{{\rm Ly}\alpha}^{g}(\alpha,\delta)\,{\rm d}\alpha\,{\rm d}\delta = \frac{\int_0^{\infty} f_{{\rm Ly}\alpha}(z,\alpha,\delta)\,{\rm d}\alpha\,{\rm d}\delta R_g((1+z)\cdot\lambda_{{\rm Ly}\alpha}){\rm d}z}{\int_0^{\infty} R_g((1+z)\cdot\lambda_{{\rm Ly}\alpha})\,{\rm d}z}, \tag{5}$$

where R_g is the response function of the g band displayed in the upper panel of Fig. 1. The integration limits have been simply been set to $z = 0$ and ∞, as the redshift interval is naturally bounded by R_g.

This $f_{{\rm Ly}\alpha}^{g}(\alpha,\delta)$ flux is all the information we receive from the IGM Lyα emission in g-band images; proper integration over angular coordinates will yield the patch fluxes f_i^{g} expressed in equation (1). Each one of these patches can be seen as the redshift projection of a narrow cylinder of Lyα emission centred on the QSO (see Fig. 3), plus all the noise from different sources. Given that we expect noise to be dominant over the Lyα emission signal, the Lyα emission contained in patches should be impossible to observe on an individual patch basis, so a very large sample and/or a signal to cross-correlate with are needed for a detection.

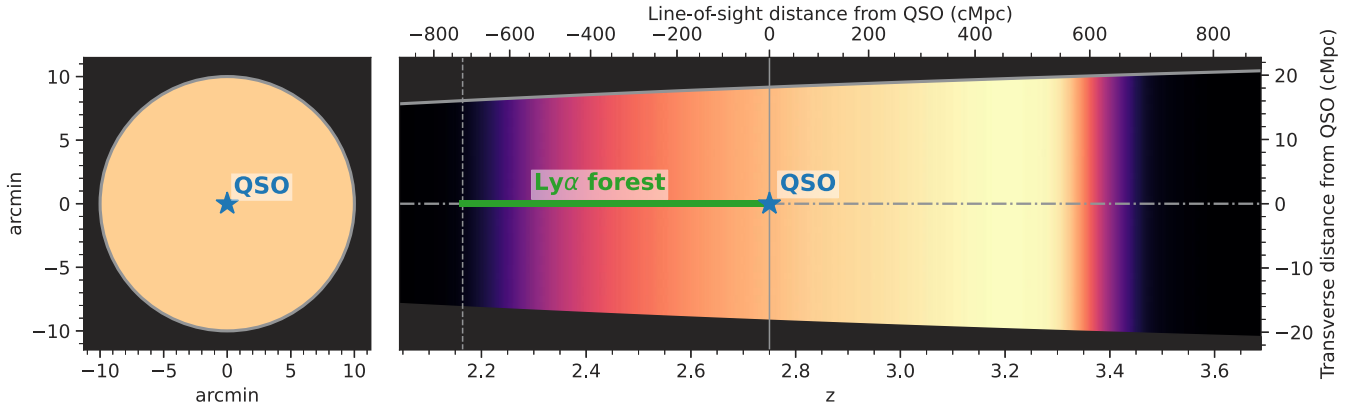

Figure 3. Left panel: cartoon of a g-band QSO patch with a radius of 10 arcmin, centred around a $z = 2.75$ QSO (not to scale). Right panel: deprojection of the patch along the redshift direction (g-band deconvolution). The lower x-axis shows the redshift of the Lyα signal in linear scale, and the upper x-axis, the corresponding comoving line-of-sight distance from the QSO. The y-axis shows the comoving transverse distance from the line of sight, and the solid grey lines, the comoving limits of the patch, which evolve with redshift (not to scale with the line-of-sight distance). The coloured region inside the patch cylinder is the sensitivity of the g-band response to Lyα emission, with brighter regions inside the volume being more sensitive, and black regions not observing any Lyα emission. Finally, both the QSO line of sight (dash–dotted line) and the redshift range covered by its Lyα forest (green solid line) are represented along the central axis of the patch.

3.2 Spectroscopic data: convolved forest probabilities

The ideal data set to cross-correlate the QSO patches with is the Lyα forest, as it contains high-SNR information of the Lyα absorption field in the line of sight of the QSO (centre of the patch). Moreover, its redshift range totally or partially overlaps with the Lyα emission observed by the g band (Fig. 3). Nevertheless, to properly correlate both quantities, the absorption field of the Lyα forest must undergo the same transformation imprinted by the g band on the Lyα emission field (equations 3–5). First, we define the flux absorption contrast (i.e. $\delta_{\rm F}$) for the Lyα forest in a given QSO spectrum i as:

$$\delta(z)_{{\rm F}\,i} \equiv \frac{e^{-\tau_{{\rm QSO}\,i}(z)}}{\langle e^{-\tau}\rangle}, \tag{6}$$

where $\tau(z)$ is the optical depth of the Lyα forest at redshift z. We note that $\delta_{\rm F}$ is a cosmological scalar field in spherical coordinates, similar to $L_{{\rm Ly}\alpha}$ in equation (3). Therefore, the redshift z in equation (3) acts as radial coordinate along the sightline; angular coordinates are omitted since these are fixed by the respective QSO whose spectra contains the Lyα forest.

We can apply to $\delta(z)_{{\rm F}\,i}$ from equation (6) the same transformation we have applied to convert Lyα luminosity to observed g-band Lyα flux (equations 3–5) as follows:

$$\delta^{g}_{{\rm F}\,i} = \frac{1}{\langle\delta^{g}_{\rm F}\rangle}\frac{\int_{z_{{\rm min}\,i}}^{z_{{\rm max}\,i}} \frac{\delta(z)_{{\rm F}\,i}}{4\pi\, D_L(z)^2\,\Delta\lambda^{\rm obs}} R_g((1+z)\cdot\lambda_{{\rm Ly}\alpha})\,{\rm d}z}{\int_{z_{{\rm min}\,i}}^{z_{{\rm max}\,i}} R_g((1+z)\cdot\lambda_{{\rm Ly}\alpha})\,{\rm d}z}, \tag{7}$$

where we have divided the convolved $\delta^{\rm g}_{{\rm F}\,i}$ by its average for all QSOs, $\langle\delta^{\rm g}_{\rm F}\rangle$, to keep it dimensionless.[1] We also note that we adopted a definition of the $\delta_{\rm F}$ flux contrast in equation (6) without the -1 term (which is the standard convention, e.g. Cisewski et al. 2014; Ozbek, Croft & Khandai 2016; de Sainte Agathe et al. 2019). This choice has been made to prevent $\langle\delta^{g}_{\rm F}\rangle \sim 0$ and keep the $\langle\delta^{g}_{\rm F}\rangle^{-1}$ factor finite and numerically stable.

Notably, the integration ranges in equation (7) have been left specified as $z_{{\rm min}\,i}$ and $z_{{\rm max}\,i}$, instead of going from 0 to ∞, since the redshift range in which the Lyα forest can be observed is limited. The observational redshift limits are given by

$$z_{{\rm max}\,i} = z_{{\rm QSO}\,i}\ , \tag{8}$$

$$z_{{\rm min}\,i} = \max(2.1, z_{{\rm Ly}\beta\,i}); \quad z_{{\rm Ly}\beta} \equiv \frac{\lambda_{{\rm Ly}\beta}}{\lambda_{{\rm Ly}\alpha}}(1+z_{{\rm QSO}\,i}) - 1, \tag{9}$$

where the numerical value 2.1 is simply the lowest redshift at which Lyα can be observed with the DESI spectrograph, while $z_{{\rm Ly}\beta}$ is the redshift at which the Lyβ forest starts. Since $\lambda_{{\rm Ly}\beta} = 1025.722$ Å and $\lambda_{{\rm Ly}\beta} < \lambda_{{\rm Ly}\alpha}$, $z_{{\rm Ly}\beta}$ is necessarily a lower redshift limit for the observation of the Lyα forest. At $z < z_{{\rm Ly}\beta}$, the observed forest will be the superimposition of Lyα and Lyβ lines, which are produced by largely uncorrelated structure, given the large redshift offset. While some works use the Lyβ-contaminated zone of the Lyα forest (e.g. to measure BAO in Blomqvist et al. 2019), or even perform power spectrum measurements on the Lyβ forest itself (e.g. Wilson, Iršič & McQuinn 2022), we will be conservative and exclude the Lyβ regions from our forecast.

Knowing that the redshift range covered by the Lyα forest data is only a fraction of the g-band redshift coverage, and that it will also vary for each QSO, a question arises: what is the optimal way to make use of the limited redshift range? To address this question, we first introduce a nomenclature distinction. We define $\delta^{g\,{\rm obs}}_{\rm F}$ as the $\delta_{\rm F}$ of the Lyα forest convolved over the observable redshift range (equation 7 with the integration limits defined in equations 8 and 9), and $\delta^{g\,{\rm full}}_{\rm F}$ as the convolution with the g band over the whole redshift range covered by the g band (i.e. equation 7 with $z_{\rm min} = 0$ and $z_{\rm max} = \infty$, as in equation 5). Following this definition, $\delta^{g\,{\rm full}}_{\rm F}$ can not be directly observed. We could just cross-correlate $\delta^{g\,{\rm obs}}_{\rm F}$ with $f^{\rm g}$ (the patch fluxes), but $f^{\rm g}$ samples a significantly larger redshift range, and thus it will only be partially correlated with $\delta^{g\,{\rm obs}}_{\rm F}$. Moreover, due to the $D_L(z)^{-2}$ term in equation (7), the value of $\delta^{g\,{\rm obs}}_{\rm F}$ has a strong correlation with $z_{\rm QSO}$ (the higher $z_{\rm QSO}$, the smaller the $D_L(z)^{-2}$ term, and thus the average $\delta^{g\,{\rm obs}}_{\rm F}$). This is a trend uncorrelated to $f^{\rm g}$ (since it should be independent of $z_{\rm QSO}$) that would be an undesired systematic in our analysis.

We may instead adopt a Bayesian approach and determine the probability of having a certain $\delta^{g\,{\rm full}}_{\rm F}$ on a QSO sightline, having

[1] Without the $\langle\delta^{g}_{\rm F}\rangle^{-1}$ factor, $\delta^{\rm g}_{\rm F}$ would have dimensions of cm^{-2}Å^{-1}, and an extremely small order of magnitude ($\sim 10^{-61}$).

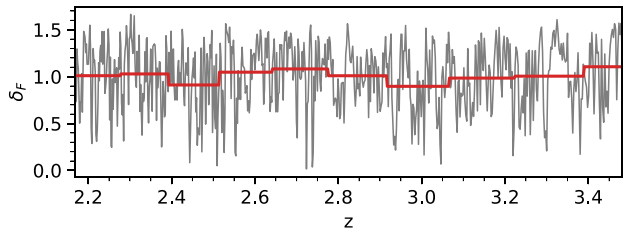

Figure 4. $\delta_F(z)$ from a full Lyα forest from our simulation (see Section 4) in grey thin line, together with the same forest, smoothed in comoving bins of width 150 cMpc, in red thick line. The bins are not uniform in wavelength/redshift space.

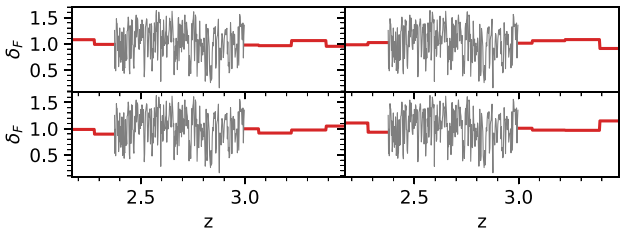

Figure 5. Four examples of generated forests (in red thick line) with the insertion of an observed forest with $z_{QSO} = 3$ (in grey thin line).

observed a given $\delta_F^{g\,obs}$. This can be done in a model-independent way by resampling observational data, if we assume that all $\delta_{F\,i}(z)$ values (i.e. all redshift bins/spectral pixels) for a given forest in a QSO spectrum i are uncorrelated. This approximation holds for bins of $\delta_F(z)$ above the scale of homogeneity, χ_h; to do so, we must smooth the Lyα forest extracted from the spectra in bins much larger than the pixels of the DESI spectrograph (DESI Collaboration 2022) (or the resolution of the simulation in a forecast like this work, see Section 4).

For this work, we will assume $\chi_h = 150$ cMpc ($\sim$105 cMpc h^{-1}), a scale similar to values reported in the literature (e.g. Pandey & Sarkar 2015; Gonçalves et al. 2018). It is worth noting that the homogeneity scale itself depends on the criterion used to measure it, as the correlation tends to zero asymptotically. For example, Pandey & Sarkar (2015) find a transition to homogeneity between 120 and 140 cMpc h^{-1} applying Shannon entropy to SDSS galaxy samples at $z < 0.2$, while Gonçalves et al. (2018) find $\chi_h < 100$ cMpc h^{-1} for most redshift bins of SDSS QSOs at $z > 0.8$ using the fractal dimension (with a threshold for homogeneity $D_2 > 2.97$). They also note that, when correcting for QSO bias, χ_h shows a clear decreasing trend with redshift; hence our assumption of $\chi_h = 150$ cMpc at $z > 2.2$ is quite conservative.

Consequently, if we smooth the forest data into wide redshift bins of comoving size $\Delta\chi = \chi_h$, (as in Fig. 4), the correlation of one bin with their adjacent bins will be close to negligible. By performing this smoothing over all our Lyα forest sample, we can generate new smoothed forests by randomly drawing the $\delta_F(z)$ values of each redshift bin from all observed values of said bin. While any information on scales smaller than the smoothing scale $\Delta\chi$ is lost, the scale of the g-band convolution is one order of magnitude larger than χ_h ($\sim$1 cGpc, see upper x-axis on Fig. 3), so the effect of this forest smoothing is close to negligible for our study (see Appendix A for further proof of the validity of this assumption).

With this method, we can generate an indefinite amount of smoothed forests covering the whole g-band redshift range just by resampling from observed forest spectra, provided that they have been previously smoothed in comoving bins of size $\Delta\chi_h = \chi_h$. This procedure allows us to integrate these generated forests using equation (7) (with $z_{min} = 0$ and $z_{max} = \infty$) to draw likely values of $\delta_F^{g\,full}$ based on real data, despite $\delta_F^{g\,full}$ not being directly observable. For a large enough sample of generated smoothed forests, the empirical distribution function (EDF) of $\delta_F^{g\,full}$ will be akin to $P(\delta_F^{g\,full})$.

Observed forest data can then be inserted into the generated smoothed forest sample by replacing the randomly drawn value of δ_F by the real observed δ_F, in the redshift range of the observed forest (dependent on z_{QSO}, see equations 8 and 9). Fig. 5 shows the distribution of $\delta_F^{g\,obs}$ for four generated smoothed forests with the insertion of the same observed forest at $z_{QSO} = 3$. We note that the difference in spectral resolution/comoving bin size of the generated and inserted forest in Fig. 5 does not affect significantly the resulting $\delta_F^{g\,full}$ value (see Appendix A for details).

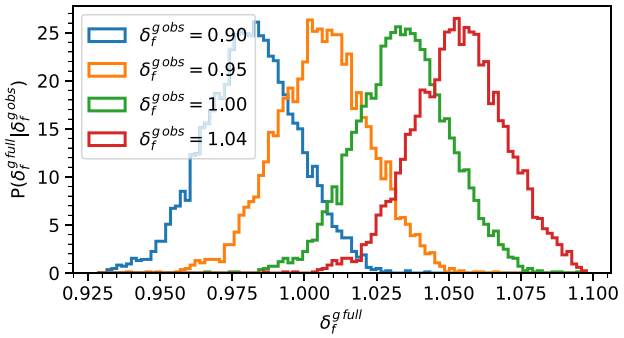

Figure 6. The EDF $P(\delta_F^{g\,full}|\delta_F^{g\,obs})$ for four different QSOs of different $\delta_F^{g\,obs}$, all of them at approximately $z_{QSO} \sim 2.7$. All data extracted from our full simulation of DESI Lyα forest data Section 4.3.

If we compute the EDF of $\delta_F^{g\,full}$ for the generated sample of smoothed forests, all of them with an insertion of a given observed forest, we obtain the distribution of $\delta_F^{g\,full}$ having observed a given $\delta_F^{g\,obs}$. In a Bayesian sense, when we insert the observed forest data we update a prior that contains no information of a specific QSO ($P(\delta_F^{g\,full})$) to a posterior with the observed information of a specific QSO ($\delta_F^{g\,obs}$), which yields $P(\delta_F^{g\,full}|\delta_F^{g\,obs})$, i.e. the probability distribution of $\delta_F^{g\,full}$ updated with the information observed for a given QSO ($\delta_F^{g\,obs}$). Fig. 6 displays an example of four different $P(\delta_F^{g\,full}|\delta_F^{g\,obs})$ for four QSOs at $z_{QSO} \sim 2.7$ with different $\delta_F^{g\,obs}$ values. We can see that the EDFs shift towards higher values of $\delta_F^{g\,full}$ when $\delta_F^{g\,obs}$ increases.

3.3 Correlation estimator: Lyα average fluctuation

For each Lyα QSO in DESI, we have defined two observable quantities. First, the g-band patch flux surrounding the QSO, f^g, which contains the information of the LSS Lyα emission field. Second, the probability distribution of δ_F convolved through the whole g band, having a given observed Lyα forest: $P(\delta_F^{g\,full}|\delta_F^{g\,obs})$. This is a Bayesian estimation of the LSS Lyα absorption field based on the redshift range we can observe. Both these quantities must be correlated, as these Lyα emission and absorption fields trace the same LSS. Now we want to define a correlation estimator with maximal SNR, which also maximizes our possibilities of detecting the Lyα emission in the g-band images, despite the very low SNR expected from the diffuse Lyα emission.

First, let us define a $\delta_{\rm F}^{g\,\rm full}$ bin j, comprising all values $\delta_{{\rm F}\,j\,\rm min}^{g\,\rm full} < \delta_{\rm F}^{g\,\rm full} < \delta_{{\rm F}\,j\,\rm max}^{g\,\rm full}$. If we consider a QSO i, with a patch flux f_i^g, and an associated forest observed value $\delta_{{\rm F}\,i}^{g\,\rm obs}$, we can express the probability of such patch i having a $\delta_{{\rm F}\,i}^{g\,\rm full}$ value inside the bin j as:

$$w_{\rm ij} = \int_{\delta_{{\rm F}\,j\,\rm min}^{g\,\rm full}}^{\delta_{{\rm F}\,j\,\rm max}^{g\,\rm full}} P\left(\delta_{\rm F}^{g\,\rm full} | \delta_{{\rm F}\,i}^{g\,\rm obs}\right) {\rm d}\delta_{\rm F}^{g\,\rm full}. \tag{10}$$

Since this probability w_{ij} is a degree of belief, we can use it as a weight in the determination of the average patch flux in the forest bin j, $\langle f^g \rangle_j$. In this way, for each ith patch w_{ij} will be proportional to the probability of $\delta_{{\rm F}\,i}^{g,\rm full}$ being within the j-th bin. The average patch flux for a given $\delta_{\rm F}^{g\,\rm full}$ bin j will be

$$\langle f^g \rangle_j = \langle f^g \rangle \left(\delta_{{\rm F}\,j\,\rm min}^{g\,\rm full}, \delta_{{\rm F}\,j\,\rm max}^{g\,\rm full}\right) = \frac{\sum_i w_{ij} f_i^g}{\sum_i w_{ij}}. \tag{11}$$

This average patch flux $\langle f^g \rangle_j$ can be understood as a Bayesian stacking of observed flux f_g in bins j of the Lyα forest absorption convolved with the entire g band ($\delta_{\rm F}^{g\,\rm full}$).

If the g-band images contain unresolved Lyα emission tracing LSS (and this is correlated to the Lyα forest), the average patch fluxes $\langle f^g \rangle_j$ must evolve for different $\delta_{\rm F}^{g\,\rm full}$ bins j. In other words, integrating the weights in equation (10) for different ranges of $\delta_{\rm F}^{g\,\rm full}$ should yield different values of $\langle f^g \rangle$. If this evolution of $\langle f^g \rangle_j$ versus $\delta_{{\rm F}\,j}^{g\,\rm full}$ is statistically significant and physically meaningful (e.g. pointing towards more overdense regions having brighter Lyα emission), then we will have detected LSS in Lyα emission. Such a detection would not be a cross-correlation in the strict sense usually applied in cosmology (i.e. the 2PCF computed using two different tracers), but it will still be a cross-correlation in the broadest sense (i.e. an estimator that proves that a correlation exists between two different data sets).

We will not assume a given functional form to fit to the $\langle f^g \rangle_j$ versus $\delta_j^{g\,\rm full}$ relation. Instead, we will start from the most agnostic assumption possible: on average, overdense regions will have brighter integrated Lyα emission than underdense regions. Therefore, our correlation estimator will be the difference in average Lyα patch flux between underdense and overdense regions, i.e.,

$$\Delta\langle L_{\rm Ly\alpha} \rangle \equiv \langle f^g \rangle_{\rm over} - \langle f^g \rangle_{\rm under} = \langle f^g \rangle(0, 1) - \langle f^g \rangle(1, \infty), \tag{12}$$

where $\langle f^g \rangle_{\rm over}$ is the average patch flux for overdense regions (for $\delta^{g\,\rm full} < 1$, i.e. regions where the transmitted flux fraction in the Lyα forest is smaller than average), and $\langle f^{\rm g} \rangle_{\rm under}$ is its counterpart for underdense regions. Hence, we expect $\Delta\langle L_{\rm Ly\alpha} \rangle > 0$, and the significance of $\Delta\langle L_{\rm Ly\alpha} \rangle > 0$ will be the significance of the detection of LSS in Lyα emission. The behaviour of the estimator versus properties of the Lyα emission field, as well as its reasonable values, will be explored and discussed in Sections 5.1 and 6.2.2. Since the estimator in equation (12) does not make any assumption of underlying physics (other than overdense regions being brighter than underdense), we expect it to effectively detect LSS in Lyα regardless of the true nature of the Lyα emission field, as long as SNR is high enough. However, given the large redshift convolution of the estimator (Fig. 1), we do not expect it to be able to detect very over/underdense local features in specific regions, such as protoclusters or large voids.

The error of $\Delta\langle L_{\rm Ly\alpha} \rangle$ in equation (12) is not trivial to propagate, since the probability distributions of $\langle f^g \rangle_{\rm over}$ and $\langle f^g \rangle_{\rm under}$ are not independent. While both are expected to have Gaussian distributions simply by being the weighted average of a large number of independent patch fluxes, the weights in equation (11) for $\langle f^g \rangle_{\rm over}$ and $\langle f^g \rangle_{\rm under}$ are correlated. For any QSO, the sum of the two weights used for $\langle f^g \rangle_{\rm over}$ and $\langle f^g \rangle_{\rm under}$ must be 1, as that sum would be the full integral of $P(\delta_{\rm F}^{g\,\rm full} | \delta_{{\rm F}\,i}^{g\,\rm obs})$ over the set of all real numbers $\mathbb{R}$ (equation 10). Therefore, we will compute the errors of $\Delta\langle L_{\rm Ly\alpha} \rangle$ via bootstrapping, resampling our data by randomly drawing QSOs subsamples from our simulation with replacement (see Section 5 for more details).

4 FORECAST SIMULATION

In this section, we will briefly explain how both the DECaLS/BASS images and the DESI Lyα forest spectra are simulated for our forecast. First, the Lyα emission and absorption field (i.e. our signal) is modelled by employing a hydrodynamic simulation specifically designed for Lyα forest studies. This simulation was performed with the P-GADGET code (Springel 2005; Di Matteo et al. 2012), using the cosmology specified in Introduction, for a comoving box size of 400 cMpc $\rm h^{-1}$ (571.4 cMpc). This is a volume larger than the current publicly available, state-of-the-art hydrodynamic simulations (e.g. IllustrisTNG, Nelson et al. 2019, with a maximum comoving size of 300 Mpc), so some approximations had to be made to reduce computation time.

The simulation contains 2×4096^2 particles, with masses of 1.19×10^7 and $5.92 \times 10^7\,\rm M_\odot\,h^{-1}$ for baryons and dark matter respectively, calculated with a gravitational softening length of 3.25 kpc $\rm h^{-1}$. A density threshold for star formation lower than usual (e.g. Springel & Hernquist 2003; Pillepich et al. 2018) was adopted (just 1000 times the average gas density), to allow for gas particles to be more easily turned into collisionless star particles and hence reduce the computational cost of following the gas hydrodynamics for longer periods. Moreover, black hole formation and stellar feedback were not applied. All these approximations simplify calculations and reduce computation time, but result in inaccurate stellar properties for galaxies. While active galactic nucleus (AGN) feedback (Bird et al. 2023) or star formation (Sorini, Davé & Anglés-Alcázar 2020) may have a limited effect on the the IGM, and thus its associated Lyα forest, we consider these effects subdominant and thus do not model them in our study, as in Viel, Haehnelt & Springel (2004).

This simulation box was used in Cisewski et al. (2014) and Ozbek et al. (2016) in order to test different approaches to interpolate the three-dimensional IGM distribution based on Lyα forest observations. The same simulation was again used by Croft et al. (2018) to place constraints on Lyα emission models by comparing it to observational data, and by Renard et al. (2021) to perform a Lyα IM forecast with narrow-band photometry over a smaller footprint. A voxel plot of the simulation snapshot at $z = 3$ (the one used for this work) is displayed in Fig. 7, for both the Lyα emission (with intermediate parameters of our Lyα emission model, see Section 4.1) and Lyα absorption fields. For our analysis, this snapshot of the simulation box is divided into 256 equal bins per side, which results in cubic voxels of 1.56 cMpc $\rm h^{-1}$ (2.23 cMpc) per side (613 pkpc at $\langle z \rangle = 2.64$).

4.1 Lyα emission model

The Lyα forest is a well-understood and extensively studied cosmological probe, and thus, its simulation has already been addressed in a multitude of works (e.g. Peirani et al. 2014; Ramírez-Pérez et al. 2022). In this work, the optical depth of the Lyα absorption field was computed along sightlines of the hydrodynamic simulation box, following Hernquist et al. (1996).

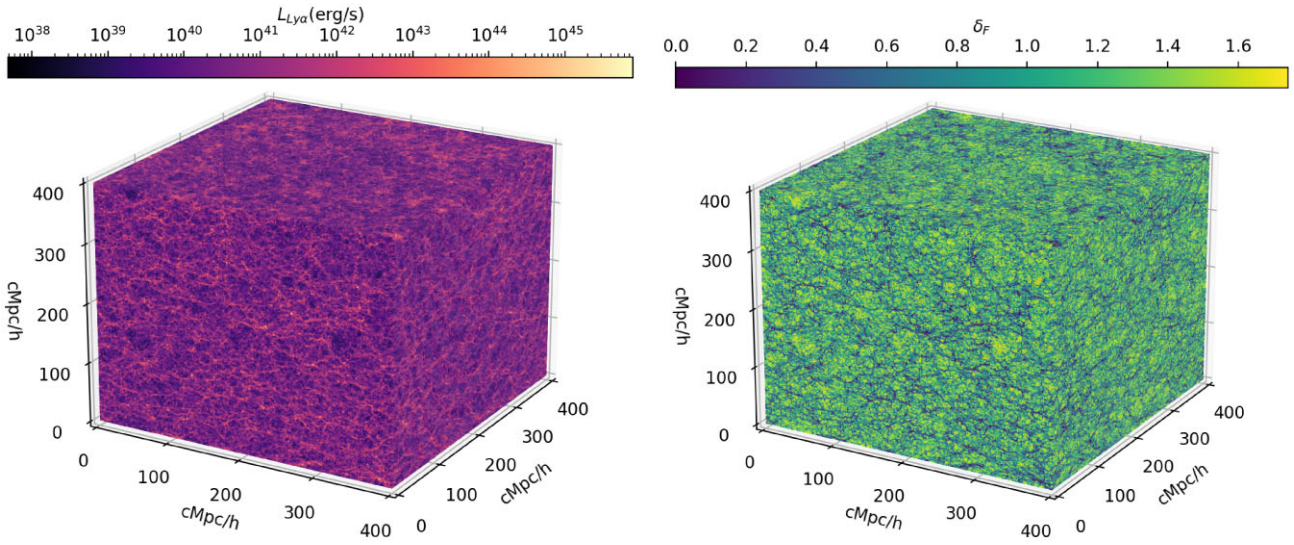

Figure 7. Left panel: simulation box for the Lyα emission field, with the parameters of an intermediate case [$\langle L_{Ly\alpha}\rangle = 7.5 \cdot 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$, $b_e = 2.5$]. Right panel: simulation box for the Lyα absorption field.

In contrast, the Lyα emission is far less constrained, lacking straightforward and standardized simulation methods apart from computationally expensive radiative transfer implementations (e.g. Cantalupo et al. 2005; Byrohl & Nelson 2023). Given the modelling constraints and the observational limitations, we have opted for a simple analytical model that determines the Lyα luminosity at any given simulation voxel as a function of the baryonic density. The model has two free parameters which we will explore in our forecast: $\langle L_{Ly\alpha}\rangle$, the average Ly$\alpha$ luminosity per unit of volume [in erg s^{-1} (cMpc^3)$^{-1}$], and b_e, the linear bias of the power spectrum of Lyα emission. In detail, below a given luminosity threshold $L_{Ly\alpha\,cut} \sim 10^{45}$ erg s^{-1}, the Lyα luminosity follows the power law

$$L_{Ly\alpha}(\delta) = C_L \cdot (1+\delta)^{b_e}, \tag{13}$$

where δ is the dark matter overdensity field, b_e the bias, and C_L a normalization constant computed numerically so that the average luminosity per unit volume in our simulation box matches the input value, $\langle L_{Ly\alpha}\rangle$. While this simple model successfully reproduces any desired average luminosity and bias in the linear regime, at $\langle L_{Ly\alpha}\rangle > 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$ the voxels with highest δ may result in excessively bright Lyα luminosities (a problem already discussed in Renard et al. 2021). It is also worth noting that no physics are assumed in this Lyα emission model: just a normalization constant derived from $\langle L_{Ly\alpha}\rangle$, and a linear bias with respect to the dark matter LSS (Bernardeau et al. 2002). More sophisticated alternatives implementing the physics behind the generation and scattering of Lyα emission will be discussed in Section 6.2.2.

To explore the wide range of average Lyα luminosities and biases described in the literature, we impose an ad hoc power-law cut-off for $L_{Ly\alpha} > L_{Ly\alpha\,cut}$. This observationally motivated cut-off ensures that the brightest voxels have $L_{Ly\alpha}$ values within approximately 1 dex of the observed luminosity functions (LFs) for bright QSOs (e.g. Spinoso et al. 2020; Liu et al. 2022; Torralba-Torregrosa et al. 2023). Additionally, it sets the maximum total $L_{Ly\alpha}$ in a voxel to around $10^{45.5}$ erg s^{-1}, consistent with the integrated luminosity of the largest observed Lyα nebulae (Cantalupo et al. 2014; Cai et al. 2017). The details of this bright-end cut-off and the whole Lyα emission model are thoroughly discussed in Appendix B.

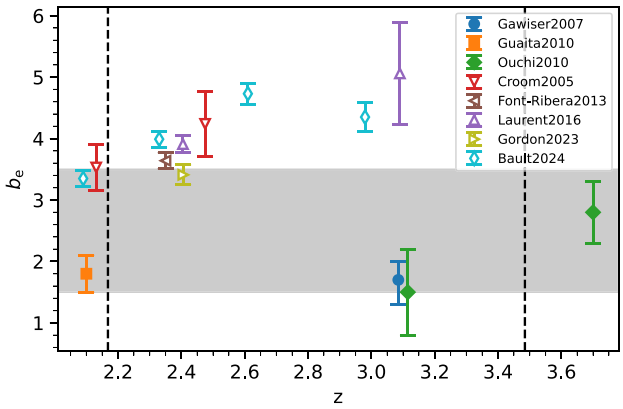

Figure 8. Bias values found in the literature for QSOs (empty markers) and LAEs (solid markers). The grey interval represents the range considered in our model. Dashed vertical line are the sensitivity limits of the g band to Lyα (5 per cent of maximum sensitivity).

Our model encompasses a range of bias values from $1.5 \leq b_e \leq 3.5$. For values outside this range, the model's inner parameters diverge (see Appendix B). Within this bias range, the model can adopt any value of $\langle L_{Ly\alpha}\rangle$. However, we consider realistic values of $\langle L_{Ly\alpha}\rangle$ to lie within the interval 1.5×10^{40} erg s^{-1} (cMpc^3)$^{-1}$ $\leq \langle L_{Ly\alpha}\rangle \leq 1.5 \times 10^{41}$ erg s^{-1} (cMpc^3)$^{-1}$, based on current literature estimates (e.g. Ouchi et al. 2008; Drake et al. 2017; Lin et al. 2022). We will explore larger values of $\langle L_{Ly\alpha}\rangle$ in Section 5 [up to 5×10^{41} erg s^{-1} (cMpc^3)$^{-1}$], but we consider these unlikely to represent the average luminosity of the real Lyα emission field. We also caution that in our Lyα emission model, such high $\langle L_{Ly\alpha}\rangle$ values may yield non-physically high luminosities in the brightest Lyα voxels, even if only by a factor of a few. However, we do not consider this an issue since it only happens for $\langle L_{Ly\alpha}\rangle$. values we already deem unphysical. These ranges of b_e and $\langle L_{Ly\alpha}\rangle$ are supported by the literature, as shown in Figs 8 and 9.

Fig. 8 displays different values of bias found in the literature for LAEs (Gawiser et al. 2007; Guaita et al. 2010; Ouchi et al. 2010) and QSOs (Croom et al. 2005; Font-Ribera et al. 2013; Laurent et al.

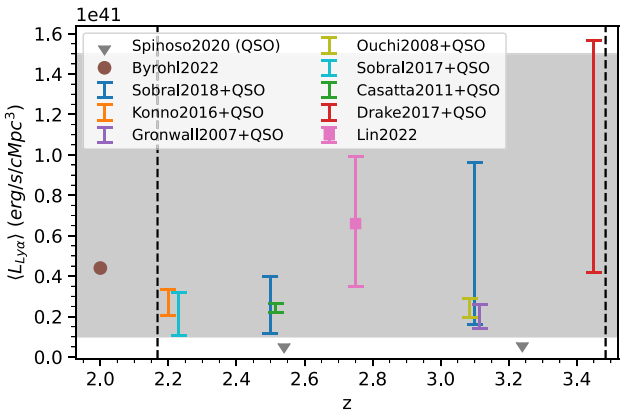

Figure 9. $\langle L_{\mathrm{Ly}\alpha} \rangle$ values found in the literature. The grey triangles are the lower limits obtained by integrating the bright QSO LFs from Spinoso et al. (2020), recomputed after applying the $g < 19$ foreground cut from Section 4.2.1. All luminosity intervals without central point correspond to integrals of LAE LFs + Spinoso et al. (2020) QSO LFs. The brown point (Byrohl & Nelson 2023) is derived from simulations, and the pink point (Lin et al. 2022) from QSO-galaxy spectra cross-correlation. The grey interval represents the interval considered as realistic in our model. Dashed vertical line are the sensitivity limits of the g band to Lyα (5 per cent of maximum sensitivity).

Table 1. Percentage of Lyα QSOs remaining in the Spinoso et al. (2020) sample for different g-band magnitude cuts, together with σ, the standard deviation of the observed pixel flux in our BASS g-band intensity map (Fig. 10), after foreground subtraction with the same g-band magnitude cuts.

Foreground cut	$g < 19$	$g < 20$	$g < 21$
QSO fraction (per cent)	93.353	49.245	3.001
σ (10^{-16} erg s^{-1} cm^{-2} Å^{-1})	1.606	1.559	1.540

2016; Gordon et al. 2023; Bault et al. 2024), in the the approximate redshift range of the g band for Lyα. The upper limit of the bias parameter in our model is close to the lowest value of the QSO bias measured to date (Croom et al. 2005, at $z \sim 2.1$). On the other hand, the lower limit corresponds to the lowest LAE bias (Ouchi et al. 2010, at $z \sim 3.1$). Since the g-band IM comprises all Lyα emission, from unresolved sources to all QSOs (except the ones removed by our foreground cut, see Table 1), it is reasonable to assume that the bias b_e of the whole Lyα emission field lies in-between that of QSOs and that of LAEs, as show with the grey area in Fig. 8. Even if unresolved LAEs or diffuse IGM emission were significant components of the total Lyα emission field, according to Byrohl & Nelson (2023), the smallest dark matter haloes with significant LAEs have $M_{\mathrm{halo}} = 10^9\,\mathrm{M}_\odot$. Such haloes have a bias $b_e \sim 1$ at $z = 3$ (see e.g. fig. 7 of Ouchi et al. 2018). Given this absolute lower boundary to Lyα emission bias, and given that bright QSOs may reach $b_e \sim 5$ and represent approximately from 3 per cent to 50 per cent of total Lyα emission in our model (Fig. 9), it is highly unlikely that the true bias of the integrated Lyα emission field is significantly lower than 1.5.

Fig. 9 shows the range of $\langle L_{\mathrm{Ly}\alpha} \rangle$ of our study, together with several values from the literature. First, the grey triangles are the $\langle L_{\mathrm{Ly}\alpha} \rangle$ obtained by integrating the QSO LFs from Spinoso et al. (2020, recomputed after applying the foreground cut $g \leq 19$ to their sample, as specified in Section 4.2.1). All the luminosity intervals without central points (Gronwall et al. 2007; Ouchi et al. 2008; Cassata et al. 2011; Drake et al. 2017; Sobral et al. 2017, 2018) correspond to integrated LAE LFs. For all cases, the lower limit is the respective LF integrated to the observational limit specified in each work, plus the bright QSO LF integral from Spinoso et al. (2020, as none of these works sample sources as bright as Spinoso et al. 2020). The upper limit is the LF integrated to the absolute lower limit $L_{\mathrm{Ly}\alpha} = 10^{37}$ erg s^{-1} proposed in Bacon et al. (2021) and Herrero-Alonso et al. (2023), plus the contribution from the QSO LF integral. The remaining points are derived from a theoretical model of Lyα emission applied to the IllustrisTNG simulation (Byrohl & Nelson 2023), and a cross-correlation study in eBOSS of residuals of galaxy spectra with QSOs (Lin et al. 2022).

It is worth commenting briefly on the highest estimate of $\langle L_{\mathrm{Ly}\alpha} \rangle$ in Fig. 9, namely: the upper limit presented by Drake et al. (2017). According to Herenz et al. (2019), most studies in the literature underestimate the faint end of the Lyα LF by a factor of 2.5 due to correcting for completeness by treating LAEs as point sources rather than extended emitters. Notably, Drake et al. (2017) is the sole study in Fig. 9 that accounts for LAEs as extended sources, thereby avoiding the underestimation of LAE density at the faint end. Consequently, even though the upper limit reported by Drake et al. (2017) is 50 per cent higher than other estimates in Fig. 9, recent literature supports its credibility. This justifies the upper limit of our $\langle L_{\mathrm{Ly}\alpha} \rangle$ range.

4.2 Image simulation: QSO patches

The integrated g-band flux of the patches around each QSO has two uncorrelated components: the Lyα signal and the noise. Here, we explain how both are simulated and combined in our forecast. The steps we follows are: (i) the construction of a noise intensity map using real BASS g-band images, (ii) the generation of a Lyα intensity map from our simulation, and (iii) the final combination of these two elements, plus the simulation of all relevant systematics.

4.2.1 Photometric noise: BASS intensity map

Any signal in the g-band images that does not correspond to Lyα emission of cosmic origin is noise for the purposes of our study. This noise encompasses various sources such as cosmic background/foregrounds, zodiacal light, airglow, moonlight, and thermal noise. In Renard et al. (2021), cosmic foregrounds and all other noise sources were modelled as two distinct components (a mock galaxy catalogue and noise extrapolations from real PAUS images, respectively). However, here we do model all noise sources with one single element. In order to do so, we employ real BASS g-band images to model both cosmic foregrounds and other noise sources, following the same image reduction strategy we propose for an eventual observational study with full imaging surveys. This procedure ensures that cosmic foregrounds are real and not simulated, as they are already contained in the BASS images, and also allows us to develop and test an image reduction strategy.

It is of utmost importance that the diffuse Lyα signal is preserved throughout the entire data reduction of the g-band images. This imposes a hard limit on how aggressive can the foreground and noise removal procedures be when applied to these images. Although in this forecast the Lyα signal is simulated, and we only use the real g-band images as noise, we still make sure that the real Lyα emission contained in them is preserved. To do so, we take all BASS reduced g-band science images in an arbitrary patch of 20 deg^2 centred on RA $= 235°$, Dec. $= 45°$. Since the standard BASS pipeline only performs background subtraction when computing the photometric

catalogue (Zou et al. 2017b), these images do still contain the diffuse Lyα emission, and can be used as provided for our forecast (and an eventual observational study). However, if any sort of background subtraction by performing a fit to the images themselves was part of the standard image reduction, a custom reduction without it would be required, as the unresolved Lyα emission that constitutes the bulk of our signal would have been removed.

We apply additional reduction steps on the selected images as follows. First, we subtract the SKY value of each observation to all CCD images in the camera. For each observation, we take all CCD images (four for BASS), compute the mean of their median SKY values, and subtract it from the four CCD images. This procedure greatly homogenizes the g-band intensity map, mitigating the variability due to atmospheric and moon conditions, which is one of the main sources of noise in our study. However, by subtracting the median SKY value at the camera level, we effectively erase any signal of structure in Lyα at scales larger than the FOV of the camera itself. 90Prime, the camera used by BASS, has an effective approximately square FOV of 1.08×1.03 deg^2 (Williams et al. 2004), which corresponds to 120 cMpc at $\langle z \rangle = 2.64$. This imposes a hard limit to the transversal scales we can sample with this study. For DECaLS, the DECam has a significantly larger FOV. Thus, by assuming the BASS specifications we are being again conservative in our forecast. If larger transversal scales needed to be sampled, more sophisticated methods of sky subtraction could be implemented. For instance, modelling sky emission without averaging out all observed flux over a given area (e.g. Kimeswenger et al. 2015; Han, Song & Zhao 2023).

Second, we crop the image borders (12 pixels) to avoid edge effects, and co-add all selected images in a low-resolution intensity map with SWARP (Bertin et al. 2002), with the same pixel size as our Lyαg-band simulated image (66.125×66.125 arcsec^2, see Section 4.2.2). We then deproject the resulting intensity map into a simple equal-area projection (Sanson-Flamsteed projection, see e.g. Calabretta & Greisen 2002). This projection is adequate for the 20 deg^2 used in this forecast. However, a study with the full DECaLS/BASS surveys may require either the computation of several small intensity maps, or the use of more sophisticated projections such as HEALPIX (Gorski et al. 2005).

Third, we apply extra reduction steps on the intensity map to remove noise sources. In particular, we employ σ-clipping on the image at the 3σ level; this mostly removes very bright stars, mimicking the effect of a survey mask (in a eventual observational study, the DECaLS/BASS survey masks could be applied instead). Then, we perform a foreground cut by removing all sources with $g < 19$. Since photometry has already been performed over the entire BASS survey, we sum the flux of all sources with $g < 19$ for each pixel of the intensity map and subtract it, instead of masking the foregrounds in each individual image. The $g < 19$ threshold has been chosen to remove the brightest sources, while keeping the great majority of bright QSOs from Spinoso et al. (2020, and thus, the great majority of DESI Lyα QSOs). Indeed, considering a pessimistic scenario, these bright QSOs may represent up to 50 per cent of the total Lyα luminosity (Fig. 9). In Table 1, we show the σ of the BASS g-band intensity map for different foreground cuts, together with the per cent of QSOs in Spinoso et al. (2020) after applying the same foreground cut. We checked that cuts more stringent than $g < 19$ would remove a significant fraction of the Lyα QSOs while decreasing the σ of the intensity map by less than 5 per cent.

Finally, we apply once more a σ-clipping at 3σ to the intensity map instead of masking very bright sources. Indeed, at this stage, the latter may be contiguous-pixels areas in the map image, so the effect of this last σ-clipping is to mask very bright isolated pixels (hence removing outliers and making the intensity map smoother) without significantly increasing the fraction of masked pixels.[2] The resulting intensity map, with the mean flux subtracted, is shown in Fig. 10.

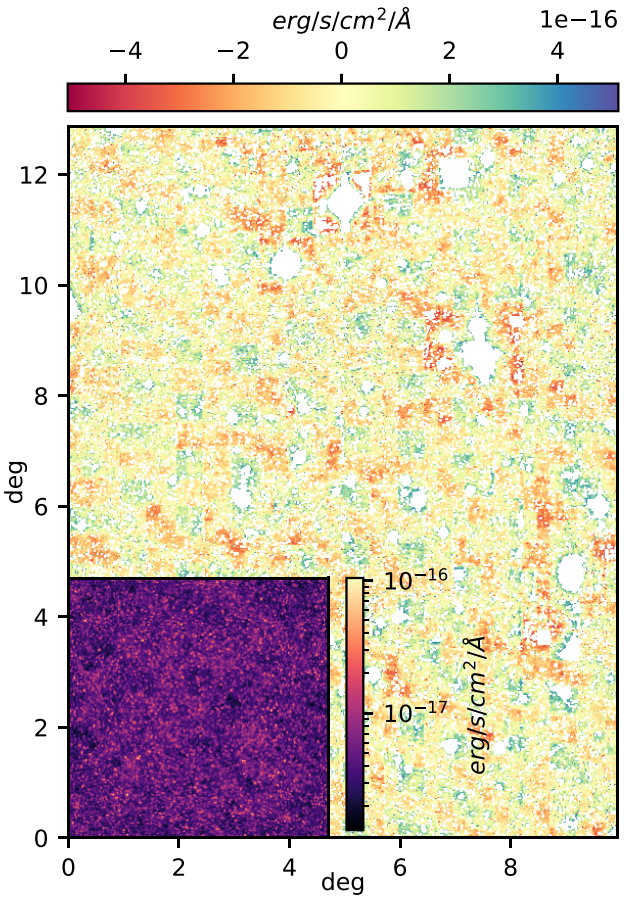

Figure 10. Intensity map of 20 deg^2 of BASS g-band images, overlaid with the simulated g-band Lyα emission intensity map, to scale (lower left corner). The Lyα emission model parameters were $\langle L_{\mathrm{Ly}\alpha} \rangle = 7.5 \times 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$ and $b_e = 2.5$. The white regions on the BASS intensity map represent masked areas.

It is worth noting that, despite the image reduction, a chequered pattern on the BASS intensity map is still clearly noticeable by visual inspection. This pattern is composed of squares of approximately 0.5×0.5 deg^2; roughly the size of the BASS camera CCDs (Williams et al. 2004).Thus, the primary source of noise for our forecast is not foregrounds of cosmic origin, but rather variations between CCDs and exposures. These variations likely arise from either the instrumental noise of the camera or changes in flux due to atmospheric or lunar sources (e.g. moonlight and airglow). We will further discuss the ramifications of this for our forecast and future Lyα IM studies in Section 6.

4.2.2 *Photometric signal: Lyα intensity map*

To simulate the Lyα emission intensity map, we apply the Lyα emission model described in Section 4.1 to our simulation box

[2]Only ∼0.3 per cent of pixels are masked, since it is a 3σ-clipping and large contiguous areas were already masked with the previous 3σ-clipping.

(Fig. 7). This allows us to obtain a Lyα emission simulation box from which we then compute a light-cone. For this, we replicate and rotate twice (by 90°) the resulting luminosity box in the redshift direction, to fully cover the Lyα redshift range associated to the g band. The simulation voxels are converted from transversal comoving to angular coordinates, interpolating to the grid at lowest redshift, i.e. the lowest angular resolution. This angular resolution limit constrains the pixel size of our forecast: 66.125×66.125 arcsec^2. Finally, the Lyα luminosity within each voxel is converted to observed flux following equation (3). The result is a light-cone of Lyα observed flux, which is convolved with the g band to obtain the Lyα intensity map displayed in the inset of Fig. 10.

4.2.3 *Photometric signal and noise combination*

The simulated Lyα intensity map plus the real g-band BASS intensity map need to be combined into a single data set simulating real g-band patches around DESI Lyα QSOs. There are more projected DESI Lyα QSOs (700 000) than pixels in our intensity maps (65 536 for the Lyα simulation and 379 241 for the BASS intensity map), so some repetition of Lyα emission/noise patches is unavoidable. To solve this issue, we separately select pixels from the Lyα emission and BASS noise intensity maps as QSOs, in a random and independent fashion.

For each DESI Lyα QSO, we randomly draw with replacement a pixel from the Lyα simulated map and the BASS real map. In other words, a pixel that has been drawn once can be drawn again; the probability of drawing a given pixel remains constant regardless of previous outcomes. We integrate the fluxes for a Lyα patch (from the Lyα map) and noise patch (from the BASS map) centred around the drawn pixels, and add them to simulate the g-band patch flux around a given QSO, containing both the Lyα emission signal and noise. By drawing random pairs of both intensity maps, we ensure that no artificial correlations arise from the periodicity of the hydrodynamic simulation combined with the periodic CCD-sized patterns in the BASS intensity map (Fig. 10). Moreover, the amount of possible combinations of Lyα patches + BASS noise patches largely outnumbers the expected number of DESI Lyα QSOs (by 4 orders of magnitude). This implies that the impact of the repetition of Lyα signal + noise combinations is negligible, even if we include all expected DESI Lyα QSOs in our study.

The most important systematics affecting the observed flux are modelled before combining the Lyα + noise patches. Zero-point calibration error is included in the Lyα patch fluxes by converting the fluxes to magnitudes, adding a random magnitude offset drawn from a Gaussian of $\mu = 0$ and $\sigma = 7$ mmag (Dey et al. 2019), and then converting back to fluxes. Since BASS noise patches are computed on real data, the calibration error is already present, so no extra offset is added. It is worth noting that calibration techniques with zero-points varying across a given CCD exist (e.g. J-PLUS/J-PAS, López-Sanjuan et al. 2019), instead of per-CCD calibration as in DECaLS/BASS (Dey et al. 2019). These calibration methods may significantly reduce our calibration error, since when computing patch fluxes we integrate relatively large fractions of a given CCD.

Fluxes in the noise patches are integrated by summing the flux in all pixels inside each patch, barring masked pixels (according to the mask computed in Section 4.2.1 and displayed as white regions in Fig. 10). For the Lyα patches, however, we do not replicate the same mask, as this would require generating a single Lyα cutout with the respective mask for every QSO patch we need to simulate, dramatically increasing computation time. Instead, we first pre-compute all possible Lyα patches of the desired radius in our simulated Lyα image, and determine the flux percentiles for each possible patch, $P_{\mathrm{Ly}\alpha}^{\mathrm{per\,cent}}$. Then, we simulate the error stemming from masking by shifting the flux percentile as follows:

$$P_{\mathrm{Ly}\alpha\ \mathrm{masked}}^{\mathrm{per\,cent}} = P_{\mathrm{Ly}\alpha}^{\mathrm{per\,cent}} + U(-\mathrm{per\,cent}_{\mathrm{masked}}/2\ \ \mathrm{per\,cent}_{\mathrm{masked}}/2), \quad (14)$$

where U is the uniform distribution, and $\mathrm{per\,cent}_{\mathrm{masked}}$ is the percentage of area masked in the respective BASS noise patch. For example, a Lyα patch with a flux in the 50th percentile, that is superimposed with a BASS noise patch that has 10 per cent of its area masked, will have a Lyα flux after making between the 45th and 55th percentile (if the result was to be below the 0th or above the 100th percentile, it is capped at 0 or 100, respectively). This method results in a flux scatter similar to what would be obtained by the Lyα image with the BASS mask, at a negligible computational cost. After the zero-point calibration error and the mask offset following equation (14) is added to the Lyα patch fluxes, and the BASS noise patch fluxes have been integrated with their actual mask, we sum the BASS noise patch fluxes to the Lyα patch fluxes. The result is a set of simulated g-band patches for each DESI QSO, each one of them with a (very likely) unique Lyα signal + noise combination.

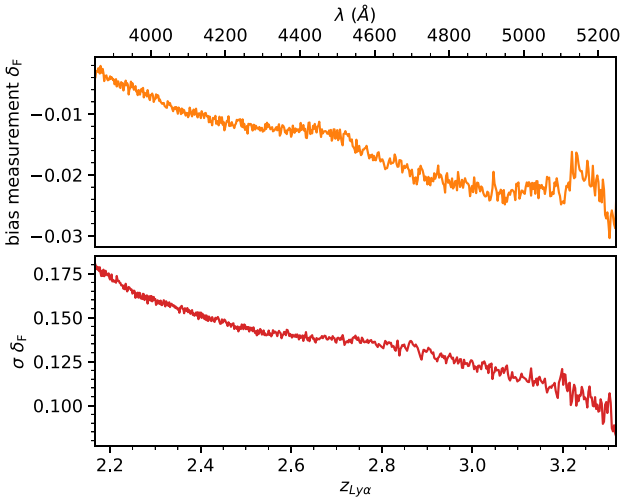

Figure 11. Upper panel: bias resulting from applying the QSO continuum-subtraction algorithm QFA (Sun et al. 2023) to the Lyα forest mocks of Bautista et al. (2015). Lower panel: average observational error of the Bautista et al. (2015) Lyα forest mocks.

4.3 Spectra simulation: Lyα forest probabilities

To simulate the Lyα forest data, first the Lyα absorption simulation box (Fig. 7) is replicated and rotated in the redshift direction, as well as interpolated to angular coordinates, following the same procedure carried out for the Lyα emission box (Section 4.2.2). We add observational noise to this Lyα forest light-cone and continuum-subtraction bias as a function of redshift/observed wavelength. The observational noise is extracted from the SDSS BOSS mocks in Bautista et al. (2015), and the bias is derived from the continuum-subtraction algorithm presented in Sun, Ting & Cai (2023). Fig. 11 displays both the bias and error as a function of Lyα redshift/observed wavelength.

After adding noise and bias, we duplicate the light-cone, and smooth the duplicate in redshift bins of width 150 cMpc. This duplicate, smoothed version of the light-cone will be used to

randomly generate smoothed forests to fill the redshift range not covered by the observed forests, as described in Section 3.2.

We simulate the observed forests from the unsmoothed light-cone by randomly drawing a QSO redshift for each angular bin (i.e. each possible sightline), following the DESI QSO redshift distribution (Fig. 1). Indeed, the latter sets the observed redshift/wavelength range in each light-cone sightline. DLAs (Wolfe et al. 2005) are added to the observed forests approximately following the redshift distribution and width of the DLAs in the Bautista et al. (2015) mocks. Since Lyα absorption is saturated in DLAs, the underlying H I density field is significantly more difficult to measure, thus we mask them. However, we fill the masked DLA regions with the average Lyα forest absorption $\delta_{\rm F}$ at its redshift. This is done because in the convolution of $\delta_{\rm F}$ with the g band ($\delta_{\rm F}^{\rm g}$, as expressed in equation 7), not all the redshift pixels contribute equally (because of the $D_L(z)^{-2}$ factor). Therefore, omitting lower redshift regions from integration results in lower values of $\delta_{\rm F}^{\rm g}$, introducing an undesired bias in our sample.

After computing a simulation of full smoothed forests and another of unsmoothed observed forests, we follow the procedure outlined in Section 3.2 to obtain, for each sightline in our simulation, the posterior distribution of full convolved forests given an observed forest, $P(\delta_{\rm F}^{g\,\rm full}|\delta_{\rm F}^{g\,\rm obs})$. To do so, for each observed light-cone sightline, we generate 5000 smoothed forests over the whole g-band redshift range by drawing with replacement values for each redshift bin from all the smoothed light-cone sightlines. We inject the observed forests into the 5000 generated smoothed forests, convolve all of them with equation (7), and compute their EDF, which is analogous to $P(\delta_{\rm F}^{g\,\rm full}|\delta_{\rm F}^{g\,\rm obs})$.

Therefore, every pixel in our simulated Lyα intensity map (Fig. 10) has an associated posterior of the full Lyα forest convolution, $P(\delta_{\rm F}^{g\,\rm full}|\delta_{\rm F}^{g\,\rm obs})$. When randomly selecting a given pixel in the Lyα intensity map as a DESI Lyα QSO to compute the g-band patch around it, its correspondent $P(\delta_{\rm F}^{g\,\rm full}|\delta_{\rm F}^{g\,\rm obs})$ is also selected.

5 FORECAST RESULTS

5.1 Estimator behaviour and SNR optimization

Before simulating the actual forecast for DESI–DECaLS/BASS, we need to evaluate the behaviour of the $\Delta\langle L_{\rm Ly\alpha}\rangle$ estimator and, more importantly, its error. We acknowledge that we have made the conservative choice of using BASS images to model our noise intensity map (Section 4.2.1); however, we compute our forecast for the combined DECaLS/BASS footprint (i.e. the entirety of DESI). Therefore, we will refer to the cross-correlation simulated in this work as DESI–DECaLS/BASS. The expected changes we may expect in our results by using actual DECaLS images to simulate its respective fraction of DESI are discussed in Section 6.1; however these anticipated changes do not modify our overarching conclusions.

Since we expect the SNR to be low, and a detection of Lyα LSS to be uncertain, it is crucial that the errors provided for $\Delta\langle L_{\rm Ly\alpha}\rangle$ are properly justified, and the observational parameters are chosen to maximize the SNR. Unless specified otherwise, in this subsection we will use a 'fiducial' Lyα emission model with $\langle L_{\rm Ly\alpha}\rangle = 7.5 \times 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$ and $b_{\rm e} = 2.5$.

We will assume that all results can be extrapolated to any other combination of $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ of our model. Since $\langle L_{\rm Ly\alpha}\rangle$ is just a multiplicative constant of our model (equation 13), we have no reason to believe the optimal parameters yielding the highest SNR will be sensitive to $\langle L_{\rm Ly\alpha}\rangle$. Moreover, since our estimator $\Delta\langle L_{\rm Ly\alpha}\rangle$ evolves symmetrically with respect to $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ (Fig. 16), we assume the same invariance with respect to $b_{\rm e}$.

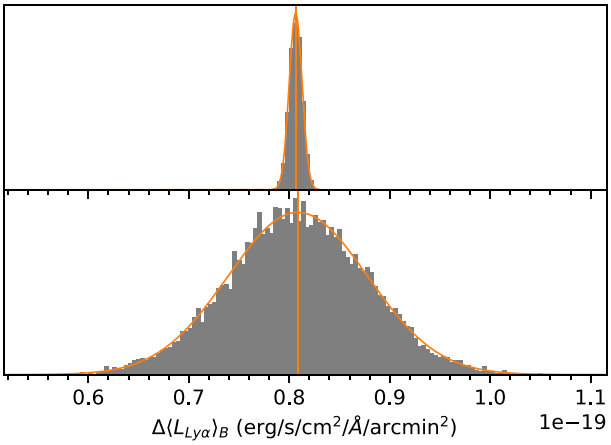

Figure 12. Histograms of $\Delta\langle L_{\rm Ly\alpha}\rangle_B$ for 10 000 bootstrap resamplings of our forecast (grey histogram), together with a Gaussian distribution with the mean and σ of the bootstrap resampling (orange lines). Upper panel for noiseless forecast, lower panel for forecast with 5 per cent of expected noise.

5.1.1 *Error estimation*

As specified in Section 3.3, we will compute the errors via bootstrapping. Having a given number of QSOs, $N_{\rm QSO}$, we randomly draw with replacement $N_{\rm QSO}$ QSOs N_B times, creating N_B bootstrap resamplings. For each one of these resamples, we compute $\Delta\langle L_{\rm Ly\alpha}\rangle_B$, the value of the estimator in a given bootstrap resampling. The value and error of $\Delta\langle L_{\rm Ly\alpha}\rangle$ will be the average and standard deviation of all $\Delta\langle L_{\rm Ly\alpha}\rangle_B$, respectively.

Fig. 12 shows the distribution of $\Delta\langle L_{\rm Ly\alpha}\rangle_B$ for 10 000 bootstrap resamplings, for a noiseless case and for a case with 5 per cent of the expected BASS noise. In detail, for the former, the BASS noise patches were not combined with the Lyα patches from the Lyα image (i.e. the BASS intensity map was not used in the forecast at all); for the latter, we added the BASS noise patches, after multiplying their fluxes by 0.05. Additional systematics such as masking and flux calibration errors were added only in the 5 per cent noise case. For both cases, the histogram of $\Delta\langle L_{\rm Ly\alpha}\rangle_B$ values closely follows a Gaussian distribution; adding the 5 per cent BASS noise and systematics widens the Gaussian and increases the error, but it does not significantly alter the average value of $\Delta\langle L_{\rm Ly\alpha}\rangle_B$ (approximately at 8.16×10^{-20} erg s^{-1} cm^{-2} Å^{-1} arcmin^{-2}). Hence, Fig. 12 shows that the error estimated via bootstrapping is well behaved, and the introduction of noise and systematics from a real imaging survey like BASS does not bias the results.

Beyond the intrinsic cosmic variance in the hydrodynamic simulation and the noise/systematics broadening the $\Delta\langle L_{\rm Ly\alpha}\rangle_B$ distribution in Fig. 12, we must consider an additional source of variability. Our forecast involves randomly selecting pixels from the Lyα simulated image as QSOs (patch centres) and pairing them with randomly chosen pixels from the BASS intensity map. In other words, we randomly pair each patch of integrated signal with a patch of integrated noise. Depending on how these signal-noise patches are paired, the value of $\Delta\langle L_{\rm Ly\alpha}\rangle$ may artificially be increased (e.g. if areas with high Lyα emission happen to be paired with exposures with a particularly bright sky) or decreased (e.g. if a large masked area caused by a bright star overlaps with a bright Lyα region). In the following, we will call a given set of randomly chosen signal-noise patches a realization of our forecast.

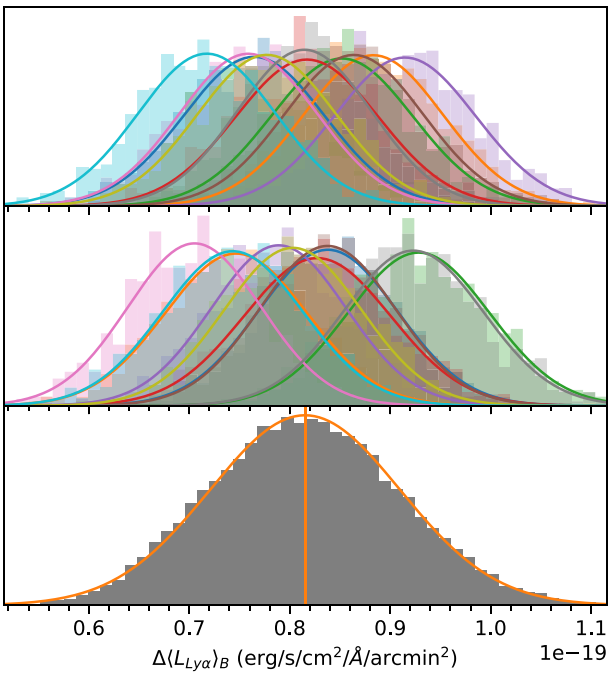

Figure 13. Upper and middle panels: histograms of $\Delta\langle L_{Ly\alpha}\rangle_B$ for 20 realizations of Lyα patches–noise patches, with 1000 bootstrap resamplings each, together with their respective derived Gaussian distributions. The realizations have been randomly divided in two panels, each one displaying 10. Lower panel: histogram and Gaussian distributions for all the $\Delta\langle L_{Ly\alpha}\rangle_B$ from the 20 realizations combined. The x-axis has the same scale as Fig. 12.

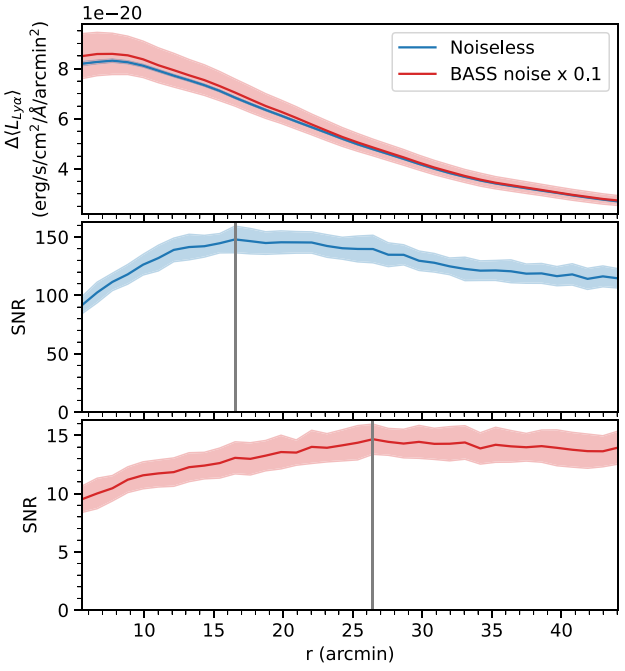

Figure 14. Upper panel: evolution of $\Delta\langle L_{Ly\alpha}\rangle$ versus patch radius, for a noiseless simulation (blue) and a forecast with 10 per cent of BASS noise (red). Middle panel: SNR versus patch radius, for the noiseless case; vertical line indicates the maximum SNR radius. Lower panel: SNR versus patch radius, for the 10 per cent noise case.

In Fig. 13, we show how $\Delta\langle L_{Ly\alpha}\rangle$ varies for 20 different realizations of our forecast; each one obtained with 1000 bootstrap resamplings and 5 per cent of BASS noise. The values of $\Delta\langle L_{Ly\alpha}\rangle$ (the mean of $\Delta\langle L_{Ly\alpha}\rangle_B$) vary between realizations, the error (the σ of $\Delta\langle L_{Ly\alpha}\rangle_B$) remains approximately constant. For example, in Fig. 13, the average error and its standard deviation for all 20 realizations is $6.89 \pm 0.15 \times 10^{-21}$ erg s^{-1} cm^{-2} Å^{-1} arcmin^{-2} (i.e. the σ of the error is $\sim$2 per cent of its average value).

If all the realizations are combined into a single one (i.e. combining the 20 realizations of 1000 bootstraps into a single one with 20 000), the result is a wider Gaussian distribution for $\Delta\langle L_{Ly\alpha}\rangle_B$ than the distributions for separate realizations (lower panel of Fig. 13). The standard deviation for this combined distribution is 9.48×10^{-21} erg s^{-1} cm^{-2} Å^{-1} arcmin^{-2}; an increase of almost 40 per cent compared to individual realizations. In fact, some of the Gaussian distributions corresponding to different realizations in Fig. 13 are separated by more than 1σ, indicating that this variation between realizations might be significant in some cases (larger than the error of $\Delta\langle L_{Ly\alpha}\rangle$).

This variability of results between realizations is not a limitation of our forecast, but an additional systematic that is also expected to arise when carrying out the Lyα IM study with real data. Indeed, just as in our forecast, the correlated noise (chequered pattern in Fig. 10) and masking over the full DECaLS/BASS g-band intensity map may artificially enhance or worsen the structure of the real Lyα emission field, resulting in a shift of the true value of $\Delta\langle L_{Ly\alpha}\rangle$. Moreover, in the real Universe there is only one realization of the sky positions of the QSOs targeted by DESI; hence the variability between realizations is not the error of the $\Delta\langle L_{Ly\alpha}\rangle$ measurement, but the variability of outcome for a measurement of $\Delta\langle L_{Ly\alpha}\rangle$ with real DESI–DECaLS/BASS data.

Consequently, the variability between forecast realizations imposes a hard limit on the precision with which we can determine $\Delta\langle L_{Ly\alpha}\rangle$, in a similar way cosmic variance limits any cosmological statistics (e.g. Marra et al. 2013; Ma et al. 2024). In fact, the intrinsic variability of $\Delta\langle L_{Ly\alpha}\rangle$ can be seen as the combined effect of cosmic variance, the limited space sampling of the Lyα forest, and the correlated noise in the g-band intensity map. Hence, all error intervals and regions shown hereafter, both for $\Delta\langle L_{Ly\alpha}\rangle$ and its SNR, correspond to their variability between forecast realizations.

5.1.2 *Optimal* $r_{\rm patch}$

The most relevant free parameter in our Lyα IM methodology is $r_{\rm patch}$; in principle, we can choose any $r_{\rm patch}$ as long as it remains below the angular scales of the BASS camera. Hence, we should evaluate how our $\Delta\langle L_{Ly\alpha}\rangle$ (and more importantly, its SNR), evolve versus $r_{\rm patch}$, with the aim of finding its optimal value. By optimal $r_{\rm patch}$, we exclusively refer to the $r_{\rm patch}$ that yields the maximum $\Delta\langle L_{Ly\alpha}\rangle$ SNR, not the $r_{\rm patch}$ that may provide tighter constraints for other parameters (e.g. $\langle L_{Ly\alpha}\rangle$ or b_e).

In Fig. 14, we display how the value of $\Delta\langle L_{Ly\alpha}\rangle$ and, more importantly, its SNR evolve versus patch radius. We show both a noiseless forecast and one with 10 per cent of BASS noise (for both cases, 100 forecast realizations with 100 bootstrap resamplings were computed). The value of $\Delta\langle L_{Ly\alpha}\rangle$ monotonically decreases with radius, as would be expected from integrating regions further away from the Lyα forest sightline. The SNR, however, increases with radius up to a given maximum, and then smoothly decreases. This SNR maximum is caused by two opposing trends: the integration

of more Lyα emission versus the smaller correlation between Lyα emission and Lyα forest absorption as the patch radius increases. The SNR maximum is localized at $r_{\rm patch} = 16.5$ arcmin for the noiseless forecast, but when we add noise the optimal radius shifts to $r_{\rm patch} = 26.4$ arcmin (approximately 33 cMpc h^{-1} at $\langle z \rangle = 2.64$). We have verified that this radius remains constant for increasing noise levels, so we will adopt $r_{\rm patch} = 26.4$ arcmin for our forecast. Moreover, we have also verified that smaller trends such as the apparent minimum at $r_{\rm patch} \sim 35$ arcmin in the lower panel of Fig. 14 or the apparent SNR increase at the largest $r_{\rm patch}$ on the same panel are not significant. They stem from the variability between realizations, and they do not consistently appear if we recompute Fig. 14 with a different set of realizations.

Moreover, by choosing such a large $r_{\rm patch}$, we ensure that all Lyα emission inside the cosmic volume defined by the DESI footprint and the g-band redshift range ($2.2 < z < 3.4$) is integrated in our study. A circular patch of radius 26.4 arcmin has an area of 0.61 deg^2; according to the DESI QSO density distribution (Fig. 1), on average such an area should contain at least one Lyα QSO (and thus a patch should be integrated around it) up to $z = 3.75$. Since this redshift is higher than the maximum Lyα redshift observed by the g band, we can state that on average all Lyα emission in the cosmic volume defined by the DESI footprint and the g-band redshift convolution is sampled by at least one patch.

Three additional remarks are worth mentioning regarding Fig. 14. First, at $r_{\rm patch} > 26.4$, the SNR decreases with radius very slowly, to the point that the decrease until $r_{\rm patch} = 45$ is barely significant taking into account the aforementioned variability between forecast realizations (shaded regions). Hence, the SNR is almost insensitive to patch radius at $r_{\rm patch} > 26.4$, which means that a real observational study does not necessarily have to use $r_{\rm patch} = 26.4$ arcmin, and that a significant detection of $\Delta\langle L_{\rm Ly\alpha}\rangle > 0$ would very likely imply a detection at larger patch radii, and thus larger scales.

Second, it is possible that the SNR peak at $r_{\rm patch} = 26.4$ and the slow decline at larger radii are partially or totally caused by the limited size of our simulation box. The maximum scales at which the 2PCF/power spectrum can be sampled in a given simulation are limited by the simulation size (Bagla & Ray 2005). In Renard et al. (2021), using the same hydrodynamic simulation (with a box size of 400 cMpc h^{-1}, Section 4) a sharp decrease in the 2PCF and its respective SNR at scales of > 35 cMpc h^{-1} could be seen. Direct extrapolations of Renard et al. (2021) results cannot be made here to understand the impact of the box size. This is because Renard et al. (2021) used fine redshift slices to compute the 2PCF while here we define a custom estimator over a much larger light-cone. Nevertheless, given their results, it is likely that the limited size of the simulation box artificially decreases the SNR of our forecast after a given $r_{\rm patch}$. Regardless, this limitation of our forecast would only mean that the $r_{\rm patch}$ for true maximum SNR is larger than 26.4 arcmin, and thus by choosing $r_{\rm patch} = 26.4$ arcmin we might be being conservative on our SNR estimates. Based on this reasoning, while we suggest $r_{\rm patch} = 26.4$ arcmin for an actual observational study based on Fig. 14, we warn that repeating said study with larger $r_{\rm patch}$ may yield higher SNR.

Third, the value of $r_{\rm patch} = 26.4$ is close to the CCD size of 90Prime, the BASS camera. It is also possible that the slow SNR decrease at $r_{\rm patch} > 26.4$ is due to the chequered CCD-sized pattern in Fig. 10 at similar scales. Regardless, even if we used a DECaLS intensity map for the forecast instead, the transversal scales we can sample are still limited by the size of our hydrodynamic simulation; without a significantly larger simulation box we cannot disentangle those effects. Hence, our conclusions remain the same: it is possible

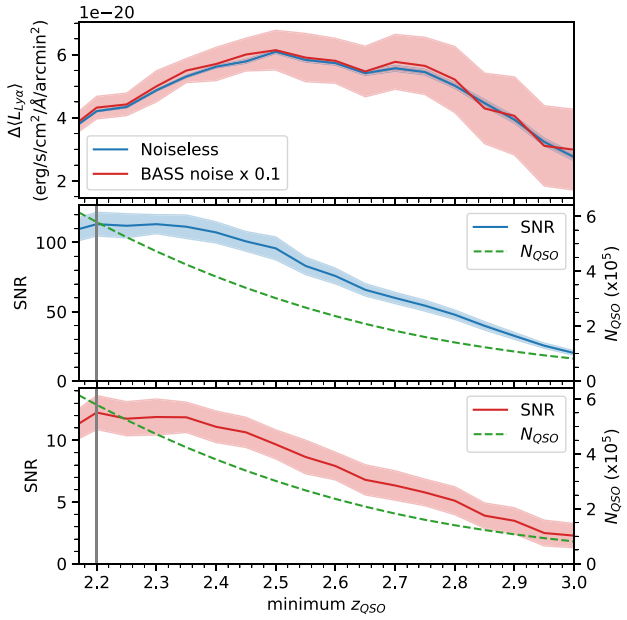

Figure 15. Upper panel: evolution of $\Delta\langle L_{\rm Ly\alpha}\rangle$ versus minimum QSO redshift in our study, for a noiseless case (blue) and 10 per cent of BASS noise case (red). Middle panel: SNR versus minimum QSO redshift (left y-axis) for the noiseless case, together with the total DESI QSOs for a given minimum redshift (right y-axis, green dashed line). The vertical line indicated the minimum QSO redshift of maximum SNR. Lower panel: analogous plot, for the 10 per cent noise case.

that a larger $r_{\rm patch}$ yields a higher SNR with a real data study, and it is also possible that the real optimal $r_{\rm patch}$ is limited by CCD or FOV size. However, since the $\Delta\langle L_{\rm Ly\alpha}\rangle$ SNR versus $r_{\rm patch}$ in Fig. 14 fluctuates by less than 50 per cent, we expect the predicted SNR to have the same order of magnitude as our forecast even if the optimal $r_{\rm patch}$ for real data ends up being significantly different than the $r_{\rm patch}$ we employ.

5.1.3 *Optimal* $z_{\rm min\,QSO}$

In addition to $r_{\rm patch}$, another parameter we may explore to optimise SNR is $z_{\rm min\,QSO}$; the minimum redshift at which we take the Lyα QSOs from DESI into our Lyα IM study. Following DESI Collaboration (2016), all QSOs with $z > 2.1$ have Lyα forest that can be observed with DESI, but the Lyα redshift range of the g band is approximately $2.2 < z < 3.4$ (Fig. 1). Hence, we can choose a higher value of $z_{\rm min\,QSO}$ to increase the minimum redshift overlap between the Lyα forest data and the integrated Lyα emission in the g-band patches. However, by choosing a higher $z_{\rm min\,QSO}$ cut, we also select a smaller total number of QSOs, reducing the size of our sample.

This trade-off between number of QSOs and redshift overlap is explored in Fig. 15, which shows the evolution of $\Delta\langle L_{\rm Ly\alpha}\rangle$ and its SNR versus $z_{\rm min\,QSO}$, for our fiducial Lyα emission model and 100 realizations with 100 bootstrap resamplings in each case. The QSO number has been computed following DESI Collaboration (2016), which assumes a total of 700 000 QSOs[3] for the full DESI

[3]This number of QSOs is smaller than the number of Lyα QSOs used for the first DESI BAO analysis by almost 10 000 (DESI Collaboration 2024), although these comprise a lower redshift range ($1.77 < z < 4.16$).

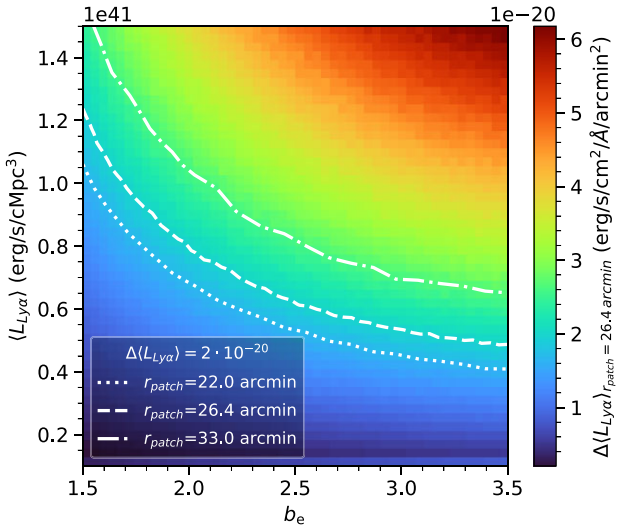

Figure 16. Colour map of $\Delta\langle L_{\rm Ly\alpha}\rangle$ values for the ranges of $b_{\rm e}$ of our model and $\langle L_{\rm Ly\alpha}\rangle$ we deem realistic, for $r_{\rm patch} = 26.4$ arcmin. The isolines are for the value $\Delta\langle L_{\rm Ly\alpha}\rangle = 2 \times 10^{-20}$ erg s^{-1} cm^{-2} Å^{-1} arcmin^{-2} for $r_{\rm patch} =$ 22 arcmin (dotted line), $r_{\rm patch} = 26.4$ arcmin (dashed line), and $r_{\rm patch} = 33$ arcmin (dash–dotted line).

survey at $z_{\rm min\,QSO} = 2.1$. The $\Delta\langle L_{\rm Ly\alpha}\rangle$ values in Fig. 15 increase with $z_{\rm min\,QSO}$ up until $z_{\rm min\,QSO} = 2.5$ (since we remove the QSOs with the least redshift overlap with the g band, and thus the least correlated to patch flux). At higher $z_{\rm min\,QSO}$, the $\Delta\langle L_{\rm Ly\alpha}\rangle$ value steadily decreases (because at that point, increasing $z_{\rm min\,QSO}$ removes the most overlapping/correlated QSOs).

The SNR, however, peaks just at $z_{\rm min\,QSO} = 2.2$, both in the noiseless case and for 10 per cent BASS noise. At $z_{\rm min\,QSO} > 2.2$, removing the least overlapping/correlated QSOs in exchange for reducing our sample does not increase $\Delta\langle L_{\rm Ly\alpha}\rangle$ SNR. This behaviour is to be expected, since the QSO redshift distribution is monotonically decreasing in our redshift range, and by removing QSOs starting by the lowest redshift we significantly reduce our sample, and thus the SNR of $\Delta\langle L_{\rm Ly\alpha}\rangle$. Hence, we choose $z_{\rm min\,QSO} = 2.2$ for our study, which results in 579 183 DESI QSOs (DESI Collaboration 2016).

5.1.4 *Evolution of* $\Delta\langle L_{\rm Ly\alpha}\rangle$ *versus* $\langle L_{\rm Ly\alpha}\rangle$ *and* $b_{\rm e}$

After determining the optimal $r_{\rm patch}$ and $z_{\rm QSO\,min}$ to maximize SNR, we show in Fig. 16 how our statistic $\Delta\langle L_{\rm Ly\alpha}\rangle$ evolves in the $\langle L_{\rm Ly\alpha}\rangle$, $b_{\rm e}$ parameter space of our Lyα emission model. Here, for every pixel on the colour map we computed our forecast without BASS noise nor systematics, using the $\langle L_{\rm Ly\alpha}\rangle$, $b_{\rm e}$ parameter values corresponding to the pixel coordinates (shown on Fig. 16 axis), instead of the 'fiducial' model used so far. For all $\langle L_{\rm Ly\alpha}\rangle$, $b_{\rm e}$ pairs we used the same QSO realization, by choosing the exact same pixels of our Lyα and noise intensity maps for all cases. This was done to control that source of variability. Furthermore, we did not resample via bootstrapping because the SNR is high enough (SNR$\sim$ 100) for the bootstrapping variability to become negligible.

From Fig. 16, we can see that $\Delta\langle L_{\rm Ly\alpha}\rangle$ increases for both higher values of $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$. These trends are in line with our Lyα emission model (equation 13): $\langle L_{\rm Ly\alpha}\rangle$ is simply a multiplicative factor, and the bias $b_{\rm e}$ increases the contrast between underdense and overdense regions. Fig. 16 colour map also exhibits an almost radial symmetry centred at the highest $\langle L_{\rm Ly\alpha}\rangle$, $b_{\rm e}$ corner (upper right). Therefore, a given measured value of $\Delta\langle L_{\rm Ly\alpha}\rangle$ just constrains a curve in $\langle L_{\rm Ly\alpha}\rangle$, $b_{\rm e}$ space. Nevertheless, it is worth examining if we can break the $\langle L_{\rm Ly\alpha}\rangle$–$b_{\rm e}$ degeneracy by measuring the value of $\Delta\langle L_{\rm Ly\alpha}\rangle$ for different patch radii.

This question can be answered by examining the isolines plotted over the colour map in Fig. 16. These correspond to the arbitrary value of $\Delta\langle L_{\rm Ly\alpha}\rangle = 2 \times 10^{-20}$ erg s^{-1} cm^{-2} Å^{-1} arcmin^{-2}, for the optimal $r_{\rm patch}$ of 26.4 arcmin, as well as for patch radii of 22 and 33 arcmin. Increasing (decreasing) patch radius just offsets the $\langle L_{\rm Ly\alpha}\rangle$–$b_{\rm e}$ curve away (towards) the upper right corner (the 'centre' of radial symmetry). Interestingly, curves corresponding to different $r_{\rm patch}$ do not intersect. Consequently, modifying $r_{\rm patch}$ only offsets our $\Delta\langle L_{\rm Ly\alpha}\rangle$ by a given value, without providing additional constraints for either $\langle L_{\rm Ly\alpha}\rangle$ or $b_{\rm e}$.

Regardless, this degeneracy is not an issue, since we have developed the $\Delta\langle L_{\rm Ly\alpha}\rangle$ estimator assuming a low SNR scenario with a significant detection for a single $r_{\rm patch}$ at most. If the $\Delta\langle L_{\rm Ly\alpha}\rangle$ SNR with an eventual observational study was high enough to constrain $\Delta\langle L_{\rm Ly\alpha}\rangle$ in several radial bins, the angular power spectrum could be computed instead, which could be used to constrain $b_{\rm e}$ alone [as galaxy bias has been constrained with e.g. weak lensing in Dvornik et al. 2017, or cosmic microwave background (CMB) lensing in Alonso et al. 2021].

Therefore, our forecast for DESI in the following subsection will only focus on the $\Delta\langle L_{\rm Ly\alpha}\rangle$ SNR we expect for the optimal $r_{\rm patch}$ and $z_{\rm min\,QSO}$, as well as the upper bounds that may be placed on $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ in the case of a non-detection. It is worth noting that, following the isolines in Fig. 16, smaller $r_{\rm patch}$ values result in a $\Delta\langle L_{\rm Ly\alpha}\rangle$ estimator more sensitive to $b_{\rm e}$: a given variation in $b_{\rm e}$ results in a larger variation in $\Delta\langle L_{\rm Ly\alpha}\rangle$. Hence, if the SNR is high enough for a confident detection with the optimal $r_{\rm patch} = 26.4$ arcmin, exploring smaller $r_{\rm patch}$ may yield better constraints in the $\langle L_{\rm Ly\alpha}\rangle$–$b_{\rm e}$ space.

5.1.5 *Additional remarks*

Before moving on, here we shortly comment on two relevant validations we performed while evaluating the behaviour of $\Delta\langle L_{\rm Ly\alpha}\rangle$. First, for a number of QSOs smaller than the number of pixels of our Lyα simulated image (Fig. 10), selecting adjacent QSO pixels (in a 'compact' configuration') or sparse pixels from the whole image (which would be closer to a real QSO distribution) did not make a significant difference for the $\Delta\langle L_{\rm Ly\alpha}\rangle$ value or SNR. Therefore, an arbitrarily wide survey with an arbitrary large number of QSOs can be simulated by just drawing enough Lyα patch-noise patch combinations, with the only limit being the cosmic variance intrinsic to the hydrodynamic simulation.

Second, the BASS intensity map we use to model all noise sources together might actually lead to an SNR underestimation. The real Lyα LSS signal is already contained in that map, but uncorrelated to the simulated Lyα signal of our hydrodynamic simulation, so it is an additional source of noise that will not be present in an eventual observational study. To evaluate this effect, we computed $\Delta\langle L_{\rm Ly\alpha}\rangle$ and its SNR by adding the flux of an uncorrelated Lyα patch to each Lyα patch (i.e. from a randomly drawn QSO that does not correspond to the QSO used to extract the Lyα fores data). This uncorrelated Lyα signal decreased SNR by 30 per cent–40 per cent when no extra noise sources were added. Nevertheless, when a fraction of the BASS noise was added to our forecast, the SNR decrease became negligible, i.e.: the SNR was virtually the same for the 10 per cent BASS noise cases in Figs 14 and 15. Hence, while having a real, uncorrelated Lyα signal in our noise model makes our forecast more conservative, the

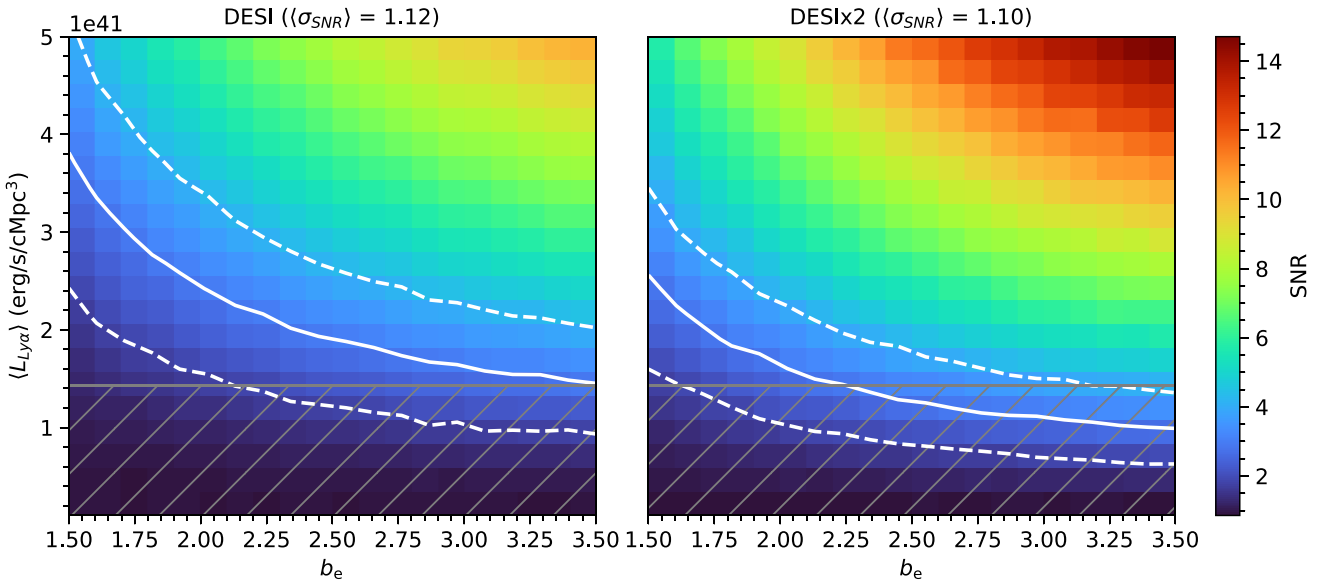

Figure 17. Average SNR of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ for our DESI forecast (left) and a hypothetical DESI×2 extension (right), versus $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_{e} of our Lyα emission model. The SNR is derived from 100 bootstrap resamplings, and for each $\langle L_{\mathrm{Ly}\alpha}\rangle$–$b_{\mathrm{e}}$ case 100 QSO realizations were computed. Solid white line is the isoline for average SNR = 3 (the average detection limit), while dashed white lines are the 1σ confidence interval of the detection limit. Grey hatched area is the $\langle L_{\mathrm{Ly}\alpha}\rangle$ range we deem realistic.

effect we expect in our SNR predictions with real BASS noise is negligible.

5.2 Results for DESI/DESI×2

We computed our final forecast for two surveys: DESI, as originally laid out in DESI Collaboration (2016, with 579 183 QSOs at $z > 2.2$ considered in our study), and a hypothetical DESI phase-II extension which doubles the amount of observed Lyα QSOs (1 158 366 for our study). The exact characteristics of DESI phase-II are not disclosed/decided at the time of writing (to the best of the authors knowledge), hence we will simply dub our hypothetical extension as DESI×2. Simply increasing the number of QSOs does not consider other factors that influence our forecast. For example, the characteristics of the overlapping imaging surveys, or a likely change in the QSO redshift distribution towards higher redshift stemming from deeper target selection. Regardless, doubling the number of QSOs is a reasonable order-of-magnitude estimate of what an eventual DESI extension may yield. For example, Schlegel et al. (2022) propose observing an additional extra 1.1 million LAEs at $z > 2.3$, although it does not specify how well the Lyα forest could be extracted from the spectra of those sources.

Fig. 17 shows the $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ average SNR versus the parameters of our Lyα emission model ($\langle L_{\mathrm{Ly}\alpha}\rangle$, b_{e}). We have explored a larger range of $\langle L_{\mathrm{Ly}\alpha}\rangle$ than the limit we considered realistic [up to 5×10^{41} erg s^{-1} (cMpc^3)$^{-1}$ instead of 1.5×10^{41} erg s^{-1} (cMpc^3)$^{-1}$]: the credible $\langle L_{\mathrm{Ly}\alpha}\rangle$ range [up to 1.5×10^{41} erg s^{-1} (cMpc^3)$^{-1}$] is hatched in grey. For all simulated cases, the 100 realizations of the forecast were computed, with 100 bootstrap resamplings each.

The solid white line in Fig. 17 represents the detection limit (i.e. average SNR = 3), while the two dashed white lines are its $\pm\sigma$ contours. By examining the colour maps on Fig. 17, the SNR dependence with Lyα tracer bias b_{e} becomes evident: for both DESI and DESI×2, a detection seems very unlikely for $b_{\mathrm{e}} = 1.5$ (the SNR = 3 contour lies above the grey-hatched area by more than 1σ). On the other hand, for $b_{\mathrm{e}} = 3.5$, a detection appears plausible for DESI (the SNR = 3 contour sits just at the limit of the realistic $\langle L_{\mathrm{Ly}\alpha}\rangle$ region). For DESI×2, a detection becomes highly likely, with the SNR = 3 line 1σ below the limit of the grey hatched area.

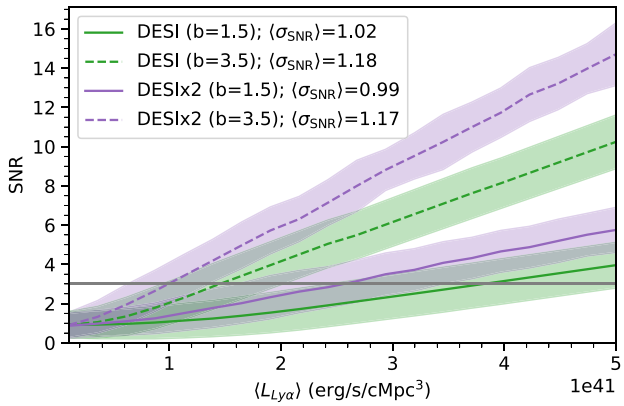

Figure 18. Mean SNR and the 1σ confidence interval (coloured area), versus $\langle L_{\mathrm{Ly}\alpha}\rangle$, for DESI (green) and DESI×2 (purple). Solid lines represent the cases with the minimum bias of our model ($b_{\mathrm{e}} = 1.5$), and dashed lines, the maximum bias ($b_{\mathrm{e}} = 3.5$). Horizontal grey line represents the SNR = 3 detection threshold.

To better understand the SNR variability between forecast realizations, in Fig. 18 the average SNR and its $\pm\sigma$ range are displayed for both DESI and DESI×2, for the highest and lowest bias ($b_{\mathrm{e}} = 3.5$ and 1.5, respectively). For all cases, the SNR variability (i.e. σ_{SNR}) remains approximately constant versus $\langle L_{\mathrm{Ly}\alpha}\rangle$, except in the lowest SNR regime (SNR $\sim$ 1), which stems from the SNR distribution becoming progressively more skewed (as SNR is strictly positive). Higher bias b_{e} increases the SNR and its variability (from $\sigma_{\mathrm{SNR}} \sim 1$ at $b_{\mathrm{e}} = 1.5$ to $\sigma_{\mathrm{SNR}} \sim 1.2$ at $b_{\mathrm{e}} = 3.5$), but a larger survey footprint or number of QSOs seems to increase the SNR while leaving σ_{SNR}

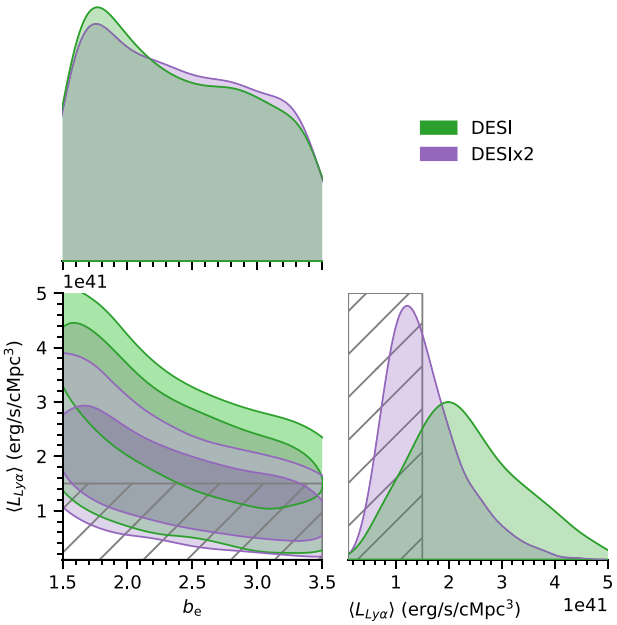

Figure 19. Triangle plot of the upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_e placed by a non-detection ($\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR < 3) for DESI (green) and DESI×2 (purple). Contours in the bottom left panel correspond to 1σ and 2σ constraints, while the hatched areas indicate the range of $\langle L_{\mathrm{Ly}\alpha}\rangle$ we deem realistic.

largely unaffected. Regardless, the most relevant result of Fig. 18 is that the SNR variability is only significant in the low SNR regime, namely: when a detection seems uncertain. For an average SNR $= 3$ we find $\sigma_{\mathrm{SNR}} \sim 1$, which means that a detection is plausible, yet unsure. Also, any upper bounds placed by a non-detection will have significant errors intervals. On the other hand, for a hypothetical $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR $= 10$, which could happen if the total $\langle L_{\mathrm{Ly}\alpha}\rangle$ was heavily underestimated in the literature, or if a deeper imaging survey was used, we still find $\sigma_{\mathrm{SNR}} \sim 1$. On such a high SNR regime, the intrinsic variability between forecast realizations (and between QSO positions and imaging data with real observations) becomes subdominant.

Hence, for a given value $\langle L_{\mathrm{Ly}\alpha}\rangle$, there are two sources of variability for the expected SNR of our forecast: the (unknown) real value of the bias b_e, and the intrinsic variability between forecast realizations of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$. Nevertheless, we can still place upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ for a non-detection of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$. For each $(\langle L_{\mathrm{Ly}\alpha}\rangle_n, b_{e\,n})$ case evaluated in Fig. 17 (i.e. each pixel n on the colour map), the fraction of forecast realizations with SNR < 3 is the probability of not detecting $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ for that given $(\langle L_{\mathrm{Ly}\alpha}\rangle_n, b_{e\,n})$. We can also interpret it as the probability of the real value of $\langle L_{\mathrm{Ly}\alpha}\rangle$, b_e being equal or lower to $\langle L_{\mathrm{Ly}\alpha}\rangle_n$, $b_{e\,n}$, given a non-detection. In other words, the cumulative distribution function $P(\langle L_{\mathrm{Ly}\alpha}\rangle \leq \langle L_{\mathrm{Ly}\alpha}\rangle_n, b_e \leq b_{e\,n} \,|\, \mathrm{SNR} < 3)$. Therefore, by differentiating this cumulative distribution with respect of $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_e, we obtain $P(\langle L_{\mathrm{Ly}\alpha}\rangle_n, b_{e\,n} \,|\, \mathrm{SNR} < 3)$, the probability distribution of the upper bounds of $(\langle L_{\mathrm{Ly}\alpha}\rangle, b_e)$.

Fig. 19 displays this probability distribution of the upper bounds, together with the marginalized probability distributions for $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_e. Table 2 contains the mean and σ of the marginalized upper bounds, together with the probability of detection for an actual observational study ($\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR > 3). This probability of detection was computed by integrating $P(\langle L_{\mathrm{Ly}\alpha}\rangle, b_e \,|\, \mathrm{SNR} < 3)$ over the whole bias range of our Lyα emission model, and over the credible range of $\langle L_{\mathrm{Ly}\alpha}\rangle$ values (assuming a flat prior) and then normalizing by the integral of $P(\langle L_{\mathrm{Ly}\alpha}\rangle, b_e \,|\, \mathrm{SNR} < 3)$ over the whole $(\langle L_{\mathrm{Ly}\alpha}\rangle, b_e)$ parameter space sampled in Fig. 19.

Table 2. Marginalized upper bounds for $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_e for a non-detection in our forecast, together with probability of detecting Lyα LSS (SNR $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle > 3$).

	$\langle L_{\mathrm{Ly}\alpha}\rangle$ [10^{40} erg s^{-1} (cMpc^3)$^{-1}$]	b_e	P_{detect} (per cent)
DESI	23 ± 10	2.39 ± 0.56	23.95
DESI×2	15.5 ± 7.0	2.40 ± 0.56	54.93

From both Fig. 19 and Table 2, a clear conclusion can be drawn: our methodology can impose increasingly tight upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$, but it has little constraining power for the tracer bias b_e. Indeed, compared to DESI, the marginalized $\langle L_{\mathrm{Ly}\alpha}\rangle$ distribution for DESI×2 is significantly narrower and shifted towards lower $\langle L_{\mathrm{Ly}\alpha}\rangle$ values, while the marginalized b_e distribution remains effectively the same, namely: a rather flat distribution across the sampled b_e range, with some preference for lower b_e values (given that a detection is less likely for lower b_e).

Regarding the $\langle L_{\mathrm{Ly}\alpha}\rangle$ upper bounds in Table 2, a non-detection in DESI would predict them to be approximately a factor of $\sim$ 1.5 times larger than the most optimistic literature estimates [$2.3 \pm 1.0 \times 10^{41}$ erg s^{-1} (cMpc^3)$^{-1}$ versus 1.5×10^{41} erg s^{-1} (cMpc^3)$^{-1}$]. On the other hand, for the hypothetical DESI×2, the upper bounds are extremely similar to said estimates [$1.55 \pm 0.73 \cdot 10^{41}$ erg s^{-1} (cMpc^3)$^{-1}$]. This is also reflected in the probability of a detection: going from just 23.95 per cent in DESI to 54.93 per cent in DESI×2. Consequently, we can claim that even with DESI alone the upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ are of the same order of magnitude of the state-of-the-art LFs of faint LAEs (Drake et al. 2017), and with DESI×2 a proper detection of Lyα LSS in g-band images becomes quite plausible. In the following section, we will discuss more in-depth the implications of these results, and in Appendix C, we show the results of a extrapolation of this forecast to the cross-correlation of DESI with the Large Synoptic Survey Telescope (LSST, Ivezić et al. 2019).

6 DISCUSSION

The results of our forecast are moderately optimistic for DESI: while a detection of Lyα LSS is uncertain, the upper bounds that we derive from a non-detection are competitive. Here, we discuss the interpretation of the results presented here, as well as the possible ways to improve our analysis and results in future works, either by improving the SNR of an eventual study with real DESI–DECaLS/BASS data, or tightening the constraints on $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_e. We also comment on the results and science that can be expected from upcoming imaging/spectroscopic surveys.

6.1 Extrapolation to number of observed QSOs

The forecast presented in this work is intentionally conservative at every stage; whenever a decision was necessary, we consistently opted for a reasonable and cautious approach. For example, imaging data were modelled using BASS instead of DECaLS, which is slightly shallower, and thus noisier [23.95 5σ g-band point spread function (PSF) depth versus 23.65, Dey et al. 2019], and public SDSS BOSS mocks were used to model the Lyα forest noise (Bautista et al. 2015), which are bound to be noisier than DESI observations. In fact, DECaLS has a g-band magnitude limit 0.3 mag deeper than BASS (Dey et al. 2019); extrapolating the imaging noise map with a very

simple approximation (equation C1) we can expect a reduction in noise (and thus SNR increase) of ∼25 per cent. Given that DECaLS covers 9000 deg^2 of the DESI footprint, and BASS 5000 deg^2, a total SNR increase of ∼15 per cent can be expected (assuming the total SNR is the average of DECaLS and BASS, weighted by footprint size). This simple extrapolation confirms that our estimations are conservative, although an SNR increase of 15 per cent does not change the general conclusions of this work.

The image reduction procedure to mitigate photometric noise was also the most rudimentary we could perform while ensuring that the Lyα emission signal was preserved (subtraction of average sky per pointing and resolved foreground subtraction, Section 4.2.1). We did not assume more sophisticated image reduction or foreground removal methods that might be developed in future work. Therefore, the results presented throughout Section 5.2, and especially in Table 2, are to be taken as conservative estimates.

However, we can easily update our estimates for the actual number of Lyα QSOs contained in an eventual DESI Y5 data release. This number will exceed the baseline projection from DESI Collaboration (2016). Moreover, a extension/second phase of DESI will not be an exact duplicate of its currently ongoing first phase, as assumed here; an extension reaching a total footprint of 18 000 deg^2 is is currently more likely (instead of 28 000 deg^2 assumed for DESI×2).

On the other hand, for the first phase of DESI there is reason for optimism: current QSO densities observed with DESI are approximately 20 per cent higher than the requirements in DESI Collaboration (2016; Chaussidon et al. 2023). Regardless, the $\langle L_{\mathrm{Ly}\alpha}\rangle$ upper bounds scale well with the inverse of the square root of the number of QSOs: in Table 2 both the upper bound derived for DESI×2 and its error are approximately the DESI upper bounds divided by $\sqrt{2}$. Hence, the upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ for a given number of observed DESI QSOs $N_{\mathrm{QSO\,obs}}$ can be rescaled from our DESI forecast simply with

$$\langle L_{\mathrm{Ly}\alpha}\rangle_{\mathrm{UB\,obs}} = (2.3 \pm 1.0) \times \sqrt{\frac{579,183}{N_{\mathrm{QSO\,obs}}}} \times 10^{41}\ \mathrm{erg\,s^{-1}\,(cMpc^3)^{-1}} \quad (15)$$

where the $\langle L_{\mathrm{Ly}\alpha}\rangle_{\mathrm{UB\,obs}}$ is the observational upper bound, and the numerical values correspond to the upper bound and number of QSOs of our work. Following this extrapolation, just the increase in 20 per cent of QSO density shown in Chaussidon et al. (2023) already lowers the $\langle L_{\mathrm{Ly}\alpha}\rangle$ upper bounds of our forecast by ∼9 per cent. If we consider a DESI phase-II with a total footprint of 18 000 deg^2 in addition to the QSO density improvement from Chaussidon et al. (2023), the total QSO number would increase by 54 per cent compared to our DESI forecast, lowering the $\langle L_{Ly\alpha}\rangle$ upper bounds by approximately 20 per cent.

The extrapolation presented in equation (15) would only work as long as observations were to be performed in an area with g-band data of similar depth to DECaLS/BASS. In case a DESI phase-II target selection was to be carried out with a significantly deeper survey, our forecast would have to be recomputed with images from such survey.

6.2 Future work

This work demonstrates that detecting LSS in Lyα emission within a g-band intensity map is feasible with DESI, particularly with a potential phase-II. Even if no LSS is detected, the upper bounds on Lyα emission that can be established are comparable to the most optimistic estimates in the literature. Nevertheless, two areas of further research could substantially enhance this projections: improved image reduction/foreground subtraction and more sophisticated simulations of the Lyα emission and absorption fields. We also expect that future work will use larger simulation boxes, computed with state-of-the-art cosmological models (e.g, MillenniumTNG, Hernández-Aguayo et al. 2023).

6.2.1 *Foreground removal and image reduction*

Foreground subtraction has arguably been the most important observational challenge for 21 cm IM studies, as the resulting intensity maps are not dominated by 21 cm emission of H I LSS, but synchrotron and free–free emission of galactic origin (Santos, Cooray & Knox 2005; Jelić et al. 2010). Several methods have been developed to subtract this unwanted signal, from polynomial fitting (e.g. Ansari et al. 2012) to more sophisticated statistical methods such as principal component analysis (Alonso et al. 2015; Zuo, Chen & Mao 2023), or machine learning approaches like Gaussian Processes (Soares et al. 2022). All these approaches rely on the foreground emission being smoother in frequency than the 21 cm cosmic signal; the foreground-contaminated region in Fourier space is usually called the 'foreground wedge' (Liu, Parsons & Trott 2014).

In this study, it is evident that the Lyα intensity map derived from g-band images is predominantly dominated by components other than the Lyα signal. The chequered pattern in Fig. 10 roughly matches the CCD size of the BASS camera, which clearly points to the dominant component of the intensity map being of instrumental/atmospheric origin. Moreover, in Renard et al. (2021), it was already proven that cosmic foregrounds are subdominant for similar imaging depths: for a Lyα forest-narrow-band imaging cross-correlation on 100 deg^2 (eBOSS/DESI–PAUS, with magnitude limit $i = 23$), cosmic foregrounds alone would still allow a clear detection of the 2PCF in several radial bins, while adding instrumental noise would make a detection unfeasible.

Hence, we are faced with a fundamentally different problem than foreground subtraction for 21 cm: instead of modelling and removing a foreground intensity map of mostly galactic origin (i.e. a scalar field defined in RA, Dec., and z coordinates), we have to model and subtract a series of signals that will not depend on cosmic coordinates, but on observation time, atmospheric, and instrumental conditions. It is worth noting that observations of the 21 cm line is also contaminated by noise of atmospheric origin (ionospheric effects, Shen et al. 2021) or even man-made, dependant on observation time (radio-frequency intereference, Wilensky et al. 2023), but these remain subdominant.

The mitigation of instrumental systematics (mainly the CCD-to-CCD variability in Fig. 10) and undesired emission of atmospheric/lunar/zodiacal origin (e.g. sky gradients remaining in Fig. 10) could be tackled separately. For example, detailed sky emission models already exist for certain observatories (e.g. Noll et al. 2012), albeit they are more focused on modelling sky spectra. The development and subtraction of mock sky images based on similar models could remove part of the undesired sky foregrounds. Reducing the CCD-to-CCD variability, on the other hand, should be mostly attained via the improvement of image reduction pipelines.

Therefore, we can confidently state that there is a margin for improvement on the SNR side and on the general previsions of our forecast. This is especially true if research is carried out on image reduction/foreground subtraction. However, we cannot quantify the extent of this improvement without further work.

6.2.2 *Modelling the Lyα emission field*

In this work, we have used a simple two-parameter Lyα emission model that allows tuning the total average cosmic Lyα emission, $\langle L_{\mathrm{Ly}\alpha}\rangle$, and its bias as a cosmological tracer, b_{e}, while respecting within $\sim$1 dex the constraints imposed by the brightest QSOs and Lyα nebulae on the bright end (Section 4.1 and Appendix B). While this model generates a Lyα emission field with the desired features, it is agnostic when it comes to the relationship between $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_{e}: it allows for any combination of both, as long as b_{e} is within the range it has been tuned to work.

Nevertheless, not all combinations of $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_{e} are plausible or allowed by a physically motivated model. In Fig. 8, we showed that QSOs are significantly more biased ($b_{\mathrm{e}} \gtrsim 3.5$) than observed LAE samples ($b_{\mathrm{e}} \sim 1.5$). However, the total Lyα luminosity contribution of QSOs is fairly well constrained, given their brightness (e.g. Spinoso et al. 2020; Liu et al. 2022; Torralba-Torregrosa et al. 2023), while the faint of the LAE LF is just hypothesized from ultradeep observations of cosmic filaments (Bacon et al. 2021). Hence, if the real value of $\langle L_{\mathrm{Ly}\alpha}\rangle$ corresponds to the most optimistic estimates [1.5×10^{41} erg s^{-1} (cMpc^3)$^{-1}$, Fig. 9], the reason for it will be that the bulk of Lyα luminosity is produced by the currently undetected faint end of the LAE LF of Drake et al. (2017), plus diffuse fluorescent/resonant emission like the one sampled by Croft et al. (2016) and Lin et al. (2022).

Therefore, a high value of $\langle L_{\mathrm{Ly}\alpha}\rangle$ will push the tracer bias of Lyα emission to the low end ($b_{\mathrm{e}} \sim 1.5$). Conversely, a low value of $\langle L_{\mathrm{Ly}\alpha}\rangle$ will necessarily imply a higher bias, since QSOs, the most biased tracers, are making a significant contribution to the total Lyα luminosity budget. From this reasoning, it follows that there must be a 'forbidden' high-$\langle L_{\mathrm{Ly}\alpha}\rangle$ high-b_{e} zone in our parameter space, namely: a region where $\langle L_{\mathrm{Ly}\alpha}\rangle$ is high enough to be mostly produced by faint LAEs and diffuse emission, but a b_{e} so high to also require a significant contribution from QSOs.

Such 'forbidden zone' would restrict our $\langle L_{\mathrm{Ly}\alpha}\rangle$–$b_{\mathrm{e}}$ parameter space, and would be represented by a masked area in the upper right corner of Fig. 17. This restriction has remarkable implications for our results, since we would not integrate over the whole $\langle L_{\mathrm{Ly}\alpha}\rangle$–$b_{\mathrm{e}}$ space sampled in Fig. 19. First, the upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ displayed in Table 2 would tighten, as we marginalize over a smaller b_{e} range. Second, the detection probabilities in Table 2 would be likely to decrease, given that we are removing the highest SNR region of our parameter space. Regardless, a physically motivated Lyα emission model that considers all possible sources (from QSOs to fluorescent IGM emission) is out of the scope of this work, and without such a model, any speculation of the shape of this 'forbidden zone' would be just a guess.

With a physically motivated (or at least more developed) Lyα emission model, we may also improve the SNR of an eventual Lyα IM study. In this work, we have integrated the g-band patches with uniform radial weighting – equation (1) –, but for a given Lyα emission model and a given radial length, there must be a radial kernel ($W(r)$ in equation 2) that maximizes the SNR of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$. Convolution kernels that increase the SNR of a given statistic are already used in cosmology, i.e. in CMB lensing by weighting the redshift convolution by the inverse of the variance (e.g. Qu et al. 2024), or in CMB polarization by weighting more the redshift ranges where a signal is expected (Montero-Camacho & Hirata 2018). However, since we have performed our forecast for the r_{patch} with maximum $\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR, the gains we may expect from a more sophisticated $W(r)$ will be limited.

Moreover, we do not expect a different $W(r)$ to break the $\langle L_{\mathrm{Ly}\alpha}\rangle$–$b_e$ degeneracy displayed in Fig. 16. The isolines of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ for different r_{patch} display a radial symmetry in the $\langle L_{\mathrm{Ly}\alpha}\rangle$–$b_{\mathrm{e}}$ space (which is the cause of the degeneracy). A different $W(r)$ would be a weighted average of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ using step kernels of different r_{patch}, i.e. a linear combination that would not break the radial symmetry of isolines in Fig. 16.

There are already examples in the literature of physically motivated Lyα emission models for the cosmic web, computed by post-processing hydrodynamic simulations with radiative transfer (e.g. Cantalupo et al. 2005; Goerdt et al. 2010; Behrens et al. 2019; Byrohl & Nelson 2023). Less computationally expensive approaches, such as lognormal mocks, have also been explored (Niemeyer 2024). However, the box sizes of these simulations are generally too small for a study on the scales we sample ($>$ 1 Gpc in redshift direction, and $>$ 30 cMpc in transverse direction for our patch size). Some of these simulations (e.g. Cantalupo et al. 2005; Behrens et al. 2019) are of halo scales or slightly above, and even the largest ones (to the best of the author's knowledge) are not large enough for the proper study of LSS. For example, in Byrohl & Nelson (2023), the simulation box size is of 50 cMpc, as it is post-processed on the Illustris TNG50 simulation (Nelson et al. 2019). Moreover, not all possible sources of Lyα emission are always considered (e.g. in Behrens et al. 2019 only stellar emission as an ionizing source is considered, and in Byrohl & Nelson 2023 ionization from local stellar sources is not considered).

Regardless, the implementation of a radiative transfer model, instead of the simple Lyα emission model presented in this work, would significantly improve the accuracy of our forecast, and tighten the $\langle L_{\mathrm{Ly}\alpha}\rangle$ upper bounds. State-of-the-art hydrodynamical simulations with larger sizes than our simulation box already exist (e.g. MillenniumTNG, Hernández-Aguayo et al. 2023, with a box size of 500 cMpc h^{-1}). Running radiative transfer simulations with different sets of parameters (e.g. UV background, ionization from QSOs and stellar emission, cooling mechanisms) may result in different values of $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_{e} for the Lyα emission field, which would allow the Lyα IM studies like the one presented here to directly constrain physical properties of the IGM and/or cosmological parameters.

Simulating the Lyα emission field for different radiative transfer models may prove computationally too expensive; however, recent developments in machine learning may dramatically reduce computational costs. Moreover, the exploration of the cosmological parameter space with Lyα IM amplifies the need for inexpensive Lyα emission simulations, as the Lyα emission field will have to be computed on several boxes with different cosmologies.

The application of deep-learning models to extrapolate the results of small-scale radiative-transfer codes to larger simulated volumes is a very active field of research. Baryon inpainting generates baryon distributions from much faster N-body simulations without the need for hydrodynamical computations (Wadekar et al. 2021; Dai et al. 2024), and even H I density fields can be generated with similar methodologies (Bernardini et al. 2022). Physical properties of the IGM gas (e.g. temperature, pressure, and electron density) can also be simulated with neural networks (Mohammad et al. 2022; Andrianomena, Hassan & Villaescusa-Navarro 2023); a similar neural network could be trained to generate Lyα emission fields from hydrodynamical or N-body simulations for different sets of IGM physical parameters. Another possible approach would be the use of emulators to interpolate statistics such as the power spectrum between simulations with different parameters. This approach has been widely explored for the matter power spectrum (e.g. Heitmann et al. 2013; Aricò et al. 2021), and tracers such as galaxies (Kwan et al. 2015) or the Lyα forest (Bird et al. 2023).

In fact, if next-generation spectroscopic and imaging surveys allow for Lyα IM studies with higher SNR than our expectations for DESI/DESI×2 (i.e. enough SNR to constrain the angular power spectrum in several bins of distance), a very clear synergy between simulations and observations may emerge. More sophisticated Lyα emission models than those currently available will give more constraining power to Lyα IM observations, and observations themselves will help rule out IGM physical models or place cosmological constraints. Moreover, resolved sources could be selectively removed from Lyα intensity maps before coadding (i.e. by applying colour cuts to remove QSOs, or general magnitude cuts), to better understand how diffuse emission versus different kinds of resolved emitters trace the cosmic web.

6.3 Expectations for upcoming surveys

The Lyα IM methodology presented in this work requires both imaging data to generate the Lyα emission intensity maps, and spectroscopic data to extract the Lyα forest and cross-correlate it with the patches integrated around QSOs. Hence, future Lyα IM studies beyond DESI and its Legacy Imaging Surveys are contingent on the development of both imaging surveys in the optical spectrum, and spectroscopic surveys targeting QSOs at redshifts high enough to observe the Lyα forest. Moreover, if the SNR provided by future surveys is high enough to have significant detections of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ in several r_{patch}, the 2PCF could be used instead of our estimator to cross-correlate Lyα emission with Lyα forest absorption. A more 'conventional' IM study with the 2PCF would constrain the shape of the power spectrum in Lyα emission, which would have a much more direct link to the existing LSS theory (e.g. Peebles & Yu 1970; Bernardeau et al. 2002) than our custom estimator.

Two types of upcoming imaging surveys hold great promise for not only detecting Lyα IM, but even constraining the Lyα power spectrum: space telescopes and ground-based time-domain surveys.

6.3.1 *Broad-band imaging surveys*

As we have discussed and shown in Fig. 10, the dominant pattern in our BASS intensity map, and thus our main source of noise, is from instrumental and atmospheric origin. Observations from space telescopes are free of any undesired emission of atmospheric origin, so an optical intensity map constructed from such images should have the chequered pattern displayed in Fig. 10 significantly mitigated, even with the exact same image reduction procedure carried out in this paper. The already-launched *Euclid* mission (Scaramella et al. 2022) will observe over 15 000 deg^s of the sky significantly overlapping with the DESI footprint (over 9000 deg^2, Naidoo et al. 2023). Its filter system consists of a very broad optical band I_E (between 5500–9000 Å), expected to reach a depth of $I_E = 26.2$, as well as three near-infrared bands up to similar depths. The Euclid I_E band observes Lyα in a redshift range higher than most of DESI Lyα forest data ($3.5 \leq z_{\mathrm{Ly}\alpha} \leq 6.4$ versus $2.1 \leq z_{\mathrm{Ly}\alpha} \leq 4$ for the great majority of DESI Lyα QSOs), so deeper, next-generation spectroscopic surveys may be needed to cross-correlate with *Euclid* images (e.g. Schlegel et al. 2022).

Another space telescope projected to launch in the next years that holds even more promise for Lyα IM is the China Space Station Telescope (CSST, Miao et al. 2022; Liu et al. 2023). Planned to carry out a imaging survey covering at least the entire *Euclid* footprint, CSST is designed to observe at bluer wavelength, using standard NUV + *ugrizy* filters, with a approximate 5σ magnitude limit of 25.5 for point sources (Liu et al. 2023). Reaching almost two magnitudes deeper than DECaLS/BASS, and free from any undesired atmospheric emission, Lyα IM with g-band CSST images may not only yield a very significant detection of our estimator $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$, but even allow to constrain the angular power spectrum in Lyα in several angular bins.

Regarding ground-based imaging surveys, the largest planned for the upcoming years is the Vera C. Rubin Observatory (Rubin, formerly LSST, Ivezić et al. 2019), planned to observe at least 18 000 deg^2 in the *ugrizy* filter system, and reaching a g-band 5σ point-source magnitude depth of 25.6 and 26.84 in the first and tenth year of operations, respectively (LSST Collaboration 2018). Rubin is designed to be a time-domain survey, observing variable and transient phenomena (such as supernovae or AGN) over an unprecedented volume. Consequently, for each survey pointing, approximately 1000 exposures are planned along its 10-yr programme (LSST Collaboration 2018). For the interest of IM studies, averaging out ∼1000 exposures instead of the 3 exposures per pointing taken by DECaLS/BASS (Dey et al. 2019) will greatly mitigate any exposure-dependent systematic in the g-band intensity map, be it instrumental or atmospheric. Although Rubin will be based on the southern hemisphere, at least 3200 deg^2 of overlap are expected with DESI (Bolton et al. 2018). In Appendix C, we perform a simple extrapolation of our forecast to this minimal DESI–LSST overlap and show that, just after a year of operations of LSST, a detection of the cosmic web in Lyα with our estimator seems extremely likely. Given this promising preliminary forecast, we expect future work to investigate the insight on IGM physical properties and cosmological parameters that can gained from Lyα IM with LSST.

6.3.2 *Spectroscopic surveys*

The deeper and cleaner images from upcoming imaging surveys will be the main observational drive moving Lyα IM forward, but we can also briefly comment on upcoming spectroscopic surveys. While very preliminary plans have been laid out for a Stage-V successor to DESI (MegaMapper, Schlegel et al. 2022), and telescopes with similar characteristics and science goals are under construction (MUltiplexed Survey Telescope, Zhang et al. 2024), the improvement they may bring over DESI for Lyα IM is less certain. First, because their target selection is entirely contingent on available imaging data, and second, because only the targets that provide Lyα forest data can be used for Lyα IM. Regardless, we can state that increases in Lyα forest sightline densities can be expected from future spectroscopic surveys, due to both increased completeness and depth. An increase in the average redshift sampled by their Lyα forest and a shift towards higher redshifts of the redshift kernel for Lyα IM studies (Fig. 1) is also to be expected, due to increased target selection depth.

6.3.3 *Narrow-band imaging surveys*

Finally, we also have to mention multi-narrow-band photometric surveys, with filters of FWHM ∼ 100 Å, instead of the typical FWHM $\gtrsim$ 1000 Å, of standard broad-band filters. The increase of spectral resolution provided by these narrow-band surveys results in integration of the Lyα line in fine redshift bins of $\Delta z \sim 0.1$ instead of $\Delta z \sim 1$. This finer binning results in a much stronger correlation of the underlying Lyα emission field than broad-band surveys, given that the smoothing in the redshift direction is smaller than the scale of homogeneity (just ∼50 cMpc h^{-1}, Renard et al. 2021). While this would make narrow-band imaging ideal for Lyα IM, narrow-band surveys are significantly scarcer, and most of them present important

caveats. For example, some are devoted to studying specific emission lines in fields of few deg^2 (e.g. SILVERRUSH, Ouchi et al. 2018), while wider ones reach relatively shallow depths in a limited set of filters (e.g. $r < 21$ in J-PLUS and S-PLUS, Cenarro et al. 2019; De Oliveira et al. 2019). Narrow-band imaging surveys with depths similar to DECaLS/BASS (r or i magnitude up to 23) covering a significant fraction of the optical spectrum are generally not wide enough for Lyα IM, e.g. PAUS (Benítez et al. 2009) covers $\sim$ 50 deg^2 (Navarro-Gironés et al. 2024), and Renard et al. (2021) already showed it was insufficient for Lyα IM.

However, the recently started J-PAS survey (Benitez et al. 2014) will be deep (Bonoli et al. 2021) and wide enough for Lyα IM: with a projected total footprint of 8000 deg^2, reaching an approximate magnitude limit of $i = 23$. Approximately $\sim$6000 deg^2 are expected to overlap with the spectroscopic survey WEAVE (Jin et al. 2024), reaching almost 100 per cent completeness in its Lyα QSO observation program (Pieri et al. 2016). Cross-correlation of J-PAS narrow-band intensity maps with WEAVE Lyα forest data may yield a detection of the transversal Lyα power spectrum in several redshift bins (Renard et al., in preparation).

7 CONCLUSIONS

In this work, we propose a novel approach for IM using the Lyα line, involving the cross-correlation of broad-band imaging data with Lyα forest observations. While this method is applicable to various imaging-spectroscopic surveys, we focus on a detailed forecast for DESI and its associated Legacy Imaging Surveys, DECaLS and BASS. In our analysis, we have especially focused on how image reduction and co-adding of the intensity maps should be carried out to minimize noise while preserving most of the Lyα emission signal. Furthermore, we introduce a simple analytical model of Lyα emission, enabling us to explore a wide range of key parameters characterizing Lyα emission as a tracer of LSS, including its average emission per volume unit ($\langle L_{\rm Ly\alpha}\rangle$) and its linear bias ($b_{\rm e}$) relative to the dark matter distribution.

The results of our forecast show that, while a detection of LSS in Lyα emission is not certain, competitive upper bounds on the average Lyα emission can still be placed for DESI–DECaLS/BASS and for a hypothetical DESI extension with twice the number of Lyα QSOs (DESI$\times$2). For DESI, this upper bound on $\langle L_{\rm Ly\alpha}\rangle$ is $23 \pm 10 \times 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$ [$\sim$ 50 per cent higher than the most optimistic literature estimate of 15×10^{40} erg s^{-1} (cMpc^3)$^{-1}$ from integration of the Lyα LF in Drake et al. 2017]. In contrast, for DESI$\times$2, the upper bounds are lowered to $15.5 \pm 7.0 \times 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$ (very similar to the most optimistic literature estimates). Moreover, a simple extrapolation of our forecast to the overlap between DESI and LSST (Appendix C) points to an almost certain detection of Lyα LSS, even after just a year of LSST observations.

Consequently, in this work we provide one of the very first practical examples and guidelines for performing IM with the Lyα line using optical imaging. We prove that currently ongoing surveys will provide, at the very least, valuable constraints in the absence of a detection. Moreover, our discussion of upcoming surveys and extrapolation to LSST points at next generation surveys ushering the beginning of Lyα IM as a fully fledged cosmological/astrophysical probe, that would help disentangling how baryons distribute with respect of the underlying dark matter LSS, and how the IGM physical properties determine the scattering of Lyα emission across the cosmic web. Its redshift range ($2.2 < z < 3.4$ for the g band) is higher than typical galaxy surveys, providing a complementary cosmological probe to the Lyα forest that may help constrain dark matter candidates, probe the matter power spectrum, or even study the impacts of reionization at later redshifts (e.g. Long et al. 2023).

ACKNOWLEDGEMENTS

PR and DS acknowledge the support by the Tsinghua Shui Mu Scholarship. PR, DS, and ZC acknowledge funding of the National Key R&D Program of China (grant no. 2023YFA1605600), the science research grants from the China Manned Space Project with no. CMS-CSST2021-A05, and the Tsinghua University Initiative Scientific Research Program (no. 20223080023). PR acknowledges additional funding from the National Science Foundation of China (grant no. 12350410365). HZ acknowledges the supports from National Key R&D Program of China (grant nos 2023YFA1607800, 2023YFA1607804, and 2022YFA1602902) and the National Natural Science Foundation of China (NSFC; grant nos 12120101003 and 12373010). PM-C was supported by the Major Key Project of PCL. We acknowledge the fruitful scientific discussions with Rupert A.C. Croft, Enrique Gaztanaga, and Richard Grumitt. We would also like to thank the referee for their insightful comments.

DATA AVAILABILITY

The data and code underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Adelberger K. L., Steidel C. C., Pettini M., Shapley A. E., Reddy N. A., Erb D. K., 2005, ApJ, 619, 697
Ahumada R. et al., 2020, ApJS, 249, 3
Alonso D., Bull P., Ferreira P. G., Santos M. G., 2015, MNRAS, 447, 400
Alonso D., Bellini E., Hale C., Jarvis M. J., Schwarz D. J., 2021, MNRAS, 502, 876
Andrianomena S., Hassan S., Villaescusa-Navarro F., 2023,arXiv:2303.07473
Ansari R. et al., 2012, A&A, 540, A129
Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021, MNRAS, 506, 4070
Bacon R. et al., 2021, A&A, 647, A107
Bagla J. S., Ray S., 2005, MNRAS, 358, 1076
Battaia F. A. et al., 2022, ApJ, 930, 72
Bault A. et al., 2024, Impact of Systematic Redshift Errors on the Cross-correlation of the Lyman-α Forest with Quasars at Small Scales Using DESI Early Data , preprint (arXiv:2402.18009), https://ui.adsabs.harvard.edu/abs/2024arXiv240218009B/abstract
Bautista J. E. et al., 2015, J. Cosmol. Astropart. Phys., 2015, 060
Bautista J. E. et al., 2017, A&A, 603, A12
Behrens C., Pallottini A., Ferrara A., Gallerani S., Vallini L., 2019, MNRAS, 486, 2197
Benítez N. et al., 2009, ApJ, 691, 241
Benitez N. et al., 2014, J-PAS: The Javalambre-Physics of the Accelerated Universe Astrophysical Survey, Technical Report preprint (arXiv:1403.5237), http://arxiv.org/abs/1403.5237
Bernal J. L., Kovetz E. D., 2022, A&A Rev., 30, 5
Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rep., 367, 1
Bernardini M., Feldmann R., Anglés-Alcázar D., Boylan-Kolchin M., Bullock J., Mayer L., Stadel J., 2022, MNRAS, 509, 1323
Bertin E., Mellier Y., Radovich, Mario Missonnier G., Didelon P., Morin B., 2002, in Bohlender D. A., Durand D., Handley Th. H., eds, ASP Conf. Ser. Vol. 281, Astronomical Data Analysis Software and Systems XI. Astron. Soc. Pac., San Francisco, p. 228
Bird S. et al., 2023, J. Cosmol. Astropart. Phys., 2023, 037
Blomqvist M. et al., 2019, A&A, 629, A86

Bolton A. S. et al., 2018, Maximizing the Joint Science Return of LSST and DESI, Technical Report, https://docushare.lsstcorp.org/docushare/dsweb/Get/Document-30603/bolton_desi_overlap_mini.pdf.
Bonoli S. et al., 2021, A&A, 653, A31
Breysse P. C., Kovetz E. D., Kamionkowski M., 2014, MNRAS, 443, 3506
Byrohl C., Nelson D., 2023, MNRAS, 523, 5248
Cai Z. et al., 2017, ApJ, 837, 71
Cai Z. et al., 2019, ApJS, 245, 23
Calabretta M. R., Greisen E. W., 2002, A&A, 395, 1077
Cantalupo S., Porciani C., Lilly S. J., Miniati F., 2005, ApJ, 628, 61
Cantalupo S., Arrigoni-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, Nature, 506, 63
Cassata P. et al., 2011, A&A, 525, A143
Cenarro A. J. et al., 2019, A&A, 622, A176
Chabanier S. et al., 2019, J. Cosmol. Astropart. Phys., 2019, 17
Chaussidon E. et al., 2023, ApJ, 944, 107
Chiang Y.-K., Ménard B., Schiminovich D., 2019, ApJ, 877, 150
Cisewski J., Croft R. A. C., Freeman P. E., Genovese C. R., Khandai N., Ozbek M., Wasserman L., 2014, MNRAS, 440, 2599
Croft R. A. et al., 2016, MNRAS, 457, 3541
Croft R. A. C., Miralda-Escudé J., Zheng Z., Blomqvist M., Pieri M., 2018, MNRAS, 481, 1320
Croom S. M. et al., 2005, MNRAS, 356, 415
DESI Collaboration, 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design, Technical Report, preprint (arXiv:1611.00036), http://arxiv.org/abs/1611.00036
DESI Collaboration, 2022, AJ, 164, 207
DESI Collaboration, 2024, DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations , preprint (arXiv:2404.03002), https://ui.adsabs.harvard.edu/abs/2024arXiv240403002D/abstract
Dai Z., Moews B., Vilalta R., Davé R., Dave R., 2024, MNRAS, 527, 3381
Dawson K. S. et al., 2013, AJ, 145, 10
De Oliveira C. M. et al., 2019, MNRAS, 489, 241
Dey A. et al., 2019, AJ, 157, 168
Di Matteo T., Khandai N., DeGraf C., Feng Y., Croft R. A. C., Lopez J., Springel V., 2012, ApJ, 745, L29
Diener C. et al., 2017, MNRAS, 471, 3186
Dijkstra M., 2017, Saas-Fee Lecture Notes: Physics of Lyman Alpha Radiative Transfer, Technical Report, preprint (arXiv:1704.03416), https://arxiv.org/abs/1704.03416v1
Doré O. et al., 2014, Cosmology with the SPHEREX All-Sky Spectral Survey, Technical Report, preprint (arXiv:1412.4872), https://arxiv.org/abs/1412.4872v3
Drake A. B. et al., 2017, A&A, 608, A6
DuMasDesBourboux et al., 2020, ApJ, 901, 153
Dvornik A. et al., 2017, MNRAS, 468, 3251
Eisenstein D. J. et al., 2011, AJ, 142, 72
Flaugher B. et al., 2015, AJ, 150, 150
Font-Ribera A. et al., 2013, J. Cosmol. Astropart. Phys., 18, 2013
Gallego S. G. et al., 2021, MNRAS, 504, 16
Gawiser E. et al., 2007, ApJ, 671, 278
Goerdt T., Dekel A., Sternberg A., Ceverino D., Teyssier R., Primack J. R., 2010, MNRAS, 407, 613
Gonçalves R. S., Carvalho G. C., Bengaly C. A. P., Carvalho J. C., Alcaniz J. S., 2018, MNRAS, 481, 5270
Gong Y., Cooray A., Silva M., Santos M. G., Bock J., Bradford C. M., Zemcov M., 2012, ApJ, 745, 49
Gordon C. et al., 2023, J. Cosmol. Astropart. Phys., 2023, 045
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gronwall C. et al., 2007, ApJ, 667, 79
Guaita L. et al., 2010, ApJ, 714, 255
Han B., Song Y., Zhao Y., 2023, MNRAS, 526, 5520
Heitmann K., Lawrence E., Kwan J., Habib S., Higdon D., 2013, ApJ, 780, 111
Herenz E. C. et al., 2019, A&A, 621, A107
Hernández-Aguayo C. et al., 2023, MNRAS, 524, 2556
Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51
Herrero Alonso Y. et al., 2021, A&A, 653, A136
Herrero-Alonso Y., Wisotzki L., Miyaji T., Schaye J., Pharo J., Krumpe M., 2023, A&A, 677, A125
Hu W., Wang X., Wu F., Wang Y., Zhang P., Chen X., 2020, MNRAS, 493, 5854
Ihle H. T. et al., 2019, ApJ, 871, 75
Iršič V. et al., 2017, MNRAS, 466, 4332
Iršič V. et al., 2024, Phys. Rev. D, 109, 043511
Ivezić Ž. et al., 2019, ApJ, 873, 111
Jelić V., Zaroubi S., Labropoulos P., Bernardi G., De Bruyn A. G., Koopmans L. V., 2010, MNRAS, 409, 1647
Jin S. et al., 2024, MNRAS, 530, 2688
Keating G. K., Marrone D. P., Bower G. C., Keenan R. P., 2020, ApJ, 901, 141
Kimeswenger S., Kausch W., Noll S., Jones A. M., 2015, EPJ Web Conf., 89, 01001
Kooistra R., Lee K.-G., Horowitz B., 2022, ApJ, 938, 123
Kovetz E. D., Viero M. P., Lidz A., Newburgh L., Rahman M., Switzer E., Kamionkowski M., 2017, Line-Intensity Mapping: 2017 Status Report, Technical Report, preprint (arXiv:1709.09066v1), https://arxiv.org/abs/1709.09066
Kunz M., Nesseris S., Sawicki I., 2016, Phys. Rev. D, 94, 023510
Kwan J. et al., 2015, ApJ, 810, 35
LSST Collaboration, 2018, The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document, Technical Report, preprint (arXiv:1809.01669), https://arxiv.org/abs/1809.01669v2
Laurent P. et al., 2016, J. Cosmol. Astropart. Phys., 2016, 060
Li T. Y., Wechsler R. H., Devaraj K., Church S. E., 2016, ApJ, 817, 169
Lin X., Zheng Z., Cai Z., 2022, ApJS, 262, 38
Liu A., Parsons A. R., Trott C. M., 2014, Phys. Rev. D, 90, 023018
Liu C. et al., 2022, ApJ, 935, 132
Liu D. Z. et al., 2023, A&A, 669, A128
Loeb A., Zaldarriaga M., 2004, Phys. Rev. Lett., 92, 211301
Long H., Morales-Gutiérrez C., Montero-Camacho P., Hirata C. M., 2023, MNRAS, 525, 6036
López-Sanjuan C. et al., 2019, A&A, 631, A119
Ma K. et al., 2024, ApJ, 961, 102
Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429
Maksora Tohfa H., Bird S., Ho M.-F., Qezlou M., Fernandez M., 2024, Phys. Rev. Lett., 132, 231002
Marra V., Amendola L., Sawicki I., Valkenburg W., 2013, Phys. Rev. Lett., 110, 241305
Matsuda Y. et al., 2004, AJ, 128, 569
Miao H., Gong Y., Chen X., Huang Z., Li X. D., Zhan H., 2022, MNRAS, 519, 1132
Mohammad F. G., Villaescusa-Navarro F., Genel S., Anglés-Alcázar D., Vogelsberger M., 2022, ApJ, 941, 132
Montero-Camacho P., Hirata C. M., 2018, J. Cosmol. Astropart. Phys., 2018, 040
Montero-Camacho P., Mao Y., 2021, MNRAS, 508, 1262
Naidoo K. et al., 2023, A&A, 670, A149
Navarro-Gironés D. et al., 2024, MNRAS, 534, 1504
Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2
Niemeyer M. L., 2024, Effect of Lyman-α Radiative Transfer on Intensity Mapping Power Spectra , preprint (arXiv:2407.03060), https://arxiv.org/abs/2407.03060v1
Noll S., Kausch W., Barden M., Jones A. M., Szyszka C., Kimeswenger S., Vinther J., 2012, A&A, 543, A92
Ouchi M. et al., 2008, ApJS, 176, 301
Ouchi M. et al., 2010, ApJ, 723, 869
Ouchi M. et al., 2018, PASJ, 70, 13
Ouchi M., Ono Y., Shibuya T., 2020, ARA&A, 58, 617
Ozbek M., Croft R. A. C., Khandai N., 2016, MNRAS, 456, 3610
Padilla C. et al., 2019, AJ, 157, 246
Palanque-Delabrouille N. et al., 2013, A&A, 559, A85

Palanque-Delabrouille N. et al., 2015, J. Cosmol. Astropart. Phys., 2015, 011
Palanque-Delabrouille N., Yèche C., Schöneberg N., Lesgourgues J., Walther M., Chabanier S., Armengaud E., 2020, J. Cosmol. Astropart. Phys., 2020, 038
Pandey B., Sarkar S., 2015, MNRAS, 454, 2647
Parsons J., Mas-Ribas L., Sun G., Chang T.-C., Gonzalez M. O., Mebane R. H., 2022, ApJ, 933, 141
Peebles P. J. E., Yu J. T., 1970, ApJ, 162, 815
Peirani S., Weinberg D. H., Colombi S., Blaizot J., Dubois Y., Pichon C., 2014, ApJ, 784, 11
Peterson J. B. et al., 2009, astro2010 Astron. Astrophys. Decad. Surv., 2010, 234
Pieri M. M. et al., 2016, in Proc. Annu. Meet. French Soc. Astron. Astrophys.. Lyon, p. 259 , preprint (arXiv:1611.09388), https://arxiv.org/abs/1611.09388v1
Pillepich A. et al., 2018, MNRAS, 473, 4077
Planck Collaboration, 2020, A&A, 63
Pullen A. R., Doré O., Bock J., 2014, ApJ, 786, 111
Qu F. J. et al., 2024, ApJ, 962, 112
Ramírez-Pérez C., Sanchez J., Alonso D., Font-Ribera A., 2022, J. Cosmol. Astropart. Phys., 002
Ramírez-Pérez C. et al., 2024, MNRAS, 528, 6666
Rauch M., 1998, ARA&A, 36, 267
Renard P. et al., 2021, MNRAS, 501, 3883
Santos M. G., Cooray A., Knox L., 2005, ApJ, 625, 575
Sarkar P., Yadav J., Pandey B., Bharadwaj S., 2009, MNRAS, 399, L128
Sato-Polito G., Bernal J. L., 2022, Phys. Rev. D, 106, 103534
Scaramella R. et al., 2022, A&A, 662, A112
Schlegel D. J. et al., 2022, A Spectroscopic Road Map for Cosmic Frontier: DESI, DESI-II, Stage-5, Technical Report, preprint (arXiv:2209.03585), https://arxiv.org/abs/2209.03585v1
Selsing J., Fynbo J. P., Christensen L., Krogager J. K., 2016, A&A, 585, A87
Shen E., Anstey D., De Lera Acedo E., Fialkov A., Handley W., 2021, MNRAS, 503, 344
Shull J. M., Smith B. D., Danforth C. W., 2012, ApJ, 759, 23
Silva M. B., Santos M. G., Gong Y., Cooray A., Bock J., 2013, ApJ, 763, 132
Soares P. S., Watkinson C. A., Cunnington S., Pourtsidou A., 2022, MNRAS, 510, 5872
Sobral D. et al., 2017, MNRAS, 466, 1242
Sobral D., Santos S., Matthee J., Paulino-Afonso A., Ribeiro B., Calhau J., Khostovan A. A., 2018, MNRAS, 476, 4725
Songaila A., Cowie L. L., 2010, ApJ, 721, 1448
Sorini D., Davé R., Anglés-Alcázar D., 2020, MNRAS, 499, 2760
Spinoso D. et al., 2020, A&A, 643, A149
Springel V., 2005, MNRAS, 364, 1105
Springel V., Hernquist L., 2003, MNRAS, 339, 289
Sun Z., Ting Y.-S., Cai Z., 2023, ApJS, 269, 30
Taniguchi Y., Shioya Y., Kakazu Y., 2001, ApJ, 562, L15
Torralba-Torregrosa A. et al., 2023, A&A, 680, A14
Vanden Berk D. E. et al., 2001, AJ, 122, 549
Viel M., Haehnelt M. G., Springel V., 2004, MNRAS, 354, 684
Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502
Visbal E., McQuinn M., 2023, ApJ, 956, 84
Wadekar D., Villaescusa-Navarro F., Ho S., Perreault-Levasseur L., 2021, ApJ, 916, 42
Wang B. et al., 2022a, ApJS, 259, 28
Wang Y., Yang H.-Y., He P., 2022b, ApJ, 934, 77
Weinberg D. H., Hernquist L., Katz N., Croft R., Miralda-Escudé J., 1997, in Proc. 13th IAP Astrophys. Colloq. Structure and Evolution Intergalactic Medium from QSO Absorption Line System. Editions Frontieres, Paris, p. 133, Patrick Petitjean, Stephane Charlot, preprint(arXiv:9709303), http://arxiv.org/abs/astro-ph/9709303
Wilensky M. J. et al., 2023, ApJ, 957, 78
Williams G. G., Olszewski E., Lesser M. P., Burge J. H., 2004, Proc. SPIE, 5492, 787
Wilson B., Iršič V., McQuinn M., 2022, MNRAS, 509, 2423
Wisotzki L. et al., 2018, Nature, 562, 229
Wolfe A. M., Gawiser E., Prochaska J. X., 2005, ARA&A, 43, 861
Wright E. L. et al., 2010, AJ, 140, 1868
Yang S., Pullen A. R., Switzer E. R., 2019, MNRAS, 489, L53
Yèche C., Palanque-Delabrouille N., Baur J., Du Mas Des Bourboux H., 2017, J. Cosmol. Astropart. Phys., 2017, 047
York D. G. et al., 2000, AJ, 120, 1579
Yue B., Ferrara A., Pallottini A., Gallerani S., Vallini L., 2015, MNRAS, 450, 3829
Zhang Y. et al., 2024, MNRAS, 530, 1235
Zou H. et al., 2017a, PASP, 129, 064101
Zou H. et al., 2017b, AJ, 153, 276
Zuo S., Chen X., Mao Y., 2023, ApJ, 945, 38
de Sainte Agathe V. et al., 2019, A&A, 629, A85

APPENDIX A: VALIDATION OF FOREST SMOOTHING ASSUMPTIONS

The probability of a given QSO sightline having a Lyα forest absorption field convolved with the whole g band $\delta_{\rm F}^{g\,\rm full}$ for a given observed Lyα forest $\delta_{\rm F}^{g\,\rm obs}$ is a key component of our correlation estimator $\Delta\langle L_{\rm Ly\alpha}\rangle$ (Section 3.3). The determination of this probability $P(\delta_{\rm F}^{g\,\rm full}|\delta_{\rm F}^{g\,\rm obs})$ in a model-independent manner, as laid out in Section 3.2, relies on a key assumption: we can smooth the forest data on scales above the scale of homogeneity $\chi_{\rm h}$, without altering the statistical distribution of $\delta_{\rm F}^{g\,\rm full}$.

Fig. A1 shows the distribution of $\delta_{\rm F}^{g\,\rm full}$ values for our simulation (grey), as well as for resampled pixels of our simulation, without smoothing (blue), and resampled pixels, with smoothing at a homogeneity scale $\chi_{\rm h} = 150$ cMpc (red). The resampled sample without smoothing is significantly narrower than the original distribution of $\delta_{\rm F}^{g\,\rm full}$, while the resampled sample with smoothing is almost identical to the original. For a more quantitative assessment of this claim, Table A1 shows the values of $\delta_{\rm F}^{g\,\rm full}$ for the percentiles corresponding to the 1σ and 2σ intervals. The percentiles for the random smoothed forests agree with the real forest distribution one order of magnitude better than the random forests without smoothing.

When resampling at the scales of our simulation pixels (2.23 cMpc), any correlation is broken, which results in a Lyα absorption field without over/underdensities at scales significantly larger than 2.23 cMpc, but a random succession of equally likely pixel values. The result is a distribution of $\delta_{\rm F}^{g\,\rm full}$ with significantly less variability than the original simulation. On the other hand, if the simulation is smoothed in redshift direction to $\Delta\chi = \chi_{\rm h} = 150$ cMpc, most of the correlation (large over/underdensities) is preserved inside these smoothed pixels. Hence, when we resample from these smoothed pixels to generate new Lyα forests, these pixels display a much larger variability per length unit than the 2.23 cMpc pixels, to the point where the distribution of $\delta_{\rm F}^{g\,\rm full}$ closely resembles the original. Therefore, from Fig. A1, we can safely rely on the assumption that smoothing the Lyα forest data to $\Delta\chi = 150$ cMpc and resampling generates new forests that are statistically identical after applying the g-band smoothing, and thus valid for our study.

The $\delta_{\rm F}^{g\,\rm full}$ from the unaltered simulation forests and the $\delta_{\rm F}^{g\,\rm full}$ from the smoothed and resampled forests follow a virtually identical distribution. However, we also need to examine if the smoothing on scales $\Delta\chi = 150$ cMpc significantly affects $\delta_{\rm F}^{g\,\rm full}$ for the same forest: we do so in Fig. A2. The left panel shows that the values $\delta_{\rm F}^{g\,\rm full\,smooth}$ from smoothed forests are unbiased with respect to the unsmoothed counterpart $\delta_{\rm F}^{g\,\rm full}$, with a small scatter. The right panel shows that this scatter introduces a relative error well within ± 0.5 per cent in $\delta_{\rm F}^{g\,\rm full\,smooth}$ with respect to $\delta_{\rm F}^{g\,\rm full}$. Consequently, the error introduced in

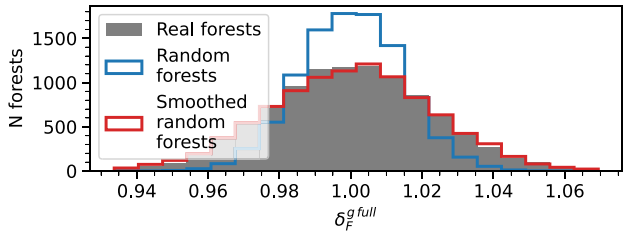

Figure A1. Distribution of $\delta_F^{g\,full}$ for 10 000 simulated forests. In grey, forest data as extracted from our simulation. In blue, randomly generated forests by resampling forest pixels, without smoothing (so the correlation at scales $\chi < \chi_h$ is broken). In red, randomly generated smoothed forests, resampling forests pixels smoothed to $\Delta\chi = \chi_h$.

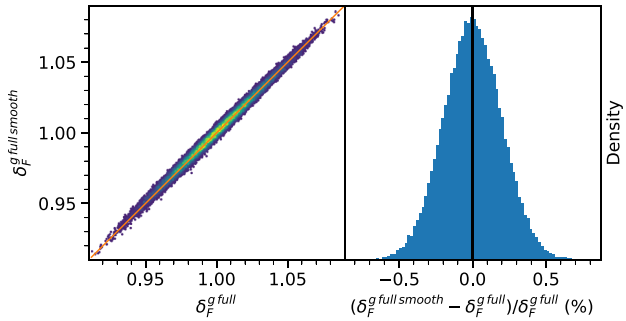

Figure A2. Left panel: scatter plot of $\delta_F^{g\,full}$ versus $\delta_F^{g\,full\,smooth}$. Right panel: histogram of relative error of $\delta_F^{g\,full\,smooth}$ with respect to $\delta_F^{g\,full}$, in per cent.

Table A1. 1σ (16th–84th) and 2σ (2.5th–97.5th) percentiles for the distributions of real, random, and random smoothed forests in Fig. A1.

Percentile	2.5th	16th	84th	97.5th
Real	0.9565	0.9780	1.0225	1.0439
Random	0.9717	0.9858	1.0144	1.0283
Random smoothed	0.9538	0.9770	1.0233	1.0458

$\delta_F^{g\,full}$ by smoothing the forests is negligible. (Note that in Section 3.2, $\delta_F^{g\,full}$ refers to a smoothed forest, since smoothing is necessary before resampling. Only in the context of Fig. A2, we refer to $\delta_F^{g\,full}$ as the g-band convolution of an unsmoothed forest).

APPENDIX B: LYα EMISSION MODEL CUT-OFF

As specified in Section 4.1, if $L_{Ly\alpha} \leq L_{Ly\alpha\,cut}$, our Ly$\alpha$ emission model follows the power law

$$L_{Ly\alpha}(\delta) = C_L \cdot (1+\delta)^{b_e}, \tag{B1}$$

which results in a cosmological field with a given bias b_e (equation 13). C_L in equation (B1) is a normalization constant so the total average luminosity in our simulation matches the input value $\langle L_{Ly\alpha}\rangle$, i.e.

$$\frac{1}{N_{voxels}\,V_{voxel}} \sum_{i=0}^{N} L_{Ly\alpha\,i} = \langle L_{Ly\alpha}\rangle, \tag{B2}$$

where N_{voxels} and V_{voxel} are the number of voxels and the voxel volume in the simulation box, respectively, and $L_{Ly\alpha\,i}$ the Lyα luminosity of voxel i. However, for voxels with $L_{Ly\alpha} > L_{Ly\alpha\,cut}$, the Ly$\alpha$ luminosity is computed with another power law,

$$L_{Ly\alpha}(r) = \frac{L_{Ly\alpha\,max} - L_{Ly\alpha\,cut}}{(\delta_{max} - \delta_{cut})^{\eta}} (\delta - \delta_{cut})^{\eta} + L_{Ly\alpha\,cut}, \tag{B3}$$

which has been determined ad hoc to ensure that the bright end of the Lyα luminosity distribution matches within $\sim$1 dex the LFs for bright QSOs presented in Spinoso et al. (2020). In other words, we assume that the dominant component of Lyα emission for the densest voxels ($\delta > \delta_{cut}$) comes from very bright, resolved QSOs, as well as their surrounding nebulae. Given that the physical volume of our cells is $\sim$613 pkpc at $\langle z\rangle = 2.64$, and that the largest Lyα nebulae detected so far have similar sizes (e.g. 442 pkpc in Cai et al. 2017), this is a perfectly reasonable assumption. The exponent η and maximum voxel luminosity $L_{Ly\alpha\,max}$ of equation (B3) are

$$\eta = 1.55 - 0.3 b_e, \tag{B4}$$

$$L_{Ly\alpha\,max} = 10^4 \langle L_{Ly\alpha}\rangle \cdot V_{voxel}, \tag{B5}$$

These parameters have been chosen to ensure that (1) the bright end of the $L_{Ly\alpha}$ histogram is within $\sim$1 dex of Spinoso et al. (2020) for the range of $L_{Ly\alpha}$ specified in Section 4.1, and (2) the model converges for the range of bias in Section 4.1 ($1.5 \leq b_e \leq 3.5$). Regarding the cut-off luminosity, $L_{Ly\alpha\,cut}$, it is derived from the condition of continuity in the $L_{Ly\alpha}$ distribution, i.e. forcing the derivatives of equations (B1) and (B3) to be equal,

$$L_{Ly\alpha\,cut} = L_{Ly\alpha\,max} - \frac{b_e}{\eta} C_L \cdot (1+\delta_{cut})^{b_e - 1} \cdot (\delta_{max} - \delta_{cut})^{\eta}, \tag{B6}$$

and δ_{cut} is derived from evaluating equation (B1) at $L_{Ly\alpha\,cut}$,

$$\delta_{cut} = \left(\frac{L_{Ly\alpha\,cut}}{C_L}\right)^{1/b_e} - 1. \tag{B7}$$

Consequently, our Lyα emission model has just two entry parameters: $\langle L_{Ly\alpha}\rangle$ and b_e. For a given $(\langle L_{Ly\alpha}\rangle, b_e)$ pair, C_L and $L_{Ly\alpha\,cut}$ are solved numerically from equations (B2) and (B7). A numerical solution is needed because equation (B2) is evaluated over all simulations voxels, and not only over those voxels under $L_{Ly\alpha\,cut}$ that follow equation (B1).

Fig. B1 displays histograms of our Lyα emission model applied to our simulation box, for the four possible maximum/minimum combinations of $\langle L_{Ly\alpha}\rangle$, b_e, together with a comparison with the QSO LFs of Spinoso et al. (2020) at the bright luminosity end, recomputed after removing all QSOs with $g < 19$ (as this is the foreground cut we apply in Section 4.2.1). For this comparison, we have considered each voxel as a point source emitting in Lyα, and compared their distributions to the QSO LFs. We find that, except for the faintest, less biased case [$\langle L_{Ly\alpha}\rangle = 1.5 \times 10^{40}$ erg s^{-1} (cMpc^3)$^{-1}$, $b_e = 1.5$], and the brightest, most biased case [$\langle L_{Ly\alpha}\rangle = 1.5 \times 10^{41}$ erg s^{-1} (cMpc^3)$^{-1}$, $b_e = 3.5$], the bright end of the histograms is at most $\sim$1 dex and $\sim$0.5 mag brighter than the LFs. This difference is consistent with the QSO LFs only considering the resolved Lyα QSO emission in a small aperture, but our simulation including all Lyα luminosity enclosed in a $\sim$560 pkpc^3 volume around a bright QSO (e.g. in the extremely large Lyα nebula described in Cai et al. 2017, only 4 per cent of detected Lyα emission comes from a resolved source). The extreme cases of $\langle L_{Ly\alpha}\rangle$, b_e show a luminosity distribution which may be deemed unrealistic, as it deviates from the QSO LFs of Spinoso et al. (2020) by being approximately half an order of magnitude fainter or an order of magnitude brighter. However, since these represent the extremes of our parameter space, they are the most likely to be far away from the $\langle L_{Ly\alpha}\rangle$, b_e values of

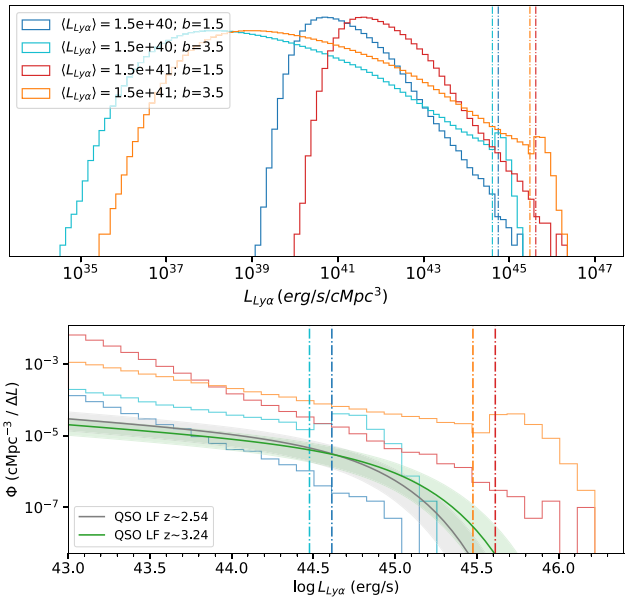

Figure B1. Upper panel: histogram of Lyα luminosity per volume unit of our simulation box, for the four possible combinations of maximum/minimum $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$. The dash–dotted lines are the values of $L_{\rm Ly\alpha\,cut}$ for each case. Lower panel: detail of the histogram for the bright end, plotted together with two of the QSO LFs of Spinoso et al. (2020, i.e. those compatible in redshift with our analysis). Units have been modified to match those of LFs (number density per unit volume versus total luminosity per voxel).

the real Lyα emission field, so a larger deviation from the real QSO LFs is to be expected.

One interesting feature of the model that can be observed in Fig. B1 is that $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ are independent entry parameters. Increasing/decreasing $\langle L_{\rm Ly\alpha}\rangle$ is equivalent to shifting right/left the histogram in Fig. B1, while increasing/decreasing $b_{\rm e}$ just modifies the shape of the histogram (making it become wider/narrower). $L_{\rm Ly\alpha\,max}$ is proportional to $\langle L_{\rm Ly\alpha}\rangle$ in equation (B5), and η only depends on $b_{\rm e}$ in equation (B4), this means that the convergence of the model only depends on $b_{\rm e}$. If the model has converged for a given value of $b_{\rm e}$ (i.e. $L_{\rm Ly\alpha}(\delta)$ is a continuous function for a given histogram shape), the histogram shape will be the same regardless of the value of $\langle L_{\rm Ly\alpha}\rangle$, as $L_{\rm Ly\alpha\,max}$ will shift accordingly.

The independent behaviour of $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ has some useful consequences. First, even though we restrict our study to the values of $\langle L_{\rm Ly\alpha}\rangle$ specified in Section 4.1, our model can simulate any value of $\langle L_{\rm Ly\alpha}\rangle$. Likewise, the model can be solved (i.e. the values of C_L and $L_{\rm Ly\alpha\,cut}$ can be found) for a single pair of ($\langle L_{\rm Ly\alpha}\rangle$, $b_{\rm e}$), and then simply renormalized to any other value of $\langle L_{\rm Ly\alpha}\rangle$. Moreover, the model can be extrapolated to different bias ranges, or even to other hydrodynamic simulations with similar voxel sizes, simply by slightly tuning equation (B4).

APPENDIX C: FORECAST EXTRAPOLATION TO DESI–LSST

Here, we present a simple extrapolation of our forecast for a hypothetical DESI–LSST Lyα IM study. The area for of the DESI–LSST overlap has been assumed to be 3200 deg^2 (the baseline in Bolton et al. 2018). By using the same QSO density distribution used throughout this work (from DESI Collaboration 2016), we expect a total of 132 384 QSOs in the overlap in the g-band redshift range.

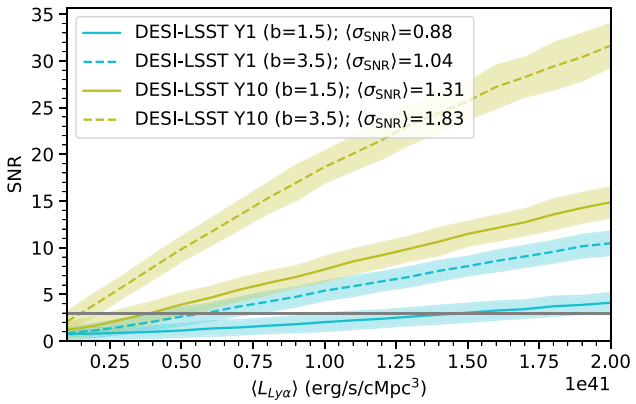

Figure C1. Mean SNR and the 1σ confidence interval (coloured area), versus $\langle L_{\rm Ly\alpha}\rangle$, for DESI-LSST Y1 (cyan) and DESI-LSST Y10 (olive). Solid lines represent the cases with the minimum bias of our model ($b_{\rm e} = 1.5$), and dashed lines the maximum bias ($b_{\rm e} = 3.5$).

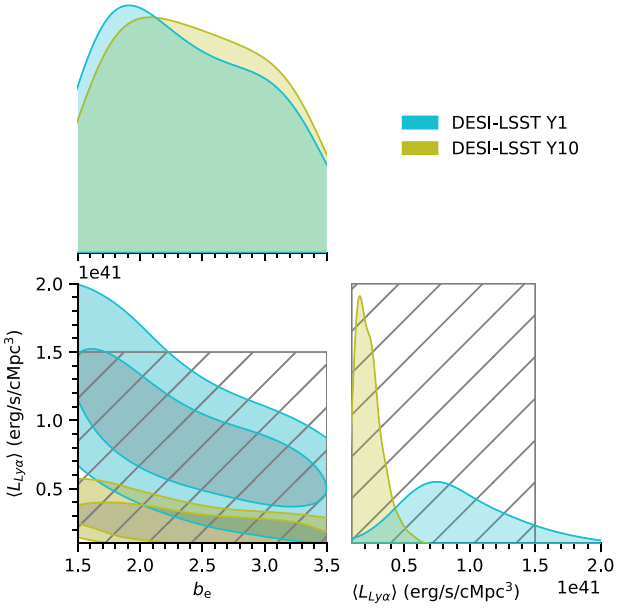

Figure C2. Triangle plot of the upper bounds on $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ placed by a non-detection ($\Delta\langle L_{\rm Ly\alpha}\rangle$ SNR < 3), for DESI–LSST Y1 (cyan) and DESI–LSST Y10 (olive). Contours on the bottom left panel correspond to 1σ and 2σ constraints, hatched areas to the range of $\langle L_{\rm Ly\alpha}\rangle$ we deem realistic.

Table C1. Marginalized upper bounds for $\langle L_{\rm Ly\alpha}\rangle$ and $b_{\rm e}$ for a non-detection in our forecast extrapolated to DESI–LSST, together with probability of detecting Lyα LSS (SNR$_{\Delta\langle L_{\rm Ly\alpha}\rangle} > 3$).

	$\langle L_{\rm Ly\alpha}\rangle$ [10^{40} erg s^{-1} (cMpc^3)$^{-1}$]	$b_{\rm e}$	$P_{\rm detect}$ (per cent)
DESI–LSST Y1	8.8 ± 3.6	2.38 ± 0.56	93.30
DESI–LSST Y10	2.8 ± 1.0	2.36 ± 0.55	100.00

The same BASS g-band IM as in Fig. 10 has been used for imaging noise, but multiplying its flux by a factor $C_{\rm noise}$ determined by the magnitude limits as

$$C_{\rm noise} = 10^{(g_{\rm lim\,BASS} - g_{\rm lim\,LSST})/2.5}, \quad \text{(C1)}$$

where $g_{\mathrm{lim\,BASS}}$ and $g_{\mathrm{lim\,LSST}}$ are the 5σ PSF detection limits for the respective surveys. With this simple scaling, and assuming that all flux contained in our BASS intensity map is noise from non-cosmic origin, the SNR of any hypothetical source of magnitude $g_{\mathrm{lim\,LSST}}$ in LSST would be the same of a hypothetical source of magnitude $g_{\mathrm{lim\,BASS}}$ in BASS. We simulated two different stages of LSST with two different depth limits (LSST Collaboration 2018): LSST Y1, with $g_{\mathrm{lim\,LSST}} = 25.5$ ($C_{\mathrm{noise}} = 0.182$) and LSST Y10, with $g_{\mathrm{lim\,LSST}} = 26.9$ ($C_{\mathrm{noise}} = 0.050$).

By applying this scaling factor C_{noise}, however, we are also artificially reducing the emission of cosmic foregrounds (i.e. the flux of all real galaxies in the intensity map is also downscaled). Nevertheless, in Renard et al. (2021), the cross-correlation between Lyα forest and narrow-band imaging ($i < 23$) could still clearly be detected for 100 deg^2 with cosmic foregrounds alone. Hence, it seems unlikely that cosmic foregrounds on their own could hide the cross-correlation between DESI Lyα forest data and LSST g-band images. Moreover, since we are just rescaling the flux of the BASS intensity map, we also neglect the mitigation of the sky and instrument variability noise components by the very high number of exposures of LSST, as discussed in Section 6.3.

In an analogous manner to the results presented in Section 5.2, Fig. C1 displays the evolution of $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR versus $\langle L_{\mathrm{Ly}\alpha}\rangle$ for $b_{\mathrm{e}} = 1.5$ and 3.5, for both DESI–LSST Y1 and DESI–LSST Y10. Fig. C2 shows the triangle plot of the upper bounds on $\langle L_{\mathrm{Ly}\alpha}\rangle$ and b_{e}, and Table C1 shows the marginalized upper bounds, together with the probability of detection. However, for this case we have only sampled $\langle L_{\mathrm{Ly}\alpha}\rangle$ up to 2×10^{41} erg s^{-1} (cMpc^3)$^{-1}$, since at this level the $\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR is already so high that a detection seems certain ($\Delta\langle L_{\mathrm{Ly}\alpha}\rangle$ SNR $= 4\pm1$ for DESI–LSST Y1 and $b_{\mathrm{e}} = 1.5$).

This paper has been typeset from a TEX/LATEX file prepared by the author.