Chronicles of cosmic dawn: high redshift quasars as probes of supermassive black holes and the intergalactic medium

Onorato, S.

3 Homogeneous measurements of proximity zone sizes for 59 quasars in the Epoch of Reionization

Work publicly available in **S. Onorato**, J. F. Hennawi, E. Pizzati, B. P. Venemans, A.-C. Eilers, 2025, arXiv:2505.09676, currently under review for publication on *Monthly Notices of the Royal Astronomical Society*. Reprinted here in its entirety.

Abstract

The overionized regions surrounding high-redshift quasars, known as proximity zones, provide a window into the interaction between supermassive black holes (SMBHs) and the intergalactic medium (IGM) during the epoch of reionization (EoR). We present new homogeneous measurements of proximity zone sizes ($R_{\rm p}$) for a sample of 59 quasars spanning redshifts $5.77 \leq z \leq 7.54$ (median $z = 6.59$). For 15 of these sources, we measure $R_{\rm p}$ for the first time. The quasars in our catalog have absolute magnitudes at rest-frame 1450 Å in the range $-29.13 \leq M_{1450} \leq -25.20$ (median $M_{1450} \simeq -26.49$), providing one of the most extensive data sets for exploring $R_{\rm p}$ at these epochs. The distribution of $R_{\rm p}$ values shows a large scatter at fixed redshift and luminosity, likely reflecting variations in quasar lifetimes ($t_{\rm Q}$), IGM density fluctuations, and IGM neutral fraction. We fit a bivariate power-law model to a large sample of 100 objects to study the dependence of $R_{\rm p}$ with both M_{1450} and z: we find that the evolution of $R_{\rm p}$ with luminosity is in agreement with the models ($R_{\rm p} \propto 10^{-0.4 M_{1450}/2.87}$), while the evolution of $R_{\rm p}$ with z is steeper than previous works ($R_{\rm p} \propto (1+z)^{-2.44}$). We identify 13 quasars with small proximity zone size, defined using the residuals of our fit. In all cases, except for J2211−6320, we rule out the presence of associated dense absorbers that prematurely truncate $R_{\rm p}$, and suggest a short $t_{\rm Q}$ ($\lesssim 10^4$ yr) as a possible explanation for their small proximity zone sizes.

3.1 Introduction

One of the fundamental goals of modern observational cosmology is to study the epoch of reionization (EoR), during which the intergalactic medium (IGM) transitioned from a predominantly neutral state to the highly ionized state observed today. While considerable progress has been made in constraining the timeline and evolution of this process, critical questions remain regarding the morphology of reionization and the driving physical mechanisms. A principal observational probe of this epoch is the study of Lyα absorption in the spectra of high-redshift quasars. The rapid increase in Lyα optical depth beyond $z \gtrsim 5.5$ and the associated scatter in absorption measurements provide strong evidence for substantial changes in the IGM, consistent with the final stages of reionization (Fan et al. 2006c; Becker et al. 2015b; Eilers et al. 2018a; Yang et al. 2020b; Bosman et al. 2022). The absence of extended Gunn-Peterson troughs (Gunn & Peterson 1965) at $z \lesssim 5.5$ further supports the conclusion that hydrogen reionization was largely complete by this time (McGreer et al. 2015; Bosman et al. 2022).

At $z \gtrsim 6$, constraints on the reionization history are still relatively weak. One of the main ways to probe reionization at these redshifts is by analyzing quasar proximity zones (or near zones), regions of enhanced transmission in quasar spectra immediately blueward of the Lyα emission line. These zones arise from the intense ionizing radiation of the quasar, which produces a local ionized bubble within the surrounding IGM (Cen & Haiman 2000; Haiman & Cen 2001; Bolton & Haehnelt 2007a,b). The proper size of this ionized region ($R_{\rm ion}$) depends on the quasar's ionizing photon emission rate ($\dot{N}_\gamma$), its lifetime ($t_{\rm Q}$), and the ambient neutral hydrogen fraction ($x_{\rm HI}$; Haiman & Cen 2001). In the simplest scenario of a homogeneous, predominantly neutral IGM with a uniform neutral fraction throughout and negligible recombination on quasar timescales, $R_{\rm ion}$ is given by:

$$R_{\rm ion} \approx \left(\frac{3\dot{N}_\gamma t_{\rm Q}}{4\pi n_{\rm H} x_{\rm HI}} \right)^{1/3}, \tag{3.1}$$

where $n_{\rm H}$ is the hydrogen number density. For a highly ionized IGM, instead, the proximity zone reflects the region where the IGM has had time to respond to the quasar's radiation field. In this regime, the size of the proximity zone is governed by the equilibration timescale of the IGM ($t_{\rm eq} \simeq 3{\cdot}10^4$ yr), and the proximity zone becomes sensitive to $t_{\rm Q}$ only when the quasar age is shorter than this timescale (Davies et al. 2020a). On top of that, Equation 3.1 does not fully account for the complexity of the IGM, including pre-existing ionization effects by local galaxies, small-scale density fluctuations, and the presence of overlapping ionized regions (Bolton & Haehnelt 2007a; Lidz et al. 2007; Khrykin et al. 2016). Furthermore, the measured proximity zone size in the spectrum ($R_{\rm p}$; Fan et al. 2006c) may differ from

the physical size $R_{\rm ion}$ due to the presence of residual neutral hydrogen within the ionized bubble, which can produce significant absorption before the light reaches the ionization front (Bolton & Haehnelt 2007a). Accounting for most of these effects is possible using semi-analytical models and radiative transfer simulations (Davies & Furlanetto 2016; Davies et al. 2018b, 2020a). These theoretical efforts offer a way to model and disentangle the effects of the average neutral hydrogen fraction, the quasar luminosity, and the quasar lifetime.

The first systematic measurements of proximity zones were carried out by Fan et al. (2006c), who analyzed 16 quasars at $z \sim 6$ and found that luminosity-scaled proximity zone sizes decrease with increasing redshift, providing evidence for an evolving IGM neutral fraction. Subsequent studies refined these measurements using different, but comparable, resolution spectra (Willott et al. 2007; Carilli et al. 2010; Mortlock et al. 2011; Venemans et al. 2015; Reed et al. 2015, 2017; Eilers et al. 2017, 2020; Mazzucchelli et al. 2017; Bañados et al. 2018; Davies et al. 2018b; Ishimoto et al. 2020; Greig et al. 2022; Satyavolu et al. 2023b; Bigwood et al. 2024). These works confirmed a general decline in the proximity zone size with z, but also reported an intrinsically large scatter, indicating a diversity in quasar properties and environments and highlighting the need for large statistical samples.

In addition to proximity zones, damping wings (Miralda-Escudé 1998) offer a complementary probe of reionization from quasar spectra. They arise due to the presence of neutral regions at $z \gtrsim 6.5$, which imprint a specific off-resonant absorption feature on the red side of the Lyα emission line. As a result, the strength of the damping wing features in $z \gtrsim 6.5$ quasar spectra can provide a direct constraint on the neutral fraction of hydrogen in the IGM (Bañados et al. 2018; Davies et al. 2018b; Greig et al. 2022). Jointly analyzing proximity zone and damping wing effects in high-z quasar spectra offers a promising future avenue to map out the entire evolution of the average hydrogen neutral fraction (Ďurovčíková et al. 2024; Hennawi et al. 2024).

Aside from reionization, proximity zones and damping wings also offer an indirect probe of past quasar activity. For a simple lightbulb model (in which quasars shine continuously for a time $t_{\rm Q}$), and in the case of a predominantly neutral IGM, the size of the ionized region created by a quasar is directly related to the quasar lifetime $t_{\rm Q}$, as described by Equation 3.1. In a highly ionized IGM, the proximity zone size becomes sensitive to quasar lifetimes mainly for $t_{\rm Q} \lesssim t_{\rm eq} \simeq 3 \cdot 10^4$ yr, as described above. Nonetheless, several studies have revealed that some $z \gtrsim 6$ quasars exhibit unexpectedly small proximity zones, which may indicate short quasar lifetimes of $t_{\rm Q} \lesssim 10^4 - 10^5$ yr (Eilers et al. 2017, 2020). A global average of a larger set of high-z quasars suggests quasar activity timescales of $t_{\rm Q} \approx 10^6$ yr (Eilers et al. 2021b).

These findings have major consequences for our understanding of supermassive black holes (SMBH) growth: current theories (e.g., Fan et al. 2023)

suggest that high-z black holes grow from smaller seeds via accretion of material, and this accretion is directly related to quasar activity via a radiative efficiency parameter ($\varepsilon \simeq 0.1$). A short quasar lifetime $t_{\rm Q} \lesssim 10^6 - 10^7$ yr implies that the accretion timescale of SMBHs is not long enough to explain their large $\gtrsim 10^9\,{\rm M}_\odot$ masses inferred from quasar spectra (Davies et al. 2019; Eilers et al. 2021b). This conclusion assumes a simple light-bulb scenario for the radiative history of quasars. Whether more complex quasar lightcurves can account for the small observed proximity zone sizes while being compatible with the constraints coming from SMBH growth is currently unknown (Satyavolu et al. 2023a).

Interestingly, quasar clustering measurements can offer independent constraints on the *total* time quasars are active, on average, across cosmic time (or, equivalently, on the *duty cycle* of quasars, $t_{\rm dc}$). Measurements of quasar clustering at low to intermediate-z generally predict $t_{\rm dc} \sim 10^7 - 10^9$ yr (Haiman & Hui 2001; Martini & Weinberg 2001; Martini 2004; Shen et al. 2007, 2009; White et al. 2012; Pizzati et al. 2024a). Recent measurements of the quasar-galaxy cross-correlation function at $z \gtrsim 6$, however, suggest a lower timescale for quasar activity of $t_{\rm dc} \sim 10^6 - 10^7$ yr (Eilers et al. 2024; Pizzati et al. 2024b), broadly compatible with constraints from proximity zone sizes and in tension with the timescale required for SMBH growth in a standard, Eddington-limited scenario (Eilers et al. 2024).

In summary, the study of proximity zones in high-z quasar spectra offers key insight into the reionization process and the early evolution of SMBHs. However, current measurements are limited by small sample sizes and heterogeneous assumptions between different data sets. In this work, we overcome these limitations by providing uniform proximity zone size measurements for a sample of 59 quasars at $5.77 \leq z \leq 7.54$, the largest to date. We combine three different existing data sets, providing updated measurements of proximity zone sizes for most quasars and adding 15 new measurements, which expand the number of proximity zone sizes available. We then fit a bivariate power-law model to our results and the literature to explore the dependence of $R_{\rm p}$ on quasar luminosity and redshift.

This paper is organized as follows: in Section 3.2, we describe our data set and its properties; in Section 3.3, we explain the method used for the quasar continuum normalization and proximity zone size measurements; in Section 3.4, we detail our measurements, compare them to the literature, fit our data and the literature with a bivariate power-law model to study the evolution of $R_{\rm p}$ with M_{1450} and z, and investigate the cases of small $R_{\rm p}$; in Section 3.5, we present our summary and main conclusions.

3.2 The Data Set

We assemble our high-z quasar data set by selecting a sub-sample of objects from three existing catalogs (Onorato et al. 2025b; D'Odorico et al. 2023; Ďurovčíková et al. 2024), each of which is described in detail in the following sections.

3.2.1 The ENIGMA sub-sample

This work involves a sub-sample of 35 out of the 45 quasars analyzed in Onorato et al. 2025b (hereafter, O25) that we will refer to as the ENIGMA sub-sample. O25 contains 12 Broad Absorption Lines (BAL) quasars (sources with absorption lines with FWHM $\gtrsim$ 2000 km s^{-1}) which in principle we want to exclude from this study since their features could make it difficult to determine the continuum level and cannot be disentangled from IGM absorption (Eilers et al. 2020). However, for two of these BALs (J2348−3054 and J1526−2050; flagged in Table 3.1) the proximity zone sizes were already measured from other works from the literature (see Mazzucchelli et al. 2017 and Bigwood et al. 2024), so we keep them in the sub-sample to make a comparison with previous measurements, but we exclude them from our bivariate power-law fit (see Section 3.4.2). Thus, this newly defined sample contains 35 quasars, where 33 are non-BAL sources and two are classified as BALs in the literature. It spans the redshift range $6.50 < z \leq 7.54$ ($z_{\rm median} = 6.70$) and a range of absolute magnitudes at 1450 Å of $-27.19 \leq M_{1450} \leq -25.20$ ($M_{1450,\rm median} \simeq -26.0$, see Fig. 3.1 and Table 3.1). We include two out of the three unpublished high-z quasars reported in O25; where J0410−0139 is now published in Bañados et al. 2025, and J1917+5003 is included in Belladitta et al. (2025). We exclude the one classified as a BAL (J0430−1445, Belladitta et al. 2025).

The main facilities used to collect the data are optical and near-infrared (VIS and NIR) echelle spectrographs, such as Gemini/GNIRS (Gemini Near-Infrared Spectrograph; Elias et al. 2006a,b), Keck/NIRES (Near-Infrared Echellette Spectrometer; Wilson et al. 2004), and VLT/X-Shooter (Vernet et al. 2011), both VIS and NIR arms. However, for some quasars, the Lyα region was not adequately covered by these instruments. To address this, we supplemented our data set with observations from additional long-slit instruments, including Gemini/GMOS (Gemini Multi-Object Spectrographs; Hook et al. 2004), Keck/LRIS (Low-Resolution Imaging Spectrometer; Oke et al. 1995; Rockosi et al. 2010), and LBT/MODS (Multi-Object Double Spectrographs; Pogge et al. 2010).

The spectra are consistently reduced with the open-source Python-based spectroscopic data reduction pipeline `PypeIt`[1], using versions between 1.7.1

[1] `https://github.com/pypeit/PypeIt`

and 1.14.1 (Prochaska et al. 2020b). More details on the data reduction and quasar properties can be found in O25, which presents the largest medium-to-moderate resolution sample of $z > 6.5$ quasar spectra from ground-based instruments.

3.2.2 The E-XQR-30 sub-sample

An analysis of quasar proximity zones on a sub-sample of the enlarged XQR-30 (E-XQR-30) data has already been performed and described in Satyavolu et al. (2023b). However, all the spectra are publicly available from D'Odorico et al. (2023), so we perform our own analysis including two more objects than Satyavolu et al. (2023b), and compare it with their results. We refer to D'Odorico et al. (2023) for a full description of the data reduction and quasar details, but we summarize here the main information relevant to this work. From the 42 quasars in the E-XQR-30 sample, only 21 are considered, as four sources are already included in the ENIGMA sub-sample, and we exclude all those labeled as BALs, proximate Damped Lyα systems (pDLAs), reddened, and some of the mini BALs in Table 1 from Greig et al. (2024a). The mini BALs included in our sub-sample and bivariate power-law fit are indicated in Table 3.1 (i.e., J0842+1218, J0408−5632, and J1213−1246). The 21 analyzed quasars span the redshift range $5.76 < z < 6.34$ ($z_{\rm median} = 6.03$), and the magnitude range $-29.13 < M_{1450} < -26.40$ ($M_{1450,\rm median} \simeq -26.77$, see Fig. 3.1 and Table 3.1), and all are collected with VLT/X-Shooter.

3.2.3 The FIRE sub-sample

Taking advantage of the public data release from Ďurovčíková et al. (2024) at doi: `10.5281/zenodo.11402934`, we make use of their quasars to enlarge our collection of proximity zone size measurements. We refer to Ďurovčíková et al. (2024) for more details, while we summarize here the relevant information to this analysis. From the 18 quasar spectra they provide, we exclude those already present in the ENIGMA and the E-XQR-30 samples described above, leading to a final sub-sample (FIRE) of three objects in the redshift range $6.37 < z < 6.74$ and magnitude range $-27.65 < M_{1450} < -26.31$ (see Fig. 3.1 and Table 3.1). Their spectra come from Magellan/FIRE (Folded port InfraRed Echellette; Simcoe et al. 2013), VLT/X-Shooter, Keck/MOSFIRE (Multi-Object Spectrograph for Infrared Exploration; McLean et al. 2010, 2012) and Keck/ESI (Echellette Spectrograph and Imager; Sheinis et al. 2002).

3.2.4 Final catalog

To summarize, our final catalog consists of 59 objects coming from the three samples described above (ENIGMA, E-XQR-30, and FIRE). It spans the redshift range $5.77 \leq z \leq 7.54$ and the magnitude range $-29.13 \leq M_{1450} \leq -25.20$ (with $z_{\rm median} = 6.59$ and $M_{1450,\rm median} \simeq -26.49$, see Fig. 3.1). Many of these quasars already have documented analyses of proximity zone sizes in the literature (i.e., Bigwood et al. 2024; Satyavolu et al. 2023b; Greig et al. 2022; Ishimoto et al. 2020; Eilers et al. 2020, 2017; Bañados et al. 2018; Davies et al. 2018b; Mazzucchelli et al. 2017; Venemans et al. 2015). For those in common with other works, we perform new measurements and then compare the different results (see Section 3.4.1). In addition, there are 15 objects that do not have previous proximity zone measurements. Here we provide for the first time measurements of their proximity zones. They are flagged with an asterisk in Table 3.1, where we report all the sources grouped by sample (first ENIGMA, then E-XQR-30, and finally FIRE), and sorted by decreasing redshift. The properties of the quasars are listed in the table as follows: the number of the quasar, the short version of its name, its redshift, the method adopted to determine the redshift, the reference for this measurement, its magnitude (M_{1450}), the reference for the discovery of each quasar, the proximity zone size measurement with associated uncertainties in proper Mpc (pMpc), and notes. The references are listed at the bottom of the table.

We interpret our results along with those coming from several other samples (Bigwood et al. 2024; Ishimoto et al. 2020; Eilers et al. 2020, 2017; Mazzucchelli et al. 2017; Reed et al. 2017, 2015), described in Section 3.4.1. We do not re-analyze these data, but rather we use the published literature measurements, which adopt a methodology similar to this work. For these sources, we quote the median values of redshift and magnitude which are, respectively, $z_{\rm Literature} = 6.10$ and $M_{1450,\rm Literature} \simeq -26.02$ (see Fig. 3.1).

3.2.5 Redshift uncertainties

As already discussed in O25, we address uncertainties in the systemic redshift of each quasar depending on the emission line used for the redshift determination. Systemic redshifts are challenging to obtain due to factors like the broad widths of emission lines, Gunn-Peterson absorption (Gunn & Peterson 1965), and discrepancies between different ionization lines (Gaskell 1982; Tytler & Fan 1992; Vanden Berk et al. 2001; Richards et al. 2002; Shen et al. 2016). Additionally, because of the winds and significant internal motions that many quasars exhibit, the emission lines deviate substantially from the systemic redshift of the host galaxy.

Thus, following Eilers et al. (2017), we conservatively assign a redshift error of $\Delta v = 100$ km s^{-1} for the most accurate determinations based on

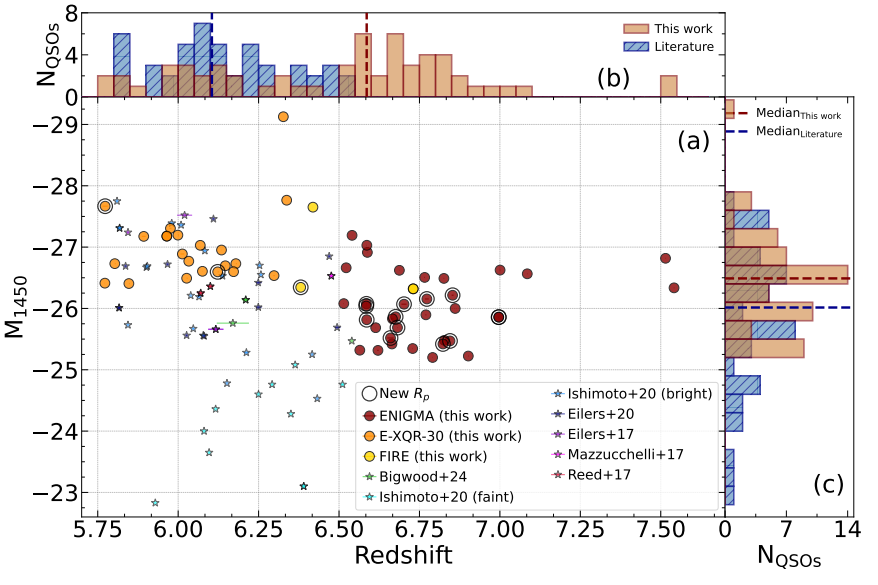


Figure 3.1: *Panel (a)*: Distribution of M_{1450} as a function of z for the subsamples of quasars considered in this study (35 from ENIGMA in red, 21 from E-XQR-30 in orange, and three from FIRE in yellow), and for others from the literature in which the proximity zone sizes have also been measured, i.e., Bigwood et al. 2024 (green), Ishimoto et al. 2020 (both faint and bright samples, respectively in cyan and light blue), Eilers et al. 2020 (dark blue), Eilers et al. 2017 (purple), Mazzucchelli et al. 2017 (magenta), and Reed et al. 2015, 2017 (crimson). The error bars show the uncertainties on z. All the quasars in common between this work and the literature or among themselves have been included only once, always prioritizing this work in the first case and the most recent work in the second one. The big black circles surrounding some points indicate the quasars from this work that did not have previous proximity zone size measurements from the literature. *Panel (b)*: Histogram of the redshift distribution of the quasars considered in this study (light brown) and of the quasars from the literature (blue), with bins of size 0.05. The dashed lines represent the median redshift for this sample ($z_{\text{This work}} = 6.59$, in dark red) and the literature ($z_{\text{Literature}} = 6.10$, in blue). *Panel (c)*: Histogram of the M_{1450} distribution of the quasars considered in this study (light brown) and of the quasars from the literature (blue), with bins of size 0.3. The dashed lines represent the median M_{1450} for this work ($M_{1450,\text{This work}} \simeq -26.49$, in dark red) and the literature ($M_{1450,\text{Literature}} \simeq -26.02$, in blue).

emission lines from the host galaxy's atomic gas reservoir ([C II] lines). However, we state clearly that obtaining this measurement error is not trivial: the typical velocity dispersion (σ_v) measured in quasar host galaxies is approximately 300 km s^{-1} for detections with a signal-to-noise ratio (SNR) greater than 5. The statistical uncertainty on the mean redshift is given by σ_v/SNR $\approx$ 60 km s^{-1}. Consequently, our adopted uncertainty of 100 km s^{-1} represents a conservative estimate. This assumption relies on the expectation that the quasar is embedded within the gravitational potential traced by the cool interstellar medium. We adopt the same uncertainty of Δv = 100 km s^{-1} also for measurements inferred from the narrow component of the [O III] line λ5008.24 Å observed with mid-IR JWST WFSS (*James Webb Space Telescope* Wide Field Slitless Spectroscopy; Kashino et al. 2023; Wang et al. 2023; Yang et al. 2023).

For quasars with redshift measurements from low-ionization lines like Mg II, we assume a redshift error of Δv = 390 km s^{-1} to accommodate the discrepancy between the Mg II line redshift and that of the host galaxy, measured by [C II] line (e.g., Schindler et al. 2020; Mazzucchelli et al. 2017). This value is consistent with previous findings and is further validated by our analysis of a sub-sample of quasars with both Mg II and [C II] redshifts, for which we measure a median velocity offset of $\Delta v \simeq -388$ km s^{-1}, in excellent agreement with Schindler et al. (2020). In Table 3.1 we report redshifts, methods, and references for all the quasars. We use these Δv as uncertainties on the redshift, and consequently, on the position of the Lyα line in the analysis of the proximity zone sizes (see Section 3.3.2.2). Estimates of Δz for every quasar are also available from the literature (see $z_{\rm Ref}$ in Table 3.1), but we do not report them in this work. They are used in Fig. 3.1 as uncertainties on z, and available online in the GitHub repository[2] referenced in O25.

3.2.6 Absolute magnitudes at 1450 Å

We want to carefully determine the quasars' absolute magnitudes at 1450 Å (M_{1450}) as they will be crucial for evaluating the evolution of the proximity zone sizes with luminosity (see Section 3.4.2). The values of M_{1450} are calculated from the flux-scaled spectra of the quasars with J, Y, or K_p-band photometry, as already described in O25. Briefly, we compute the rest-frame wavelengths for every spectrum based on the redshift estimates. From each median flux between 1445 and 1455 Å converted to Jansky ($f_{\nu,1450}$), we determine the luminosity per unit frequency (L_ν) at 1450 Å as:

$$L_\nu = \frac{4\pi {d_L}^2 \cdot f_{\nu,1450}}{1+z}, \tag{3.2}$$

[2] https://github.com/enigma-igm/onorat025_hiz_qsos

where d_L is the luminosity distance to the object at a redshift z. Finally, we find M_{1450} from the luminosity per unit frequency L_ν via:

$$M_{1450} = -2.5 \cdot \log_{10}\left(\frac{L_\nu}{4\pi {d_0}^2 \cdot 3631\mathrm{Jy}}\right), \tag{3.3}$$

where d_0 is the reference distance (10 pc), and 3631 Jy is the zero-point flux density in the AB system. We show the absolute magnitude values M_{1450} as a function of redshift z for this sample in panel (a) of Fig. 3.1, with a comparison to other literature works in which the study of proximity zone size measurements has also been performed. To avoid repetitions among the different samples, all quasars in common between this work and the literature, or among themselves, have been considered only once, always prioritizing this work in the first case and the most recent work in the second one. We use big black circles surrounding some points to indicate quasars from this work that did not have previous proximity zone size measurements from the literature. In panel (b) of Fig. 3.1 we plot the histogram of the redshift distribution of the quasars for both this work and the quasars from the literature, with bins of size 0.05. In panel (c) of Fig. 3.1 we plot the histogram of the M_{1450} distribution for the quasars in this work and for the quasars considered from the literature, with bins of size 0.3.

3.3 Methods

3.3.1 Quasar continuum estimates using PCA

To obtain proximity zone size measurements, we first require a prediction for the unabsorbed quasar continuum in the Lyα forest. The detailed procedure is described in Davies et al. (2018c), but we summarize here the key aspects of the process. For estimating the quasar continuum, we adopt *principal component analysis* (PCA), a statistical technique used to simplify a data set by reducing its dimensionality while retaining most of the variation in the data (Suzuki et al. 2005; Pâris et al. 2011; Davies et al. 2018c). To construct the PCA model, we process 12764 quasar spectra with $2.09 < z < 2.51$ and SNR > 7 from the BOSS DR14 database. The logarithm of each quasar spectrum (F_λ) can be described as:

$$\log F_\lambda \approx \langle \log F_\lambda \rangle + \sum_{i=0}^{N_{\mathrm{PCA}}} a_i A_i, \tag{3.4}$$

where $\langle \log F_\lambda \rangle$ is the mean of the natural logarithm of the spectrum, N_{PCA} is the number of PCA components, A_i are the PCA components, and a_i the coefficients weighting them. This logarithmic space is useful because it accounts for variations between power-law quasar continua. We adopt the

Table 3.1: Information on the 59 quasars in this work, sorted by sub-sample and decreasing redshift. The details on the columns are explained in Section 3.2.4.

N_{QSO}	Name	z	z_{method}	z_{Ref}	M_{1450}	Discovery	R_p [pMpc]	σ_{up}	σ_{low}	Notes
1	J1342+0928	7.5413	[C II]	1	−26.336	2	1.41	0.11	0.11	
2	J1007+2115	7.5149	[C II]	3	−26.818	3	1.40	0.11	0.11	
3	J1120+0641	7.0851	[C II]	1	−26.565	4	1.41	0.17	0.12	
4	J0252−0503	7.0006	[C II]	5	−26.625	6	1.47	0.12	0.12	
5*	J0410−0139	6.9964	[C II]	23	−25.858	23	2.17	0.12	0.12	
6[a]	J2348−3054	6.9018	[C II]	9	−25.224	10	2.26	0.13	0.13	BAL quasar.
7	J0020−3653	6.861	[C II]	32	−25.999	11	2.23	0.12	0.13	
8*	J1917+5003	6.853	Mg II	24	−26.208	24	1.28	0.47	0.47	
9*	J2211−6320	6.8449	[C II]	5	−25.470	6	0.19	0.12	0.12	
10	J0319−1008	6.8275	[C II]	5	−25.480	6	1.77	0.13	0.12	
11	J0411−0907	6.8260	[C II]	5	−26.490	8	3.27	1.18	0.13	
12*	J1129+1846	6.823	Mg II	7	−25.421	12	4.25	0.47	0.47	
13	J0109−3047	6.7909	[C II]	9	−25.200	10	1.33	0.13	0.13	
14*	J0829+4117	6.773	Mg II	7	−26.154	8	3.36	0.48	0.48	
15	J0218+0007	6.7700	[C II]	5	−25.896	7,13	1.01	0.12	0.12	
16	J1104+2134	6.7662	[C II]	5	−26.506	8	4.03	0.12	0.12	
17	J0910+1656	6.7289	[C II]	5	−25.346	8	0.49	0.12	0.13	
18*	J0837+4929	6.702	Mg II	7	−26.069	8	3.10	0.48	0.48	
19	J2002−3013	6.6876	[C II]	5	−26.622	7	2.19	0.12	0.13	
20*	J0923+0753	6.6817	[C II]	5	−25.687	7	1.87	0.12	0.13	
21*	J1048−0109	6.6759	[C II]	14	−25.864	15	2.23	0.12	0.12	
22	J2232+2930	6.666	[C II]	25	−25.831	16	5.02	0.13	2.98	
23	J2102−1458	6.6645	[C II]	5	−25.421	8	2.39	0.12	0.13	
24*	J1216+4519	6.66	Mg II	7	−25.518	8	1.35	0.49	0.49	
25	J0024+3913	6.6210	[C II]	18	−25.316	19	3.26	0.13	0.13	
26	J0305−3150	6.6139	[C II]	26	−25.690	10	3.13	0.13	0.33	
27	J2132+1217	6.5881	[C II]	18	−26.914	18	6.11	0.13	0.13	
28[a]	J1526−2050	6.5864	[C II]	14	−27.030	18	3.98	0.13	0.13	BAL quasar.
29*	J2338+2143	6.586	[C II]	32	−25.816	7	0.53	0.13	0.13	
30*	J1135+5011	6.5851	[C II]	5	−26.075	8	4.18	0.13	0.13	
31*	J1058+2930	6.5846	[C II]	22	−26.039	7	1.94	0.13	0.13	
32	J0921+0007	6.5646	[C II]	5	−25.319	17	2.09	0.13	0.13	
33	J0226+0302	6.5405	[C II]	26	−27.192	16	3.61	0.13	0.13	
34	J0224−4711	6.5222	[C II]	5	−26.663	21	5.97	0.22	0.19	
35	J1110−1329	6.5148	[C II]	14	−26.079	16	1.40	0.13	0.16	
36	J0142−3327	6.3373	[C II]	26	−27.764	30	7.81	0.13	0.14	
37	J0100+2802	6.3269	[C II]	26	−29.126	31	6.99	0.16	0.14	
38	J1030+0524	6.298	[O III]	33	−26.538	34	5.27	0.14	0.19	
39	J0402+2451	6.1793	[C II]	35	−26.730	29	4.13	0.15	0.15	
40	J2356−0622	6.1722	[C II]	36	−26.599	29	2.73	0.17	0.15	
41	J1428−1602	6.1466	[C II]	14	−26.695	29	2.69	0.17	0.17	
42	J1319+0950	6.1347	[C II]	26	−26.952	37	4.96	0.15	0.14	
43*	J1509−1749	6.1228	[C II]	35	−26.596	38	4.49	0.14	0.15	
44	J0842+1218	6.0754	[C II]	26	−26.605	39	6.76	0.15	0.15	Mini BAL quasar (see 46).
45	J1034−1425	6.0687	[C II]	36	−27.029	40,41	1.92	0.14	0.15	
46	J1306+0356	6.033	[C II]	26	−26.770	34	6.33	0.15	0.14	
47	J0408−5632	6.0264	[C II]	35	−26.493	21	2.58	0.15	0.15	Mini BAL quasar (see 46).
48	J0159−3633	6.013	Mg II	42	−26.888	30	3.79	0.56	0.55	
49	J0818+1722	5.9991	[C II]	35	−27.196	43	5.10	0.14	0.15	
50	J0158−2905	5.976	Mg II	42	−27.305	29	4.38	0.56	0.56	
51	J0713+0855	5.9647	[C II]	35	−27.179	29	2.51	0.47	0.14	
52	J1213−1246	5.893	Mg II	42	−27.175	44	1.53	0.57	0.57	Mini BAL quasar (see 46).
53	J1609−1258	5.8468	[C II]	35	−26.406	29	5.93	0.15	0.15	
54	J2033−2738	5.803	[C II]	35	−26.730	29	3.23	2.43	0.16	
55*	J0836+0054	5.773	Mg II	42	−27.667	34	2.76	0.59	0.58	
56	J0927+2001	5.7722	[C II]	45	−26.412	43	4.66	0.16	0.15	
57	J0244−5008	6.7306	[C II]	25	−26.320	11	2.28	0.40	0.13	
58	J1148+5251	6.4189	[C II]	27	−27.651	28	4.47	0.14	0.14	
59*	J1036−0232	6.3809	[C II]	14	−26.344	29	4.83	0.13	0.14	

Notes. Ref: 1 - Venemans et al. (2017); 2 - Bañados et al. (2018); 3 - Yang et al. (2020a); 4 - Mortlock et al. (2011); 5 - Wang et al. (2021b); 6 - Yang et al. (2019); 7 - Yang et al. (2021); 8 - Wang et al. (2019); 9 - Venemans et al. (2016); 10 - Venemans et al. (2013); 11 - Reed et al. (2019); 12 - Bañados et al. (2021); 13 - Matsuoka et al. (2022); 14 - Decarli et al. (2018); 15 - Wang et al. (2017); 16 - Venemans et al. (2015); 17 - Matsuoka et al. (2018); 18 - Mazzucchelli et al. (2017); 19 - Tang et al. (2017); 20 - Bañados et al. (2015); 21 - Reed et al. (2017); 22 - Wang et al. (2024); 23 - Bañados et al. (2025); 24 - Belladitta et al. (2025); 25 - Yang et al. (2023); 26 - Venemans et al. (2020); 27 - Maiolino et al. (2005); 28 - Fan et al. (2003); 29 - Bañados et al. (2016); 30 - Carnall et al. (2015); 31 - Wu et al. (2015); 32 - Bouwens et al., in prep; 33 - Kashino et al. (2023); 34 - Fan et al. (2001); 35 - Private comm. with Sarah Bosman; 36 - Eilers et al. (2021a); 37 - Mortlock et al. (2009); 38 - Willott et al. (2007); 39 - Jiang et al. (2015); 40 - Bañados et al. (2023); 41 - Chehade et al. (2018); 42 - Bischetti et al. (2022); D'Odorico et al. (2023); 43 - Fan et al. (2006b); 44 - Bañados et al. (2014); 45 - Carilli et al. (2007); 46 - Greig et al. (2024a).
*New $R_{\rm p}$ measurement. [a]Excluded from the bivariate power-law fit.

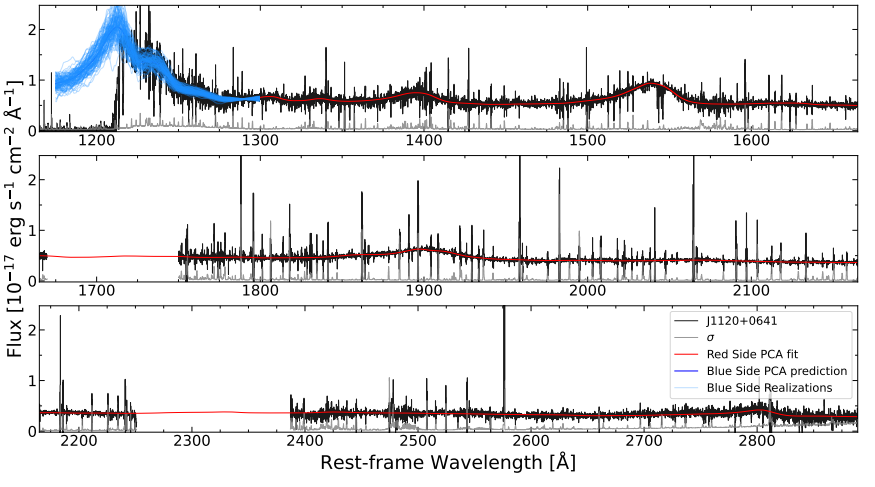


Figure 3.2: Example case for the quasar spectrum of J1120+0641 (black), its associated noise vector σ (grey), and its best-fitting continuum model determined using principal component analysis (PCA). We estimate the continuum across the full spectral range by fitting 10 red principal components to the region $1280 < \lambda_{\rm rest}$ [Å] < 2890 (red curve), and projecting onto a set of 6 blue principal components covering $1175 < \lambda_{\rm rest}$ [Å] < 1280 (blue curve). The 100 realizations of the interpolated PCA projection errors for the blue side are shown in light blue. The spectrum is smoothed for visual purposes, and we have masked regions of strong telluric absorption. The 3-panel plot lets us evaluate the fit at every wavelength in detail. The continuum model reconstructions for all the quasars included in this sample are shown in Fig. 3.10, 3.11, and 3.12.

same parameter choices as in Davies et al. (2018c), decomposing the red side ($1280 < \lambda_{\rm rest}$ [Å] < 2890) and the blue side ($1175 < \lambda_{\rm rest}$ [Å] < 1280) of the spectra independently, keeping 10 red (R_i) and 6 blue (B_i) PCA components. We compute the best estimate for the red PCA coefficients (r_i) by fitting them to the red side of the quasar spectra. Finally, the best estimated r_i are projected onto a set of blue coefficients (b_j) for the blue PCA components using a *projection matrix* (P_{ij}) determined from the training spectra, following:

$$b_j = \sum_{i=1}^{N_{\rm PCA},r} r_i P_{ij}. \tag{3.5}$$

The only difference in our approach is the use of iterative sigma clipping and outlier rejection during the fitting process to mask absorption lines and bad pixels. The errors on the continuum prediction are derived by generating multiple realizations of the PCA model using the covariance matrix of the PCA components.

An example case of a quasar spectrum and its PCA continuum model is shown in Fig. 3.2 with J1120+0641. The spectra and estimated continuum models for all the sources in this work are reported in Fig. 3.10, 3.11, and 3.12. We see that the predicted continua match the data with an overall good agreement. While we could have used other continuum fitting methods, the differences are shown to have a negligible impact on the proximity zone size measurements (Eilers et al. 2017; Greig et al. 2024b).

3.3.2 Proximity zone sizes

3.3.2.1 Measurements

To determine the proximity zone sizes of the quasars in this work, we adopt the definition introduced by Fan et al. (2006c) and widely used in the literature referenced above. First, we normalize the quasar spectra by dividing the observed flux by the estimated continuum reconstruction obtained as described in Section 3.3.1. We then smooth the continuum-normalized flux using a boxcar function with a 20 Å width in the observed frame.

The size of the proximity zone ($R_{\rm p}$) is defined as the proper distance at which the smoothed normalized flux first drops below 10% of the continuum level. Previous work defined that the smoothed normalized flux had to remain below the 10% threshold for at least three consecutive pixels for a proper measurement. However, such a requirement is strongly dependent on the pixel size adopted, which changes from instrument to instrument.

We generalize this condition by consistently re-binning all the spectra to a new wavelength grid with a fixed velocity interval size ($dv = 35.0$ km s^{-1}), and interpolating the smoothed normalized flux onto this new

distance grid. Then we determine the proximity zone size by identifying the distance where the flux drops below 10% of the continuum level for seven consecutive pixels, corresponding to $\gtrsim 205$ km s^{-1}. The choice of a velocity interval $dv = 35.0$ km s^{-1} arises from a compromise between having a finer wavelength grid for the spectra with the coarsest spectral resolution (to decrease the interval between two consecutive pixels) and keeping the number of consecutive pixels required to set the condition reasonably low.

The measurements for all sources in this work are listed in Table 3.1, and reported in Fig. 3.3, 3.4, and 3.5. They span the range $0.19 \leq R_{\rm p}$ [pMpc] ≤ 7.81, with a median size of $R_{\rm p} = 2.76$ pMpc.

3.3.2.2 Errors

We discuss the two main sources of uncertainty affecting our final results. The first one considered is the error on the redshift measurements, arising from the fact that the emission lines used to measure the quasar's redshift (i.e., [C II], [O III], and Mg II) have different associated velocity errors (see Section 3.2.5). It influences the calculation of the physical distance from the quasar, impacting the proximity zone size. The typical errors correspond to $\sigma_{\rm z} \simeq 0.12$ pMpc in the case of z measurements coming from [C II] and [O III] emission lines, and $\sigma_{\rm z} \simeq 0.48$ pMpc if coming from Mg II.

The second source of uncertainty is the error on the continuum fit, which is used to normalize the quasar spectrum (see Section 3.3.1). The PCA fit provides a model of the quasar continuum; any error in this fit affects the normalized flux and thus the determination of $R_{\rm p}$. To estimate this uncertainty, we use the set of PCA continuum realizations. For each of them, we measure the proximity zone sizes using the same method as for the main measurements (see 3.3.2.1). This results in a distribution of $R_{\rm p}$ values, one for each continuum realization. The $16^{\rm th}$ and $84^{\rm th}$ percentiles of this distribution are taken as the lower and upper bounds of the uncertainty due to the PCA fit ($\sigma_{\rm PCA}$). They are generally small in a relative sense, with typical values of a few percent. Most quasars exhibit relative errors $\sigma_{\rm PCA}/R_{\rm p} \lesssim 0.03$, though the most extreme case (i.e., J2033−2738) reaches up to $\sim 75\%$. Finally, we use the square root of the sum of the squares of all the individual errors to provide a total uncertainty on the proximity zone size ($\sigma_{\rm up}$ and $\sigma_{\rm low}$ in Table 3.1).

To assess the impact of the chosen number of consecutive pixels on our measurements, we repeat the analysis using alternative conditions of five ($\gtrsim 135$ km s^{-1}) and nine ($\gtrsim 275$ km s^{-1}) consecutive pixels. In the vast majority of cases, the measured proximity zone sizes remain unchanged. However, we identify a few objects (see Table 3.2) for which the measurements exhibit differences, though they remain consistent within the error bars. Despite these minor variations, the overall statistical trends and conclusions remain robust, confirming that our choice of seven pixels is a reliable

Table 3.2: Proximity zone size measurements for the objects where differences are observed when varying the number of consecutive pixels required for setting the condition. The table lists: the number of consecutive pixels considered, the quasar name, its redshift, and the measured $R_{\rm p}$, along with the associated uncertainties.

$N_{\rm pixels}$	Name	z	$R_{\rm p}$ [pMpc]	$\sigma_{\rm up}$	$\sigma_{\rm low}$
5	J2232+2930	6.666	2.04	0.13	0.12
	J2033−2738	5.803	3.23	0.25	0.16
9	J0411−0907	6.8260	4.44	0.12	1.22
	J2232+2930	6.666	5.02	0.13	0.12
	J0713+0855	5.9647	2.51	4.43	0.14
	J0836+0054	5.773	10.10	0.58	7.36

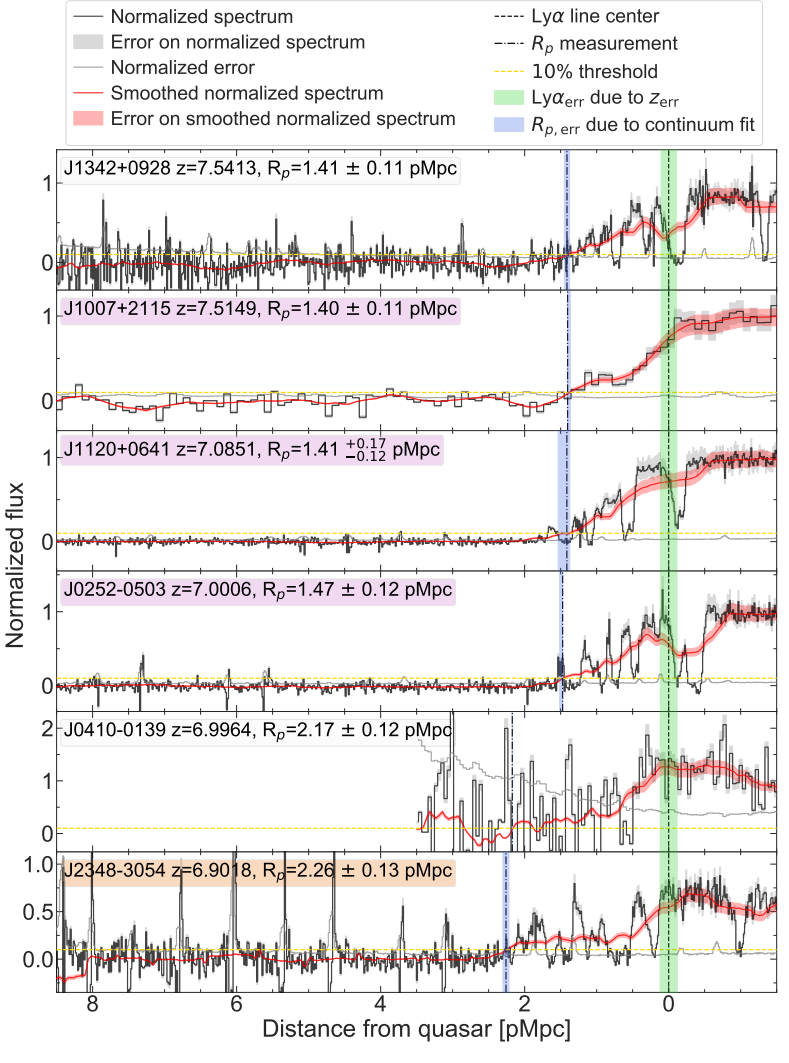
criterion for defining $R_{\rm p}$.

3.4 Results

3.4.1 Comparison with the literature

We compare the proximity zone sizes obtained in this work with the literature on other proximity zone size analyses, such as Bigwood et al. (2024); Satyavolu et al. (2023b); Greig et al. (2022); Ishimoto et al. (2020); Eilers et al. (2020, 2017); Bañados et al. (2018); Davies et al. (2018b); Mazzucchelli et al. (2017); Reed et al. (2017, 2015); Venemans et al. (2015). The most important properties of the samples are summarized in Table 3.3. Fig. 3.6 shows a scatter plot that compares the size of the proximity zones measured in this work with those reported in the literature. Overall, there is a good agreement between our measurements and those from the literature. The few sources showing the largest scatter ($\Delta R_{\rm p} \gtrsim 1$ pMpc) are listed in Table 3.4, with numbers on the first column (N) corresponding to the labels in Fig. 3.6.

From their normalized spectra (see Fig. 3.3 and 3.4), we observe that the smoothed flux remains close to the 10% threshold. This suggests that minor variations in data quality, continuum normalization, or the specific criteria used to define $R_{\rm p}$ could slightly affect the final measurement. In the case of J0024+3913, we suspect the discrepancy may stem from the fact that Mazzucchelli et al. (2017) re-binned their spectra, changing the SNR per re-binned pixel, and altering the effective resolution.



.

Figure 3.3: Proximity zone size measurements of the 35 quasars in the ENIGMA subsample, sorted by decreasing z. The continuum-normalized flux, obtained by dividing the observed flux by the PCA reconstruction, is shown in black and its associated error due to the error in the PCA continuum fit is shown as a grey band. The normalized error on the flux is shown as a grey line. The smoothed normalized flux and its uncertainties are plotted as a red curve with a shaded region. The black dashed line sets the location of the quasar, while the black dash-dotted line marks the distance where the smoothed normalized flux crosses the 10% threshold (dashed yellow line) setting the $R_{\rm p}$ measurement. The uncertainty on the position of the Lyα line due to the error on the redshift measurement is shown in green. The uncertainty on the $R_{\rm p}$ measurement due to the continuum prediction is reported in blue. The box containing name, redshift, and $R_{\rm p}$ is colored in orange to mark the BALs and in pink to highlight the small proximity zones defined in Section 3.4.2.1

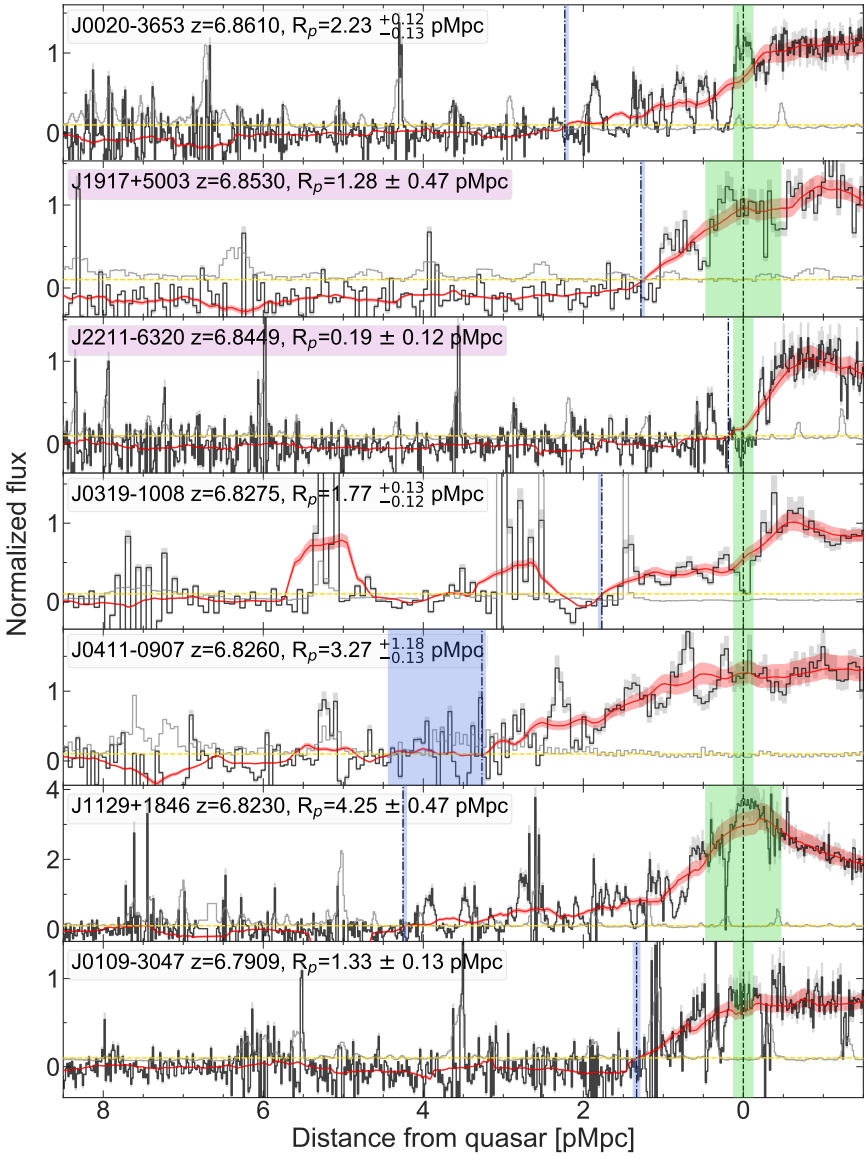


Figure 3.3: (Continued)

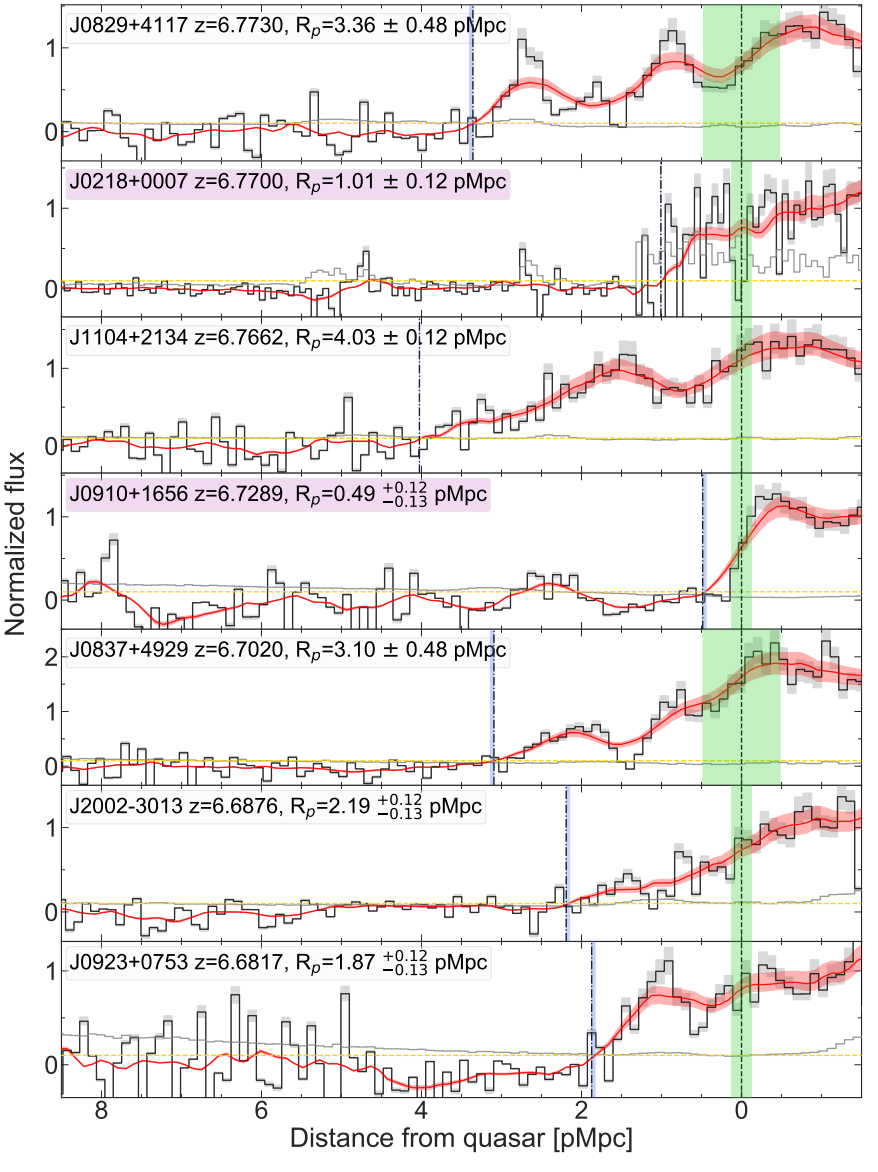


Figure 3.3: (Continued)

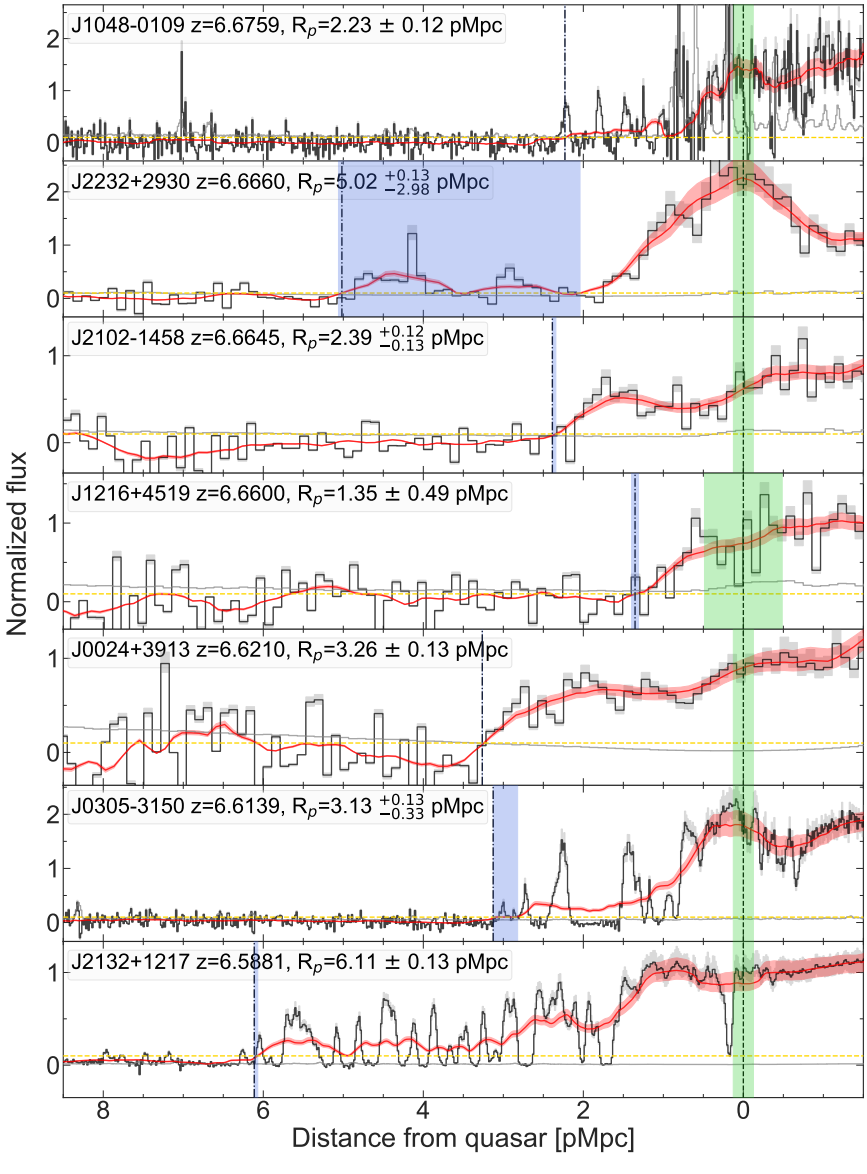


Figure 3.3: (Continued)

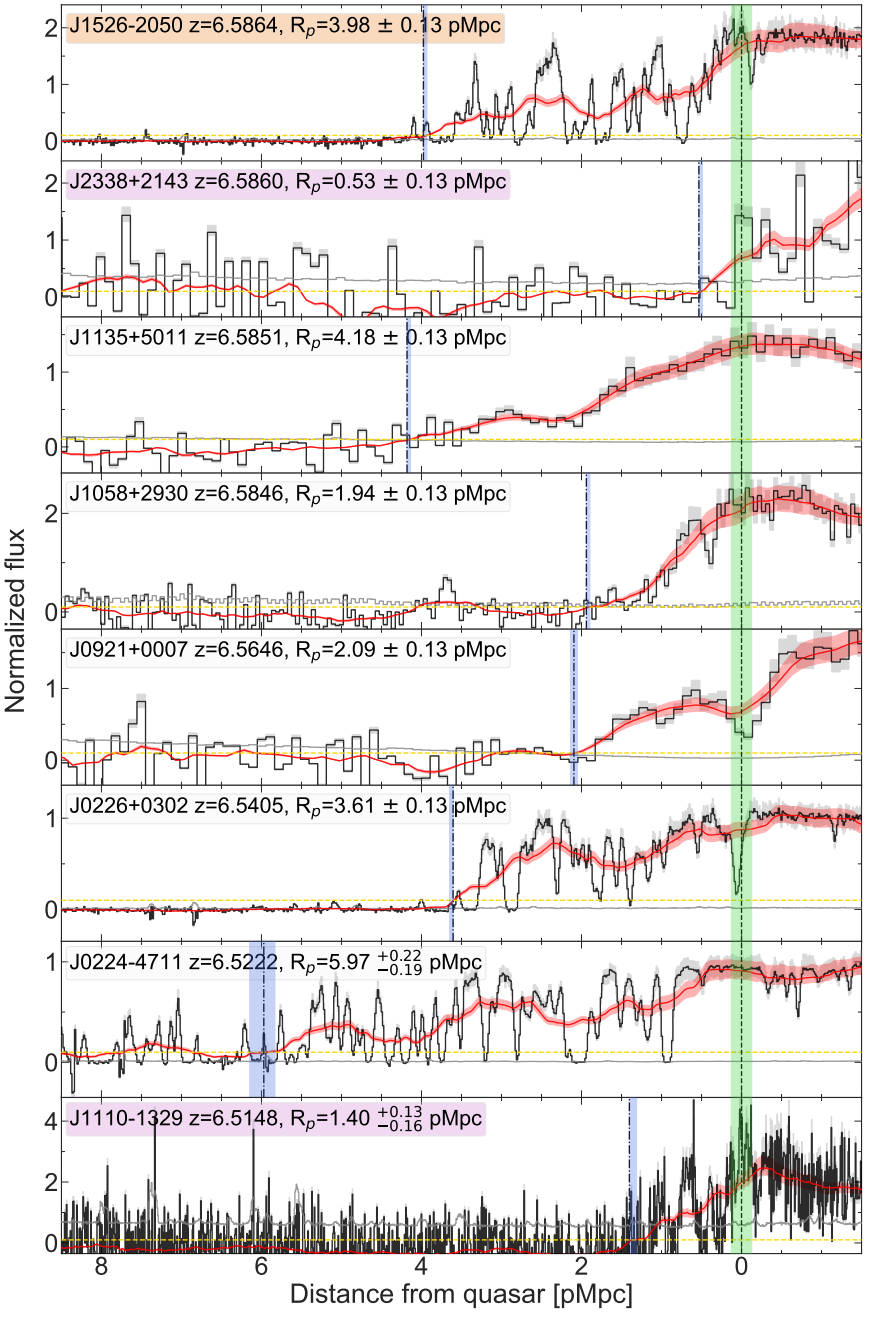


Figure 3.3: (Continued)

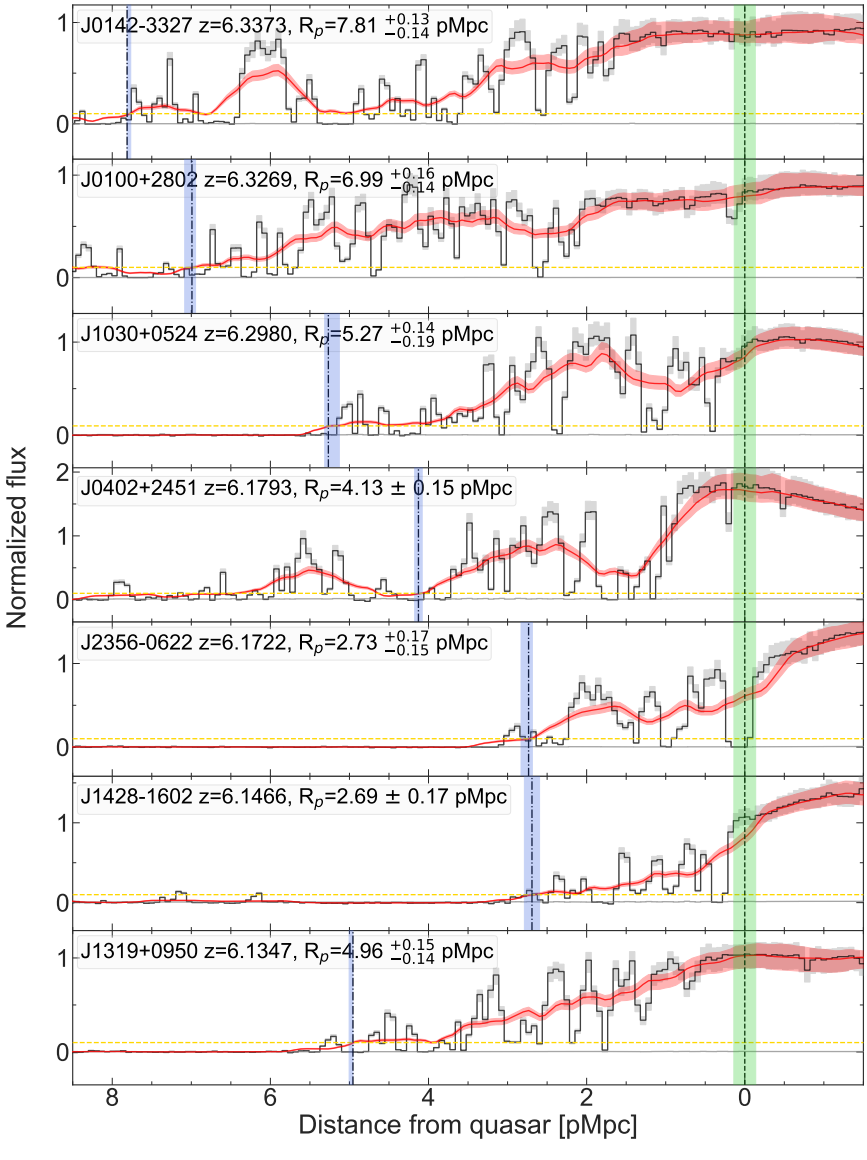


Figure 3.4: Proximity zone size measurements of the 21 quasars in the E-XQR-30 sub-sample, sorted by decreasing z, as described in Fig. 3.3.

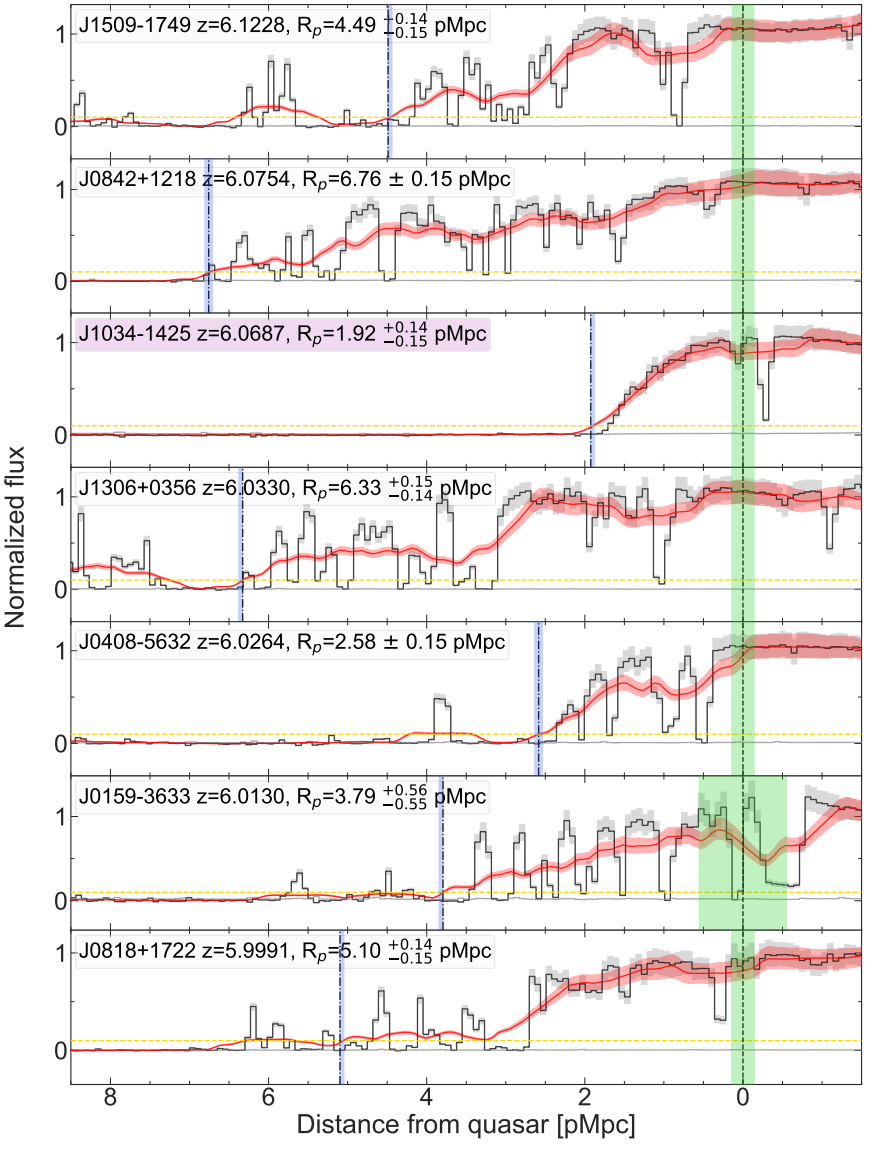


Figure 3.4: (Continued)

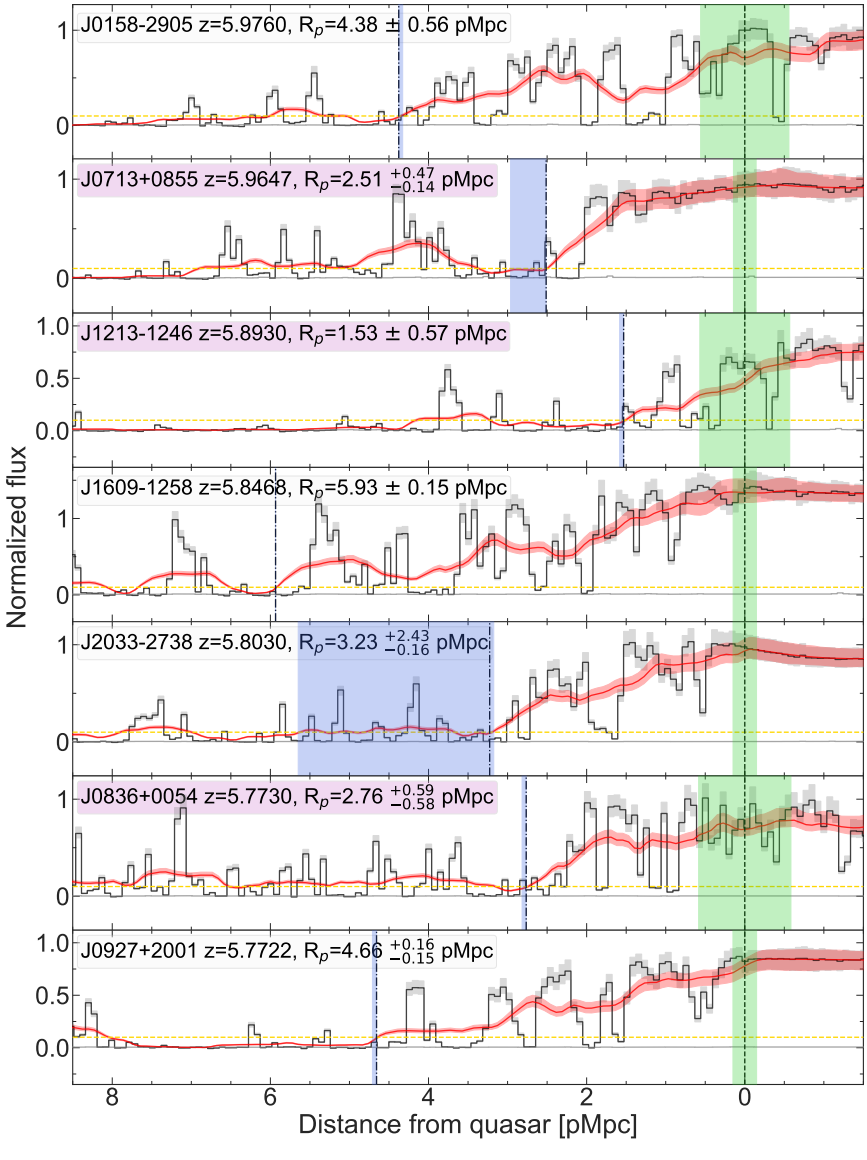


Figure 3.4: (Continued)

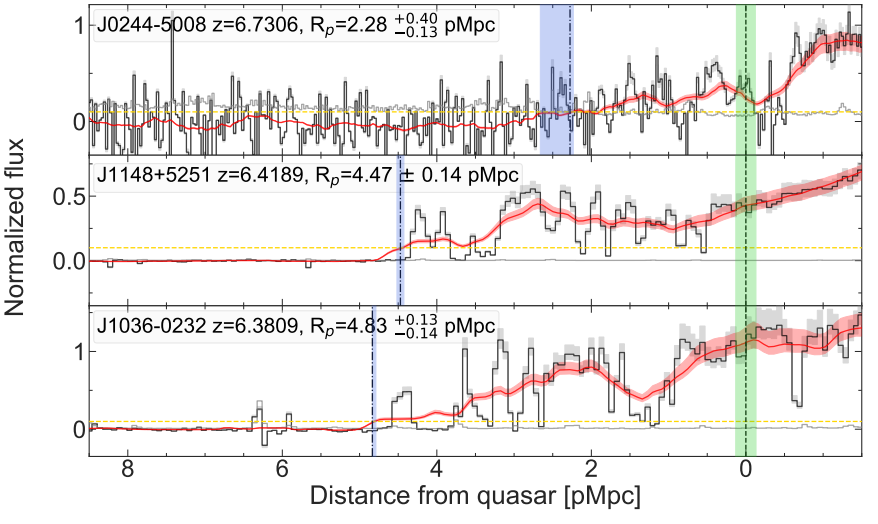


Figure 3.5: Proximity zone size measurements of the three quasars in the FIRE sub-sample, sorted by decreasing z, as described in Fig. 3.3.

Table 3.3: Main properties of the final sample in this paper, compared with those of other samples from the literature in which proximity zone analysis has also been performed. The columns show, respectively: the reference of the sample, the redshift range, the M_{1450} range, the total number of quasars in the sample (with the objects in common with this work), and the instruments used to take the spectra.

Sample	z range	M_{1450} range	N_{QSOs} (in common)	Instruments
Onorato et al. (2025 - this work)	$5.77 - 7.54$	$[-29.13, -25.20]$	59	GNIRS/NIRES/X-Shooter/GMOS/LRIS/MODS/ FIRE/MOSFIRE/ESI
Bigwood et al. (2024)	$6.17 - 6.86$	$[-27.19, -25.32]$	19[a] (14)	FIRE
Satyavolu et al. (2023b)	$5.77 - 6.59$	$[-29.13, -26.41]$	22 (22[b])	X-Shooter
Greig et al. (2022)	$7.00 - 7.51$	$[-26.82, -26.63]$	2 (2)	NIRES/GNIRS/GMOS
Ishimoto et al. (2020) (faint)	$5.93 - 6.56$	$[-25.32, -23.10]$	11 (1)	FOCAS/OSIRIS[c]
Ishimoto et al. (2020) (bright)	$5.77 - 6.54$	$[-27.75, -24.53]$	26 (7)	Spectra from `igmspec`[d] database
Eilers et al. (2020)	$5.82 - 6.49$	$[-27.46, -24.78]$	13 (2)	X-Shooter/DEIMOS[e]
Bañados et al. (2018)	7.54	-26.34	1 (1)	GNIRS/FIRE
Davies et al. (2018b)	$7.09 - 7.54$	$[-26.57, -26.34]$	2 (2)	FORS2[f]/GNIRS/FIRE
Eilers et al. (2017)	$5.77 - 6.54$	$[-29.13, -24.53]$	34[g] (10)	ESI/LRIS
Mazzucchelli et al. (2017)	$6.42 - 7.08$	$[-27.19, -24.89]$	15[h] (10)	X-Shooter/LRIS/FIRE/GNIRS/MODS FORS2/DBSP[i]/Red Channel[j]
Reed et al. (2017, 2015)	$6.07 - 6.25$	$[-26.42, -26.02]$	9[k]	EFOSC2[l]/GMOS/MagE[m]
Venemans et al. (2015)	$6.51 - 6.67$	$[-27.19, -25.83]$	3 (3)	FORS2/FIRE/EFOSC2/LRIS/Red Channel/MODS/LUCI[n]

a Only 15 out of 19 quasars have R_p measurements.
b We measure R_p for all the E-XQR-30 sources in Satyavolu et al. (2023b), and we include two more from D'Odorico et al. (2023).
c Acronyms for Faint Object Camera and Spectrograph (Kashikawa et al. 2002), and Optical System for Imaging and low-intermediate-Resolution Integrated Spectroscopy (Cepa et al. 2000), respectively.
d `https://specdb.readthedocs.io/en/latest/igmspec.html`.
e Acronym for DEep Imaging Multi-Object Spectrograph (Faber et al. 2003).
f Acronym for FOcal Reducer/low dispersion Spectrograph 2 (Appenzeller et al. 1998).
g Only 31 out of 34 quasars have R_p measurements. Also, this sample significantly overlaps with Ishimoto et al. (2020); thus, we favor comparisons with measurements from the latter being the most recent analysis with updated z measurements.
h Only 11 out of 15 quasars have R_p measurements.
i Acronym for DouBle SPectrograph (Oke & Gunn 1982).
j Schmidt et al. (1989).
k Only 5 out of 9 quasars have R_p measurements.
l Acronym for ESO Faint Object Spectrograph and Camera (v.2; Buzzoni et al. 1984).
m Acronym for Magellan Echellette Spectrograph (Marshall et al. 2008).
n Formerly known as LUCIFER (Seifert et al. 2003).

3.4.2 Bivariate power-law fit of quasar proximity zones

Observational and theoretical studies have shown that the proximity zone size depends on both the quasar's UV luminosity and redshift (e.g., Eilers et al. 2017; Davies et al. 2020a). However, the exact nature of these dependencies remains uncertain, particularly at $z > 6.5$, where constraints on quasar lifetimes and IGM neutral fractions are still debated.

In this section, we fit the dependence of R_p on M_{1450} and z, using a Bayesian framework to robustly infer the scaling relations while accounting for measurement uncertainties. Specifically, we fit a bivariate power-law model to the observed R_p values and infer the posterior distributions of the parameters via Markov Chain Monte Carlo (MCMC) sampling. This approach allows us to quantify uncertainties and assess the robustness of our results. We adopt the following parametric functional form for R_p:

$$R_\mathrm{p}(M_{1450}, z; \theta) = \mathrm{C} \cdot \left(\frac{1+z}{7}\right)^{\beta} \cdot 10^{-0.4(M_{1450}+27)/\alpha}, \tag{3.6}$$

where $\theta = (\mathrm{s}, \mathrm{C}, \beta, \alpha)$ represents the parameters to be inferred. This form simultaneously accounts for the evolution of proximity zones with redshift and luminosity and is therefore well-suited for fitting the data. Given our data set $D = \{M_{1450,i}, z_i, R_{\mathrm{p},i}, \sigma_i\}$ with measured proximity zone sizes $R_{\mathrm{p},i}$ and associated uncertainties σ_i, we introduce an intrinsic scatter term, s, to account for stochastic effects (i.e., cosmic variance due to the IGM, variations in quasar lifetimes, and spatial fluctuations in x_HI due to the reionization). We assume Gaussian-distributed deviations and define the log-likelihood function as:

$$\ln \mathcal{L}(D|\theta) = -\frac{1}{2} \sum_i \left[\frac{(R_{\mathrm{p},i} - R_\mathrm{p}(M_{1450,i}, z_i; \theta))^2}{\sigma_i^2 + s^2} + \ln(2\pi(\sigma_i^2 + s^2)) \right]. \tag{3.7}$$

This formulation accounts for both the measured uncertainties and the additional scatter parameter s, ensuring a more realistic assessment of observational errors. We impose uniform priors on the parameters: $0 < s < 10$, $0 < C < 10$, $-10 < \beta < 10$, and $0 < \alpha < 10$. The posterior probability is given by Bayes' theorem: $P(\theta|D) \propto P(D|\theta)P(\theta)$; which in logarithmic form becomes: $\ln P(\theta|D) = \ln \mathcal{L}(D|\theta) + \ln P(\theta)$.

To sample the posterior, we use the affine-invariant MCMC ensemble sampler implemented in the Python package `emcee` (Foreman-Mackey et al. 2013), initializing 32 walkers around an optimized starting point. This initial estimate is obtained via optimization, finding a plausible parameter set before MCMC sampling. We run the chains for 50000 steps per walker and discard the first 5000 steps (10%) of each chain as burn-in. We monitor the acceptance fraction of the MCMC chains to ensure proper mixing and convergence.

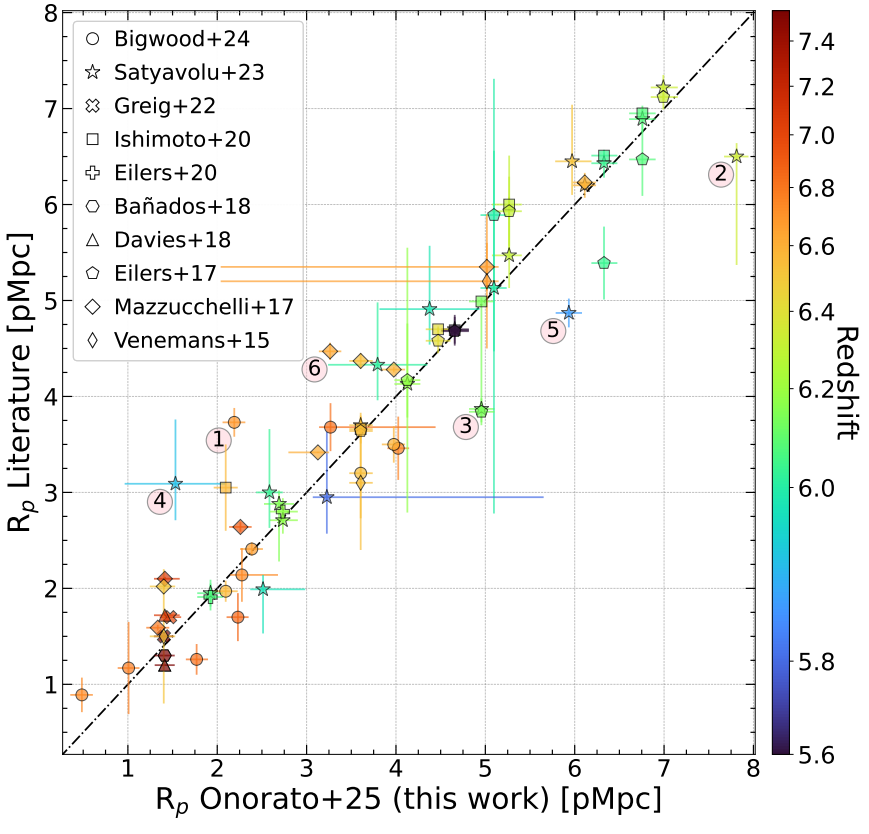


Figure 3.6: $R_{\rm p}$ measured in this work compared with those from the literature. The different markers represent data from various studies, while their color is determined by the redshift of the quasars, according to the colormap to the right. The error bars are the uncertainties on $R_{\rm p}$ estimated from the corresponding study. The 1:1 dash-dotted line indicates where the proximity zone sizes from this work and the literature would be equal. There is good agreement between $R_{\rm p}$ measured in this work and others from the literature. The six quasars where the scatter between our measurements and literature ones is $\Delta R_{\rm p} \gtrsim 1$ pMpc are marked with a number in a pink circle, corresponding to the number N in Table 3.4, with their description.

Table 3.4: Quasars for which the proximity zone size measured in this work differs by $\Delta R_{\rm p} \gtrsim 1$ pMpc from literature values. The columns show, respectively: the number N corresponding to the quasar label in Fig. 3.6, its name, its redshift, and the $R_{\rm p}$ measurements with errors, from both this work and the literature. The letters correspond to the literature samples from which the measurements are taken, listed below.

			$R_{\rm p}$ [pMpc]	
N	Name	z	This work	Literature
1	J2002−3013	6.6876	$2.19^{+0.12}_{-0.13}$	(a) 3.73 ± 0.15
2	J0142−3327	6.3373	$7.81^{+0.13}_{-0.14}$	(b) $6.50^{+0.14}_{-1.13}$
3	J1319+0950	6.1347	$4.96^{+0.15}_{-0.14}$	(b) $3.87^{+1.08}_{-0.14}$ (c) 3.84 ± 0.14 (d) 4.99 ± 0.04*
4	J1213−1246	5.893	1.53 ± 0.57	(b) $3.09^{+0.67}_{-0.38}$
5	J1609−1258	5.8468	5.93 ± 0.15	(b) 4.87 ± 0.15
6	J0024+3913	6.6210	3.26 ± 0.13	(e) 4.47 ± 0.09

Ref: a - Bigwood et al. (2024); b - Satyavolu et al. (2023b); c - Eilers et al. (2017); d - Ishimoto et al. (2020); e - Mazzucchelli et al. (2017).
* Similar to our measurement.

We run the fit for the first time on a large ensemble of 105 quasars, made by both this work and the literature, after removing the two quasars in our sample classified as BALs (i.e., J2348−3054 and J1526−2050), as they could introduce systematic effects in our $R_{\rm p}$ measurements. The resulting constraints on the model parameters are reported in Table 3.6 (No BALs) as median values with 1σ uncertainties ($16^{\rm th}$ and $84^{\rm th}$ percentiles). We evaluate our result by computing the residuals (χ) of the fit between the data (observed $R_{\rm p}$) and the model predictions. Specifically, we define:

$$\chi = \frac{\text{data} - \text{model}}{\sigma_{\rm eff}} \tag{3.8}$$

where $\sigma_{\rm eff} = \sqrt{\sigma^2 + s^2}$ represents the effective uncertainty on the data, combining the real errors on $R_{\rm p}$ and the scatter parameter inferred from the MCMC. The distribution of χ is shown in Fig. 3.13. In Section 3.4.2.1, we use these residuals to define a threshold ($\chi \leq -0.75$) to rigorously identify quasars with small $R_{\rm p}$ and investigate them for the presence of possible associated absorbers prematurely truncating the proximity zone.

3.4.2.1 Possible contaminants in proximity zones

Small proximity zone sizes (typically $R_{\rm p} \lesssim 2$ pMpc) can be interpreted as signatures of short quasar lifetimes ($t_{\rm Q} \lesssim 10^4$ yr). This interpretation depends on the ionization state of the surrounding IGM. If it is significantly neutral, a small proximity zone arises because the quasar's ionization front has not had sufficient time to propagate outward and fully ionize the surrounding gas (Eilers et al. 2017, 2020). Conversely, if the IGM is already highly ionized, $R_{\rm p}$ is sensitive to the quasar lifetime only when $t_{\rm Q} \lesssim t_{\rm eq} \sim 3 \times 10^4$ yr (see Section 3.1). These measurements are critical for constraining the growth timescales of SMBHs at early cosmic epochs, which must accrete substantial mass within $\lesssim 1$ Gyr of the Big Bang (Volonteri 2010, 2012).

However, the interpretation of a small $R_{\rm p}$ as evidence for a young quasar can be confounded by the presence of absorption systems that truncate the transmitted Lyα flux. These systems can absorb ionizing photons or Lyα photons and thus artificially shorten the observed proximity zone, independent of the quasar's age. A rigorous classification of such contaminants is necessary to avoid misidentification. Absorbers can be broadly divided into two categories:

- Intervening absorbers, which lie along the line of sight at redshifts lower than the quasar, and hence are not physically associated with its environment. These typically have limited impact on proximity zone measurements unless they are exceptionally strong metal absorption systems, with line equivalent width EW > 1 Å.

- Associated absorbers, which reside at or near the quasar redshift, often within its host halo or immediate surroundings. This is the most relevant class of contaminants for proximity zone analyses.

Both types of absorbers can, in principle, affect proximity zone measurements, but only *associated*, optically thick absorbers are considered true contaminants in the context of $R_{\rm p}$ interpretation. Below, we outline several key scenarios, previously discussed in Eilers et al. (2017, 2020):

1. BALs (and possibly mini BALs): these are intrinsic to the quasar and arise from high-velocity outflows, manifesting as broad, blue-shifted absorption features. They can absorb Lyα and associated metal-line photons, leading to suppression of flux near the systemic redshift. As they alter the continuum and transmitted flux in the proximity zone region, BAL quasars are usually excluded from proximity zone studies.

2. Optically thick associated absorbers (Lyman Limit Systems - LLSs and pDLAs): these include systems with column densities $N_{\rm HI} \simeq 10^{17-20}$ cm^{-2} (LLSs) and $N_{\rm HI} \gtrsim 10^{20}$ cm^{-2} (DLAs), located within a

few thousand km s^{-1} of the quasar redshift. These absorbers, which likely reside in the quasar's circumgalactic medium (CGM), are self-shielded and optically thick to both ionizing and Lyα photons, and they can truncate the proximity zone by absorbing the entire ionizing flux at their location. From a modeling standpoint, these systems break the assumptions of Lyα forest simulations, which assume a low-density optically thin IGM, and therefore must be excluded from samples used to infer quasar ages. Identifying these systems requires detection of saturated Lyα absorption (i.e., line-black regions) in combination with strong low-ionization metal lines (e.g., O I, Si II, C II).

3. Optically thin associated absorbers: these exhibit narrow high-ionization metal-line systems (e.g., C IV, Si IV) which are detected without associated saturated Lyα absorption, and are optically thin to ionizing photons. Quasars showing such systems are often retained in proximity zone analyses, as their impact on the observed $R_{\rm p}$ is limited. However, some caution is warranted: such systems may trace previously optically thick gas that has been photoionized away by the quasar. If so, the absorber may still alter the ionization state of the gas in a manner not captured by optically thin Lyα forest models. While these subtleties are usually neglected in practice, they highlight the importance of carefully evaluating the spectral context of each system.

It is therefore imperative to conduct a thorough assessment of each quasar with a small proximity zone to exclude the presence of these contaminants. Unlike previous studies, where small proximity zones were usually defined as $R_{\rm p} \lesssim 2$ pMpc (see Eilers et al. 2017, 2020; Satyavolu et al. 2023b), here we take advantage of our bivariate power-law fit described in Section 3.4.2 and identify such objects as those with residuals $\chi \leq -0.75$ (see Fig. 3.13). They are highlighted in Fig. 3.3 and 3.4 by coloring in pink the box containing their details.

We perform a detailed metal absorbers analysis of these quasars (see Appendix 3.8), and we summarize our main results in Table 3.5. We find that the proximity zone size of the quasar J2211−6320 (see Fig. 3.15) may be truncated by the presence of an associated absorber, and we exclude it from our fit, along with the two BALs in our sample already excluded (i.e., J2348−3054 and J1526−2050) and the other identified four contaminated sources from the literature that exhibit BAL (J1558−0724, J1743+4124), pDLA (J0346−1628) features, or $R_{\rm p}$ truncated by metal absorbers (J0840+5624; see Table 3.5). Our best-fit parameters are those obtained by excluding these seven contaminated quasars, and we report them in Table 3.6 and Fig. 3.7. In Fig. 3.8 and 3.9, these objects excluded from the fit are marked with a black cross.

Table 3.5: Quasars with small R_p, identified as those with $\chi \leq -0.75$ from the residuals of the fit in Section 3.4.2. For both this work and the literature, we list the quasar name, its redshift, the proximity zone size measurement with associated uncertainties in pMpc, and notes. In the last column, we indicate whether metal absorption systems (abs) have been identified in the quasar spectrum, whether these are likely to affect R_p, or if the spectrum suffers from other limitations (e.g., low SNR, poor continuum fit, BAL or pDLA features).

Sample	Name	z	R_p [pMpc]	σ_{up}	σ_{low}	Notes
Onorato et al. 2025 (this work)	J1007+2115	7.5149	1.40	0.11	0.11	No abs identified.
	J1120+0641	7.0851	1.41	0.17	0.12	Many abs, not affecting R_p (see 5).
	J0252−0503	7.0006	1.47	0.12	0.12	Many abs, not affecting R_p (see 6).
	J1917+5003*	6.853	1.28	0.47	0.47	Abs at $z_{abs} \simeq 3.687$, not affecting R_p.
	J2211−6320*,[a]	6.8449	0.19	0.12	0.12	Abs at $z_{abs} \simeq 6.8447$, affecting R_p.
	J0218+0007	6.7700	1.01	0.12	0.12	No abs identified.
	J0910+1656	6.7289	0.49	0.12	0.13	No abs identified.
	J2338+2143*	6.586	0.53	0.13	0.13	Low SNR to find abs.
	J1110−1329	6.5148	1.40	0.13	0.16	Low SNR to find abs.
	J1034−1425	6.0687	1.92	0.14	0.15	Many abs, not affecting R_p (see 1, 3, 7).
	J0713+0855	5.9647	2.51	0.47	0.14	Many abs, not affecting R_p (see 1, 7).
	J1213−1246	5.893	1.53	0.57	0.57	Many abs, not affecting R_p (see 7).
	J0836+0054*	5.773	2.76	0.59	0.58	Many abs, not affecting R_p (see 7).
Literature	J1406−0116	6.292	0.14	0.05	0.05	(2, faint): noisy spectrum, poor PCA fit.
	J0330−4025	6.249	1.68	0.62	0.38	(3): no abs identified.
	J2229+1457	6.1517	0.48	0.04	0.04	(2, bright; 3): low SNR to find abs.
	J1558−0724[a]	6.1097	1.32	0.14	0.14	(3): BAL, might affect R_p.
	J2100−1715	6.0806	0.37	0.14	0.14	(3): no abs identified.
	J1743+4124[a]	6.0263	1.04	0.14	0.14	(3): BAL, might affect R_p.
	J0346−1628[a]	5.967	0.79	0.14	0.14	(3): pDLA, might affect R_p.
	J1335+3533	5.9012	0.70	0.10	0.10	(2, bright; 4): no abs identified.
	J0840+5624[a]	5.8441	0.88	0.15	0.15	(4): Many abs, affecting R_p.
	J0017+1705	5.8165	1.16	0.15	0.15	(3): abs at $z_{abs} \simeq 5.8165$, not affecting R_p.

Ref: 1 - Satyavolu et al. (2023b); 2 - Ishimoto et al. (2020); 3 - Eilers et al. (2020); 4 - Eilers et al. (2017); 5 - Bosman et al. (2017); 6 - Wang et al. (2020); 7 - Davies et al. (2023). *New R_p measurement. [a]Excluded from the bivariate power-law fit.

Table 3.6: Constraints (median values, 16^{th}, and 84^{th} percentiles) on the model parameters of the bivariate power-law fit. The first row (No BALs) is obtained after excluding the two BAL quasars in the sample (i.e., J2348−3054 and J1526−2050). The second row is our best-fit, obtained after excluding all the contaminated sources identified in Section 3.4.2.1 (i.e., BALs, pDLAs, quasars whose $R_{\rm p}$ is affected by metal absorbers). These are based on the corner plots in Fig. 3.7.

	s	C	β	α
No BALs	$1.52^{+0.12}_{-0.11}$	4.06 ± 0.23	$-1.92^{+0.97}_{-1.01}$	$2.95^{+0.60}_{-0.42}$
Best-fit	$1.39^{+0.11}_{-0.10}$	4.36 ± 0.22	$-2.44^{+0.89}_{-0.92}$	$2.87^{+0.48}_{-0.35}$

3.4.2.2 Correlation with luminosity

The evolution of proximity zone sizes with the luminosity could depend on different factors, one of them being the ionization state of the surrounding IGM. From Equation 3.1 we see that the size of the ionization front expanding into a neutral IGM evolves as $R_{\rm p} \propto \dot{N}_{\gamma}^{1/3}$. However, the analytical model from Bolton & Haehnelt (2007a) shows that $R_{\rm p} \propto \dot{N}_{\gamma}^{1/2}$ in a highly ionized medium. Given the patchy topology of the IGM at these redshifts (e.g., Davies et al. 2018b), the expected scaling should fall between the two limits. We also consider the predicted evolutions inferred by the simulations in Eilers et al. (2017) for a mostly neutral (Equation 3.9) and ionized (Equation 3.10) IGM, respectively, described by the following power-laws:

$$R_{\rm p} \approx 5.03 \cdot 10^{-0.4(M_{1450}+27)/2.45}, \tag{3.9}$$

$$R_{\rm p} \approx 5.57 \cdot 10^{-0.4(M_{1450}+27)/2.35}, \tag{3.10}$$

implying that $R_{\rm p} \propto 10^{-0.4M/\alpha} \propto L^{1/\alpha} \propto \dot{N}_{\gamma}^{1/\alpha}$, with $\alpha = 2.45$ and 2.35, from Equations 3.9 and 3.10, respectively. In Fig. 3.8, we consider these predictions together with our data and the results obtained from our bivariate power-law fit, finalized in Section 3.4.2.1. We present the distribution of $R_{\rm p}$ as a function of the quasars' magnitude, color-coded with their redshift. Given the large sample size of this work and the literature, the sources considered span in a wide magnitude range $-29.13 \leq M_{1450} \leq -22.83$, in which the distribution of proximity zone sizes shows significant variation and scatter.

We point out the presence of many outliers, highlighted by black squares. These objects show small $R_{\rm p}$ and possibly have short lifetimes ($t_{\rm Q} \lesssim 10^4$ yr), except for those also marked with a black cross, identified as contaminated quasars and, as such, excluded from our final fit (see Section 3.4.2.1). We plot a distribution of curves from our fit, obtained by varying the redshift from $z = 7.55$ (the darkest red curve) to $z = 5.76$ (the darkest blue curve).

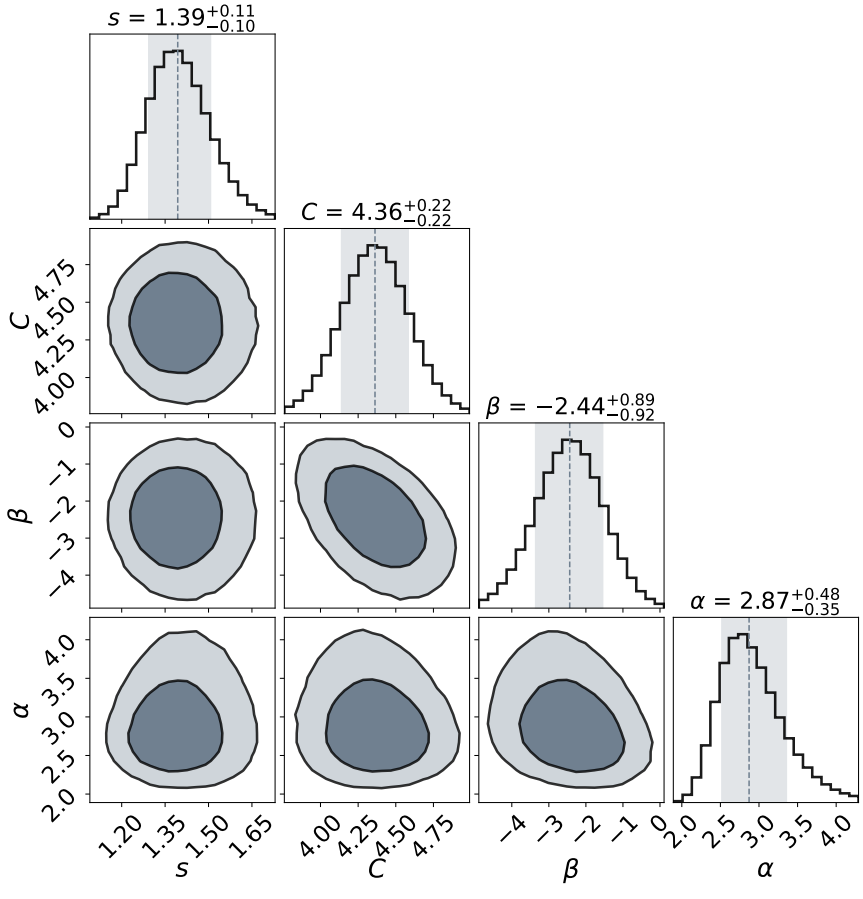


Figure 3.7: Corner plots of the 4D posterior distribution for the fit described in Section 3.4.2. They represent our best-fit, obtained after excluding all the contaminated sources identified in Section 3.4.2.1 (i.e., BALs, pDLAs, quasars whose $R_{\rm p}$ is affected by metal absorbers). The contours in the 2D histograms highlight the 1σ, and 2σ regions, whereas the dashed lines in the 1D histograms represent the median values of the parameters with the 1σ errors shown as shaded regions. These values are reported on top of each histogram.

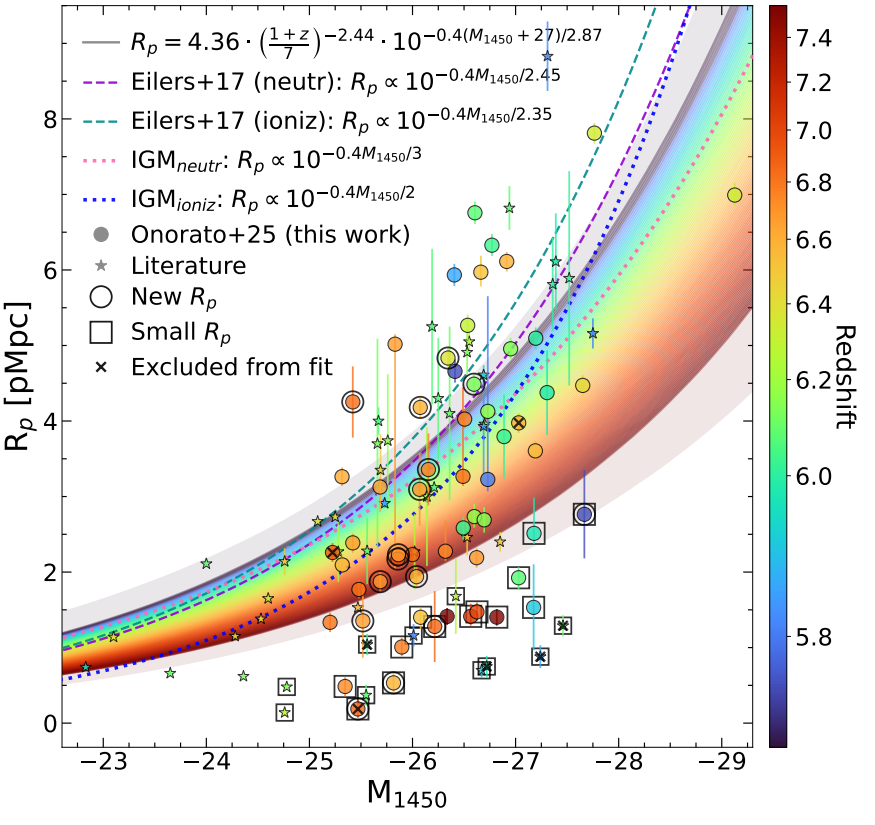


Figure 3.8: Distribution of $R_{\rm p}$ as a function of M_{1450} for the quasars included in this study (points) and for others from the literature (stars), with the corresponding error bars. The color of the symbols is determined by the redshift of the quasars, according to the colormap to the right. The big black circles surrounding some points highlight the new $R_{\rm p}$. The black squares mark the objects with small $R_{\rm p}$, identified from the residuals of the fit in Section 3.4.2 ($\chi \leq -0.75$). The quasars excluded from the plotted final fit (i.e., BALs, pDLAs, $R_{\rm p}$ truncated by metal absorbers) are marked with a black cross. We show a distribution of curves from our bivariate power-law fit, obtained by varying the redshift from $z = 7.55$ (the darkest red curve) to $z = 5.76$ (the darkest blue curve). Median values (solid lines) come from randomly sampling the MCMC for the posterior distribution 1000 times. The 1σ uncertainty regions (shaded areas) are derived from the $16^{\rm th}$ and $84^{\rm th}$ percentiles of the distribution, but we only show the lower uncertainty bound for the highest-z curve and the upper uncertainty bound for the lowest-z curve. The dashed purple and green curves come from the radiative transfer simulations for a mostly neutral and ionized IGM by Eilers et al. (2017), respectively. The dotted pink and blue curves show the theoretical expectations for a mostly neutral IGM surrounding the quasar and the expected evolution of $R_{\rm p}$ from the analytical model by Bolton & Haehnelt (2007a) in a highly ionized IGM, respectively. Our fit falls between the models' predictions, and we find many quasars with small $R_{\rm p}$, that, hence, might have short lifetimes ($t_{\rm Q} \lesssim 10^4$ yr).

Median values (solid lines) are obtained by randomly sampling the Markov chains for the posterior distribution 1000 times. The 1σ uncertainty regions (shaded areas) are derived from the $16^{\rm th}$ and $84^{\rm th}$ percentiles of the distribution; however, for clarity, we only show the lower uncertainty bound for the highest-z curve and the upper uncertainty bound for the lowest-z curve.

We compare them with the theoretical expectations for a mostly neutral IGM (dotted pink line), the analytical model described by Bolton & Haehnelt 2007a (dotted blue line) and the predictions obtained by Eilers et al. (2017) for a highly neutral and ionized medium, shown in Equations 3.9 and 3.10 (dashed purple and green lines, respectively). The normalization for the theoretical and analytical curves was arbitrarily chosen in Eilers et al. (2017), so we change and set it to our derived parameter C, for visual and comparison purposes. We also quote that our normalization corresponds to a smaller value than those inferred in the simulations by Eilers et al. (2017). Our estimated luminosity dependence falls between the limits expected for neutral and highly ionized environments.

To conclude, we try to explain the visibly increasing scatter in the $R_{\rm p}$ distribution when moving from lower to higher luminosity quasars, echoing Satyavolu et al. (2023b). They attribute this evidence to the different expansion rates of ionization fronts. In luminous objects, they rapidly expand, increasing the likelihood of finding neutral hydrogen islands along different lines of sight. In contrast, fainter quasars require more time for their ionization fronts to propagate to similar distances, limiting their interaction with neutral regions. As a result, the surrounding environment of fainter quasars appears more uniformly ionized, leading to a narrower distribution of $R_{\rm p}$.

3.4.2.3 Correlation with redshift

We study the variation in proximity zone sizes with redshift as it can provide valuable information on the reionization process. In this paper, we follow the same approach as adopted in Satyavolu et al. (2023b) and do not apply luminosity corrections to standardize proximity zones to a single magnitude, as the relationship between $R_{\rm p}$ and M_{1450} varies significantly with z, making a universal correction inapplicable. Thus, Fig. 3.9 shows the proximity zone sizes non-corrected for their luminosity, from both this work and the literature, as a function of redshift. We plot seven curves from our bivariate power-law fit (Section 3.4.2), obtained by fixing the M_{1450} to be, in the order, $M_{1450} = -29$ (dark red), $M_{1450} = -28$ (light red), $M_{1450} = -27$ (dark yellow), $M_{1450} = -26$ (lime), $M_{1450} = -25$ (turquoise), $M_{1450} = -24$ (light blue) and $M_{1450} = -23$ (dark blue). Median values (solid lines) and 1σ uncertainty regions (shaded areas) are obtained by randomly re-sampling the Markov chains for the posterior distribution 1000 times.

Our results indicate a steeper redshift dependence than found in previous studies (e.g., Satyavolu et al. 2023b: $R_{\rm p} \propto (1+z)^{-0.89}$), suggesting that the

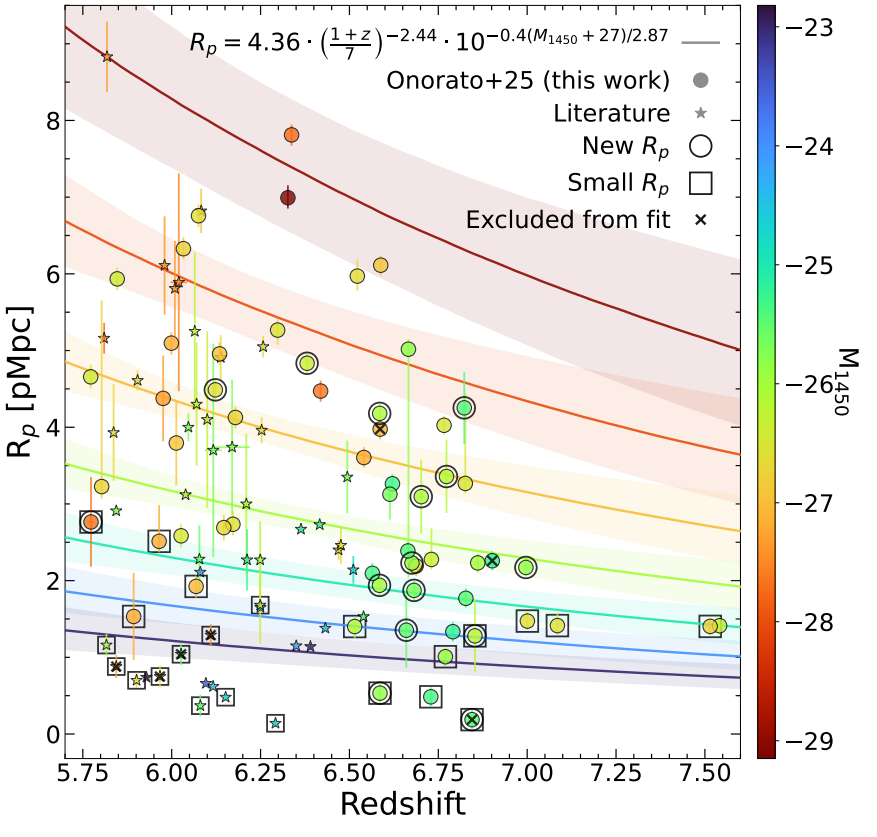


Figure 3.9: Distribution of $R_{\rm p}$ non-corrected for their luminosity as a function of z for the quasars included in this study (points) and for others from the literature (stars), with the corresponding error bars. The color of the symbols is determined by the M_{1450} of the quasars, according to the colormap to the right. The big black circles surrounding some points highlight the new $R_{\rm p}$. The black squares mark the objects with small $R_{\rm p}$, identified from the residuals of the fit in Section 3.4.2 ($\chi \leq -0.75$). The quasars excluded from the plotted final fit (i.e., BALs, pDLAs, $R_{\rm p}$ truncated by metal absorbers) are marked with a black cross. We show seven curves from our bivariate power-law fit, obtained by setting the magnitude to be $M_{1450} = -29$ (dark red), $M_{1450} = -28$ (light red), $M_{1450} = -27$ (dark yellow), $M_{1450} = -26$ (lime), $M_{1450} = -25$ (turquoise), $M_{1450} = -24$ (light blue) and $M_{1450} = -23$ (dark blue). Median values (solid lines) and 1σ uncertainty regions (shaded areas) come from randomly sampling the MCMC for the posterior distribution 1000 times. They show a steeper redshift evolution than found in previous studies, suggesting that the IGM's ionization state might play a role in shaping $R_{\rm p}$.

evolution of the IGM's ionization state may play a more significant role in shaping proximity zone sizes than previously assumed. We conclude that at fixed M_{1450}, variations in $R_{\rm p}$ can arise either from z evolution (i.e., if the quasars exhibit different redshifts in Fig. 3.8) or from differences in the quasar lifetimes. Further investigation is needed to disentangle these effects.

3.5 Summary and conclusions

In this work, we have conducted a comprehensive analysis of quasar proximity zones using a sample of 59 quasars spanning redshifts $5.77 \leq z \leq 7.54$. This data set includes new measurements for 15 quasars that had not been previously analyzed. By leveraging good quality spectra from multiple instruments, we have obtained robust estimates of proximity zone sizes ($R_{\rm p}$) and explored their dependence on quasar luminosity and redshift. Our main findings can be summarized as follows:

- Our measurements yield a range of $0.19 \leq R_{\rm p} \leq 7.81$ pMpc, with a median size of $R_{\rm p} = 2.76$ pMpc. The distribution exhibits significant scatter, reflecting potential variations in quasar lifetimes and IGM properties.

- We fit a bivariate power-law model to our data and literature. The inferred dependence on redshift and quasar luminosity follows the scaling $R_{\rm p} \propto (1+z)^{\beta} \cdot 10^{-0.4 M_{1450}/\alpha}$, with $\alpha = 2.87^{+0.48}_{-0.35}$ and $\beta = -2.44^{+0.89}_{-0.92}$, when we exclude contaminated sources (i.e., BALs, pDLAs, quasars with $R_{\rm p}$ truncated by metal absorbers). The evolution of the proximity zones with the luminosity falls between the model predictions for a highly neutral and ionized IGM. The redshift evolution is steeper than in other studies (e.g., Satyavolu et al. 2023b), suggesting that the ionization state of the IGM might play a role in shaping proximity zone sizes.

- There are 13 quasars that exhibit $R_{\rm p}$ values smaller than expected from model predictions. For all of them except J2211−6320, we rule out the presence of proximate dense absorbers as the cause, leading us to propose short quasar lifetimes ($t_{\rm Q} \lesssim 10^4$ yr) as a plausible explanation, and hence posing challenges on SMBH growth.

Our findings highlight the importance of considering both reionization and quasar lifetimes when interpreting proximity zone sizes. In a future work (Onorato et al., in prep), we will incorporate a new inference approach (Hennawi et al. 2024; Kist et al. 2024) and hydrodynamical simulations of ionization front evolution. This will enable us to rigorously test the reliability of our models by comparing the Lyα transmission probability

distribution function (PDF) derived from observed quasar spectra in our sample with that from mock spectra, generated under specific assumptions on the neutral hydrogen fraction $x_{\rm HI}$ and the quasar lifetime $t_{\rm Q}$.

Acknowledgements

We acknowledge D'Odorico et al. (2023) and Ďurovčíková et al. (2024) for their public data release, which made this work possible. We acknowledge helpful conversations with the ENIGMA group at UC Santa Barbara and Leiden University. SO is grateful to Timo Kist, Shane Bechtel, and Daming Yang for comments on an early version of the manuscript. JFH acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 885301), from the National Science Foundation (NSF) under Grant No. 2307180, and from NASA under the Astrophysics Data Analysis Program (ADAP, Grant No. 80NSSC21K1568).

Data Availability

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

3.6 Appendix: Continuum modeling for the quasars in the sample

Due to the large sample size, for displaying choices, we report here the continuum model reconstructions for the quasars in this work (see Fig. 3.10, 3.11, and 3.12). A full description of the method is provided in Section 3.3.1.

3.7 Appendix: Additional results from the fit

In Fig. 3.13, we show the residuals of the bivariate power-law fit described in Section 3.4.2, and the threshold used to identify the quasars with small proximity zone size ($\chi \leq -0.75$, see Section 3.4.2.1).

Furthermore, we compare our predictions with observational data in a three-dimensional (3D) space of $(M_{1450}, z, R_{\rm p})$. A surface plot is shown in Fig. 3.14 to visualize the functional dependence of $R_{\rm p}$ on M_{1450} and z, overlaid with observational data points and their associated error bars.

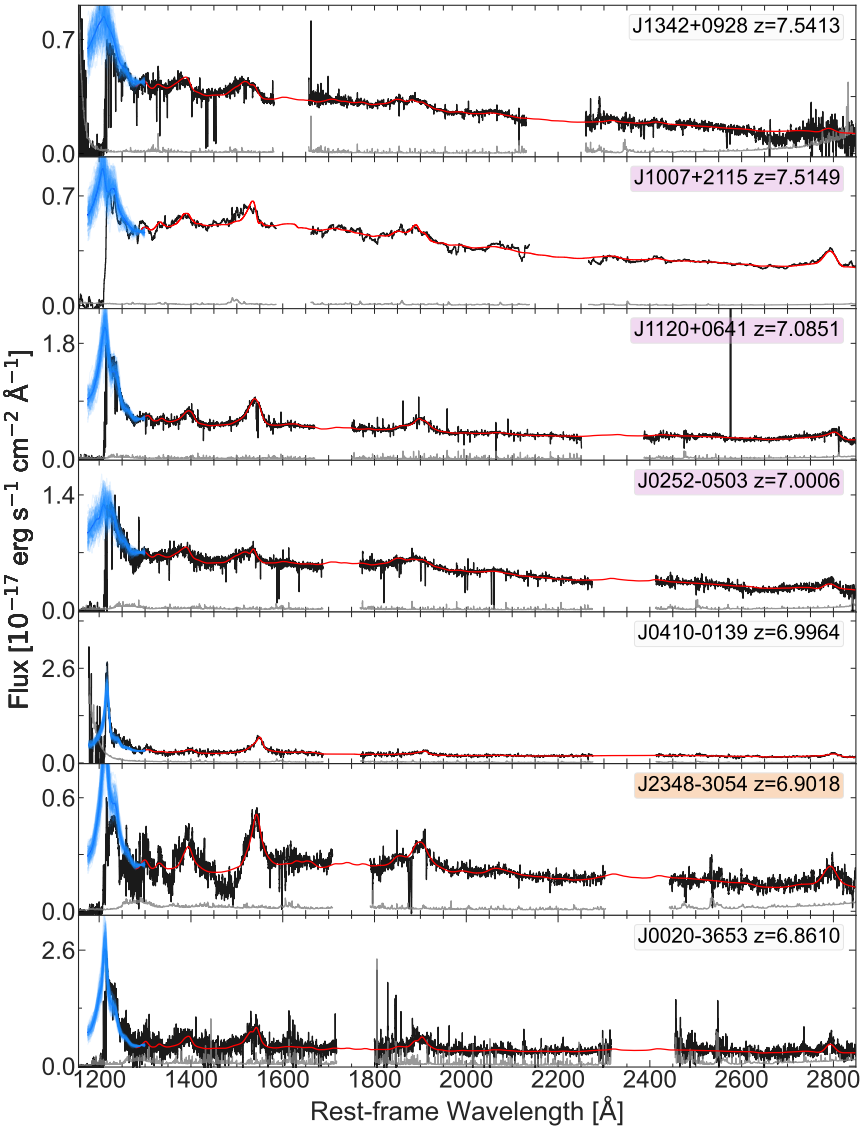


Figure 3.10: Spectra of the 35 quasars in the ENIGMA sub-sample, sorted by decreasing z, and their best-fitting continuum model using principal component analysis (PCA) as described in Fig. 3.2. All the spectra are smoothed for visual purposes and we have masked regions of strong telluric absorption. The box containing the name and the redshift of the quasars is colored in orange to mark the BALs and in pink to highlight the small proximity zones defined in Section 3.4.2.1.

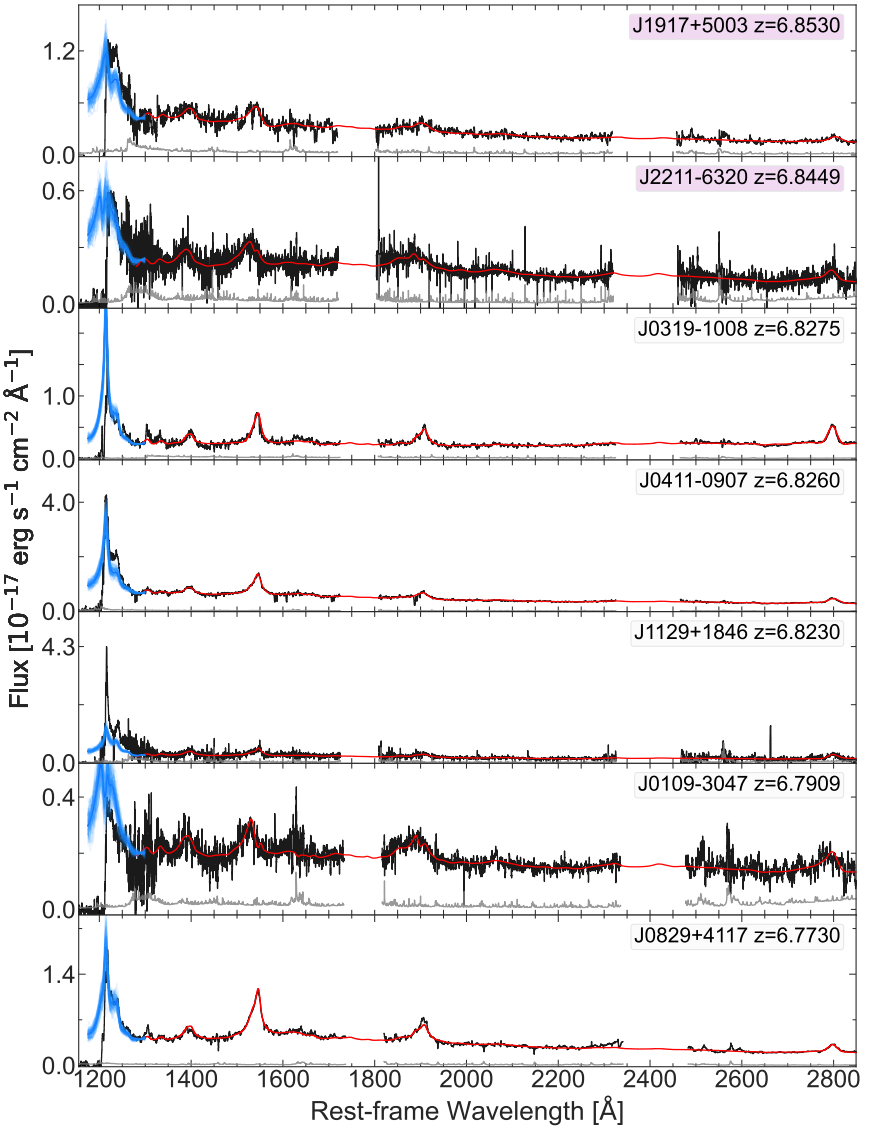


Figure 3.10: (Continued)

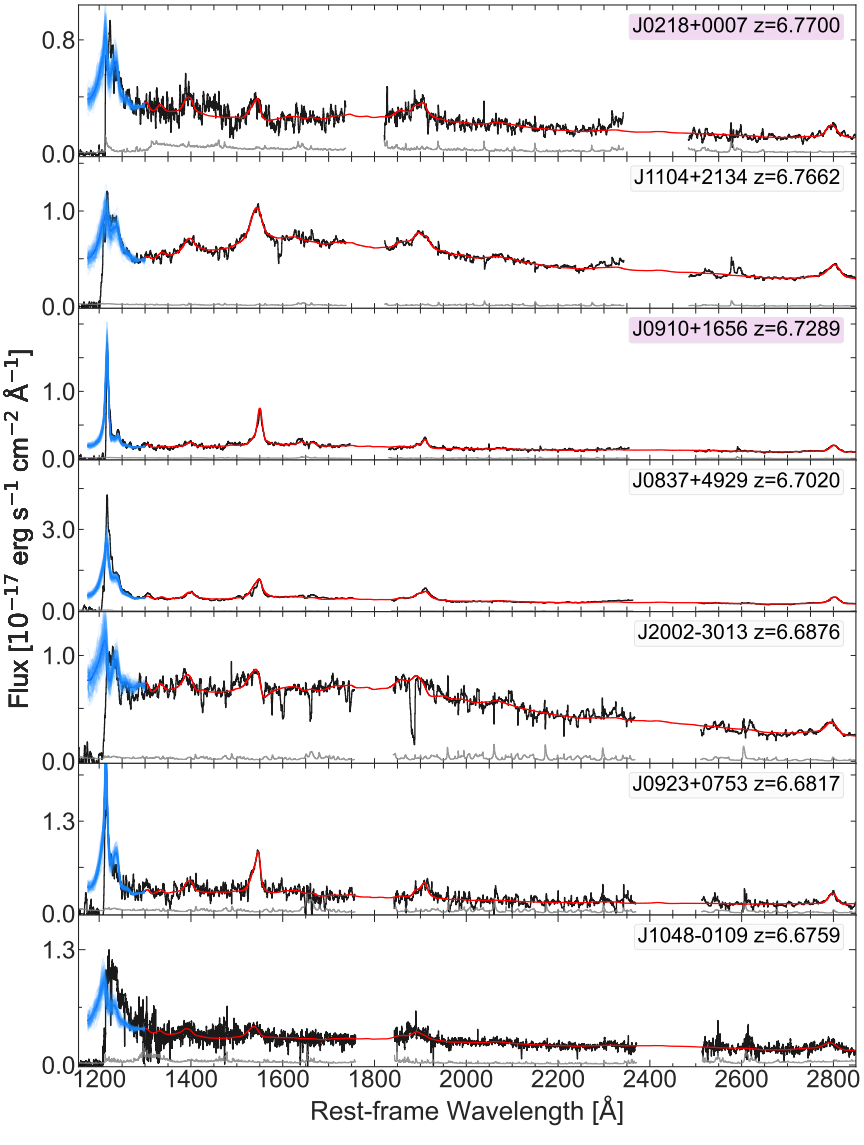


Figure 3.10: (Continued)

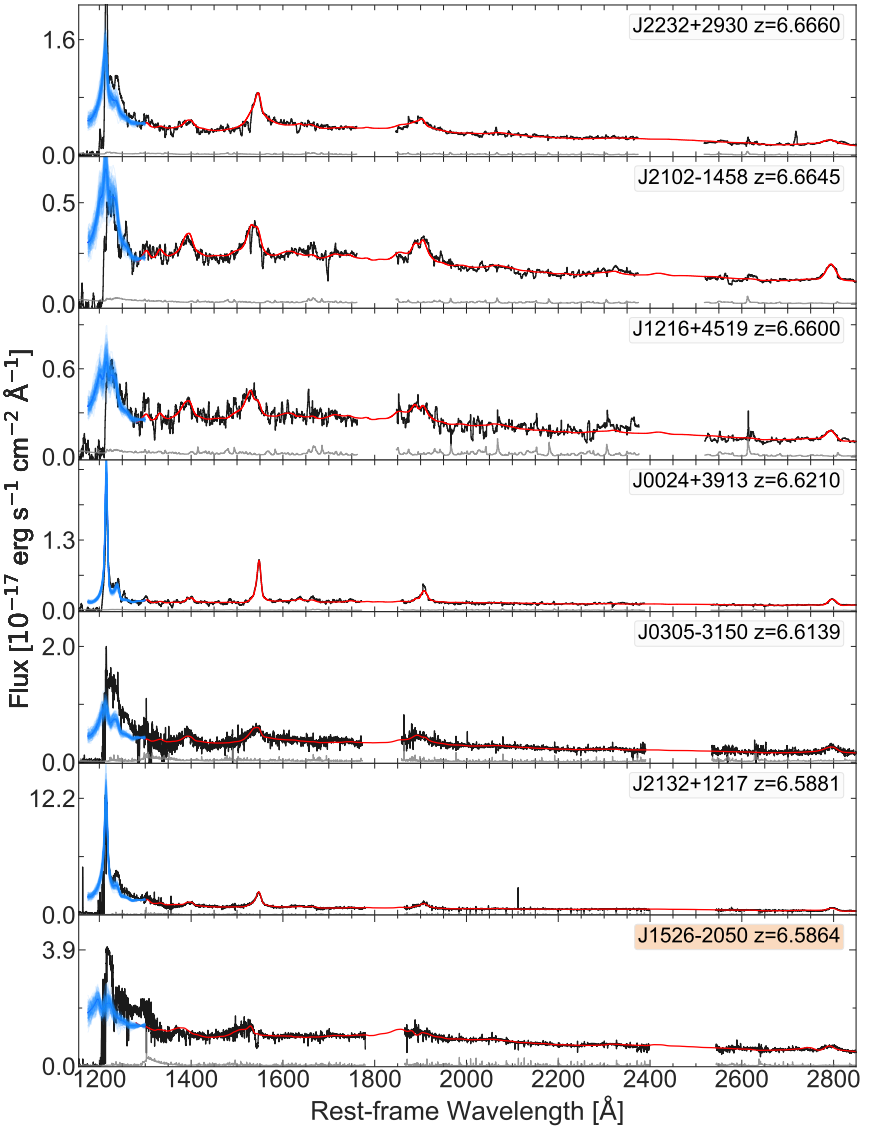


Figure 3.10: (Continued)

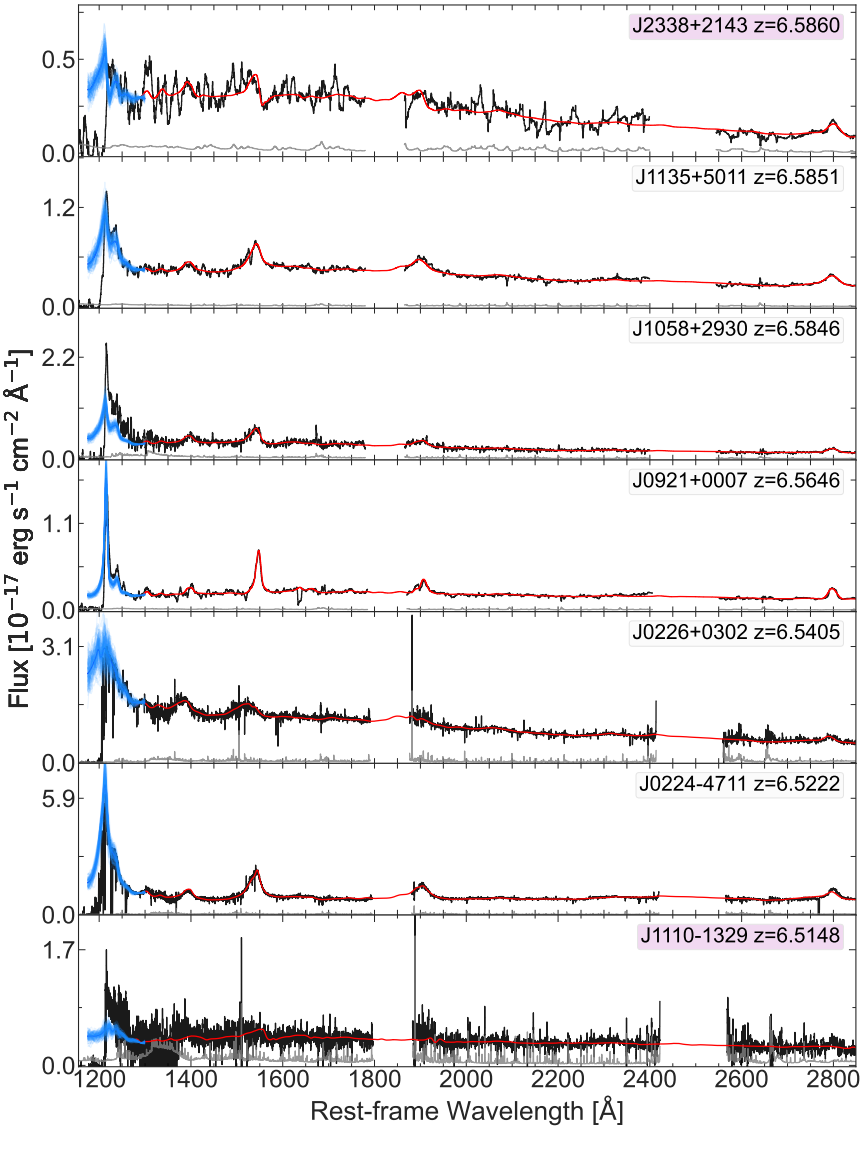


Figure 3.10: (Continued)

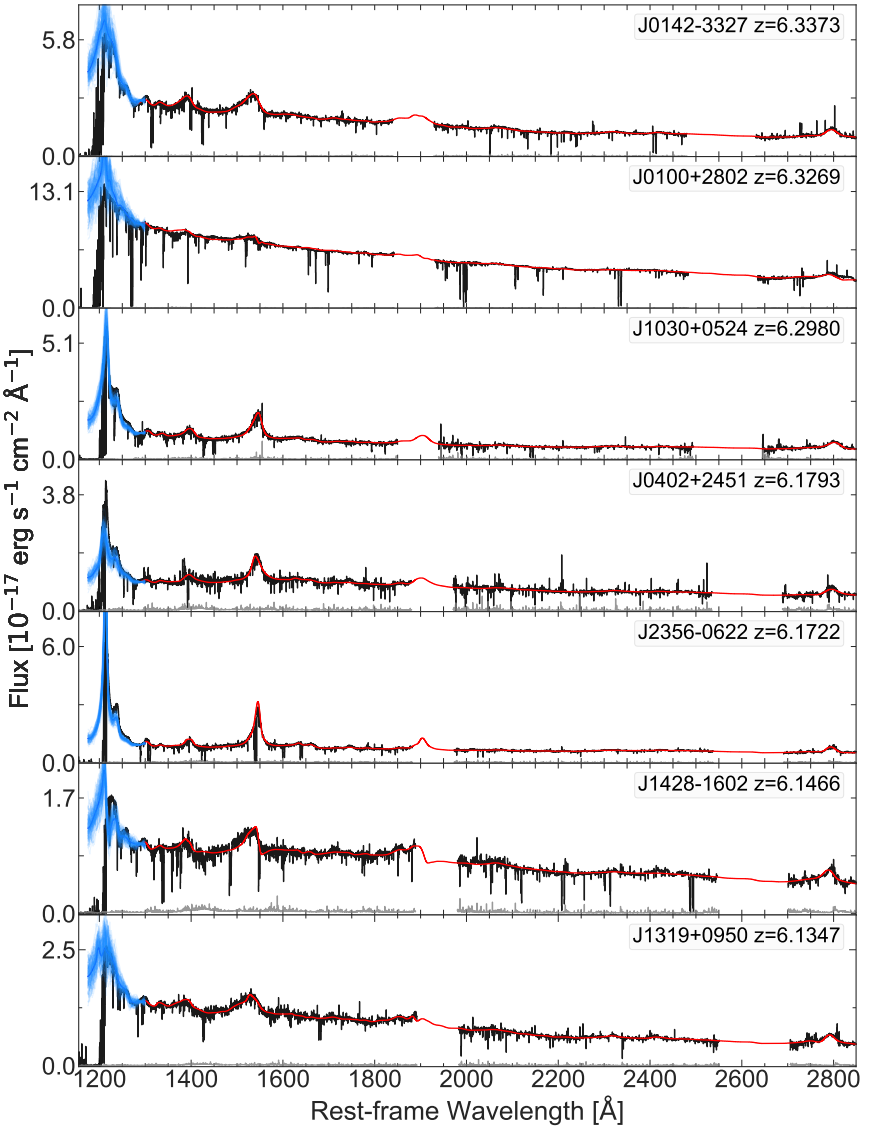


Figure 3.11: Spectra of the 21 quasars in the E-XQR-30 sub-sample, sorted by decreasing z, and their best-fitting continuum model using principal component analysis (PCA) as described in Fig. 3.2. All the spectra are smoothed for visual purposes and we have masked regions of strong telluric absorption. The box containing the name and the redshift of the quasars is colored in pink to highlight the small proximity zones defined in Section 3.4.2.1.

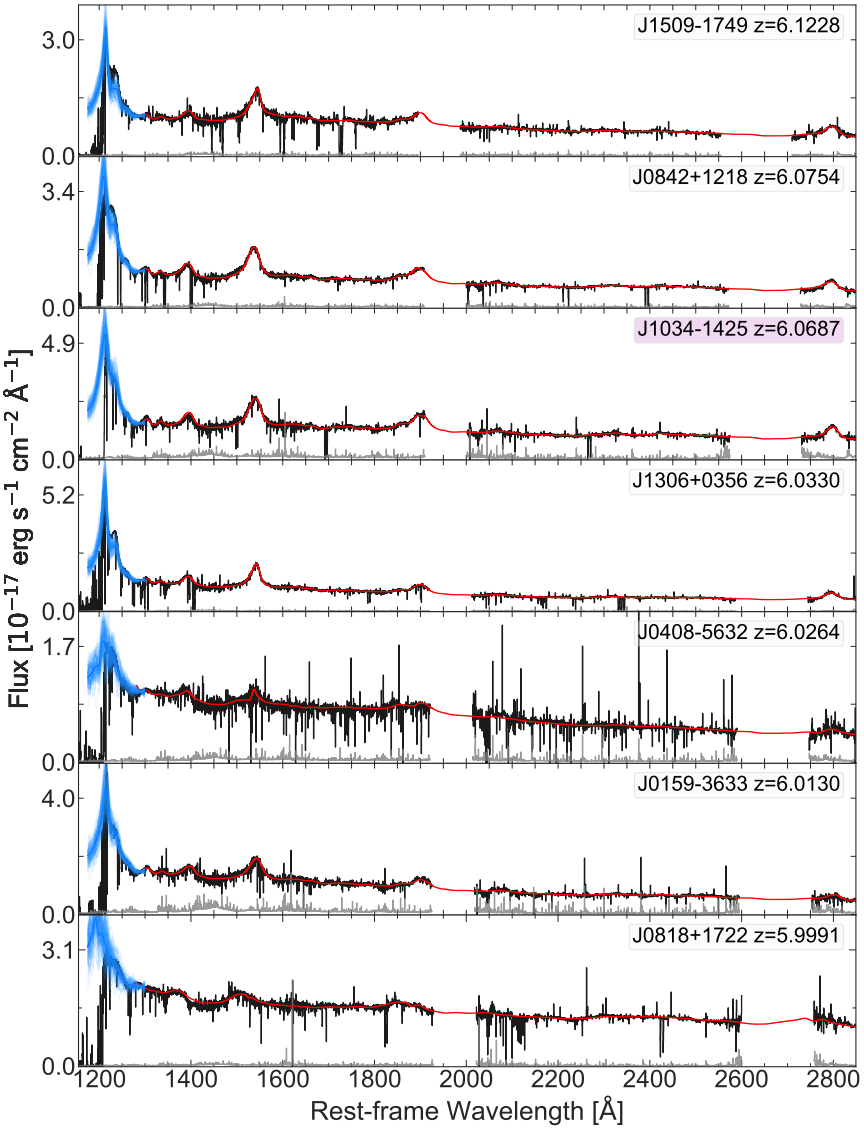


Figure 3.11: (Continued)

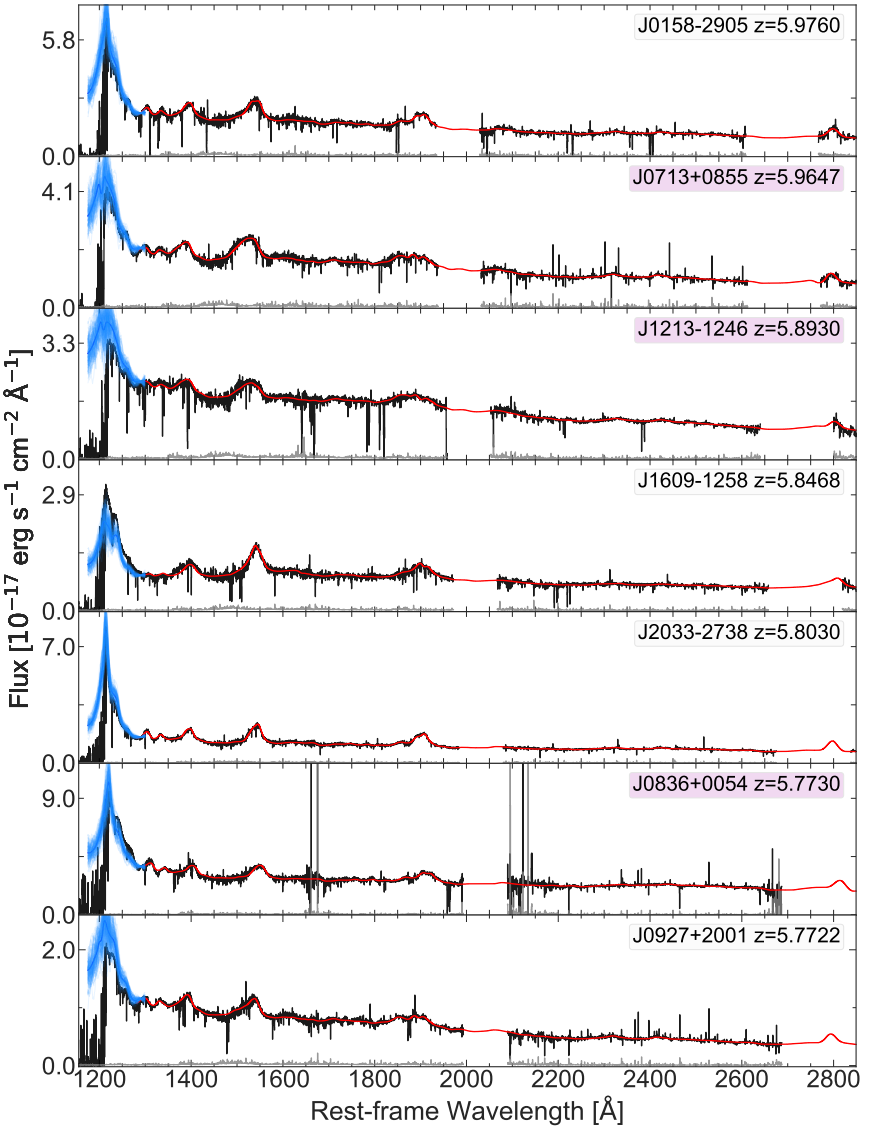


Figure 3.11: (Continued)

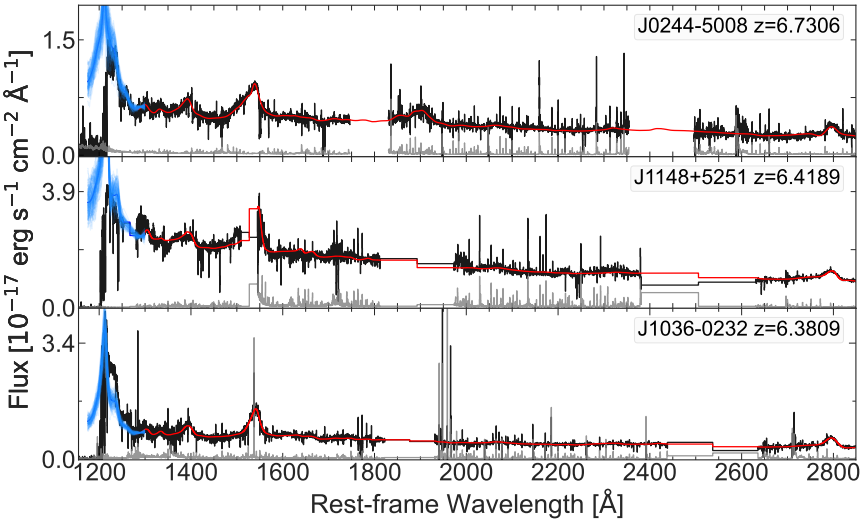


Figure 3.12: Spectra of the three quasars in the FIRE sub-sample, sorted by decreasing z, and their best-fitting continuum model using principal component analysis (PCA) as described in Fig. 3.2. All the spectra are smoothed for visual purposes and we have masked regions of strong telluric absorption.

3.8 Appendix: Metal absorbers analysis

Here we describe our metal absorbers analysis for the objects showing small proximity zone size ($\chi \leq -0.75$, from the fit in Section 3.4.2) performed via visual inspection of the spectra (see Section 3.4.2.1):

- J1120+0641: we recognize C IV, Mg II doublets and several metal lines associated with at least seven intervening absorbers at $z_{\rm abs} \simeq 7.0603$, 7.0555, 7.0167, 5.7950, 4.4725, 2.8097, 2.7952 (see also Bosman et al. 2017). They imprint several absorption features redward the Lyα line, but due to their distance from the quasar, the absorbers are unlikely to influence the Lyα region and the size of the proximity zone.

- J0252−0503: we distinguish C IV doublet associated with a possible absorber at $z_{\rm abs} \simeq 6.9633$, and Mg II doublet and Fe II absorption lines associated with another absorber at $z_{\rm abs} \simeq 4.2094$ (see also Wang et al. 2020). They appear to imprint H I and Al III absorption lines in the Lyα region, respectively, but should not alter the size of the proximity zone. We also find a Mg II doublet and many Fe II absorption lines associated with four absorbers at $z_{\rm abs} \simeq 4.8790$, 4.7148, 4.1904, and 3.5426. The first one imprints absorption features redward the Lyα

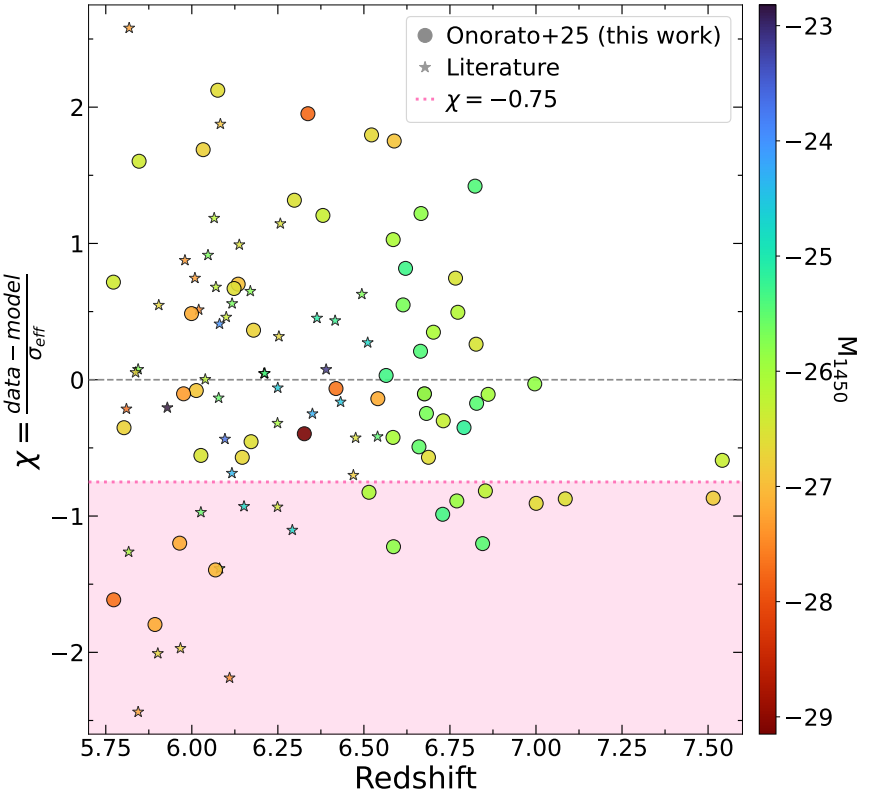


Figure 3.13: Residuals (χ) of the bivariate power-law fit performed after excluding the two BAL quasars in the sample (i.e., J2348−3054 and J1526−2050) as a function of redshift. The symbols are the data from this work (points) and the literature (stars), and they are color-coded with M_{1450}, according to the colormap on the right. The dashed gray line indicates where the model perfectly matches the data. The dotted pink line marks the threshold of $\chi = -0.75$, and the shaded region below highlights the quasars with small $R_{\rm p}$ (see Section 3.4.2.1).

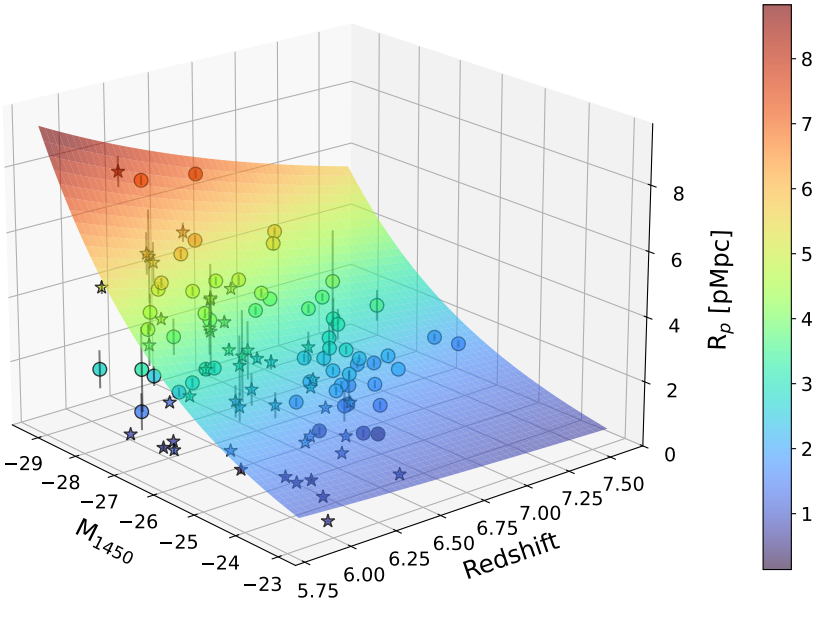


Figure 3.14: 3D surface plot of the proximity zone size as a function of redshift and absolute magnitude. The surface represents the best-fit bivariate power-law model obtained from the MCMC analysis in Section 3.4.2, with colors varying according to $R_{\rm p}$, following the colormap on the right. The symbols are the data from this work (points) and the literature (stars). Their color corresponds to the proximity zone size (also according to the colormap), with error bars indicating the uncertainties in the measurements. The plot provides a comprehensive visualization of the relationship between $R_{\rm p}$, z, and M_{1450}.

line but does not affect the proximity zone, while all the others do not have any particular absorption lines in that portion of the spectrum.

- J1917+5003: we recognize a Mg II doublet associated with an intervening absorber at $z_{\rm abs} \simeq 3.687$. As before, it is unlikely that the absorber influences the size of the proximity zone.

- J2211−6320: we find a C IV doublet at $z_{\rm abs} \simeq 6.8447$, likely associated with a metal absorber, which we suspect could be the quasar's host galaxy. However, we do not see any low-ionization lines, probably due to the spectrum's low SNR (i.e., $\langle {\rm SNR_J} \rangle = 9$, see O25). We conservatively decide to exclude this object from our final fit. Because of the closeness of the absorber to the quasar, it likely influences the Lyα region and distorts the proximity zone. The top panel of Fig. 3.15 displays the identified C IV absorption lines, while the bottom panel highlights in magenta the region where we estimate the true $R_{\rm p}$ to lie ($R_{\rm p} = 0.70 - 1.15$ pMpc). This estimate is based on visible fluctuations in the normalized flux up to $\simeq 0.70$ pMpc, which we interpret as remnants of a truncated proximity zone.

- J1007+2115, J0218+0007, J0910+1656: we do not identify any metal absorber.

- J2338+2143: because of the spectrum's very low SNR ($\langle {\rm SNR_J} \rangle = 3.4$, from O25), we cannot infer anything about the presence of metal absorbers.

- J1034−1425, J0713+0855, J1213−1246, J0836+0054: these quasars belong to the E-XQR-30 sample, and the presence of metal absorbers has already been widely studied by Davies et al. (2023). We refer to them for a complete list of the absorbers identified in the whole catalog, while here we focus on the ones that might contaminate the shape of the proximity zone in the above-mentioned sources.

  For J1034−1425, the closest metal absorber is located at $z_{\rm abs} \simeq 5.8987$. We do not see any clear low-ionization lines or fluctuations in the normalized flux of this object (see Fig. 3.4), and Satyavolu et al. (2023b); Eilers et al. (2020) have excluded the presence of an absorber ahead of the quasar. However, Ďurovčíková et al. (2025) quote the presence of a proximate Lyα emitter (LAE) that might truncate $R_{\rm p}$. We do not recognize any metal absorption lines on the quasar spectrum associated with it, except for a possible DW feature redward of the Lyα line. We decide to keep this object in our fit.

  For J0713+0855, we find Mg II and Fe II doublets associated with an intervening absorber at $z_{\rm abs} \simeq 3.5085$ that imprints Al III absorption lines in the Lyα region, but due to its distance from the quasar, it

does not contaminate the proximity zone. Furthermore, Satyavolu et al. (2023b); Davies et al. (2023) quote a high-ionization associated absorber at $z_{\rm abs} \simeq 5.9238$ that might truncate $R_{\rm p}$, but we do not see any clear Mg II or C IV absorption lines associated with it, and H I does not go line-black. We conclude that the $R_{\rm p}$ of this object should not be affected.

For J1213−1246, the absorbers at $z_{\rm abs} \simeq 5.6024$ and 4.8710 imprint Si II and Si IV absorption lines in the Lyα region, while the one at $z_{\rm abs} \simeq 3.8254$ imprints an absorption feature redward of the Lyα region. However, due to their distance from the quasar, they do not affect $R_{\rm p}$.

For J0836+0054, we identify several intervening absorbers at $z_{\rm abs} \simeq$ 5.1256, 4.9965, 3.7439, and 2.2990, but they do not contaminate $R_{\rm p}$.

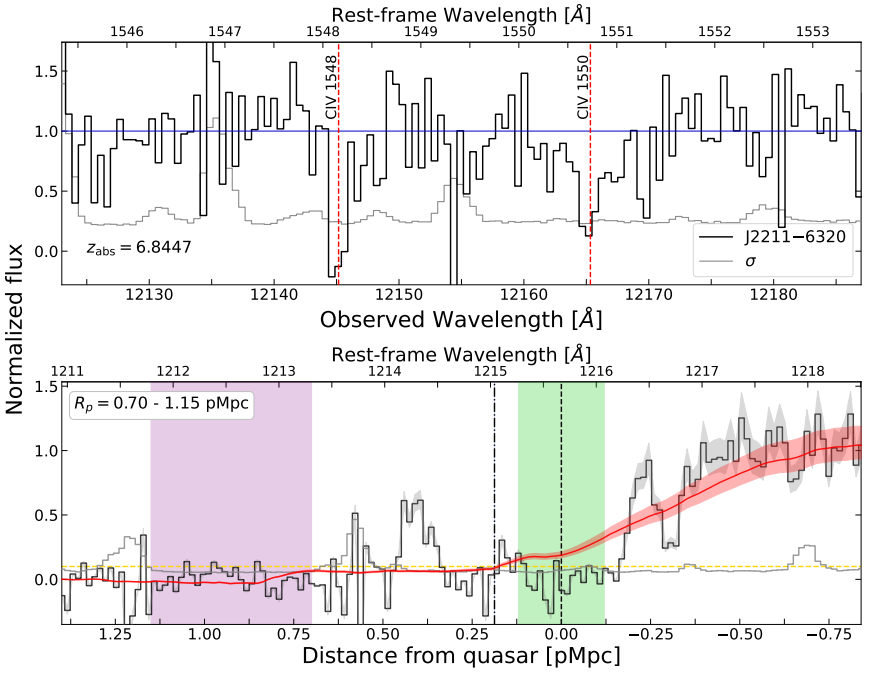


Figure 3.15: *Top:* Zoom-in of the continuum-normalized spectrum (black) and noise vector (grey) of J2211−6320. We identify C IV as the main absorption line system (dashed red lines) associated with an absorber at $z_{\rm abs} = 6.8447$, which is likely to affect the shape of the proximity zone of the quasar. *Bottom:* Proximity zone size measurement of J2211−6320, as described in Fig. 3.3. In magenta, we show the area in which we suspect $R_{\rm p}$ of J2211−6320 would fall ($R_{\rm p} = 0.70 - 1.15$ pMpc) without the truncation effect due to the metal absorber.