Effect of BALs in redshift fitting with Redrock

Luz Ángela García

Broad absorption lines (BALs) are saturated column density features in the spectra of quasars. BALs are associated with QSOs, but the astrophysical scenario that originates them is still not entirely clear. They are mostly associated to strong ion metal lines in the quasar spectra, in particular, CIV, SiIV, and MgII.
The present document has been built using London mocks in version 5.0.0, comparing Ly–α only and Ly–α + BALs spectra -DESI footprint, and at 40% downsampling- in the realizations quick-0.0 and quick-2.6, respectively. The data spans in a redshift range of 1.5 to ~ 4.0. The tests shown here include about 180000 spectra with 12% of probability to have a simulated BAL feature.

Figure 1 shows the distribution in redshift for the mock spectra: Ly–α only (in blue) and Ly–α + BALs spectra (in red). The redshift used to populate the distribution is fitted by Redrock, not z_{th}. The histograms show that, effectively, more than a tenth of the total QSO simulated spectra contains BAL features and most of the data peaks at the low-end of the sample. The number of systems decreases to the half at redshift ~ 2.5 and is almost null at high redshift. This trend is also seen in the blue histogram, which is not surprising because the redshift-distribution of spectra with BALs mimics the one with Ly–α only. Nonetheless, the decline in the number of systems is smoother in the spectra with BAL features. That is not the case in the blue histogram, where there is a rapid decline in the number of systems when moving to high redshift.

The first (and only test by now) that can be done to evaluate the goodness of the fitting by Redrock is based on the velocity dispersion dv:

$$dv = c * \frac{z_{\mathrm{th}} - z_{\mathrm{rr}}}{1 + z_{\mathrm{th}}} \tag{1}$$

with c, the speed of light, z_{th} the true redshift of the quasar and z_{rr} the redshift fitting produced by Redrock.

Figure 2 shows that adding BALs in the spectra increases the redshift estimate uncertainty, reflected directly in a large velocity dispersion, compared with the case of Ly-α only. The velocity dispersion is presented for the three cases that Redrock reported in the fitting of the spectra: QSO, galaxies and, stars in black dots, red squares, and yellow stars, respectively. In the left panel, the x-axis corresponds to the true redshift z_{th} and on the right panel, velocity dispersion is plotted vs. the estimated redshift by Redrock z_{rr}.
In both panels, there are not stars- and galaxy-type spectra present. This is due to a very large velocity dispersion in the case of spectra classified as these objects, which is above 6000 km/s. Therefore, stars and galaxies misclassified by Redrock have high uncertainty in the redshift-fitting. This result can be indicating that the introduction of BAL features in the spectra generates uncertainty in the redshift estimate, hence, causes a bad identification on the object associated with the spectrum.

Besides the implementation of a flag in case the object predicted by Redrock is not quasar (in such case, z_{th} is significantly different from z_{zz}), we also use the z_{warn} tag in the Redrock output file to distinguish between:

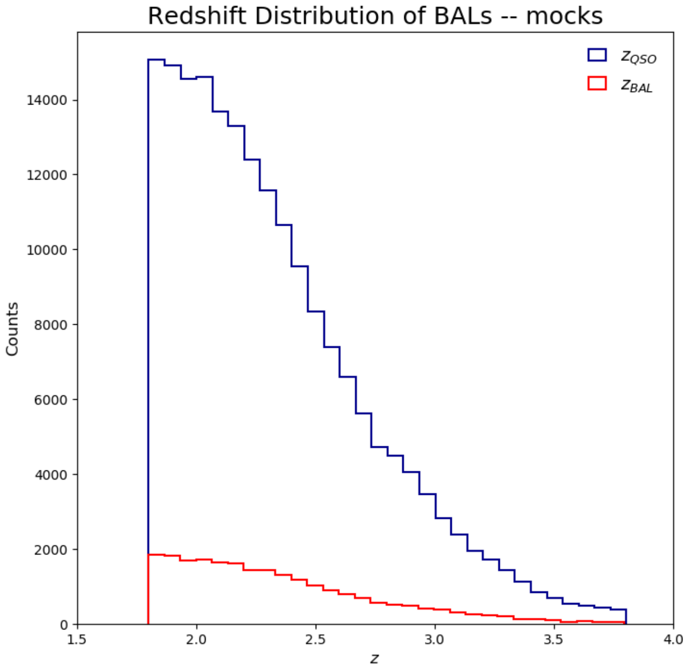

Figure 1: Distribution of redshift in Ly–α only and Ly–α + BALs spectra. The redshifts considered are predicted by Redrock (extracted from zbestrr-16-X.fits). The blue histogram represents all mocks of the QSO spectra, whereas the red histogram shows the redshift-distribution of mocks with BALs only.

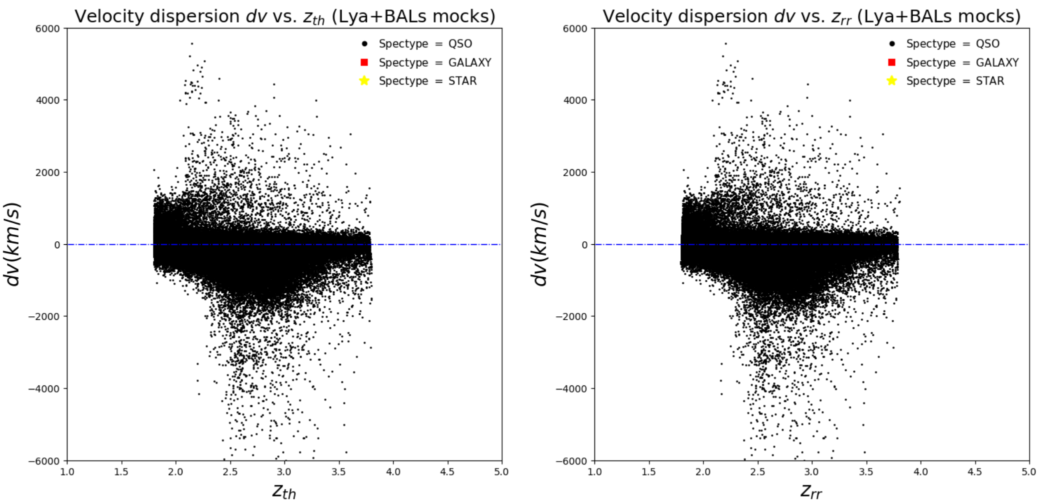

Figure 2: Velocity dispersion vs. z_{th} (from truth file) on the left, and, z_{rr} (derived from zbestrr-16-X file) on the right, respectively. Different spectra-type are predicted by Redrock: QSO (black dots), galaxies (red squares) and stars (yellow stars), but the dispersion with non-QSO objects is that large, that they do not appear in the panel.

- *good fit*: difference in redshift below a threshold (compared with the true redshift) and $z_{\text{warn}} = 0$ (**good**),
- *failed fit*: difference in redshift above a given threshold (compared with the true redshift) and $z_{\text{warn}} = 0$ (catastrophic failures; **fail**),
- *missed opportunities*: difference in redshift below a threshold (compared with the true redshift) and $z_{\text{warn}} \neq 0$ (**miss**),

- *lost:* difference in redshift above a threshold (compared with the true redshift) and $z_{\mathrm{warn}} \neq 0$ (**lost**).

The tolerance (or threshold) accepted is compared with the expression $\frac{|z_{\mathrm{th}}-z_{\mathrm{rr}}|}{z_{\mathrm{th}}}$. Its numerical value in this test was set to 0.05, but it can be changed according to the tuning in the data.

The former classifying categories can be visualize and distinguish directly with the velocity dispersion for Lyman-α + BAL features. Figure 3 shows the different classes describe above for comparing the goodness of the fitting:

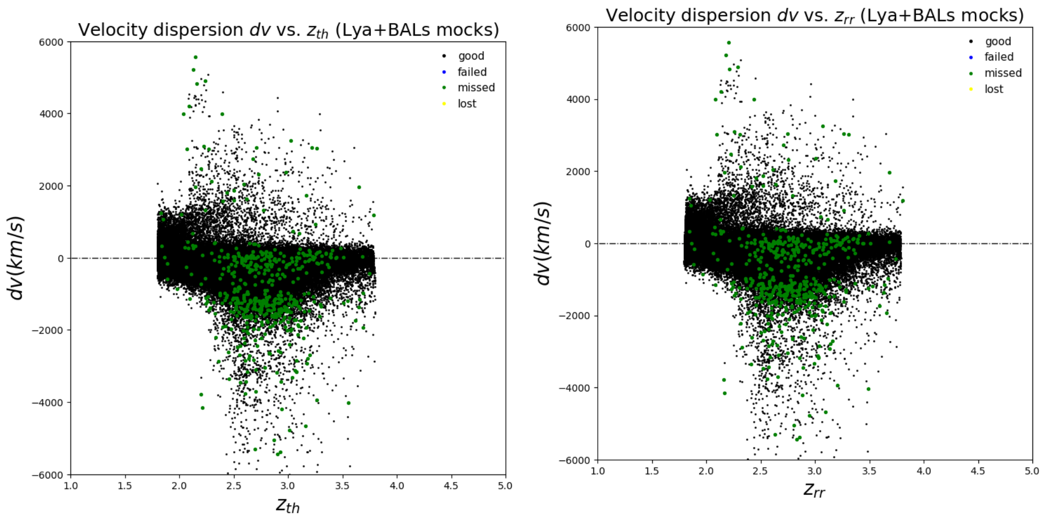

Figure 3: Velocity dispersion vs. z_{truth} (from truth file) on the left, and, z_{rr} (derived from zbestrr-16-X file) on the right, respectively. The cases considered in both plots compare the goodness of the fitting with Redrock: if the redshift is poorly estimated by the fitter, regardless of the value of z_{warn}, the cases do not appear in the boxes. Such cases are the catastrophic failures and lost chances, both with a large dispersion velocity.

Figure 3 shows the large velocity dispersion of catastrophic failures and lost chances. Neither of these cases appears in the panels, due to the poor estimate of the redshift by Redrock. Alternatively, good fits are the vast majority of the cases, which is quite promising for the analysis, and, missed opportunities are distributed in all the range displayed in velocity dispersion. It would be worth looking into these missed opportunities, to understand why their z_{warn} flag is activated.

Furthermore, confusion matrices are computed to study the goodness of Redrock fitting. By definition, all the spectra generated is associated with a QSO object (true spectra in the y-axis). Nonetheless, part of the spectra is fitted as a different object, due to features, like BALs, that add noise to the continuum and degrade the redshift estimate of the spectra. In the confusion matrix I (Figure 4), the x-axis shows the two outcomes produce by Redrock: a QSO or a non-QSO spectrum. The different objects that are associated with the fitted spectra are shown in Figure 5. Again, we are always generating quasar spectra with the routine QUICKQUASAR, but Redrock identifies three possible objects: QSOs, galaxies, and stars.

Taking into account that both, galaxies and stars are non-QSOs cases, there is a direct correspondence among the data in both confusion matrices. However, the second confusion matrix reveals that Redrock confuses a large number of spectra as being produced by galaxies, but rarely with stars, which spectra are quite distinctive from quasar spectra. This information can be also found in Table 1. Nonetheless, the table shows the percentage, not the number of occurrences.

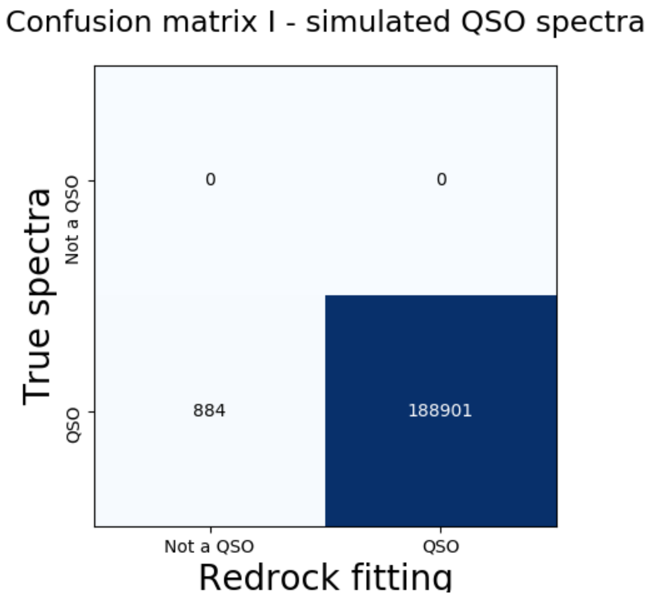

Figure 4: Confusion matrix of the fitting with Redrock. The flag used is boolean: the predicted spectra correspond to a QSO or not, regardless of the type of the spectra-object.

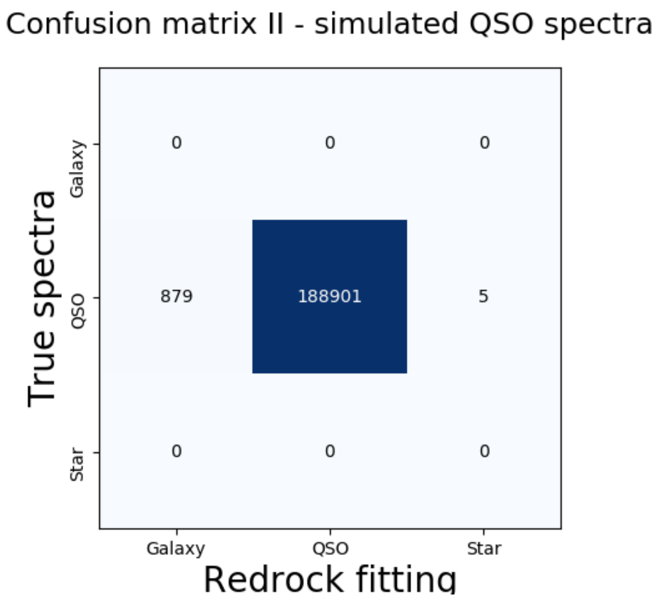

Figure 5: Confusion matrix for objects associated to the fitted spectra by Redrock: QSO-, galaxy and, star-type of spectra. The true spectra generated correspond to QSOs, but there are different spectra outputs provided by Redrock, according to the redshift predicted by the fitter.

Complementary to the confusion matrices above, Table 1 displays the percentage of predicted objects associated with the spectra delivered by Redrock.

Spectra associated with a	Percentage (%)
QSO	99.53
Galaxy	0.4632
Star	0.0026

Table 1: Normalized distribution of the objects fitted by Redrock when BALs are included in the QSO spectra. The first object in the classification is a QSO (first row), but the second and third rows are non-QSO objects: galaxies and stars. No other type of spectra were found in the Redrock fitting.

The performance of Redrock is quite significant: in less than one percent of the cases, Redrock confuses a quasar spectra with a template associated with a different object. As shown in Table 1, the spectra are confused with galactic templates in 0.46% of the cases. Lastly, hardly ever Redrock fits a spectrum as being produced by a star, with 0.0026% of the occurrences.

On the other hand, Table 2 shows the different categories established before to quantify the goodness of Redrock fitting. Good fits and missed opportunities have a redshift estimate with low uncertainty, but differ in their value of the flag z_{warn}. Instead, catastrophic failures (failed fits) and lost chances have a large difference velocity dispersion. As predicted with Figure 3, more than 99% of the cases are good fits, and then, the distribution splits into catastrophic failures, missed opportunities and lost chances.

Fit	Percentage (%)
Good	99.12
Failed	0.4916
Missed	0.2697
Lost	0.1159

Table 2: Normalized distribution of cases considered to analyse the goodness of the fitting with Redrock when BALs are included in the QSO spectra.

Finally, the number of counts in each of the cases is plotted as a function of redshift in Figure 6. The good fits are not shown since more than 99% of the fits are good estimates, and they wash-out the distribution of the less representative cases.

There is a clear bimodal distribution in redshift for catastrophic failures (histogram in magenta, Figure 6). This effect has to be revised in low and high redshift quasars, $z < 2$ and $z > 2$, respectively, because it can be an indication of displacement of the Lyman-α emission in the spectrograph with evolving redshift.
On the other hand, missed opportunities (good redshift fitting but $z_{\mathrm{warn}} \neq 0$) are centred at ~ 2.7 and the distribution spans in a redshift range of 2.0–3.5. Finally, lost chances (wrong redshift estimate and $z_{\mathrm{warn}} \neq 0$) mostly located at low redshift ($z < 2$), with a few occurrences at high redshift.

In summary, BALs in QSO spectra are a source of contamination for the redshift fitting process. These features reduce the effectiveness of Redrock in about 1% when the sample considered is significantly large ($\sim$ 180000 quasar spectra are used for this test). The expectation is that the percentage reported will reduce with a larger sample (which will be the case with DESI data).

Final remarks:

- Comparing the velocity dispersion between Ly-α only and Ly-α + BALs mock spectra, there is a larger dispersion in the latter case. This is due to the difference in the redshift

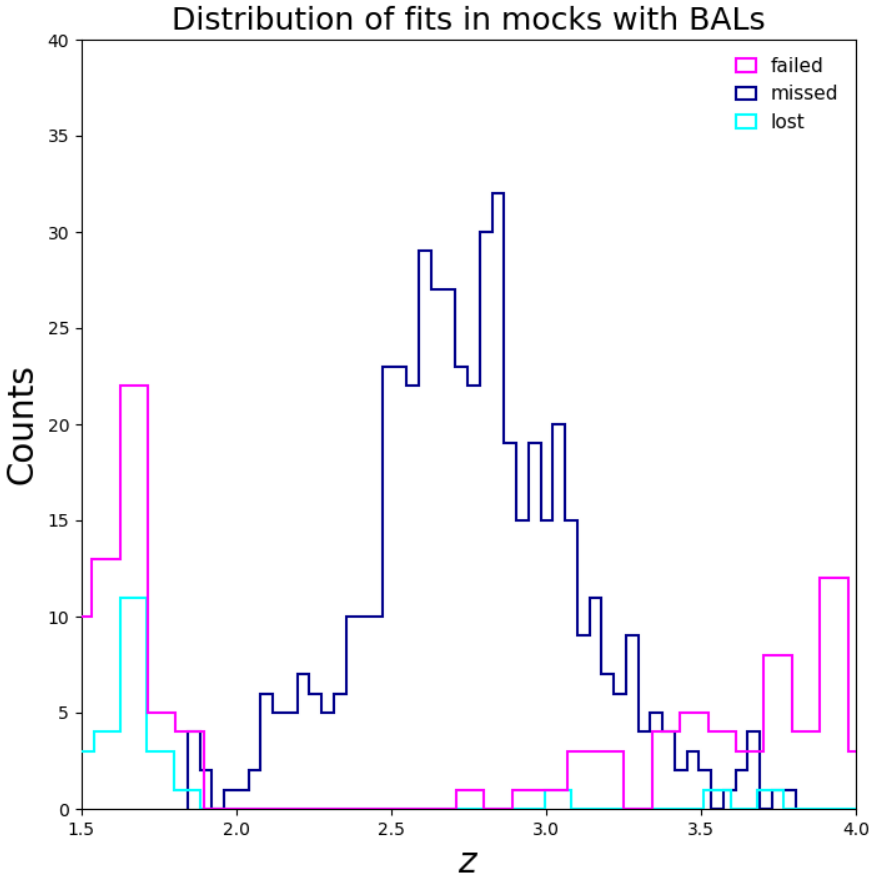

Figure 6: Distribution in redshift of catastrophic failures (failed; in magenta), missed opportunities (missed; in dark blue) and lost chances (lost; in cyan).

predicted by Redrock. As discussed before, the introduction of BAL features on the spectra implies a wider difference between the real redshift z_{th} and z_{rr}.

- Broad absorption lines introduced in the spectra have an impact in the redshift fitting pipeline. This is reflected in a misclassification in the type of object associated with a particular spectrum. Although the mocks are produced to simulate quasar spectra only, Redrock classifies some spectra as a star- or galaxy-type (see Figures). This is due to a poor estimate of the redshift for these objects.

- Misidentification of the redshift could be an issue if BALs have a low Balnicity index (BI). BI is measured with the strength of the CIV absorption (and other metal lines, for instance, SiIV and MgII). However, a low BAL index could generate confusion in the classifier because the CIV absorption line could be an intrinsic or intervening line, and not be related to the BAL feature.

- An additional test could be done once the Balnicity index (BI) is computed for each spectrum: the velocity dispersion and the fraction of catastrophic failures can be plotted as a function of BI, and, as a function of the SNR. The former dependency gives direct information on the BALs that have been fitted with Redrock, whereas the evolution with the SNR provides an estimate of the number of fits required to improve the spectrum of each quasar.