Advances in Astronomy and Space Physics, **5**, 84-88 (2015)

# A quasar sample for the Ly$\alpha$ forest studies from the Data Release 10 of the Sloan Digital Sky Survey

*O. Torbaniuk**

Main Astronomical Observatory of the NAS of Ukraine, 27 Akademika Zabolotnoho Str., Kyiv 03680, Ukraine

We present a new sample of the $z \geq 2$ quasar spectra. It contains 102 643 spectra which were visually selected from the SDSS DR10, and includes also a subsample of 65 976 spectra for the composite spectra compilation. This sample will be used for the Ly$\alpha$ forest studies, and can be used for other studies, including those of the quasar spectral properties and the spatial distribution of quasars at $z > 2$. The compiled composite spectra will be used for the determination of the intrinsic spectrum in the Ly$\alpha$ forest studies. Those objects which were not included into the main sample and rejected during the visual examination, are 11 192 quasars with the broad absorption lines, 6 804 spectra with the damped Ly$\alpha$ systems, 1 248 and 493 spectra with the absorption in the Ly$\alpha$ and Ly$\beta$ lines, respectively. The "non-quasar" objects, including 191 candidates for blazars and 30 galaxies with starburst, as well as 617 quasar spectra with wrong redshift, 417 incomplete spectra and 1 497 spectra with low S/N ratio, were also excluded.

**Key words:** quasars: general, methods: data analysis, catalogues

## INTRODUCTION

One of the main observational tools allowing to study the matter distribution in the Universe is the analysis of Ly$\alpha$ forest lines in the spectra of distant ($z > 2$) quasars. This forest is a set of the absorption Ly$\alpha$ (1215.6 Å) lines blueward of the intrinsic quasar emission Ly$\alpha$ line. It is a result of the absorption of the quasar light by the neutral intergalactic hydrogen in the filaments distributed along the line of sight, thus there is a bunch of absorption lines with different redshifts instead of one line. The distribution of the transmission (the value characterising transparency of intergalactic medium (IGM) in the Ly$\alpha$ line) fluctuations in the Ly$\alpha$ line, which is based on the two-point statistics, gives the information about the matter distribution on the intergalactic scales. Such studies are conducted with two types of data: small samples of high-resolution spectra (obtained using the echelle spectrographs at VLT etc.) and large samples of medium-resolution spectra mainly from the Sloan Digital Sky Survey (SDSS). Both have their pros and cons and should supplement each other in the overall picture.

Several steps to obtain values of the transmission fluctuations from the raw spectral data imply several sources of inaccuracy of the final result. For example, one of the main problems is the determination of the intrinsic spectrum of a quasar. This is being done using different techniques by different authors. Up to date almost all known studies of the Ly$\alpha$ forest from the SDSS have been done only by the SDSS collaboration [1, 4, 5, 6]. Consequently, it is necessary to conduct independent studies for the verification of the results. For this purpose we compiled a new sample of the quasar spectra from the Tenth Data Release of SDSS (SDSS DR10). In this paper we describe the sample selection criteria and characteristics.

Our sample will be used for two interrelated tasks: the compilation of the composite spectra and the study of the matter distribution on the intergalactic scales using the Ly$\alpha$ forest. Hence an additional sample of spectra with the relatively high S/N-ratio was compiled from the main one for the Ly$\alpha$ forest studies. The composite spectra are the averaged spectra of a certain type of objects. The main purpose of the composite spectra compilation is the increase of the S/N ratio, which allows to distinguish spectral features not seen in the individual spectra having low S/N ratios. Averaging over a sample of spectra yields also elimination of the separate intergalactic absorption lines, i. e. the Ly$\alpha$ forest, and results in a general flux reduction within the Ly$\alpha$ forest region. The mean transmission can be found from the value of the latter.

The subsample compiled for the composite spectra production will also be used for studying the spectral properties of quasars and relations between them, that can shed light on the quasar physics. For example, in our previous studies with the smaller sample of quasar spectra from SDSS DR7, we showed that there is no correlation between the spectral index $\alpha_\lambda$ and the monochromatic luminosity at 1450 Å,

*el.torbaniuk@gmail.com

$\log l_{1450}$ [2]. We also found that there is a dependence of the emission line equivalent width on spectral index (correlation or anti-correlation) for some of the lines, mostly for those for which the Baldwin effect is detected, and that there is no dependence between the virial mass of the central supermassive black hole of a quasar and its spectral index $\alpha_\lambda$ [8, 9].

## THE DATA AND SPECTRA SELECTION

We used the SDSS DR10 quasar catalogue [3] that contains 166 583 objects within the redshift range $0.053 < z < 5.855$. Firstly, all objects with redshifts $z > 2.0$ and the redshift determination confidence level $> 0.9$ were selected, since the region of Ly$\alpha$ forest is observed at these redshifts. The resulting preliminary sample contains 125 132 objects.

The second step was the visual inspection of the preliminary sample. This is needed because SDSS is an automatic survey and may contain the "non-quasar" objects (stars, supernovae, normal galaxies etc.) as well as the spectra with the too low signal-to-noise (S/N) ratio or the spectral peculiarities unwanted in the Ly$\alpha$ forest studies (e. g., BAL, DLA). Low S/N means that the noise is so high, that one can distinguish not more than one emission line; the value of S/N ratio in these cases is usually less than three.

In addition to the 30 normal or starburst galaxies and 191 blazar candidates, 1 497 spectra with very low S/N-ratio (which does not allow to classify the object type), 617 quasars with wrong redshifts, and 417 incomplete spectra, we excluded the quasar spectra with the Broad Absorption Lines (BAL) and the Damped Ly$\alpha$-systems during the visual inspection. These two latter types of spectra were excluded because of some difficulties in using them for the Ly$\alpha$ forest studies. For the BAL quasars it is not easy to use the automatic methods of the intrinsic spectrum fitting, and the presence of the intrinsic absorption in the Ly$\alpha$ forest region complicates the separation of the IGM absorption. The DLA-systems are the absorption features with the zero transparency in the Ly$\alpha$ line.

To compile a subsample for the composite spectra production we excluded the following spectra from the main one: those with the absorption in the Ly$\alpha$ and Ly$\beta$ lines (1 248 and 493, respectively) and also the spectra with the narrow absorption lines redward of the Ly$\alpha$ emission line (36 666 spectra). The latter are most likely the strong metallic absorption lines from IGM which have sporadic nature and can introduce additional uncertainties into the composite spectrum both within the Ly$\alpha$ forest and outside of it.

The following samples are obtained: the sample for Ly$\alpha$ forest studies containing 102 643 spectra (examples of spectra with $z = 2.45, 3.64, 4.34$, $5.42$ are presented in Fig. 1) and the sample for composites containing 65 976 spectra (see Fig. 2). Both samples are presented as an online table at the AASP webpage. Table 1 presents a part of this table with an example of ten quasars. The last column indicates to which sample the object belongs to (note, that the "composite" means that a given object belongs to both samples). In addition, the following objects were selected: 11 192 BAL-quasars (Fig. 3(a)), 6 804 quasars with DLA (Fig. 3(b)). Examples of the spectra with the absorption in the intrinsic Ly$\alpha$ and Ly$\beta$ emission lines are presented in Fig. 4(a) and (b), respectively.

## ADDITIONAL SAMPLE REDUCTION

Since the main purpose of the selected sample is the study of the Ly$\alpha$ forest requiring as high S/N ratio as possible, some additional reduction was applied for both samples. For this purpose some preliminary data processing was carried out first. The spectra were smoothed with a simple moving average by three points.

The next step is the normalisation of each spectrum, which is needed for the compilation of the composite spectra. Each spectrum was normalized to the mean (arithmetic) flux in all the pixels within the rest wavelength range 1450-1470Å. This range is located blueward of the C IV emission line and is usually considered to be free of emission and absorption lines. For further study we used only the spectra with the RMS of the normalisation constant $A$ less than 10%, resulting in the samples of 42 140 spectra (the sample for the Ly$\alpha$ forest study) and 21 868 spectra (for the composites). The redshift distributions of the preliminary, main and composite samples before and after the imposition of the conditions with a normalisation constant $A$ are shown in Fig. 5.

## RESULTS AND CONCLUSIONS

We compiled a new sample of quasar spectra from the SDSS DR10 sample [3] of 166 583 objects. Out of this sample, we have visually selected 102 643 "true" quasars, excluding 11 192 quasars with BAL, 6 804 spectra with DLA, 1 248 and 493 spectra with absorption in Ly$\alpha$ and Ly$\beta$ lines, respectively. Also there were excluded "non-quasar" objects, namely 191 candidates to blazars and 30 galaxies with starbursts. From the main sample we also excluded 617 quasar spectra with wrong redshifts, 417 incomplete spectra and 1 497 spectra with low S/N ratio.

From the main sample, containing 102 643 spectra, we compiled additional subsample for making the composite spectra. This subsample contains 65 976 quasar spectra and does not include the spectra with narrow absorption lines redward of the quasar Ly$\alpha$ line.

For further studies, to reduce the uncertainties we used only spectra with the RMS of the normalisation constant $A$ less than 10%, hence the number of spectra of spectra is reduced to 42 140 (the main sample)

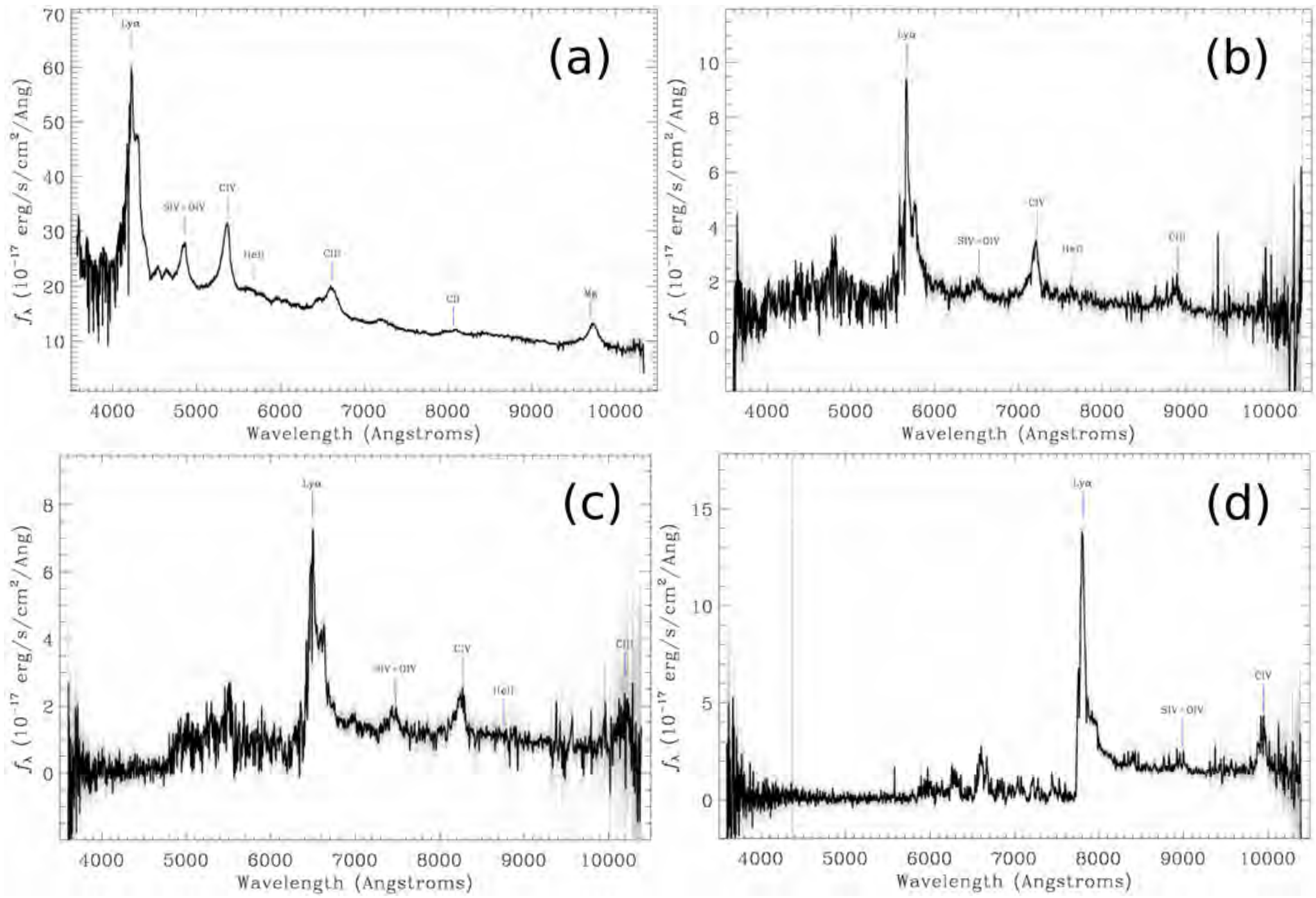


Fig. 1: The quasar spectra with the redshifts $z = 2.45$ (a), 3.64 (b), 4.34 (c), 5.42 (d), respectively.

and 21 868 (the sample for the composite spectra). From the latter sample we compiled 55 composite spectra from the subsamples with the different spectral indices $\alpha_\lambda$ and different monochromatic luminosities at 1450 Å, $\log l_{1450}$. Then we applied them for the automatic intrinsic spectrum determination in the spectra from the former sample using our own technique. Details of this method and the resulting dependence of the mean Ly$\alpha$ forest transmission on the redshift will be presented elsewhere.

## ACKNOWLEDGEMENT

The author is thankful to Dr. Ganna Ivashchenko and Dr. Irina Vavilova for their invaluable help and fruitful discussions. This work has been supported by the Target Programme of Space Research of the NAS of Ukraine for 2013-2016 and by the Swiss National Science Foundation grant SCOPE IZ7370-152581.

## REFERENCES

[1] Anderson L., Aubourg É, Bailey S. et al. 2014, MNRAS, 441, 24

[2] Ivashchenko G., Sergijenko,O. & Torbaniuk O. 2014, MNRAS, 437, 3343

[3] Pâris I., Petitjean P., Aubourg É. et al. 2014, A&A, 563, id. A54

[4] Ross A. J., Samushia L., Burden A. et al. 2013, MNRAS, 437, 1109

[5] Sánchez A. G., Montesano F., Kazin E. A. et al. 2014, MNRAS, 440, 2692

[6] Tojeiro R., Ross A. J., Burden A. et al. 2014, MNRAS, 440, 2222

[7] Torbaniuk O. & Ivashchenko G. 2014, IAU Symp., 304, 282

[8] Torbaniuk O. & Ivashchenko G. 2014, in *WDS'10 Proc. Contributed Papers – Physics*, 42

[9] Torbaniuk O., Ivashchenko G. & Sergijenko O. 2012, in *WDS'12 Proc. Contributed Papers: Part III – Physics*, 123

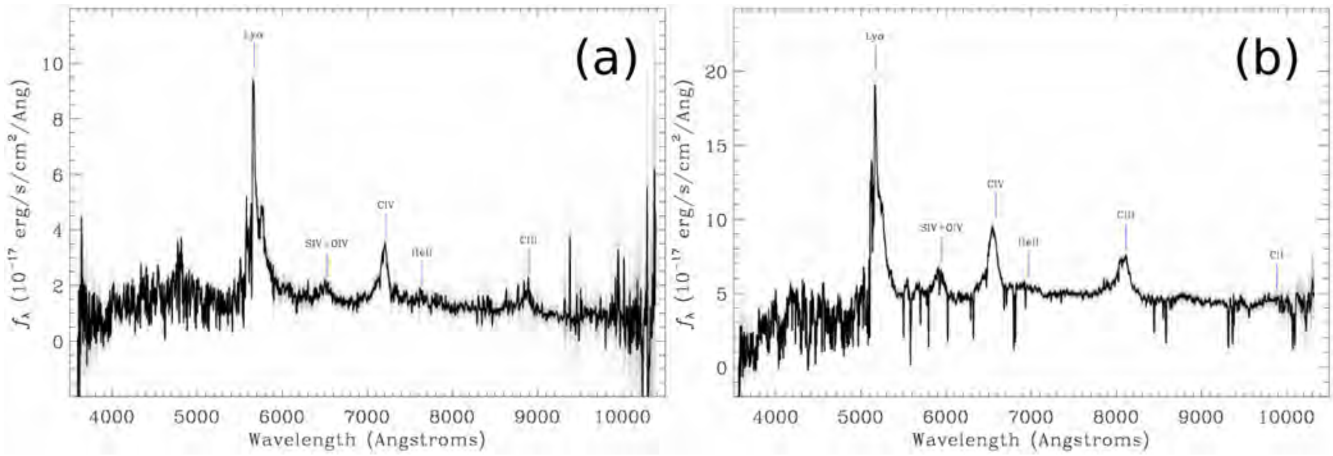


Fig. 2: Examples of the spectra from the sample for the composite spectra (a) and the spectrum with narrow absorption lines for the Ly$\alpha$ forest study (b), respectively.

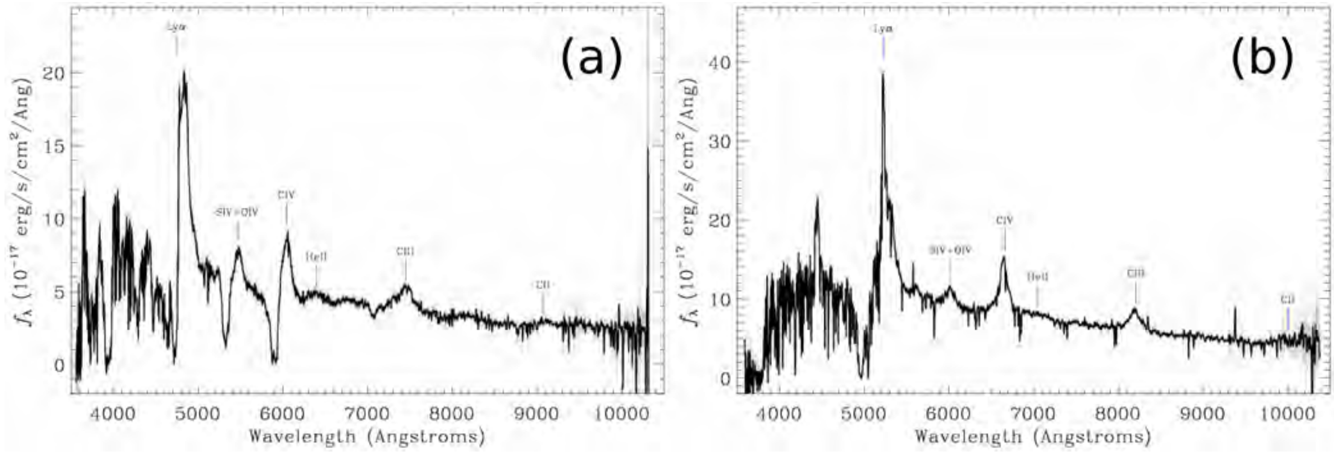


Fig. 3: Examples of quasar spectra with BAL (a) and DLA (b).

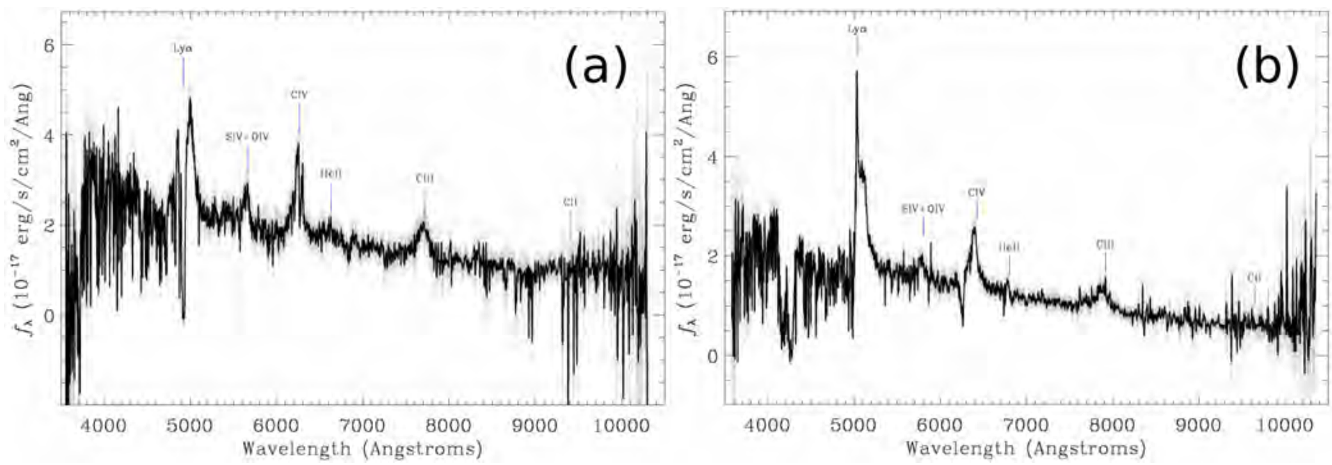


Fig. 4: Quasar spectra with absorption in Ly$\alpha$ (a) and Ly$\beta$ (b) respectively.

Table 1: Example of a part of the online table. The columns contain information on the object name, plate number ("plate"), the modified Julian date of the observation night ("MJD"), number of the fiber ("fiber"), right ascension ("RA") and declination ("DEC") of the object, the redshift ("$z$") and the value of the redshift uncertainty ("$\sigma_z$"). The last column gives the name of the sample to which the object belongs (the main or the composite).

| name | plate | mjd | fiber | RA | DEC | z | $\sigma_z$ | sample |
|---|---|---|---|---|---|---|---|---|
| 134102.26+264952.5 | 6002 | 56104 | 631 | 205.259 | 26.8313 | 2.80547 | 0.00044 | main |
| 094804.16+595050.5 | 5719 | 56014 | 436 | 147.017 | 59.8474 | 2.61632 | 0.00020 | main |
| 153209.53+130029.3 | 4891 | 55736 | 676 | 233.04 | 13.0081 | 2.46111 | 0.00034 | main |
| 075625.70+221613.8 | 4471 | 55617 | 428 | 119.107 | 22.2705 | 2.44233 | 0.00013 | main |
| 093336.98-011849.2 | 3767 | 55214 | 694 | 143.404 | -1.31369 | 3.69452 | 0.00045 | main |
| 171455.63+283520.8 | 5000 | 55715 | 16 | 258.732 | 28.5891 | 2.16432 | 0.00110 | composite |
| 171753.56+305031.3 | 4998 | 55722 | 283 | 259.473 | 30.8421 | 2.69634 | 0.00044 | composite |
| 122355.11+021518.0 | 4752 | 55653 | 534 | 185.98 | 2.25502 | 2.19234 | 0.00038 | composite |
| 160714.29+152518.3 | 4072 | 55362 | 364 | 241.81 | 15.4218 | 2.32026 | 0.00072 | composite |
| 152151.73+275914.0 | 3851 | 55302 | 858 | 230.466 | 27.9872 | 2.36072 | 0.00019 | composite |

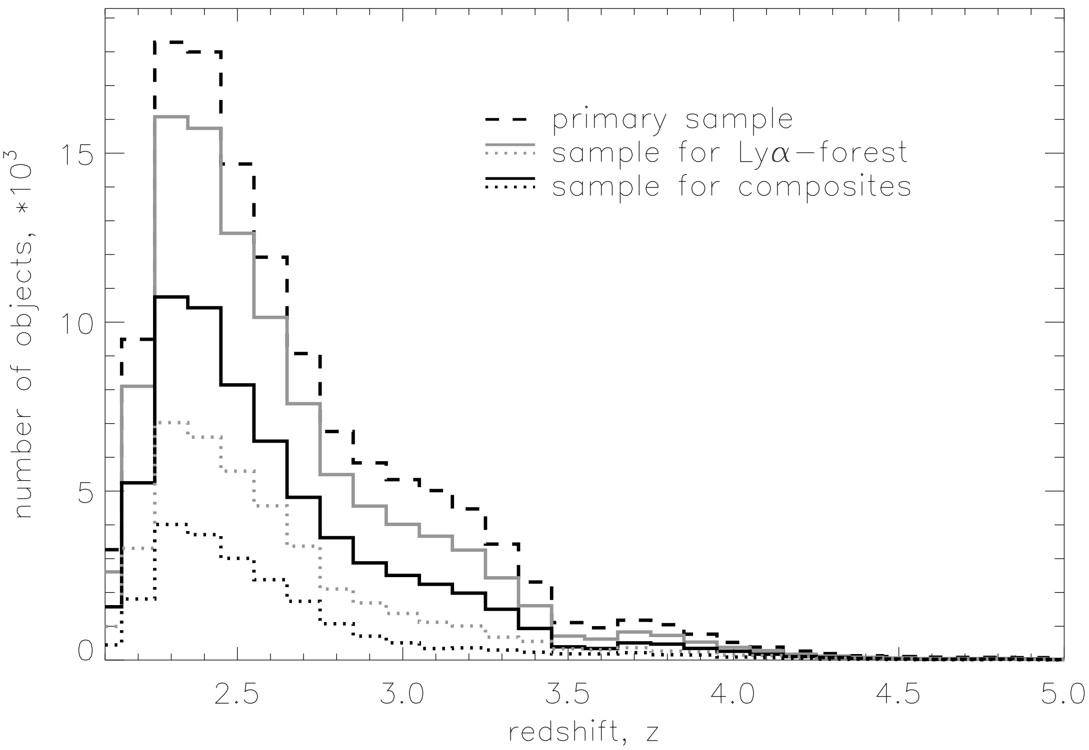


Fig. 5: The redshift distribution of quasars from the preliminary sample (dashed black line), the sample for Lya-forest studies (grey solid line), and the sample for composite compilation (black solid line). Corresponding histograms for Lya-forest and composite subsamples with RMS of the normalization constant A less than 10% are shown by grey and black dotted lines, respectively.