A SURVEY OF METAL LINES AT HIGH-REDSHIFT. I. SDSS ABSORPTION LINE STUDIES—THE METHODOLOGY AND FIRST SEARCH RESULTS FOR O\textsc{vi}

S. Frank\textsuperscript{1}, S. Mathur\textsuperscript{1}, M. Pieri\textsuperscript{1}, and D. G. York\textsuperscript{2}

\textsuperscript{1}Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210, USA; frank@astronomy.ohio-state.edu
\textsuperscript{2}Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA

Received 2007 July 11; accepted 2010 May 6; published 2010 August 11

ABSTRACT

We report the results of a systematic search for signatures of metal lines in quasar spectra of the Sloan Digital Sky Survey (SDSS) data release 3 (DR3), focusing on finding intervening absorbers via detection of their O\textsc{vi} doublet. Here, we present the search algorithm and criteria for distinguishing candidates from spurious Ly\textalpha forest lines. In addition, we compare our findings with simulations of the Ly\textalpha forest in order to estimate the detectability of O\textsc{vi} doublets over various redshift intervals. We have obtained a sample of 1756 O\textsc{vi} doublet candidates with rest-frame equivalent width (EW) $\geq 0.05$ Å in 855 active galactic nuclei spectra (out of 3702 objects with redshifts in the accessible range for O\textsc{vi} detection). This sample is further subdivided into three groups according to the likelihood of being real and the potential for follow-up observation of the candidate. The group with the cleanest and most secure candidates is comprised of 145 candidates. Sixty-nine of these reside at a velocity separation $\geq 5000$ km s$^{-1}$ from the QSO and can therefore be classified tentatively as intervening absorbers. Most of these absorbers have not been picked up by earlier, automated QSO absorption line detection algorithms. This sample increases the number of known O\textsc{vi} absorbers at redshifts beyond $z_{\text{abs}} \geq 2.7$ substantially.

Key words: intergalactic medium – quasars: absorption lines

Online-only material: color figures, machine-readable and VO tables

1. INTRODUCTION

Our understanding of the nature of the intergalactic medium (IGM) and its evolution as traced by the absorption features observed in the sightlines toward luminous objects like quasars has benefitted tremendously within the past few years both from observational and theoretical advances. High-resolution studies of the Ly\textalpha forest have been extended to very low column densities due to high-resolution echelle spectrographs on powerful 8 m class telescopes like the High Resolution Echelle Spectrometer (Keck) and UVES (VLT) (Hu et al. 1995; Lu et al. 1996; Kim et al. 1997; Kirkman & Tytler 1997). At the same time, theoretical models incorporating gas dynamics, radiative cooling, and photoionization have been developed that can reproduce the majority of the observed properties of the quasar absorption spectra (Cen et al. 1994; Miralda-Escudé et al. 1996; Hernquist et al. 1996; Petitjean et al. 1995; Davé et al. 1997). In this picture, baryonic gas can be encountered in a wide variety of physical conditions. Within the Ly\textalpha forest, at least at high redshift where the forest is the main repository for baryons, most of the gas resides in relatively cool ($T \sim 10^4$ K) low to medium overdensity structures that are not in dynamical or thermal equilibrium and are mostly governed by photoionization.

The IGM is expected to be highly ionized with a neutral fraction of hydrogen $f(\text{H}^0) \sim 10^{-4}$ to $10^{-6}$. In such a medium, metals are also highly ionized. Thus, oxygen, as the most abundant intergalactic metal, may exist in the O\textsc{vi} state. Since the O\textsc{vi} 1032/1038 Å doublet is observable from ground-based telescopes for redshifts beyond $z_{\text{abs}} \sim 2.0$, these transitions constitute a primary tool for studying the characteristics of the metals in the IGM at high redshift. The density of the IGM is sufficiently low as to allow the production of O\textsc{vi} by photoionization from the intergalactic UV and soft X-ray radiation field (Haardt & Madau 1996; Schaye et al. 2000; Simcoe et al. 2004). Indeed, O\textsc{vi} absorbers seen toward higher redshifts can usually be modeled by structures that are photoionized (Bergeron et al. 2002; Carswell et al. 2002; Levshakov et al. 2003).

Due to the different wavebands required for detections at different redshifts and the strong evolution of the Lyman forest, a variety of different techniques and observational strategies have to be applied for finding and identifying O\textsc{vi} absorbers. Detection of the O\textsc{vi} 1032/1038 Å doublet from ground-based telescopes is only possible beyond redshifts of $z_{\text{lim}} \geq 2.0$ due to the 3000 Å atmospheric cutoff, and thus studies of the absorber statistics below this threshold need to rely on space-based instruments. Burles & Tytler (1996) conducted the first systematic survey of O\textsc{vi} absorbers at $z \sim 1.0$, employing the Faint Object Spectrograph on the Hubble Space Telescope to study a sample of 11 QSOs, and found 12 suitable candidates of which 9 were expected to be real, thus establishing that the number density per redshift interval of these absorbers is similar to, if not greater than, that for C\textsc{iv} and Mg\textsc{ii} absorbers at the same redshift.

Toward higher redshifts, despite the advantage of shifting the lines into the range for ground-based instruments, the ever increasing density of the Lyman forest renders unambiguous identifications of O\textsc{vi} absorbers in its midst more difficult than detections made longward of the Ly\textalpha forest. Thus, most studies have relied upon high-resolution and high signal-to-noise ratio (S/N) spectra. Around redshifts of $z \sim 2.0–3.0$, a variety of surveys find evidence for a population of absorbers residing in low to medium overdense regions, mostly photoionized by the ambient extragalactic radiation field (Carswell et al. 2002; Bergeron et al. 2002; Aracil et al. 2004; Simcoe et al. 2004). The metal enrichment of the IGM inferred by theses surveys ranges from $10^{-2}$ to $10^{-3}$ of the solar value for structures around the mean density of the universe at $z \sim 2.0$. Using UVES spectra for two lensed quasar systems, Lopez et al. (2007) find...
size constraints for such intervening O vi absorbers with lower boundaries on the kpc scale, where no or only very little variation in different lines of sight are seen. This is broadly consistent with other studies of binary QSO sightline inferring correlation lengths of the absorbers up to several tens of kpc (Smette et al. 1997; Ferrara et al. 2000; Barkana & Loeb 2001; Madau et al. 1997; York 1974, 1977) for highly ionized metals in absorption, we anticipate to detect in such cases O vi absorbers given conservative assumptions about the metallicity of the IGM, the abundance of O vi due to photoionization and the average S/N of the QSO data sample.

Table 1 gives an overview of different surveys for direct O vi detection in a variety of environments and over a wide redshift range. Reimers et al. (2006) find six different O vi systems and a further eight potential, but blended, candidates in a single object (HS0747+4259) at 1.46 < z abs < 1.81, deriving a redshift path density of dN(Ovi)/dX ∼ 13. Simcoe et al. (2004) analyze 230 Lyα forest lines with a hydrogen column density NH1 > 10^{13.6} cm^-2 for seven QSOs with z em ∼ 2.6, and retrieve a total of about 50 accompanying O vi systems with secure identifications. Using the capabilities of VLT and UVES within the “Large Programme: The Cosmic Evolution of the IGM” survey (Bergeron & Herbert-Fort 2005) 10 bright QSOs at 2.1 < z em < 2.8, and find 136 O vi candidates with 12.7 < log N(Ovi) < 14.6 in 51 systems. Carswell et al. (2002) focus on two z em = 2 QSOs and hydrogen absorbers with log N(Hi) > 14.0. They identify seven individual O vi lines in two intervening systems in one case and find 13 individual lines in 10 intervening systems in the other. Fox et al. (2007) study the frequency of O vi absorbers associated with 35 damped and sub-damped Lyα systems, and report 12 detections of which nine are intervening. Lopez et al. (2007) present 10 intervening O vi absorbers seen toward two lensed QSO pairs at a median redshift of z abs = 2.3. Our sample, residing at z abs > 2.7, therefore greatly increases the redshift range of known O vi absorbers at high redshifts. It will allow us to test expectation based upon photoionization models like those presented in Davé et al. (1998), thereby possibly constraining the physical conditions of the IGM and the metagalactic UV/X-ray background at high-redshift.

In this paper, we have decided to apply a direct pixel-by-pixel search for signatures of strong O vi 1032/1038 Å features seen in the spectra of Sloan Digital Sky Survey (SDSS) QSOs. This is, to our knowledge, the first systematic survey at high redshifts (2.7 < z em < 5.0) at low resolution (R ∼ 1800) and low S/N. We demonstrate that there is a redshift window of opportunity where the density of the Lyman forest is not yet too high to obliterate the hope for finding and identifying relatively narrow metal absorbers, and that the pathlength covered by the combined SDSS sample is high enough to expect finding a reasonable number of such absorbers given conservative assumptions about the metallicity of the IGM, the abundance of O vi due to photoionization and the average S/N of the QSO data sample.

What type of O vi absorbers do we expect to detect via such a direct search without a priori selecting on other transitions? If a sightline passes through a galaxy and if there are analogs to local examples surveyed early by Rogerson et al. (1973) and York (1974, 1977) for highly ionized metals in absorption, we anticipate to detect in such cases O vi mixed with C iv in the gas phase inside that galaxy, like in the disk of our Galaxy. Furthermore, a recent, thorough examination of Far-Ultraviolet Spectroscopic Explorer spectra (Wakker et al. 2003) revealed the occurrence of local O vi absorbers in almost all of the more than 100 sightlines probed by extragalactic background objects, indicating a large covering fraction when a sightline passes the Galaxy. While the majority of these absorbers are probably

Table 1 Overview of Searches for Intervening O vi Absorbers at High Redshift

| Redshift Range | Number of QSO Sightlines | Number of O vi Absorbers | Reference |
|----------------|--------------------------|--------------------------|-----------|
| z abs < 0.9    | 11                       | 6 systems                | Burles & Tytler (1996) |
| 1.46 < z abs < 1.81 | 1                       | 6 systems                | Reimers et al. (2006) |
| z abs < 2.5    | 7                        | 8 additional candidates  | Smette et al. (1997) |
| 2.0 < z em < 2.5 | 10                      | 136 (component) candidates in 51 systems | Bergeron & Herbert-Fort (2005) |
| 2.0 < z abs < 2.36 | 2                       | 20 (component) candidates in 12 systems | Carswell et al. (2002) |
| 2.1 < z abs < 3.1 | 35 DLA systems          | 9 (system) candidates    | Fox et al. (2007) |
| z abs = 2.3    | 2 lensed QSO pairs       | 10 components            | Lopez et al. (2007) |
| 2.7 < z abs < 4.2 | 3702                     | 1756 candidates (systems) | This work |
|                 |                          | 145 Category 1 candidates, 69 of these intervening |           |
located within the Galaxy (Savage et al. 2003; Bowen et al. 2008), especially the high-velocity O\textsc{vi} (relative to the local standard of rest) traces a variety of different environments and phenomena, most likely including warm/hot gas interaction in an extended Galactic corona and even truly intergalactic gas within the Local Group (Sembach et al. 2003). As an aside, the gas residing in the background active galactic nucleus (AGN) host galaxy is subject to a strong UV radiation field by the central engine itself, and thus might contain more O\textsc{vi} than normal galaxies. And while we are per se not interested in such absorbers for the study of intergalactic gas, we have included “intrinsic” absorbers, as they are being found with our algorithm as well.

From hydrodynamic simulations, Cen et al. (2001) assert, however, that the majority of detectable lines with rest-frame equivalent widths, EW, (O\textsc{vi} 1032 Å) $\geq$ 0.035 Å do not reside in virialized regions such as galaxies, groups, and clusters, but trace intergalactic gas at only moderate overdensities ($\sim$10–40). Although their estimated redshift path density of about five such absorbers per redshift of unity drops rapidly by a factor of at least 10 when increasing the EW limit to 0.35 Å, even that lower rate of incidence allows us to detect an appreciable number of such strong intergalactic O\textsc{vi} absorbers, owing to the large redshift pathlength probed by the collective SDSS AGN sample.

To summarize, we expect to detect the signals of O\textsc{vi} absorbers in a variety of environments. While this first step of finding O\textsc{vi} rests upon a “blind” O\textsc{vi} doublet search plus an alignment with Ly\alpha and Ly\beta lines, in order to ensure high purity of the candidate sample, we may then, in a future effort, use independent line detections of different ions to better characterize the physical state and potentially the nature of the absorbers found here. Because there are a variety of possible origins of the O\textsc{vi} gas, including outflows, intervening galaxies, and the contribution of the AGN host galaxy, this next step is needed to arrive at a sample of IGM absorbers. There is evidence for a population of O\textsc{vi} absorbers that contain little, if any detectable hydrogen, as pointed out by Bergeron & Herbert-Fort (2005), and it is clear that our search algorithm presented here cannot detect those. However, as we anticipate the need for high-resolution follow-up of our candidates in order to confirm their nature, we might then be able to learn about this oxygen-rich subset (type 0 of Bergeron & Herbert-Fort 2005) of absorbers as well.

The paper is organized as follows. Section 2 details our preliminary analyses of whether the S/N of the SDSS spectra and the density of the Lyman forest enables our search program. In Section 3, we present our SDSS QSO sample selection, and Section 4 gives an overview of the search algorithms we applied. Before we summarize and conclude in Section 6, we present the results of our search in Section 5. The search method developed here can be easily tailored to a variety of different ionic species. We stress that follow-up observations of the candidates we retrieve with high S/N and high resolution are necessary to determine the reliability of our search algorithm by determining the nature of the features we detect.

Throughout this study, we use a cosmology with $H_0 = 71$ km s$^{-1}$ Mpc$^{-1}$, $\Omega_{\text{m}} = 0.27$, and $\Omega_{\Lambda} = 0.73$. Abundances are given by number relative to hydrogen, and vacuum wavelength are being used, if not otherwise noted.

### 2. PRELIMINARY FEASIBILITY ANALYSES

With rest-frame wavelengths below the H\textsc{i} Ly\alpha wavelength of 1215.67 Å, the O\textsc{vi} 1032 Å/1038 Å doublet absorption lines inevitably fall into the Ly\alpha forest of systems located along the same sightline with slightly lower redshifts. Thus, two effects may prohibit our attempts to find such O\textsc{vi} absorbers: first, the ubiquity of the Ly\alpha forest lines, especially at higher redshifts, can lead to blending of the oxygen lines with H\textsc{i} lines, thus rendering a correct identification of the doublet impossible. Second, two different Ly\alpha lines at different redshifts and the accidental occurrence of optical depth ratios expected for O\textsc{vi} can mimic an O\textsc{vi} doublet, leading to a false identification.

#### 2.1. The Expected Number of Falsey Identified Absorbers

In order to estimate the severity of the latter effect, we have created mock catalogs of Lyman forest lines, using the line density estimate of Kim et al. (2001) and the column density distribution function of Hu et al. (1995):

$$dn dz = 9.06 \times (1+z)^{2.19\pm0.27} \text{ for } 13.64 < \log\ N_{\text{H}_1} < 16.0,$$

(1)

$$f(N_{\text{H}_1})dN_{\text{H}_1} = 4.9 \times 10^7 \times N_{\text{H}_1}^{1.46} \times dN_{\text{H}_1},$$

(2)

where $\frac{dz}{d\lambda}$ denotes the number of Ly\alpha lines per redshift interval $dz$, and $f(N_{\text{H}_1})$ is the probability distribution of obtaining a line with column density of neutral hydrogen $N_{\text{H}_1}$ within the interval $N_{\text{H}_1}, N_{\text{H}_1}+dN_{\text{H}_1}$. We then analyzed these catalogs and retrieved all line pairs that exhibit the correct wavelength ratio (within a given tolerance) as well as the correct ratio of optical depths at the line center:

$$r_1 = \frac{\lambda_{\text{O} \text{vi} 1032 \text{ Å}}}{\lambda_{\text{O} \text{vi} 1038 \text{ Å}}} = 0.99452,$$

(3)

$$r_2 = \frac{\tau(\text{O} \text{vi} 1032 \text{ Å})}{\tau(\text{O} \text{vi} 1038 \text{ Å})} = 2.0,$$

(4)

$r_1$ is independent of redshift since both lines get shifted by the same factor $(1+z)$, and the result for $r_2$ follows from atomic physics:

$$\tau_0 = N_{\text{i}} \frac{\lambda^4}{8\pi^2} \frac{g_k}{g_j} A_{kj} b_k,$$

(5)

where $N$ is the column density, $\lambda$ is the wavelength of the transition, $c$ is the speed of light, $A_{kj}$ is the Einstein coefficient for the transition from the upper ($k$) to the lower ($j$) level, and $b_k = \lambda g_k^b$ is the line broadening factor. The statistical weights $g_k/g_j$ of the two transitions are equal to 2.

With the given resolution of SDSS spectra ($R \sim 1800$) and the conservative assumption that line centroids can be determined to about one-third of the resolution element, we expect a tolerance level for the wavelength ratio of $\Delta\lambda/\lambda = 3.0 \times 10^{-4}$. Furthermore, we can rule out all pairs that do not exhibit the range of flux decrements, expected from the ratios of the two line strengths convolved with the instrument capabilities. The efficiency of this procedure depends crucially on the ability to precisely measure the EW of lines, but even a conservative estimate of a 20% accuracy on a single pixel flux measurement, corresponding to a low S/N of about 5, leads to a reduction of the number of false detections of roughly a factor of 8 at all redshifts.

An analysis of the mock data sets spanning the complete redshift range available to us ($2.7 < z_{\text{abs}} < z_{\text{max}}\text{(SDSS)}$) and varying the thresholds on the desired precision for $r_1$ and $r_2$ yields for the number of expected interlopers:

$$n_i(z_{\text{abs}}) = n_{i,0}(z_{\text{abs}}) \left( \frac{\Delta r_1/r_1}{3.0 \times 10^{-4}} \right) \left( \frac{\Delta r_2/r_2}{0.2} \right)^{1.12},$$

(6)
where \( z_{\text{abs}} \) is the absorber redshift and \( n_i \) is the number of expected absorbers per spectrum. \( n_{i,0}(z_{\text{abs}}) \) is a strongly increasing function of the redshift:

\[
n_{i,0}(z_{\text{abs}}) = 2.75 \times 10^{-4}(1 + z_{\text{abs}})^{7.5}.
\]  

(7)

Note that the coefficients and exponents in Equations (6) and (7) are derived from analyzing the mock data set, without taking additional absorption from metal lines into account. Nor did we analyze the contribution of random noise features mimicking \( \text{O} \text{vi} \) doublets. The latter problem, however, is obviated by the fact that we are going to focus on lines that are clearly above the noise limit, as we will explain in Section 3.

We focus on \( \text{O} \text{vi} \) absorbers accompanied by strong \( \text{H} \text{i} \) Lyman absorption.\(^3\) Therefore, we rule out all lines that do not show accompanying \( \text{H} \text{i} \) transitions. Given the line density estimate of Kim et al. (2001), the lower limit for the average velocity separation of two \( \text{Ly} \alpha \) absorbers is

\[
\Delta v_{\text{Ly} \alpha} = c \times \frac{\Delta z}{1 + z_{\text{abs}}} = c \times \frac{\Delta n}{9.06 \times (1 + z_{\text{em}})^{2.19} \times (1 + z_{\text{abs}})} \geq 30,000 \text{ km s}^{-1} \times (1 + z_{\text{em}})^{-3.19},
\]  

(8)

where we used the relation \( z_{\text{abs}} \leq z_{\text{em}} \) to obtain the lower limit in the last step and equated \( \Delta n = 1.4 \).\(^4\)

Fig. 1. Number of \( \text{O} \text{vi} \) absorption features and spurious \( \text{H} \text{i} \) interlopers expected in the spectra of SDSS quasars vs. the redshift \( z_{\text{abs}} \) of the absorber. The rapid increase for the \( \text{O} \text{vi} \) lines up to \( \log (1 + z_{\text{abs}}) \sim 0.65 \) (= \( z_{\text{abs}} \sim 3.4 \)) can mainly be attributed to the nearly linearly growing redshift pathlength, whereas the rapid increase of \( \text{H} \text{i} \) interlopers is due to the redshift evolution of the line density. We assume \( \log f_{\text{Ly} \beta} = -2.0 \) to calculate the number of absorbers by integrating over the column density distribution by Hu et al. (1995) after assessing a correction factor for the fraction of lines retrievable (for details see Section 2.2). The red dashed curve represents the expected number of \( \text{H} \text{i} \) interlopers that exhibit less than a fiducial value for the deviations in both line position and line strength (for details see the text), whereas the dotted line takes the additional requirement into account that the interlopers have to show associated, strong \( \text{Ly} \alpha \beta \) absorption. It is obvious that our search efficiency peaks around a redshift of \( z_{\text{abs}} \sim 3.2 \), and rapidly deteriorates for redshifts \( z_{\text{abs}} \geq 4.0 \). Thus, we expect our “window of opportunity” to be between 2.8 \( < z_{\text{abs}} < 4.0 \), where the ratio of detections of real systems to spurious interlopers is greater than unity.

The velocity separation of one pixel in SDSS spectra, however, amounts to

\[
\Delta v_{\text{SDSS}} = c \times \frac{\Delta \lambda_{\text{pixel}}}{\lambda} \sim 70 \text{ km s}^{-1}.
\]  

(9)

Thus, the probability to have at least one \( \text{H} \text{i} \) line fall right onto the pixel corresponding to the redshift of an “\( \text{O} \text{vi} \) pixel” is

\[
P_{\text{interloper}} = \frac{\Delta v_{\text{SDSS}}}{\Delta v_{\text{Ly} \alpha}} \sim 2.3 \times 10^{-3} \times (1 + z_{\text{em}})^{3.19}.
\]  

(10)

Note that the Kim et al. (2001) line density includes \( \text{Ly} \alpha \) absorbers with column densities as low as \( \log N_{\text{Ly} \alpha} = 13.64 \). Restricting ourselves to absorbers that show at least an optical depth at line center for the \( \text{Ly} \beta \) line of \( \tau_{0}(\text{Ly} \beta) = 1.0 \), leads to another factor of \( f_{\text{strong}} \sim 0.4 \) in the exclusion of spurious interlopers. Therefore, we can further reduce the number of spurious interlopers by \( P_{\text{interloper}} \times f_{\text{strong}} \), at least, by requiring the candidate \( \text{O} \text{vi} \) feature to be accompanied by strong \( \text{Ly} \alpha \) and \( \text{Ly} \beta \) absorption. Since higher order series lines become rather weak, adding additional components probably does not follow the same simple multiplicative propagation of probabilities, and so we have considered in this feasibility study only the first two terms of the Lyman series.

Figure 1 shows in red the number of expected \( \text{H} \text{i} \) interlopers without (dashed) or with (dotted) the additional requirement of accompanying \( \text{Ly} \alpha \) and \( \text{Ly} \beta \) absorption. For these calculations, we have used the fiducial values of \( \Delta \lambda_{\text{r}} \leq 3 \times 10^{-4} \) and \( \Delta \lambda_{\text{em}} \leq 0.2 \). Introducing the additional requirement significantly reduces the number of false positives, as Figure 1 demonstrates.

Finally, we can also disqualify a potential candidate as belonging to an \( \text{O} \text{vi} \) doublet if it is clear that it is part of a...
Lyman series itself, i.e., we can check whether there are lower (or also higher) Lyman transitions at the same redshift with similar velocity structures. This cut becomes possible for all lines above 4500 Å when the Lyβ line redshifts into the lower wavelength regime of SDSS. Thus, we expect to even further reduce the number of falsely classified O vi absorbers beyond a redshift of \( z_{\text{Ovi}} \sim 3.37 \) by a large fraction. Another criterion that can be used to correctly identify a metal line doublet is the shape of the lines: both are expected to have the same velocity structure and should be rather narrow, in stark contrast to the Lyman lines that tend to be broader than 20 km s\(^{-1}\). Due to the low resolution of SDSS spectra, however, this criterion cannot be implemented here unless the H i line is so strong that its width exceeds one resolution element. Such strongly saturated or even damped lines are rare, and would have been picked up by other automatic search algorithms already, as we describe later.

We will discuss the implementation and success of these and various other criteria to distinguish real O vi lines from H i features mimicking metal lines in Section 3.

2.2. The Detectability of O vi Doublets in the Lyα Forest

In order to estimate whether blending and/or noise renders the identification of existing O vi doublets in the Lyman forest impossible, we have created 100 mock spectra of a typical Lyman forest, following the number and column density distributions from above, and adjusting the spectral resolution and pixel size to the SDSS fiducial value. For the Doppler broadening parameter, we have assumed a Gaussian distribution of width 8 km s\(^{-1}\) around a mean of \( b = 25 \) km s\(^{-1}\), the exact values for the distribution are, however, not important at this stage of modeling, as the instrumental resolution broadens the line profiles over more than the average velocity width. In this forest, we placed O vi doublets at a redshift of about 3.5, added noise according to the appropriate (worst-case) SDSS scenario (i.e., we have chosen a value of \( S/N = 3 \) here, which is below the nominal \( S/N > 4 \) SDSS limit, see Section 3.3 for details of the SDSS sample’s \( S/N \)), and checked carefully how often we could reliably retrieve the O vi doublet, when varying the strength of the absorption feature (until it completely vanished in the noise). Figure 2 shows two examples: in the first, both components are clearly identified, whereas in the second, both are severely blended. Of course, stronger lines can be identified more easily. From this analysis, we conclude that at least 15% of systems with an EW = \( \log (\text{EW}_{1038 \text{ Å obs}}) = -0.8 \) can be unambiguously identified at a redshift of \( z \sim 3.5 \), and hence higher fractions at lower \( z \) or greater strength. Because the modeling at this stage is only supposed to render an initial estimate of the line detectability, we have not included second-order effects like line clustering, or a detailed description of the effects of continuum placement errors.

How many such absorbers do we expect per sightline? The above limit for the EW lies only slightly higher than the line

![Figure 2](image-url)
The line of sight observability index (LOX) is defined properly in Hellsten et al. (1998) for a metallicity of [O/H] = −2.0.\(^6\)
Integrating the Hu et al. (1995) distribution over the interval 13.8 < log \(N_{H_1}\) < 15.4, where the LOX is of the order of 1.5 and above (cf. Figure 2 of Davé et al. 1998), and along the sightline to a QSO to a redshift \(z_{abs}\) of the absorber, we obtain the number \(n(z_{abs})\) of such lines associated with detectable O \(\text{VI}\) absorption:

\[
n(z_{abs}) = \int_{13.8}^{15.4} \int_{2.7}^{z_{em}} f(N_{H_1})r(\text{LOX}, z_{em})d\log N_{H_1} \frac{dX}{dz} dz,
\]

where \(\frac{dX}{dz} = \left(\frac{1+\frac{z_{abs}}{z_{em}}}{1+z_{abs}}\right)^2\) is the redshift pathlength, and \(r(\text{LOX}, z_{em})\) is the estimate for the fraction of lines correctly identifiable at the emission redshift of the QSO. Here we have assumed for simplicity a constant \(r(\text{LOX}) = 0.15\), a value based upon our simulation at \(z_{em} \sim 3.5\). The rapid rise of the number of expected lines in Figure 1 is a result of the almost linearly growing redshift pathlength \(\int \frac{dX}{dz} dz\). It is clear from Figure 1 that beyond redshifts of \(z_{abs} \sim 4.2\), the sheer number of interlopers will make unambiguous detections of O \(\text{VI}\) features very difficult, and the search efficiency will peak below that redshift.

2.3. Summary of the Initial Feasibility Analyses

From the simulations and estimates above, we conclude that we will be able to retrieve a large enough sample of O \(\text{VI}\) absorbers in the spectra of SDSS quasars to obtain meaningful results for a statistical analysis. Furthermore, we will be able to exclude false identifications via a variety of methods, so that we can hope to clean, to a certain degree, our sample of O \(\text{VI}\) system candidates up to redshifts \(z_{em} \sim 4.0\).

3. THE AGN SAMPLE

3.1. SDSS Imaging and Spectroscopy

The SDSS is an imaging and spectroscopic survey of the sky (York et al. 2000) using a dedicated wide field 2.5 m telescope (Gunn et al. 2006; Gunn et al. 1998) at Apache Point Observatory, New Mexico. Imaging is carried out in a drift-scan mode using five broad ur\(gri\)z bands spanning the range from 3000 to 10,000 Å (Fukugita et al. 1996). The 95% completeness limit in the \(r\) band is 22.2 mag (on the AB system). For more specific information on the imaging and quality control thereof see Hogg et al. (2001), Ivezić et al. (2004), Smith et al. (2002), and Tucker et al. (2006). Details of the astrometry can be found in Pier et al. (2003). A variety of algorithms select objects from the imaging data for spectroscopy. These targets are arranged on series of plug plates, called tiles, with radius 1.49 (Blanton et al. 2003), each containing provision for a total of 640 targets and calibration stars. Optical fibers at the focal plane feed the light from holes in aluminum plates to a pair of double spectrographs. The resulting spectra range from 3800 to 9200 Å with a resolution of \(R = \frac{\lambda}{\delta\lambda} \sim 1800\). The \(S/N\) of the SDSS spectra entering the public database is required to be above 4.0 at a fiber magnitude of 19.9 in the red SDSS spectrum and a \(g\) magnitude of 20.2 in the blue (Adelman-McCarthy 2006). Specific information on SDSS data products can be found in Stoughton (2002) and Abazajian (2003, 2004, 2005) as well as Adelman-McCarthy et al. (2007).

3.2. The SDSS Absorption-line Catalog

Quasars within the SDSS survey are mainly selected by color criteria from the imaging campaign and subsequently verified by spectroscopic follow-up studies (Richards et al. 2002). Quasar spectra with absorption lines have been published by a number of groups: Hall et al. (2002a, 2002b), Menou et al. (2001), Tolea et al. (2002), and Reichard et al. (2003), mainly focusing on broad absorption features. Several teams have developed pipelines isolating and identifying absorption lines and systems in QSO spectra (Vanden Berk et al. 2000, 2008; Richards et al. 2001; York & SDSS Collaboration 2001; York et al. 2005). Following the format of the catalog by York et al. (1991) a comparison of these pipeline-created catalogs with lists made by visual inspection reveal a completeness of above 95% for such an automation. The incompleteness results mainly from blending of the Mg \(\text{II}\) line, key to the detection of QSO absorption systems, most importantly with night sky emission.

The main catalog of York et al. (2005) presents lists of significant absorption lines, EWs, various line parameters, and possible redshifts. Discrete absorption line systems are obtained by matching lines of different transitions of redshifts within a certain range, depending on the \(S/N\) of the spectra. Different criteria such as number and quality of components allow for a grading of the systems—all Class A systems e.g., need to contain at least four, unblended lines above a \(4\sigma\) detection limit matching in redshift, free of artifacts and blends, occurring redward of the Ly\(\alpha\) forest (York et al. 2005). The QSOALS catalog used here is based on the SDSS data release 3 (DR3; Abazajian 2005).

3.3. Our SDSS QSO Sample

At rest-frame wavelengths of 1031.912 Å and 1037.613 Å, the O \(\text{VI}\) doublet starts to become visible in the spectra of SDSS quasars only at redshifts of \(z_{abs} \sim 2.7\). We have thus retrieved all normalized QSO spectra in the QSOALS database beyond this lower redshift. Our sample consists of 3702 quasars with redshifts ranging from 2.70 to 5.413 (Schneider et al. 2005). Figure 3 shows the redshift distribution of the sample. More than 60% of the sources are below a redshift of \(z_{em} = 3.5\), where we expect to maximize the search efficiency for O \(\text{VI}\) doublet signatures and the cleanliness of the O \(\text{VI}\) candidate sample. As Figure 3 demonstrates, the majority of the sources have \(i\)-band magnitudes between 19.5 and 20.5. As the reliability of an absorption line detection is a function of the local \(S/N\), we have determined the average \(S/N\) of all spectra within the region of interest for us. This average value is usually below the overall \(S/N\) of the spectrum since the flux levels within the Ly\(\alpha\) forest are much lower than beyond it. Figure 3 shows the distribution of these average \(S/N\) values with the emission redshifts of the sample sources. Especially toward higher redshifts, it is apparent that the increasing density of the forest and the general faintness of the sources lead to a severe decrease in the \(S/N\) that will hinder unambiguous identifications of absorption lines. However, below \(z_{em} \sim 3.5\) there are enough quasars with \(S/N\) above the conservative limit we used in Section 2.1 for the feasibility study: 359 spectra are of objects with \(z_{em} < 3.5\) and have an \(S/N\) within the forest of \(S/N > 5.0\).

---

5. The line of sight observability index (LOX) is defined properly in Hellsten et al. (1998). It predicts a rest EW \(W_{\gamma}\) for an absorption line of a given species by modeling gas irradiated with a given X-ray/UV background with CLOUDY (Ferland et al. 1998; Hamann & Ferland 1999). For our purposes, LOX \(\sim \log(W_{\gamma}/1\text{mA})\).

6. Here, and in the following, the bracket notation is used to indicate the logarithm of the oxygen to hydrogen abundance relative to the Sun: [O/H] = \(\log[N(\text{O})]/N(\text{H})] - \log[N(\text{O})]/N(\text{H})]\).
Figure 3. Upper panel: the redshift $z_{\text{em}}$ distribution of our SDSS QSOALS sample of 3702 quasars. Note that more than 60% of the sources are below a redshift of $z_{\text{em}} = 3.5$, where we expect to maximize the search efficiency and cleanliness of the O VI candidate sample. Three hundred fifty-nine of these have an $S/N > 5.0$. Middle panel: the $i$ mag distribution (taken from Schneider et al. 2005) of our SDSS QSOALS sample of 3702 quasars. The majority of the sources exhibit fluxes between $19.5 < i \text{mag} < 20.5$. Bottom panel: the average $S/N$ per pixel in the Ly$\alpha$ forest for each source vs. the emission redshift $z_{\text{em}}$. Due to the decreased flux in the forest, this value for the $S/N$ often falls substantially below the nominal lower limit of $S/N = 4.0$ (indicated by the dashed line) for the complete spectrum to enter the SDSS data set, especially toward higher redshifts for which the sources become fainter, in general, and also the density of the forest increases substantially.

4. SEARCH STRATEGY AND ALGORITHM FOR THE SDSS QUASAR SPECTRA

The absorption systems and line lists of the SDSS QSOALS catalogs primarily contain securely identified features, resting primarily on key absorption lines like the Mg $\text{II}$ or C IV doublets. In many cases, identification of such systems is made possible by the fact that the absorption occurs redward of the Ly$\alpha$ forest and thus in relatively clean parts of the spectra. For our search of signatures of the O VI doublet, however, it is clear that all of the candidates must be found in the forest. Furthermore, these O VI systems need not be accompanied by other metal absorption systems like C IV or Mg II that would allow unambiguous identification outside the Lyman forest (cf. Davé et al. 1998). Thus, we cannot expect the automated pipelines used to create the SDSS QSOALS catalogs to have picked up on the relatively weak features within the Lyman forest in which we are interested. Therefore, we had to develop our own algorithms to construct a complete list of O VI absorbers at intervening redshifts. After completing our search with this method, we performed a cross-check with the QSOALS database to see if the putative O VI candidates also show transitions of other ions at the same redshift. Naturally, the overlap between the redshifted wavelength range of the O VI 1032 Å component and the actual wavelength range of the pixel containing the corresponding O VI 1038 Å component is not perfect. And thus, the absorption structure of the O VI 1038 Å transition corresponding to its 1032 Å counterpart can span in principle two pixels, and hence the ratio of the absorption measured in such pairs of 1032/1038 pixels may not be a reliable

4.1. The Search Algorithm for Finding O VI Doublets

For a median Doppler broadening parameter $b_{\text{OVI}} = 16$ km s$^{-1}$, obtained by Simcoe et al. (2002), it is obvious that nearly all of the O VI lines that we expect to be present in SDSS spectra will not be resolved, as 1 Å (i.e., the rough pixel size) represents about 70 km s$^{-1}$ in velocity space at 4000 Å, where we expect to have the best search efficiency. Nonetheless, we implemented a simple pixel-by-pixel comparison$^7$ for all spectra, characterizing each single pixel within the wavelength range from the blue SDSS limit of 3800 Å up to the maximum wavelength allowed for the O VI 1032 Å component $\lambda_{\text{max}} = (z_{\text{em}} + 1) \times 1031.912$ Å by its normalized flux and thus optical depth. For each pixel, we derive its hypothetical absorption redshift $z_{\text{abs}}$ as if it were the O VI 1032 Å component. Then we identify the corresponding pixels in the spectra that would belong to the second O VI 1038 Å component as well as the Ly$\alpha$, Ly$\beta$ and, if possible, Ly$\gamma$ lines for that same redshift. Naturally, the overlap between the redshifted wavelength range of the O VI 1032 Å component pixel and the actual wavelength range of the pixel containing the corresponding O VI 1038 Å component is not perfect. And thus, the absorption structure of the O VI 1038 Å transition corresponding to its 1032 Å counterpart can span in principle two pixels, and hence the ratio of the absorption measured in such pairs of 1032/1038 pixels may not be a reliable

$^7$ One resolution element of an SDSS spectrum consists of at least 2-3 pixels, and thus even a narrow feature like a metal absorption line will spread out over a range of pixels. Our pixel-by-pixel approach here, however, works very well because we select on the EW ratio, which is not affected by this dispersion. Of course, we will thus only be able to detect strong features. For a more detailed comparison of the single pixel approach versus a search after rebinning to the resolution element size, see Appendix B.
quantity for identifying O\textsc{vi} anymore. The estimates of the line widths above, however, demonstrate that the velocity width of a line is most likely going to be only a small fraction of the width of a single pixel—and so this scenario will occur only in a small number of cases. Furthermore, by coincidence, the wavelength separation of the pixels in SDSS spectra and wavelength ratio specific to the O\textsc{vi} doublet create an overlap at the level of 0.94 for all spectra and all pixels. We did, additionally, check for potential O\textsc{vi} candidates in the O\textsc{vi} pixel pairs that do overlap by only 6\%, but the number of candidates retrieved in this way agreed well with just the expectations from random interlopers. We therefore estimate that the loss of valid detections due to this complication is minimal.

4.1.1. Criteria for Identification of O\textsc{vi} Lines and Rejection of Spurious Lyman Series Lines

In order to assess whether an absorption feature exhibits the correct ratio of optical depths while not being able to fully resolve the structure of the absorption lines, we need to derive a quantitative measure of that ratio from the observed normalized fluxes per pixel. Assume that (a) the complete profile of an O\textsc{vi} transition fits into one resolution element and (b) that this line is the only absorption line within that element (i.e., no blending with H\textsc{i} Lyman lines). Then, it is possible to extract the EW of the O\textsc{vi} in that pixel from the measured normalized flux\(^8\) as

\[
f_{\text{pixel,1032 Å}} = 1 - \frac{EW_{\lambda,1032 Å}}{\Delta \lambda},
\]

where \(EW_{\lambda,1032 Å}\) is the EW of the line in that pixel in the observer’s frame, \(\Delta \lambda\) is the pixel size in wavelength units, and \(f_{\text{pixel,1032 Å}}\) is the normalized flux of the pixel assumed to contain the O\textsc{vi} 1032 Å transition. From this estimate of the EW, \(EW_{\lambda,1032 Å}\), we can in turn derive the expected flux \(f_{\text{exp},1032 Å}\) at the corresponding pixel:

\[
f_{\text{exp},1032 Å} = 1 - \frac{1}{r_{EW}} \times (1 - f_{\text{pixel,1032 Å}})
\]

where \(r_{EW} = \frac{EW_{\lambda,1032 Å}}{EW_{\lambda,1038 Å}}\) is the ratio of EWs of the two transitions. This ratio depends on the strength and velocity width of the absorber. For unsaturated lines it remains fairly close to 2 and drops to unity for saturated lines. We have modeled the O\textsc{vi} 1032 Å and 1038 Å transitions with Voigt profiles of various Doppler parameters \(b\), and are thus able to compute:

\[
r_{EW} = \frac{EW_{\lambda,1032 Å}}{EW_{\lambda,1038 Å}} = r_{EW}(EW_{\lambda,1032 Å}, b)
\]

as a function of the estimated \(EW_{\lambda,1032 Å}\). This introduces a model dependence into our derivation as we need to assume a value for the Doppler parameter \(b\). We find, however, that even large variations of the Doppler parameter, 5.0 km s\(^{-1}\) \(< b < 20.0 \text{ km s}^{-1}\), only affect \(r\) on the 30\% level. Here we take \(b = 15 \text{ km s}^{-1}\) as a fiducial value for the Doppler parameter.

The following criteria were used to classify pixels as candidates belonging to an O\textsc{vi} doublet.

1. Both pixels of the O\textsc{vi} doublet candidate need to exhibit normalized fluxes lower than 1.0. This simply ensures that absorption features rather than pixels whose noise pushed them over the flux level of unity are selected.

2. The expected flux \(f_{\text{exp},1038 Å}\) is allowed to deviate from the actually measured flux \(f_{\text{pixel,1038 Å}}\) less than a certain threshold

\[
\Delta f \equiv |f_{\text{exp},1038 Å} - f_{\text{pixel,1038 Å}}| \leq k_1 \times \sigma,
\]

where \(\sigma\) is the standard deviation of this flux difference, estimated from the S/N of each pixel and the uncertainty on the line width mentioned above. The effects of varying \(k_1\) will be described in the next section.

3. The inferred EW of the stronger O\textsc{vi} 1032 Å component has to be above the minimum limit

\[
EW_{\lambda,1032 Å,\text{rest-frame}} \geq k_2 \times \Delta \lambda.
\]

This enables us to clear our candidate lists from very weak features, i.e., spurious Ly\textsc{α} forest lines with low column densities which we expect to be the main contaminant at high redshifts. Note that this criterion is usually automatically fulfilled for candidates in spectra with low S/N passing criterion (4) mentioned below, and thus is mostly useful for cases where the S/N is significantly above the 4.0 threshold for entering the sample.

4. The significance of absorption\(^9\) in each pixel should be above a low threshold

\[
1.0 - f_{\text{pixel}} \geq k_3 \times 1/S/N (\text{pixel}) \times f_{\text{pixel}}.
\]

While this criterion is already implicitly in place for the expected and measured flux difference and the lower limit to the EW, this formulation puts a stronger limit on the number of candidates. Requiring a high significance for each pixel allows us to focus on absorption lines with a high likelihood of being real, rather than noise artifacts. In contrast to criterion (3), this cut proves most useful in spectra with low S/N. Note that this criterion introduces a lower threshold for the EW that is S/N dependent: at a fiducial average S/N of 5.0 (8.0) and a significance level of 4.0σ (2.0σ), it corresponds to a rest-frame EW limit for the SDSS setup of \(EW_{\text{rest}} = 0.195 \text{ Å (0.06 Å)}\).

5. Since we expect O\textsc{vi} to be accompanied by H\textsc{i} Lyman features, we require

\[
f_{\text{Ly}\alpha} < f_{\lambda}(\text{limit}),
\]

where \(\lambda\) represents either Ly\textsc{α}, Ly\textsc{β} or—if applicable—Ly\gamma. Due to a lack of information regarding the metallicity and the physical state of the absorbing gas, there is, a priori, no straightforward way of calculating the H\textsc{i} absorption expected for a given value of \(f_{\text{pixel,1032 Å}}\). Therefore, we introduce these hard cuts on the flux for the associated Lyman series transitions and resort to this simple, yet an efficient method of checking for accompanying H\textsc{i} absorption.

6. We reject all candidates that show signs of being part of Lyman series absorption, i.e., we check (by eye) whether the feature in question could be a Lyman line, accompanied by higher (or lower) series members elsewhere in the spectrum with a similar velocity structure. This turns out not to be a severe cut: only four of altogether 1760 candidates passing all other tests had to be eliminated by this procedure. Thus,

\(^8\) We rely upon the flux normalization via the continuum estimate given in the QSOALS database.

\(^9\) Note that our definition for “significance” relies upon the following definitions: our signal is the strength of absorption in each pixel, i.e., \(1.0 - f_{\text{pixel}}\) in the normalized units. The ratio of this quantity to the noise in the same units, i.e., \((1/S/N) \times f_{\text{pixel}}\), is our metric for the significance.
we find it unnecessary to apply a more rigorous test than this check by eye.

It is clear that the efficiency of detecting O\textsc{vi} candidates and filtering out spurious H\textsc{i} Lyman series absorbers is a complicated function of redshift and S/N of each spectrum. Thus, a more flexible approach, tailored to each object, could result in a higher overall yield. We have, nevertheless, applied the aforementioned rigorous, yet simple strategy to the complete sample regardless of redshift and S/N of each source. This allows for a more straightforward comparison of the results with expectations based upon our initial feasibility study and certainly facilitates quantitative analyses like completeness estimates. And it is certainly the conservative approach.

We note that possible blends of the two O\textsc{vi} 1032/1038 Å transitions with other lines at very similar rest-wavelengths constitute an additional complication. Specifically, the presence of strong lines of C\textsc{ii} might have the 1036 Å line of C\textsc{ii} overlapping with O\textsc{vi} 1038 Å, and potentially shift this blend to a lower wavelength. Also, any system that contains appreciable amounts of molecular hydrogen, H\textsubscript{2}, will also produce a blend with the O\textsc{vi} 1038 Å transition, but leave the 1032 Å component untainted.\(^\text{10}\) At the redshift range of interest for us, however, the separation between the C\textsc{ii} and O\textsc{vi} 1038 Å line centers amounts to at least 6 Å (observed frame), too much to seriously affect our search criteria. Hence, we have not implemented any correction factors due to the presence of these features into our search algorithms.

Interestingly enough, none of our good candidates appear in systems that are bona fide damped-Ly\alpha absorbers (DLAs). While Fox et al. (2007) argue that potentially all DLAs contain appreciable amounts of O\textsc{vi} absorption, we caution that in their sample of 35 high-resolution spectra of DLAs and sub-DLAs covering also the O\textsc{vi} doublet, about 60% of the sightlines show blended features due to the high density of forest lines.

### 4.1.2. Effects of Varying the Selection Criteria on the Number of Candidates

Obviously, the ability to retain most real O\textsc{vi} absorbers in the candidate list, while filtering out the contaminating spurious H\textsc{i} features mimicking metal lines, depends crucially on the choice of the parameters \(k_i\) and \(f_x\) (limit) from our search criteria. We have extensively tested the effects of different choices for these parameters on the numbers and the quality of the candidates.

Figure 4 demonstrates how changing one of the cut criteria while keeping all others constant affects the average number of candidates retrieved from the 380 spectra within the redshift range 2.9 < \(z_{\text{em}}\) < 3.0. These spectra have on average about 300 usable pixels. The data points in black indicate the average number of candidates per spectrum when applying the full search strategy, while the points in red are derived by introducing a random shift of \(\delta\lambda\) (random) rather than the true \(\delta\lambda(z_{\text{obs}})\) for the doublet separation, but keeping the other selection criteria (accompanying Lyman transitions, line significance and EWs). This procedure effectively measures the frequency of interloper incidences. Note that in all cases, we retrieve more candidates for the "real" search than for the randomized case. We have chosen the following parameters for filtering our data set

1. Criterion 2 \(k_1\) : \(\Delta f \leq 0.3 \times \sigma\).
   While at first hand, it seems unnecessary to apply such a strict criterion, an inspection of the candidate lists produced with softer limits revealed that too many spurious candidates could pass. This is mainly due to the uncertainty in the estimated, expected EW of the O\textsc{vi} 1038 Å feature when not knowing the Doppler width of the absorbers. While obviously potentially eliminating good candidates in slightly more noisy spectra, the gain in clearing the sample from H\textsc{i} absorbers mimicking O\textsc{vi} clearly outweighs.

2. Criterion 3 \(k_2\) : \(\text{EW}_{1032\,\text{Å}} \geq 0.05 \times \Delta \lambda\).
   The lower limit on the EW of a line introduced by this cut occurs well before the turnover to a much lower retrieval rate (at EW ~0.3 for the case of 3.9 < \(z_{\text{abs}}\) < 4.0, see Figure 4), but still eradicates effectively a higher number of spurious very low column-density H\textsc{i} absorbers that tend to become ubiquitous at higher redshifts. We have chosen this low limit mainly to keep the (few) cases of spectra with very high S/N (for SDSS standards, i.e., S/N\textsubscript{forest} \(\geq 10.0\)). Note that the lower EW limit introduced by this cut depends on the wavelength coverage of a pixel and is thus slightly redshift dependent.

3. Criterion 4 \(k_3\) : \(1.0 - f(x) \geq 4.0 \times 1/S/N(\text{pixel}) \times f_{\text{pixel}}\).
   The number of candidates is a strongly declining function of the detection significance, as demonstrated by Figure 4. However, above a significance level of ~4.0\(\sigma\) the curve shows signs of flattening, before, finally, at even higher values one simply runs out of pixels above the required S/N. Thus, we have decided to apply the above limit, which turns out to be the strongest selector in cutting down the numbers of candidates.

4. Criterion 5: flux limit for Lyman lines.
   (a) \(f_{\text{Ly}\alpha} \leq 0.4\);
   (b) \(f_{\text{Ly}\beta} \leq 0.6\) (if applicable);
   (c) \(f_{\text{Ly}\gamma} \leq 0.8\) (if applicable).

Limiting ourselves only to candidates that are accompanied by potential Lyman features of such depth maximizes the ratio between real O\textsc{vi} absorbers and random interlopers, as can be seen from Figure 4. Requiring even lower flux values throws out mainly potential candidates in low S/N spectra, when noise pushes a Lyman pixel over the threshold, while softening the flux criterion leaves too many spurious interlopers in the sample.

5. In a final step, all of the candidates having passed the above criteria were checked for possible higher (and also lower) associated Lyman series lines. Only four (out of 1756) candidates had to be eliminated this way, which we take as a sign of having implemented already rigorous cuts to the sample with the criteria discussed above.

### 5. O\textsc{vi} DOUBLET CANDIDATE LIST

Figure 5 shows the results of applying the search algorithm with the cut criteria mentioned in Section 4 to the full data set of 3702 spectra. For each redshift bin, the black data points in the lower panel represent the average number of candidates retrieved per spectrum. This number includes both real candidates and spurious interlopers. The red dotted line is a linear fit to the number of such random interlopers derived by averaging over 15 runs of the search algorithm when applying a random redshift offset to the location of the O\textsc{vi} doublet line (properly at 1038 Å). The redshift evolution of these

\(^{10}\) In fact, there are also two H\textsubscript{2} components at 1032.191 and 1032.356 Å, but these are much weaker than the 1038 Å transitions.
Effects of varying cut criteria on the average number of candidates per spectrum retrieved from all 380 spectra of the sample within the redshift range $2.9 \leq z_{\text{em}} \leq 3.0$. For all panels, the data points in black result from the fully implemented search algorithm for O\textsc{vi} candidates, whereas data points in red are obtained by searching for candidates with random redshift separations instead of the fixed one based upon the O\textsc{vi} wavelength ratio, while still maintaining all the other criteria. The dotted, vertical lines indicate the choices for the cut criteria that minimize the spurious interloper contribution, while simultaneously maximizing the number of real candidates kept in the sample. Criterion 4 (upper left panel): the ratio of the numbers of real to spurious candidates increases up to a significance level of $\sim 4\sigma$. Criterion 2 (upper right panel): the average number of candidates retrieved is a strongly increasing function of the allowed flux deviation of the O\textsc{vi} $1038 \text{ Å}$ pixel from the expected value. The uncertainty in the Doppler width of the absorbers contributes significantly to the error allowance in the estimate for the expected flux. Criterion 3 (lower left panel): the lower limit on the rest-frame EW of the O\textsc{vi} $1032 \text{ Å}$ component appears well before the turnover to a much lower retrieval rate of real candidates ($\sim 0.3 \text{ Å}$ for $2.9 \leq z_{\text{abs}} \leq 3.0$). Criterion 4 is a stronger selector for the low S/N spectra, thus we have chosen this weaker selector mainly to avoid very weak interlopers for candidates in high S/N spectra. Criterion 5 (lower right panel): requiring a low flux value for the accompanying Lyman transitions removes effectively a large fraction of the random H\textsc{i} lines. Note that applying even stricter limits will throw out good candidates in low S/N spectra when noise can push a Lyman pixel over the threshold limit.

**(A color version of this figure is available in the online journal.)**

\[(n \propto (1 + z_{\text{abs}})^{7.3})\] agrees very well with the expectations from the initial feasibility study (cf. Equation (7) and Figure 1). Note that the number of “real candidates” plus “spurious interlopers” for the full search is always higher than in the random search process, lending further credibility to the robustness of our search strategy. The average approximate ratio of “real” to “random” candidates per redshift interval is presented in the upper panel of Figure 5. The number of “real” candidates is assumed to be the difference between the black data points and the estimate for the interloper frequency (red dotted line in the lower panel of that figure). As expected, the ability to produce a clean sample diminishes rapidly beyond redshifts of $z_{\text{abs}} \sim 4.0$.

Altogether, we retrieve 1756 candidates in 855 different AGN spectra (out of a total of 3702) passing all of the above criteria. These candidates still exhibit a wide range of properties. We examined each case manually and grouped them into three categories depending on the quality and likely interpretation.

### 1. High quality sample.

The highest quality, most unambiguous candidates for O\textsc{vi} absorbers. All lines are strong and the O\textsc{vi} doublet plus Ly\textalpha and Ly\betatransitions are not obviously blended. The putative H\textsc{i} absorption is both stronger than the O\textsc{vi} feature and exhibits the expected line ratios for Ly\alpha to Ly\betat. One hundred forty-five candidates could be placed in this group. In 76 cases, the absorption redshift is so close to the QSO’s emission redshift that these absorbers are tentatively classifiable as intrinsic (\[\Delta v = c \times \frac{z_{\text{em}} - z_{\text{abs}}}{1 + z_{\text{abs}}} \leq 5000 \text{ km s}^{-1}\])..

11 We note that in 232 cases two pixels exactly adjacent to each other were selected. Hence, the total number listed in Table 2 is 1756 + 232/2 = 1866. These 116 absorber candidates are also picked up by the search after rebinning. Such cases are exactly what is expected when the noise is low enough to allow a line that will be instrumentally broadened over 2–3 pixels to retain its shape. For details see Appendix B.

12 We allow the Ly\g line to be blended or even non-existent due to its inaccessibility in the spectra for absorbers of the lowest redshift. Maintaining a strict non-blending criterion for Ly\g, which is much weaker than Ly\alpha and Ly\betat, would have left us with too few good candidates.

13 Note that the velocity separation alone does not guarantee the systems to be intervening. See Richards et al. (1999) and Misawa et al. (2007) for cases where high velocity absorber can be regarded as intrinsic to the AGN.
Figure 5. Results of the “real” and “random” search algorithms applied to the 3702 QSO sample. This graph can be directly compared to Figure 1 where the expected values based upon theoretical estimates for the Lyα forest density and the metallicity of the IGM are presented. Lower panel: for each redshift bin, the black data points represent the average number of candidates retrieved per spectrum. This number includes both real candidates and spurious interlopers. The red dotted line is a linear fit to the number of such random interlopers derived by averaging over 15 runs of the search algorithm when applying a random redshift offset to the location of the OVI doublet line (properly at 1038 Å). The resulting redshift evolution of these interlopers agrees very well with the estimates from the feasibility study in paragraph 1. For details see the text. Upper panel: the average approximate ratio of “real” to “random” candidates per redshift interval. The number of “real” candidates is assumed to be the difference between the black data points and the estimate for the interloper frequency (red dotted line in the lower panel). Note the decreasing efficiency to produce a clean sample with increasing redshift. As expected from Figure 1, the “window of opportunity” lies roughly between 0.55 < log(1+z_{abs}) < 0.7 (= 2.7 < z_{abs} < 4.0), when the ratio real/random rises above unity.

(A color version of this figure is available in the online journal.)

Table 2
List of OVI Absorber Candidates in All 3702 SDSS QSO Spectra of the Survey

| Sequence | Plate | Fiber | JD    | \(z_{abs}\) | \(\beta_{em}\) | Quality Flag | S/N (forest) | \(\beta = \Delta v/c\) |
|----------|-------|-------|-------|------------|-------------|--------------|--------------|------------------|
| 1        | 1284  | 140   | 52736 | 2.71413    | 2.93300     | 3            | 6.5           | 0.1286           |
| 2        | 935   | 592   | 52643 | 2.71670    | 3.19500     | 3            | 6.5           | 0.1284           |
| 3        | 935   | 592   | 52643 | 2.71755    | 3.19500     | 3            | 6.5           | 0.1284           |
| 4        | 659   | 181   | 52199 | 2.71927    | 2.94200     | 3            | 4.9           | 0.0598           |
| 5        | 611   | 221   | 52055 | 2.72184    | 2.74100     | 3            | 3.4           | 0.0051           |
| 6        | 659   | 181   | 52199 | 2.72269    | 2.94200     | 3            | 4.9           | 0.0598           |
| 7        | 961   | 450   | 52615 | 2.73127    | 3.03400     | 2            | 5.1           | 0.0811           |
| 8        | 1264  | 74    | 52707 | 2.73299    | 2.97400     | 3            | 2.8           | 0.0645           |
| 9        | 291   | 612   | 51928 | 2.73557    | 2.80900     | 3            | 6.4           | 0.0196           |
| 10       | 611   | 221   | 52055 | 2.73730    | 2.74100     | 2            | 3.4           | 0.0009           |

Notes. The quality flag has the following meaning: (1) excellent, (2) mediocre, and (3) extremely poor.

(This table is available in its entirety in machine-readable and Virtual Observatory (VO) forms in the online journal. A portion is shown here for guidance regarding its form and content.)

2. Medium quality sample.
These candidates generally exhibit a secure identification of many of the features in question, but also at least one characteristic that leaves some doubt, e.g., a blending of one line, weak H\textsc{i} features in comparison to the O\textsc{vi} or excessively broad O\textsc{vi} lines. Five hundred eighty-six (30%) candidates fall into this category.

3. Low quality sample.
In this class of the lowest quality candidates, we encounter mostly examples with very low overall S/N within the
Figure 6. Rest-frame EW distribution of the 1038 Å component for the 1756 candidates found by our search. Note that we have "measured" the EWs in the normalized spectra by simply only including the pixel to the left and right of the candidate pixel into the calculation, and have made no attempt to take the often complicated structure of the forest around the central pixel of the candidate into account. Given that one resolution element should in almost all cases encompass the narrow O\text{\textit{vi}} lines ($b$ parameters of the order of 16 km s$^{-1}$) this method overestimates the width, but does provide secure upper limits.

Figures 7–9 highlight examples of each of the categories.

5.1. Cross-checking the Candidate List with the QSOALS Database

Having constructed the candidate list of O\text{\textit{vi}} absorbers with the search algorithm described above, we performed a cross-check with the QSOALS database in order to determine whether there are additional ions seen at the same redshift. Here, we have focused first on the 69 candidates for intervening absorbers of the highest quality, but will extend the cross-check fully in a forthcoming paper. We have checked for other metal lines accompanying the O\text{\textit{vi}} candidates within a range of ±800 km s$^{-1}$. For 10 of these candidates the QSOALS database lists secure identifications of absorber systems at redshifts very close to value determined from the O\text{\textit{vi}} doublet. In all of these cases, the velocity difference between the O\text{\textit{vi}} absorber determined by our search algorithm and the listed QSOALS redshift for the absorber system is less than 300 km s$^{-1}$. These absorbers show transitions in a variety of other ions, most forest or features that are probably strong Ly$\alpha$ absorbers at redshifts low enough such that we cannot see Ly$\beta$ to apply our exclusion criterion. Around 1029 of the 1756 candidates ($\sim60\%$) fall into this category, which is considerably higher than the fraction we expected to be spurious interlopers from our initial feasibility studies, and thus leads us to the conclusion that even in this group there are real O\text{\textit{vi}} absorbers.
prominently C\textsc{iv} 1548/1550 Å, C\textsc{ii} 1335 Å, Si\textsc{iv} 1394/1402 Å, Al\textsc{ii} 1671 Å, and, in two cases of low enough redshift not to push the lines beyond the SDSS spectral window, even Al\textsc{iii} 1855/1863 Å. Not only does the existence of the other transitions in these absorbers greatly increase the likelihood for the O\textsc{vi} absorbers to be real, but it will also enable us to study the physical state of the absorbing medium via the full, established arsenal of astrophysical diagnostics. The in-depth analysis of these absorbers will be the central theme of forthcoming papers.

While the remaining 59 candidates have no absorption systems listed in the QSOALS database, there are cases where the database indicates the detection of an absorption feature at the location expected from our O\textsc{vi} redshift, but does not assign an identification, probably due to a failure to securely identify the transition. We have checked for the presence of the following ions: C\textsc{iv} 1548/1550 Å, C\textsc{ii} 1335 Å, Si\textsc{iv} 1394/1402 Å, Al\textsc{ii} 1671 Å, and Al\textsc{iii} 1855/1863 Å. These are the primary transitions that are expected to be strong enough and, due to their rest wavelengths, can occur beyond the Lyman forest in the part of the spectrum where the detection is thus much easier. For 36 out of the 59 candidates, we do not find any securely detected absorption feature in the QSOALS catalogs. In 11 cases, there is one of the transitions listed, and in eight cases two absorption features are present. For another two absorbers, we found three features each, and in the two best cases, even four of the eight transitions in question could be retrieved. Given the low likelihood that these features are noise artifacts, especially for the cases where they are redward of the Lyman forest, the additional presence of up to four absorption lines lends further credibility to the reality of the O\textsc{vi} candidate.

We note that the non-detection of a transition in the QSOALS catalog cannot be equated with the non-existence of the feature in the spectrum. In order to enter the QSOALS catalog, a feature needs to pass certain threshold criteria, set by the automatic program detailed in York et al. (2005). Specifically, the requirement of a $4\sigma$ detection of a feature results in an EW sensitivity which depends on the S/N, but is at least of the order of 50–200 mÅ in the absorber restframe (York et al. 2005). We are going to manually recheck each spectrum separately at the wavelengths where we expect potential absorption features and derive upper limits for lines not detected. These can also help with the interpretation of the physical state of the absorbing gas.

A more difficult task is the assessment of the completeness level of our sample. We have noticed that, in some cases, the QSOALS database indicates the presence of absorbing systems with O\textsc{vi} which is not being picked up by our search algorithm. We have examined closely the results for both search algorithms for the 134 QSO sightlines with $3.2 \leq z_{\text{em}} \leq 3.23$. We find that we retrieve 12 out of 37 grade A, B, and C absorption systems (containing 4, 3, or 2 secure line identifications at the same redshift, respectively)\textsuperscript{14} with O\textsc{vi} doublet identification in the QSOALS, whereas none of the five grade D systems with O\textsc{vi} is detected by our algorithm. On the other hand, we obtain 24 O\textsc{vi}

\textsuperscript{14} For the exact definitions of these grades, see Lundgren et al. (2009).
Figure 9. Example of a poor quality O\textsc{vi} absorber candidate. The position of the O\textsc{vi} doublet and the associated Ly\textalpha, β lines are indicated by the dotted lines. The plate, fiber, and Julian date identification as well as the emission and absorption redshifts together with the \textit{i} magnitude and the general S/N within the Lyman forest are given at the top of the graph. These candidates generally allow only for very insecure identifications of many of the features in question, mostly because of low overall S/N in the forest (as in the case presented here), but also blending and even a few cases where a strong Ly\textalpha system could not be identified as such by our method due to the lack of accompanying higher order Lyman series transitions because of its low redshift. About 60\% of the candidates fall into this category.

candidates in that subsample that do not appear in the QSOALS database, or only show one of the two doublet components.

What is the reason for the apparently low (∼30\%) success rate for our algorithm compared to the QSOALS search program? In most of the cases where we do not detect a system, one or even both of the O\textsc{vi} doublet transitions are blended, and in some cases there was a lack of accompanying Ly\textbeta absorption. Furthermore, in a significant fraction of the QSOALS candidates, there is a velocity difference between various metal components greater than the SDSS pixel separation, as pointed out by York et al. (2005), which leads to an exclusion of the candidate in our approach. Thus, it is clear that the rigidity of our search criteria does lead to a certain fraction of lost real candidates. In order to meaningfully constrain the properties of the IGM at high redshift, the estimate of this lost fraction is absolutely crucial. We postpone a full Monte Carlo analysis of the search efficiency to the next paper in the series that will deal with this issue in depth for the absorbers with the highest likelihood of being real.

The results of this cross-check with the QSOALS database are summarized in Table 3.

6. SUMMARY AND CONCLUSIONS

We have systematically searched for signatures of metal lines in quasar spectra of the SDSS, focusing on finding intervening absorbers via detection of their O\textsc{vi} doublet. In this paper, we have presented our search algorithm and criteria for distinguishing candidates from spurious Ly\alpha forest lines. In addition, we compare our findings with simulations of the Ly\alpha forest in order to estimate the detectability of O\textsc{vi} doublets over various redshift intervals. We have obtained a sample of 1756 O\textsc{vi} doublet candidates in 855 AGN spectra (out of 3702 objects with redshifts in the accessible range for O\textsc{vi} detection). This sample is further subdivided into three groups according to the potential for follow-up of these candidates: 145 of the candidates with the cleanest signatures for O\textsc{vi} doublets with high S/N and high resolution are promising candidates for higher resolution spectroscopy in order to better constrain the physical state of the absorbers. Seventy-six of these, however, could be intrinsic absorbers as their redshift and thus velocity separation from the QSO is less than 5000 km s\textsuperscript{−1}. This leaves us with 69 strong candidates of intervening O\textsc{vi} absorbers at redshifts beyond \(z_{\text{abs}} = 2.7\).

The efficiency of detecting O\textsc{vi} candidates and filtering out spurious H\textsc{i} Lyman series absorbers is a complicated function of redshift and S/N of each spectrum, and thus more flexible search criteria, tailored to each object, could result in a higher overall yield than achieved here. We decided, however, to apply rigorous, yet simple cuts to the complete sample regardless of redshift. This will allow for a more straightforward comparison of the results with expectations based upon our initial feasibility study, and certainly facilitates quantitative analyses like completeness estimates, which we will undertake in a forthcoming paper.

To summarize, our “blind” pixel-by-pixel search for O\textsc{vi} absorbers in the over 3700 SDSS spectra of redshifts beyond...
Table 3
Details for Category 1 O vi Absorber Candidates

| Plate | Fiber | JD   | \( z_{\text{em}} \) | \( z_{\text{abs}} \) | S/N (forest) | \( \lambda \) C iv 1548 Å | \( \lambda \) C iv 1550 Å | \( \lambda \) C n 1335 Å | \( \lambda \) Si iv 1394 Å | \( \lambda \) Si iv 1402 Å | \( \lambda \) Al ii 1671 Å | \( \lambda \) Al iii 1855 Å | \( \lambda \) Al iii 1863 Å | \( z \) (QSOALS) |
|-------|-------|------|-----------------|-----------------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|--------------------|
| 819   | 530   | 52409| 3.0310          | 2.9405          | 2.1          | 6102.3         | 6114.2         | 5260.0         | 5493.7         | 5529.0         | 6587.0         | ...            | ...            | 2.9417             |
| 971   | 508   | 52644| 3.0568          | 2.9779          | 6.0          | 6160.1         | 6170.2         | 5308.6         | 5545.9         | ...            | 6647.7         | ...            | ...            | 2.9791             |
| 830   | 431   | 52293| 3.3880          | 3.3127          | 3.3          | 6678.4         | 6689.7         | ...            | 6012.5         | 6052.6         | 7205.3         | ...            | ...            | 3.3143             |
| 302   | 438   | 51688| 3.5500          | 3.4305          | 2.3          | 6866.8         | 6877.8         | ...            | 6181.4         | 6221.3         | 7409.0         | ...            | ...            | 3.4356             |
| 1165  | 391   | 52703| 3.5790          | 3.4643          | 1.9          | 6913.3         | ...            | 5957.5         | 6223.4         | ...            | 7493.3         | ...            | ...            | 3.4657             |

(This table is available in its entirety in machine-readable and Virtual Observatory (VO) forms in the online journal. A portion is shown here for guidance regarding its form and content.)
\( z_{\text{em}} = 2.7 \) complements other studies that have targeted possible O\( \text{VI} \) features by their association to already known, strong absorbers such as damped Ly\( \alpha \), Lyman limit systems and even metal-line absorbers (Simcoe et al. 2002, 2004). Traditionally, these studies have focused on very high S/N and resolution, at the price of thus being limited to relatively few individual sources due to the expensive nature of the observations. Our sample, residing at \( z_{\text{abs}} \geq 2.7 \), therefore greatly increases the redshift range of known O\( \text{VI} \) absorbers at high redshifts. It will allow us to test expectations based upon photoionization models like those presented in Davé et al. (1998), thereby possibly constraining the physical conditions of the absorbers and the metagalactic UV/X-ray background at high-redshifts.

We thank the anonymous referee for valuable comments that have improved the quality of the paper dramatically.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

**APPENDIX A**

CANDIDATE LIST OF O\( \text{VI} \) ABSORBERS AND DETAILS FOR BEST CANDIDATES

Table 2 is an excerpt of the list of all O\( \text{VI} \) candidates retrieved from the 3702 SDSS QSO spectra in our survey. Listed are the unique identifiers of the SDSS object (plate, fiber, and Julian date of observation), as well as the emission redshift of the QSO from the SDSS database. The overall quality flag for each absorber candidate at the given absorber redshift, \( z_{\text{abs}} \), has the following meaning (for details see the text): (1) excellent, (2) mediocre, and (3) extremely poor. The S/N rate in the forest is the average S/N measured for the wavelength area of interest for the O\( \text{VI} \) search, and often substantially less the nominal limit of S/N (total) > 4.0 for an object to enter the SDSS database. The velocity difference between the absorber and the QSO is calculated from the given emission line redshift estimate, \( z_{\text{em}} \). Absorbers at velocity differences less than 5000 km s\(^{-1}\) (i.e., \( \beta < 0.0167 \)) are classified here as “intrinsic” or “associated,” while all other absorbers are tentatively “intervening.” The table is sorted by increasing absorber redshift. The full table is available in the online version of the journal.

Table 3 contains additional information on the intervening O\( \text{VI} \) candidates with the highest likelihood of being real (category 1) and a comparison with the results of the QSOALS database. Listed are the unique identifiers of the SDSS object (plate, fiber, and Julian date of observation), as well as the emission redshift of the QSO from the SDSS database. The S/N rate in the forest is the average S/N measured for the wavelength area of interest for the O\( \text{VI} \) search, and often substantially less than the nominal limit of S/N (total) > 4.0 for an object to enter the SDSS database. We have also included the position of other ionic transitions at the same redshift as the O\( \text{VI} \) doublet detected in the QSOALS database. The first 10 entries in the table are the candidates where the QSOALS database lists a secure system. The following 59 sightlines have no securely detected system at the redshift of our candidates, however, in certain cases specific single transitions are detected. The full table is available in the online version of the journal.

**APPENDIX B**

SINGLE PIXEL SEARCH VERSUS SMOOTHING TO THE SDSS RESOLUTION

We have performed the search for O\( \text{VI} \) absorbers based upon the information in the single pixels of the original SDSS spectra. The SDSS detector is built in such a way that spectra are slightly oversampled compared to the resolution element achievable with the nominal \( R = 1800 \) resolution. Hence, the question arises: could we have done better by smoothing the spectra according to the oversampling before searching, hereby gaining potentially significantly in S/N? The answer to this question depends to a degree on the goals that are desired for the sample of candidates. In this appendix, we compare the pros and cons of the two possible approaches and explain why we have chosen to use the single pixel search method.

Let us begin by clearly stating our criterion for a successful search algorithm. We want to obtain a sample of good quality O\( \text{VI} \) absorber candidates. With such a sample we want to be able to perform two different actions: (1) the selection of follow-up candidates with higher resolution and S/N. Note that we have set out to construct a sample at higher redshift than has ever been tried before and (2) derive basic statistical quantities such as \( dn/dz \) and density/metallicity estimates for this sample. Hence, we need to try both to maximize the number of absorbers found and the cleanliness of the sample. It is immediately obvious that increasing the S/N of the basic search bin (pixel versus resolution element) could indeed be very helpful to increase the number of absorbers we can hope to find, given the often very low quality of the spectra with S/N per pixel of the order of a few. On the other hand, smoothing over a larger bin size increases the probability of H\( \text{I} \) interlopers falling within the bin distorting the flux measurement in a good candidate, or creating new bad candidates by mimicking an O\( \text{VI} \) doublet. Therefore, it is a priori not clear which method is favorable.

In order to decide upon a quantitative argument, we have indeed smoothed all 3702 spectra to the nominal resolution of the SDSS spectrograph by running a Gaussian filter with the appropriate FWHM over each spectrum. The oversampling factor is about 2.5 and hence we gain roughly a factor of \( \sqrt{2.5} \sim 1.6 \) in S/N for each pixel. Then we re-ran just the same search algorithm over these new spectra. The results compared to the original search can be summarized as follows.

1. In the single pixel search, we retrieve 1756 “unique” candidates for absorbers, i.e., correcting for finding double
Figure 10. Results of the search for $O\text{\scriptsize{}VI}$ candidates in the spectra that have been smoothed by the proper SDSS oversampling. The number of candidates found for each combination of the cut criteria can be directly compared to those with the single-pixel approach (Figure 4, and dotted lines here). It is evident that the smoothing yields more candidates by a factor of 5–8 for each combination, but the increase in the fraction of random interlopers (red lines) is even more dramatic.

(A color version of this figure is available in the online journal.)

It is this latter point that clearly favors the implementation of the one-pixel search strategy, given the goal of a candidate list that is as clean as possible under the given noise and resolution constraints of SDSS. Note that the number of detected candidates and random interloper estimates depends strongly on the cut criteria introduced in Section 4.1.1. The specific values we have adopted were chosen to achieve an optimal result for the single-pixel method. Is it possible to change those parameters for the new search in order to achieve a lower fraction of interlopers? We have tested this by varying the cut values over a wide range of the parameter space, just like we did for the original one-pixel search as detailed in Section 4.1.2., but could not find any combination of them that had the interloper fraction drop below 75%. This is highlighted in Figure 10 that can directly be compared to Figure 4: it is evident that the number of detections rises by a factor of about 5–8 for all cut criteria, but the fraction of random interlopers increases even more dramatically all across the board (red lines in the figure). This result is somewhat surprising given that the initial feasibility analysis showed that by increasing the $S/N$ the ratio of real-to-random candidates should also increase. A potential complication may be the clustering of (stronger) Lyman forest features that could increase the number of interlopers.

The overlap fraction between candidates found in both the single-pixel approach and the search after smoothing may appear on first glance low. Specifically, of the 1756 candidate systems from the former method only 543 remain after smoothing (31%), while our estimate of the fraction of real absorbers versus spurious interlopers and noise artifacts indicates that 69% of the candidates should show a true $O\text{\scriptsize{}VI}$ profile and hence extend over more than 1 pixel, i.e., are suitable for a successful resolution detection. In the following, we investigate how these two seemingly contradicting numbers can be reconciled.

While binning the spectra to the instrumental resolution of SDSS, i.e., combining the flux from 2 to 3 adjacent pixels, certainly reduces the noise, but since the absorption in neighboring pixels is typically weaker one may actually lower the $S/N$ of the candidate. A priori, it is difficult to assess how the detectability of lines is affected by this modification of the search routine. Hence, we have resorted to a Monte Carlo analysis, creating new data sets of mock absorbers in resolution space based upon the candidates’ specifications obtained in the single pixel search. These new sets of candidates were scrutinized by the same al-
algorithm we used to search for absorbers in the real, smoothed spectra.

The recipe for creating one such new data set is as follows: we begin by extracting for each candidate the relevant flux and noise value at each pixel in question for the search criteria (i.e., O\textsc{iv} 1032 and 1038, Ly\textsc{z}, Ly\textsc{\beta}, and Ly\textsc{\gamma} (when possible)). In order to estimate the expected flux values in adjacent pixels, we simply create a Voigt profile which needs to reproduce the flux in the central pixel and has an appropriate Doppler parameter (fixed to \( b \sim 14 \text{ km s}^{-1} \) for the O\textsc{iv}, and \( b \sim 35 \text{ km s}^{-1} \) for H\textsc{i}). To these expected fluxes in the two adjacent pixels, we add noise at the level found for that central pixel. Then we combine the information of the new set of pixels according to the method used for the smoothing of the data, i.e., adding and averaging the fluxes while adding the noise in quadrature only. Hence, we derive for each candidate of the one-pixel search a new set of flux and noise values in the relevant transitions. In this way, we construct a new sample of artificially created potential resolution candidates, based upon the assumption that the one-pixel flux values found in our search represent the depression of flux in the center of the line. The new sample is then subjected to the same search algorithm we used for the real rebinned spectra, and the fraction of candidates passing the search criteria is noted. We repeat this procedure for 10,000 versions of such mock lists, in order to estimate the expectation value and its scatter for the number of candidates remaining. From this simple procedure, we derive an average “retention rate” of 38% ± 3%, only slightly higher than the value for the overlap percentage in the real data sets. It seems plausible that allowing for more complicated scenarios (line centers not exactly at central pixel, correlated noise, range of Doppler parameters for underlying profiles, etc., ...) the fraction could be pushed even lower, and hence we conclude that it is not a sign of our search algorithms being unreliable when finding a low overlap percentage.

REFERENCES

Abazajian, K., et al. 2003, AJ, 126, 2081
Abazajian, K., et al. 2004, AJ, 128, 502
Abazajian, K., et al. 2005, AJ, 129, 1755
Adelberger, K. L., Steidel, C. C., Shapley, A. E., & Pettini, M. 2003, ApJ, 584, 45
Adelman-McCarthy, J. K., et al. 2007, VizieR Online Data Catalog, 2276, 0
Adelman-McCarthy, J. K., et al. 2006, ApJS, 162, 38
Aguirre, A., Schaye, J., & Theuns, T. 2002, ApJ, 576, 1
Aguirre, A., et al. 2001, ApJ, 560, 599
Aracil, B., Petitjean, P., Pichon, C., & Bergeron, J. 2004, A&A, 419, 811
Aracil, B., Tripp, T. M., & Bowen, D. V. 2005, in IAU Colloq. 199, Probing Galaxies through Quasar Absorption Lines, ed. P. Williams, C.-G. Shin, & B. Menard (Cambridge: Cambridge Univ. Press), 65
Barkana, R., & Loeb, A. 2001, Phys. Rep., 349, 125
Bergeron, J., Aracil, B., Petitjean, P., & Pichon, C. 2002, A&A, 396, L11
Bergeron, J., & Herbert-Fort, S. 2005, arXiv:astro-ph/0506700
Blanton, M. R., et al. 2003, AJ, 125, 2276
Bowen, D. V., et al. 2008, ApJS, 176, 59
Burles, S., & Tytler, D. 1994, ApJ, 460, 584
Carswell, B., Schaye, J., & Kim, T.-S. 2002, ApJ, 578, 43
Cen, R., Miralda-Escudé, J., Ostriker, J. P., & Rauch, M. 1994, ApJ, 437, L9
Cen, R., Tripp, T., Ostriker, J., & Jenkins, E. 2001, ApJ, 559, 5
Couchman, H. M. P., & Rees, M. J. 1986, MNRAS, 221, 53
Cowie, L. L., & Songaila, A. 1998, Nature, 394, 44
Davé, R., Hellsten, U., Hernquist, L., Katz, N., & Weinberg, D. H. 1998, ApJ, 509, 661
Davé, R., Hernquist, L., Weinberg, D. H., & Katz, N. 1997, ApJ, 477, 21