A scientific guide to Gaia Astrophysical Parameters

Day 4

organized by Coordination Unit 8

QSOC

Ludovic Delchambre

University of Liège, Belgium

Outline

- Overview
- Method
  - Quasar templates
  - Algorithm
- Performance and results
- Use of QSOC results
- Future plans

Overview

Objective

The quasar classifier (QSOC) module is designed to determine the redshift, z, of the sources that are classified as quasars by the Discrete Source Classifier (DSC) module.

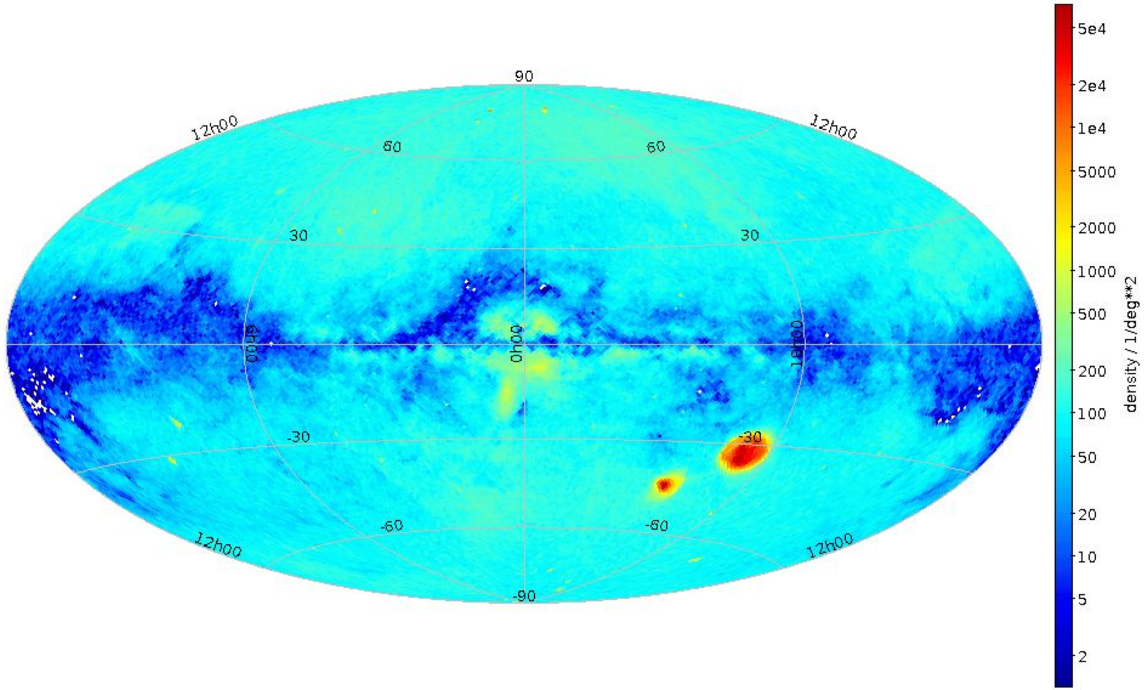
QSOC is **not** a classification module but instead processes all sources with classprob_dsc_combmod_quasar $\geq$ 0.01.

Overview

QSOC is complete but not very pure.

Reason: We can draw a pure sample from a complete sample, but not the opposite.

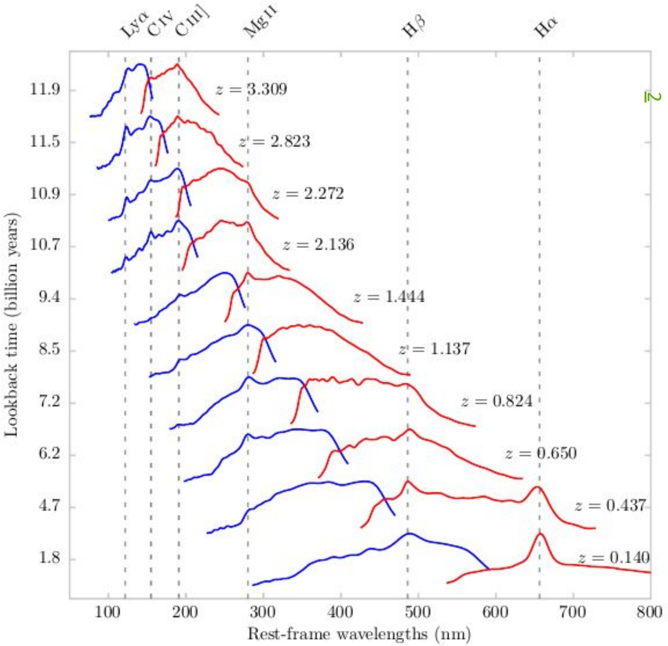
Purity can be achieved through classlabel_dsc_joint = ’quasar’ or recipes given in [Gaia Collaboration, Bailer-Jones et al. (2022)](#).



Method

QSOC is based on a chi-square approach that compares the observed BP/RP spectra sampled by SMSgen to quasar rest-frame templates in order to infer their redshift.

The predicted redshifts are in the range 0.0826 < z < 6.12295



https://www.cosmos.esa.int/web/gaia/iow_20201222

Method – Quasar templates

1. Extrapolate 297 264 quasars from the twelfth release of the Sloan Digital Sky Survey Quasar catalogue of Pâris et al. (2017) using the weighted PCA method of Delchambre (2015).

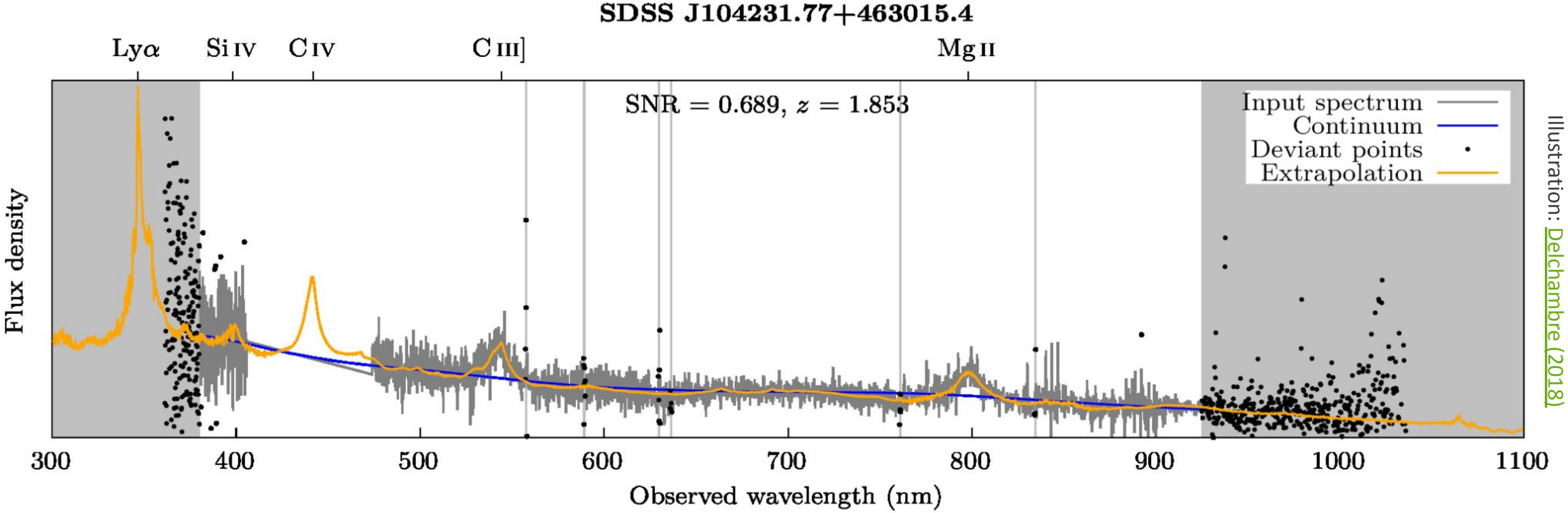


Illustration: Delchambre (2018)

Method – Quasar templates

2. Convert the extrapolated spectra into BP/RP spectra through the use of the BP/RP spectrum simulator MIOG (Montegriffo et al., 2022, but see also the GaiaXPy tool).

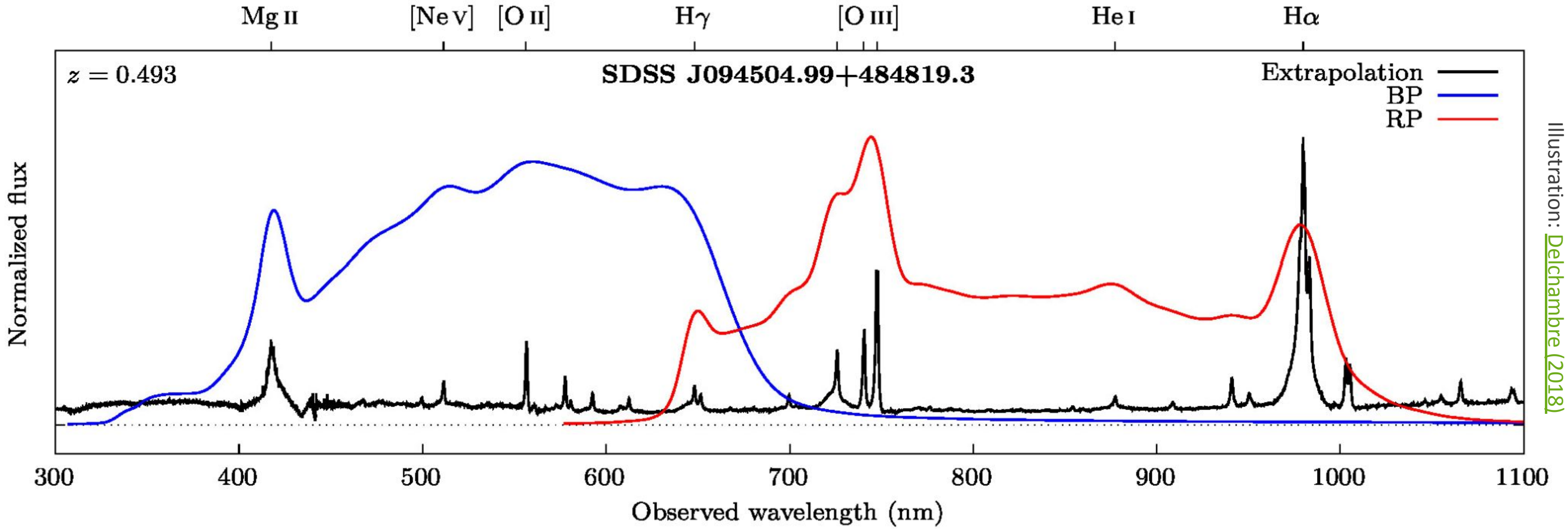


Illustration: Delchambre (2018)

Method – Quasar templates

3. Divide each of the BP/RP spectrum by the spectrum that would be obtained from a uniform SED: removal of the bell-shape + fix the different integration width of each pixel.

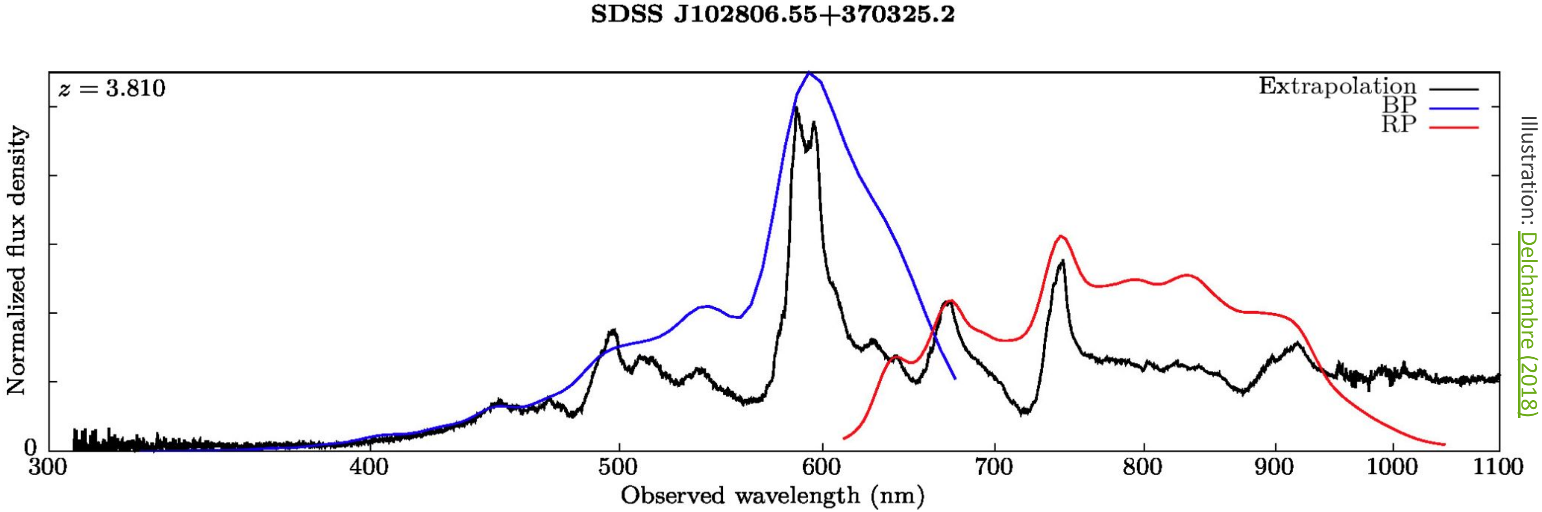


Illustration: Delchambre (2018)

Method – Quasar templates

3. Divide each of the BP/RP spectrum by the spectrum that would be obtained from a uniform SED: removal of the bell-shape + fix the different integration width of each pixel.

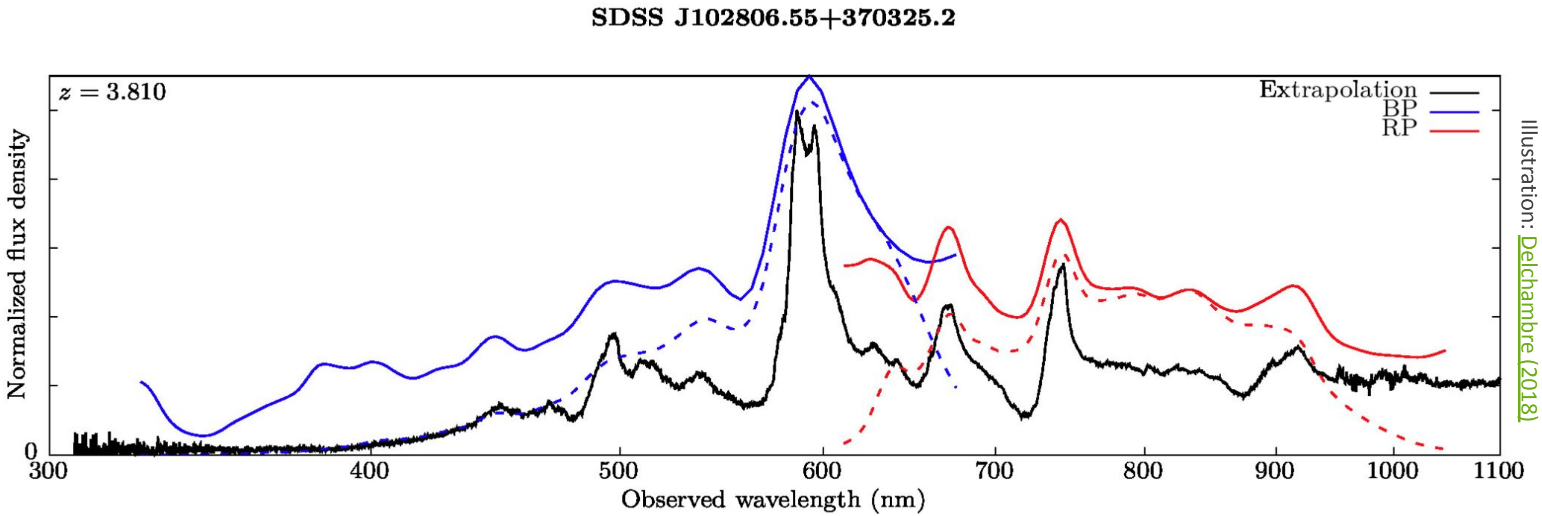


Illustration: Delchambre (2018)

Method – Quasar templates

4. Subtract continua then retrieve the dominant BP/RP components using the weighted PCA method of Delchambre (2015).

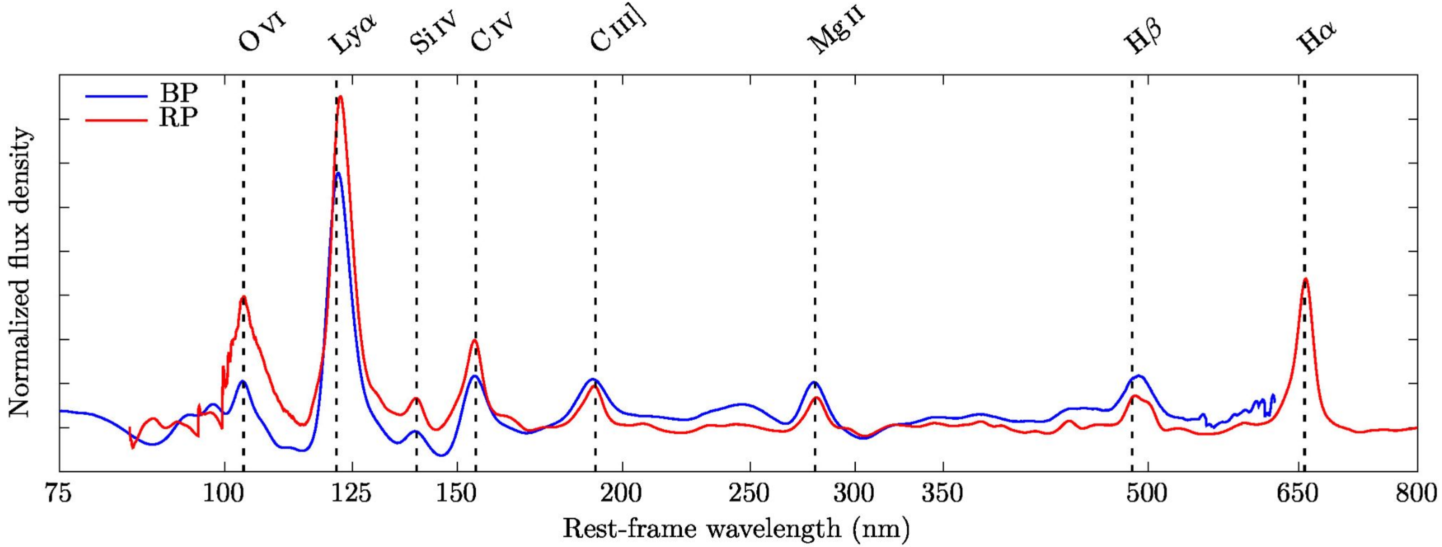


Illustration: Delchambre et al. (2022)

Method – Algorithm

The redshift

$$z = \frac{\lambda_{\mathrm{obs}} - \lambda_{\mathrm{rest}}}{\lambda_{\mathrm{rest}}} = \frac{\lambda_{\mathrm{obs}}}{\lambda_{\mathrm{rest}}} - 1$$

turns into a simple offset once considered on a logarithmic wavelength scale

$$Z = \log(z+1) = \log \lambda_{\mathrm{obs}} - \log \lambda_{\mathrm{rest}}$$

Method – Algorithm

Given a set of rest-frame templates, ***T***, and an observation vector, ***s,*** sampled on the same logarithmic wavelength scale, $\lambda_i = \lambda_0 L^i$ (here log L = 0.001), we aim to find the shift, k, that minimizes

$$\chi^2(k) = \sum_i \frac{1}{\sigma_i^2} \left(s_i - \sum_j a_{j,k} T_{i+k,j} \right)^2$$

$$z = L^k - 1$$

which can be computed in O(N log N) flops (Delchambre, 2016) through

$$\mathrm{ccf}(k) = \left(\sum_i \frac{s_i^2}{\sigma_i^2} \right) - \chi^2(k) = C - \chi^2(k)$$

In QSOC:

$$\mathrm{ccf}(k) = \mathrm{ccf}_{\mathrm{bp}}(k) + \mathrm{ccf}_{\mathrm{rp}}(k)$$

Method – Algorithm

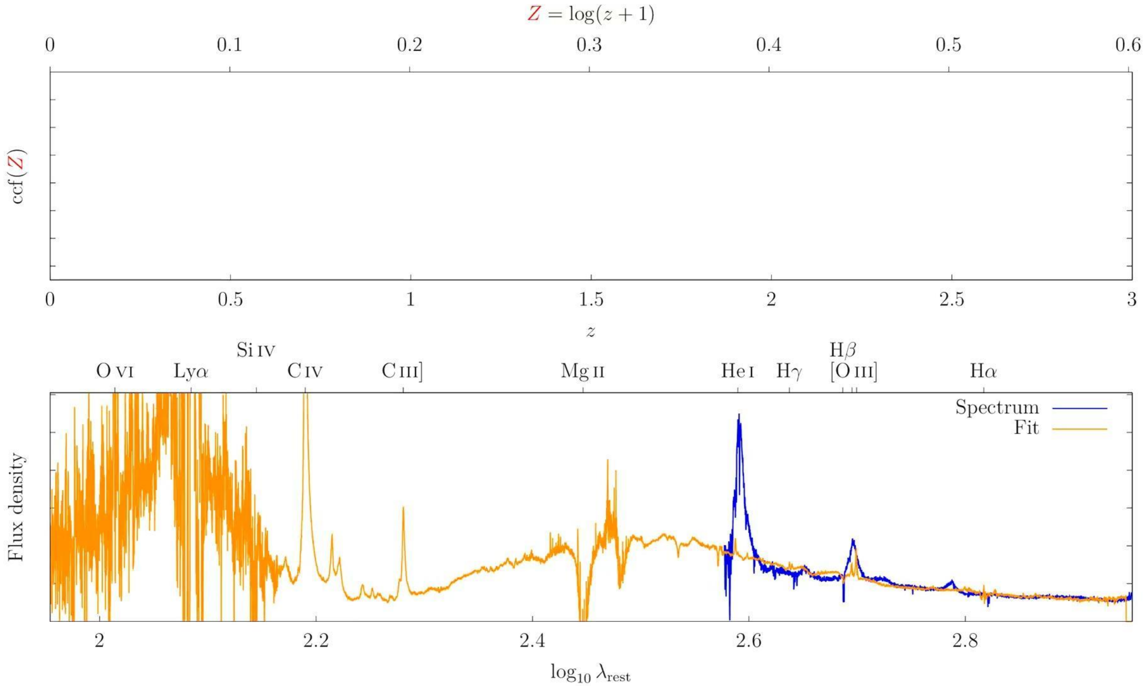
Illustration with SDSS J000313.08+274044.9, a z=2.2329 quasar.

<u>Illustration</u>

- 15 emission line templates
- 5 continuum templates

<u>Gaia DR3</u>

- 1 emission line templates
- 3 continuum templates



Source: [Delchambre (2018)]

Method – Algorithm

The maximal value of the CCF usually corresponds to the most probable redshift, but it is not always the case.

The redshift selected by QSOC is the one that maximizes

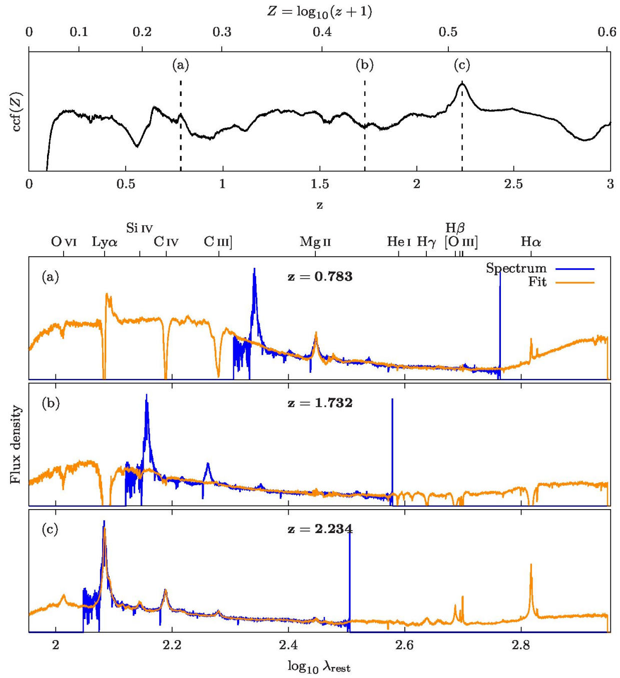
$$S(k) = w_0 \times \left[\,\chi_r^2(k)\,\right]^p + w_1 \times \left[\,Z_{\text{score}}(k)\,\right]^p$$

where $w_0 = 0.71413, w_1 = 0.28587, p = 0.24365,$

$$\chi_r^2(k) = \frac{\text{ccf}(k)}{\max_k(\text{ccf})}$$

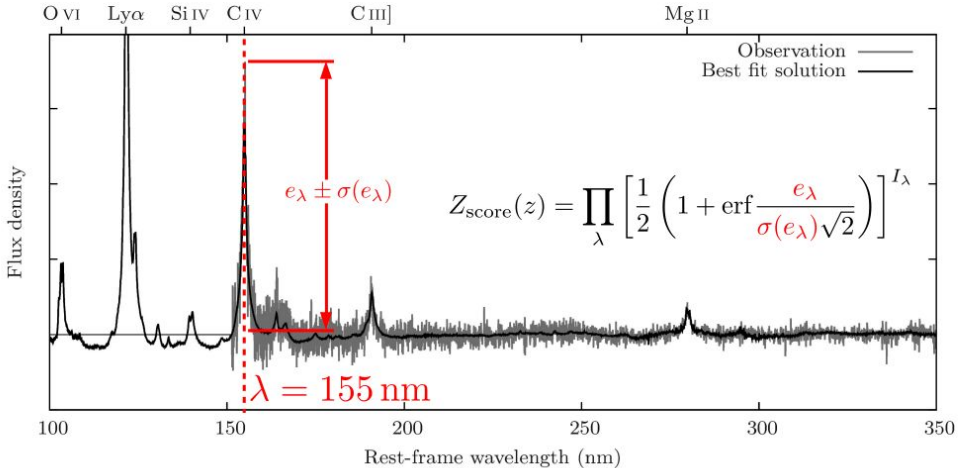
and

$$Z_{\text{score}}(k) = \prod_\lambda \left[\frac{1}{2}\left(1 + \text{erf}\,\frac{e_\lambda}{\sigma(e_\lambda)\sqrt{2}}\right)\right]^{I_\lambda},$$



Source: Delchambre (2018)

Method – Algorithm



Parameters of the Z_{score} for BP spectra

	O IV	Lyα	Si IV	C IV	C III]	Mg II	Hγ	Hβ
λ [nm]	103.202	121.896	139.349	154.658	189.957	279.259	437.904	491.899
I_λ	0.017	1.0039	0.01	0.13202	0.31359	0.94396	0.23848	0.93124

Parameters of the Z_{score} for RP spectra

	O IV	Lyα	Si IV	C IV	C III]	Mg II	Hγ	Hβ	Hα
λ [nm]	103.353	122.388	139.563	154.588	190.398	280.470	435.600	488.952	657.736
I_λ	0.062484	0.10984	0.18982	0.07023	0.1409	0.22011	0.4101	0.25137	0.59948

The zscore is a weighted geometric mean of a set of normal CDF of mean zero and standard deviations $\sigma(e_\lambda)$ evaluated at e_λ.

If zscore ≈ 1, then all emission lines we expect at redshift z are present.

If we are missing a single emission line, then zscore quickly drops to zero.

Method – Algorithm

The fit of a quadratic polynomial to the selected peak provides

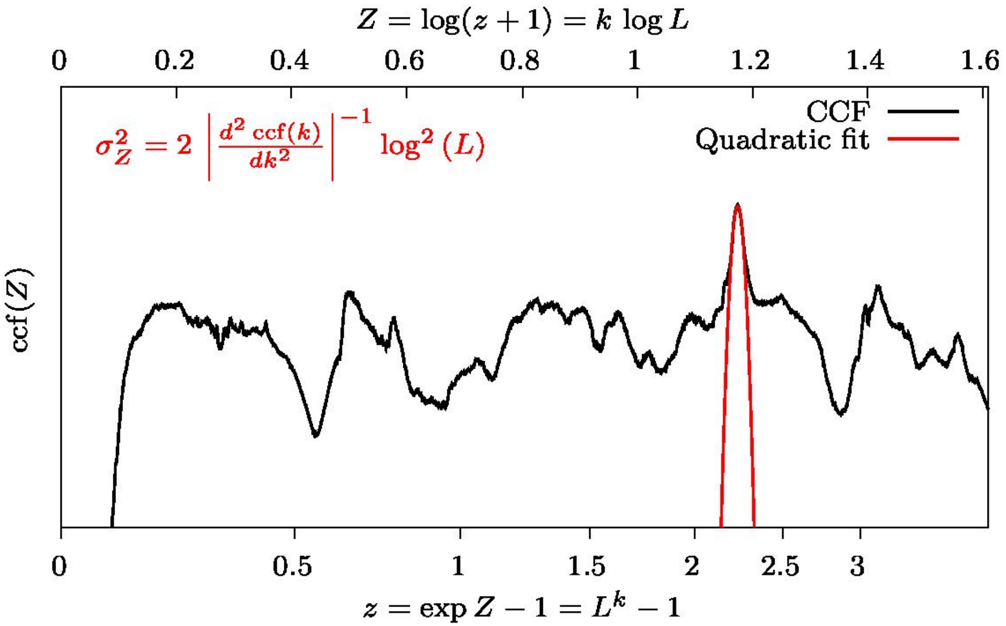
- A sub-sampling precision on the redshift
- Normal uncertainties on Z

The error on z is distributed as a log-normal distribution whose 0.16 and 0.84 quantiles are respectively given by

$$z_{\text{low}} = \exp(Z - \sigma_Z) - 1$$

and

$$z_{\text{up}} = \exp(Z + \sigma_Z) - 1$$



Method – Algorithm

Warning flag	Bit	Value	Condition(s) for rising
Z_AMBIGUOUS	1	1	The CCF has more than one maximum with $\chi_r^2(k) > 0.85$, meaning that at least two redshifts lead to a similar χ^2 and the solution is ambiguous.
Z_LOWCHI2R	2	2	$\chi_r^2(k) < 0.9$
Z_LOWZSCORE	3	4	$Z_{\text{score}}(k) < 0.9$
Z_NOTOPTIMAL	4	8	The selected solution did not correspond to the global maximum (i.e. $\chi_r^2(k) < 1$)
Z_BADSPEC	5	16	The BP/RP spectra upon which this prediction is based are considered as unreliable. An unreliable spectrum has a number of spectral transits in BP, N_{bp} or RP, N_{rp} that is lower than or equal to ten transits or $G \geq 20.5$ mag or $G \geq 19 + 0.03 \times (N_{\text{bp}} - 10)$ mag or $G \geq 19 + 0.03 \times (N_{\text{rp}} - 10)$ mag (see the online documentation for more information on the derivation of these limits).

Performance and results

QSOC results are available in the Gaia DR3 qso_candidates table

- redshift_qsoc: The quasar redshift, z
- redshift_qsoc_lower and redshift_qsoc_upper: The lower and upper confidence intervals corresponding to the 16% and 84% quantiles of z
- ccfratio_qsoc: The chi-square ratio, χ_r^2
- zscore_qsoc: The zscore
- flags_qsoc: The QSOC processing flags

Performance and results

Comparison with literature values

A 1” cross-match against the Milliquas 7.2 quasar catalogue of Flesch (2021) (type=’Q’) yielded 439 127 common sources.

- 279 850 / 439 127 = 63.73% have |Δz| < 0.1
- 89 107 / 91 320 = 97.58% have |Δz| < 0.1 if flags_qsoc = 0

The logarithmic redshift error

$$\Delta Z = \log(z+1) - \log(z_{\mathrm{true}}+1) = \log \lambda_{\mathrm{true}} - \log \lambda_{\mathrm{false}}$$

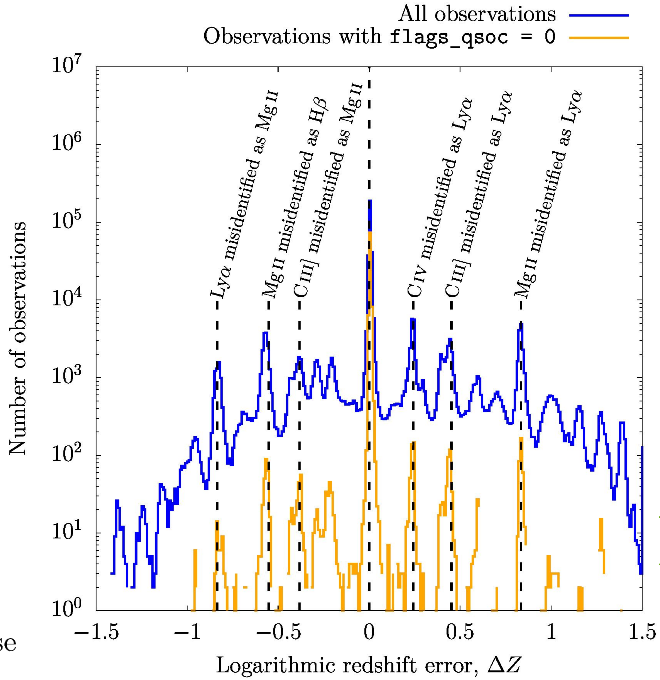


Illustration: Delchambre et al. (2022)

Performance and results

Considering sources with |Δz| < 0.1,

$$\Delta Z/\sigma_Z$$

where

$$\sigma_Z = [\log(z_{\mathrm{up}} + 1) - \log(z_{\mathrm{low}} + 1)]/2$$

approximately follows a normal distribution with

μ=0.007 and σ = 1.053

and

μ=0.002 and σ = 1.14 if flags_qsoc=0

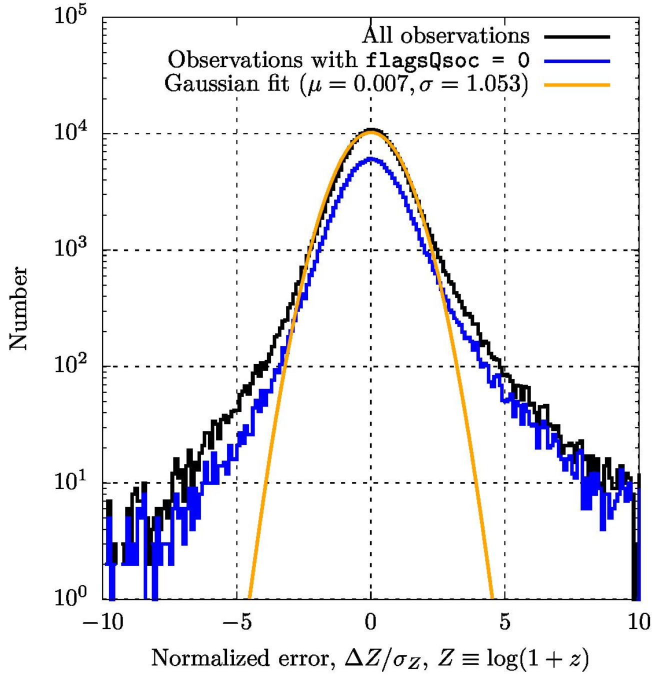


Illustration from the online documentation

Use of QSOC results

Redshift dependence on G magnitude

Around 90% of the sources with G ≤ 19 mag have |Δz| < 0.1.

The same fraction is obtained for 19.9 < G < 20 mag sources only if flags_qsoc= 0.

These correspond to a very small fraction (5.5%) of the sources in this magnitude range.

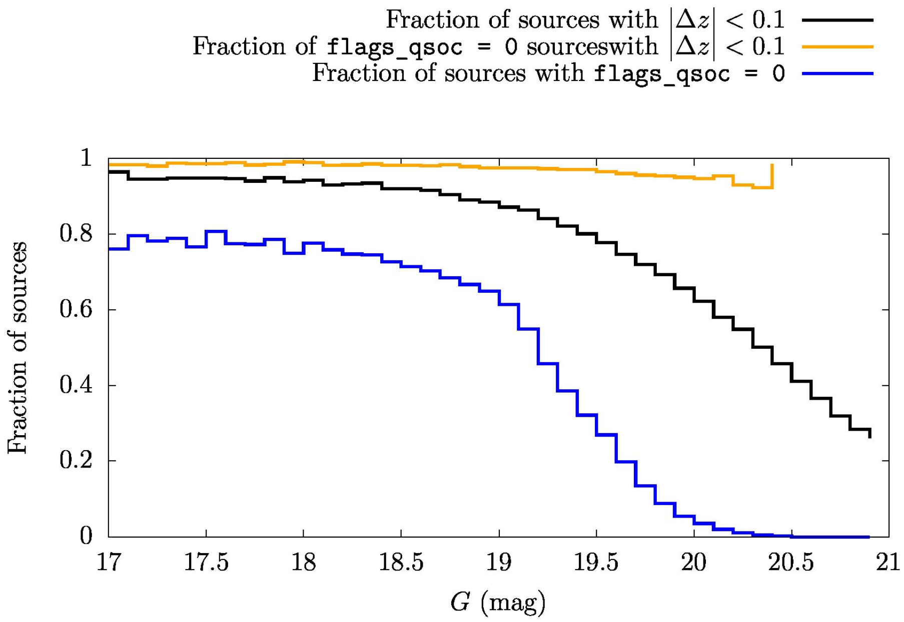


Illustration: Delchambre et al. (2022)

Use of QSOC results

Additionally consider flags_qsoc=16 sources.

Z_BADSPEC (bit 5, value 16)
Raised if one of the following conditions is met:

1. The number of BP or RP spectral transits (N_{BP} and N_{RP}), is lower than 10 transits
2. G > 20.5 mag
3. G > 19 + 0.03 × (N_{BP}−10) mag
4. G > 19 + 0.03 × (N_{RP}−10) mag

At 19.9 < G < 20 mag
92% of the sources have |Δz| < 0.1 while 36.5% of the sources are retained

At 20.4 < G < 20.5 mag
81.5% of the sources have |Δz| < 0.1 while 22% of the sources are retained

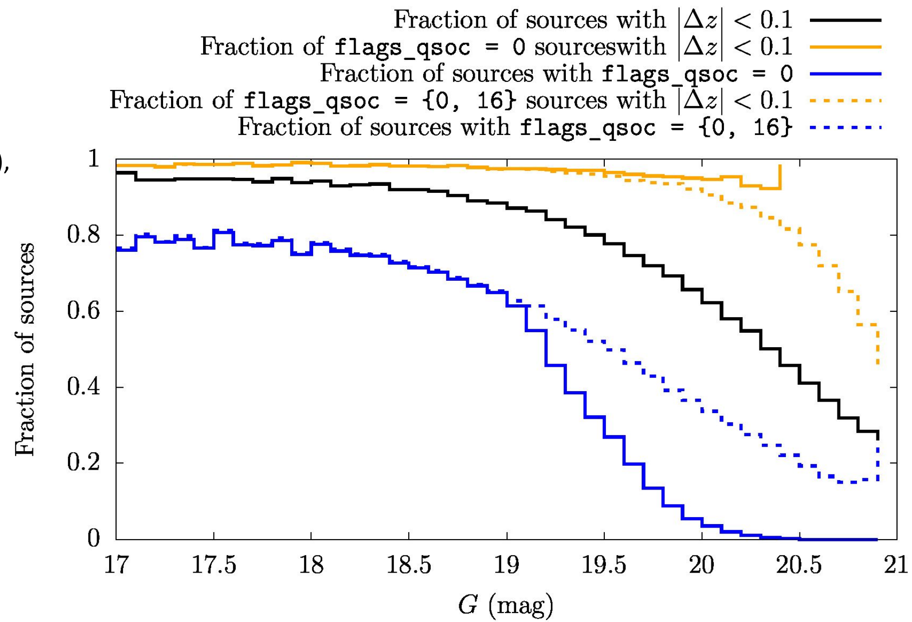


Illustration: Delchambre et al. (2022)

Use of QSOC results

Completeness

Fraction of sources that have $|\Delta z| < 0.1$ wrt. literature redshift.

Lower completeness at

- $z \approx 2$, due to $C_{IV} \leftrightarrow$ Lyα mismatches.
- $0.9 < z < 1.3$, due to the sole presence of the Mg_{II} emission line in this range

Appropriate cuts on flags_qsoc allow these shortcomings to be alleviated.

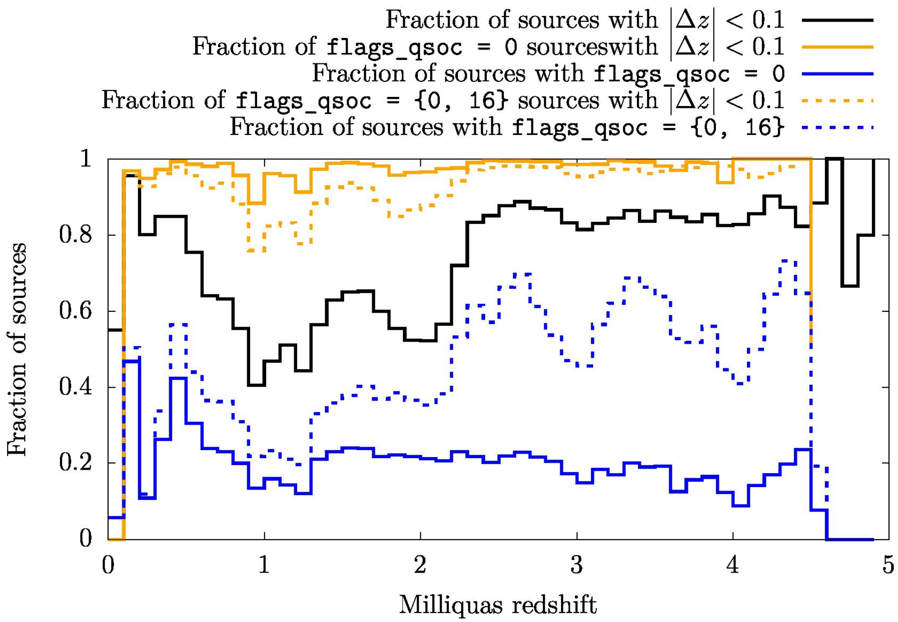


Illustration: Delchambre et al. (2022)

Use of QSOC results

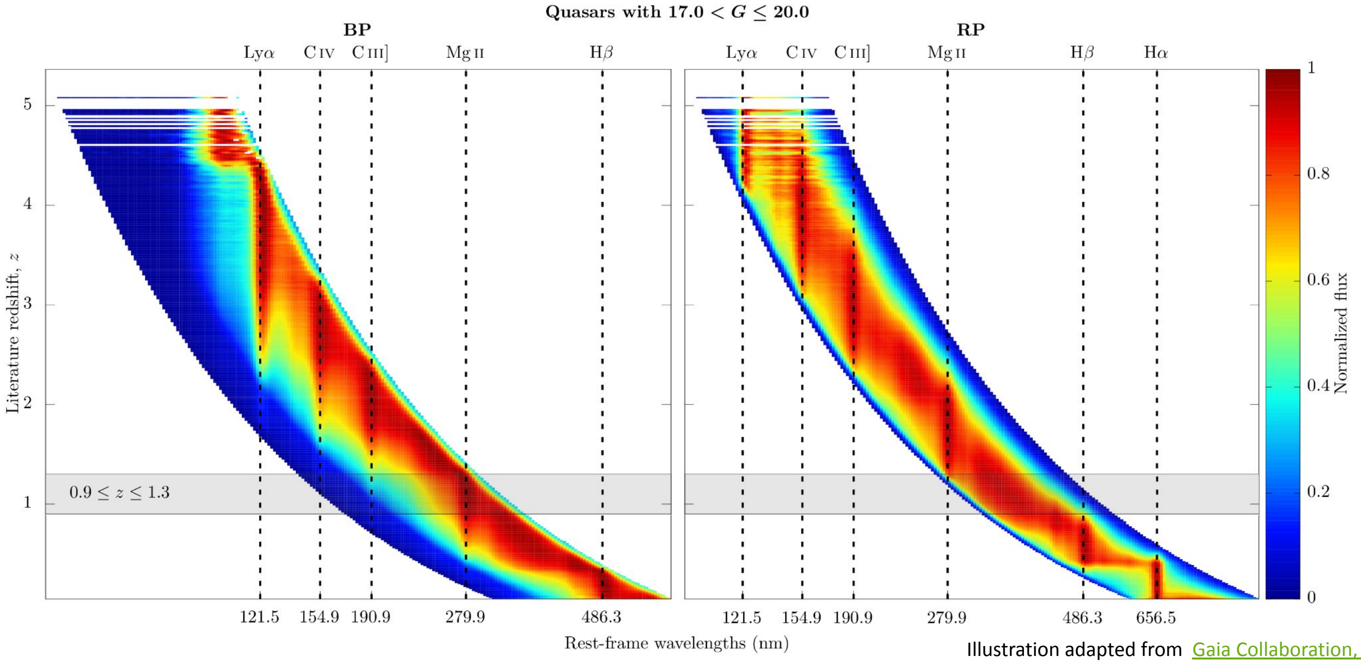
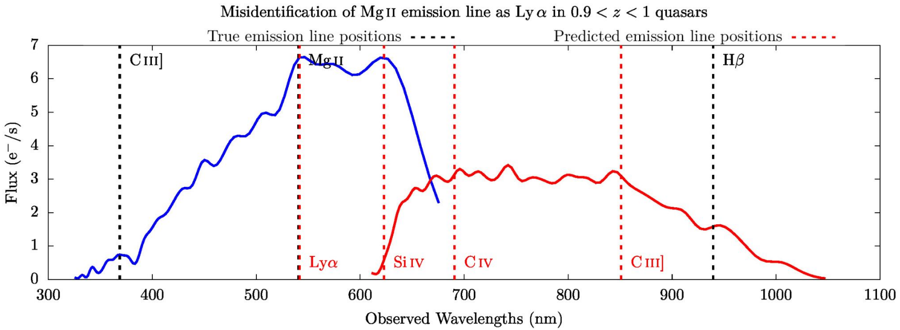


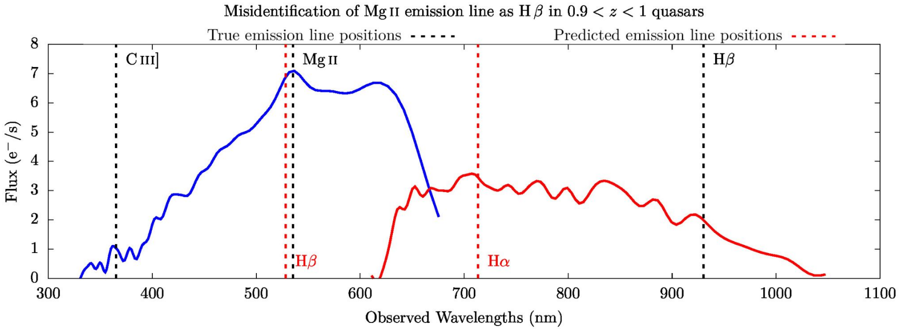
Illustration adapted from Gaia Collaboration, Bailer-Jones et al. (2022)

Use of QSOC results

The sole presence of the Mg_{II} emission line has the deleterious effect of increasing the rate of mismatches between this line and the Lyα and Hβ emission lines.

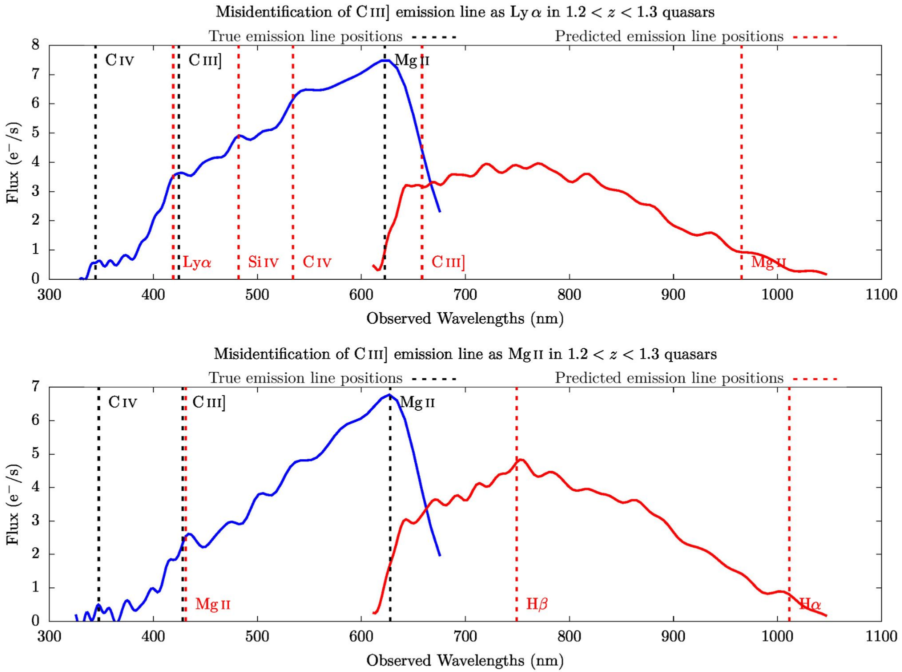
This causes many false predictions in very low ($0.08 < z < 0.32$) and very high ($3.4 < z < 4.3$) redshift regions.





Use of QSOC results

At $z \approx 1.3$, the $C_{III]}$ emission line enters the BP spectrum while the Mg_{II} line now lies on the peak of the BP spectrum, where it is wiped out in the mean spectrum representation, leading to mismatches between $C_{III]}$ and the Lyα or Mg_{II} emission lines.



Use of QSOC results

Purity

Fraction of sources that have |Δz| < 0.1 wrt. predicted redshift.

Best purity in the range 1.7 < z < 2.4 due to the combined presence of Lyα + C_{IV} + $C_{III]}$ + Mg_{II} emission lines.

Cuts on the flags_qsoc allow us to recover about 90% of sources with |Δz| < 0.1 in the range 0.2 < z < 4.

The lower purity at z < 0.2 or z > 4, comes from the rarity of these very low/high redshift quasars
→ any false predictions towards these loosely populated regions are largely reflected

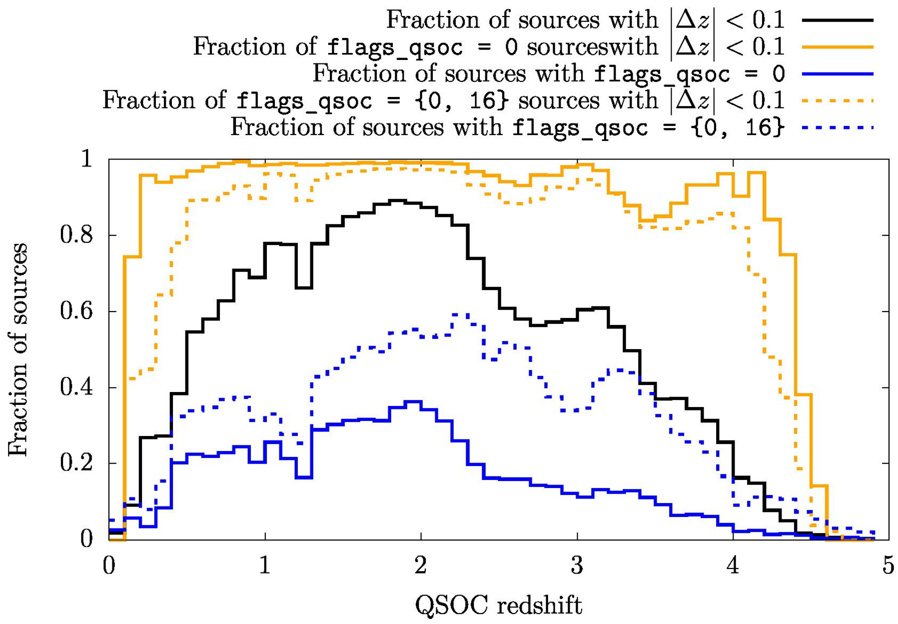


Illustration: Delchambre et al. (2022)

Use of QSOC results

Summary

- **QSOC aim to be complete rather than pure**
  Users interested in purer samples should use classlabel_dsc_joint ='quasar' or recipes given in Gaia Collaboration, Bailer-Jones et al. (2022).
- **Degeneracies exist between the predicted redshifts over some redshift ranges**
  - **$0.9 < z < 1.3$**: Due to the sole presence of the Mg_{II} emission line in this range
  - **$z \approx 2$**, due to $C_{IV} \leftrightarrow Ly\alpha$ mismatches

  More frequent amongst faint sources. Cuts on flags_qsoc can discard these degenerated cases.
- **Requiring that flags_qsoc =0 may lead to a large amount of predictions that are discarded**
  For users interested in $G \gtrsim 19$ mag quasars, we suggest to use flags_qsoc=0 or flags_qsoc=16.
- **QSOC is designed to process Type-I/core-dominated quasars**
  Galaxies, type-II AGN and BL Lacertae/Blazars objects usually have poor predictions and flags_qsoc≠0.
- **In Gaia DR3, we did not had access to the full covariance matrix on the BP/RP fluxes**
  The reduced chi-square of the solution is systematically underestimated and is consequently not published. The computed redshift and associated confidence intervals though appropriately re-scaled, might also sporadically suffer from this limitation

Future plans

1. **Use of epoch BP/RP spectra**
   Variability taken into account, sharper spectral features, uncorrelated noise on the flux (required for the correlation algorithm), fix the wiggles issues + other internal benefits.

2. **Use of priors**
   Based on a quasar luminosity function. Lift off the degeneracies between most of the predicted redshifts.

3. **Potential new astrophysical parameters**
   Continuum slope, balnicity index, emission line equivalent width, absolute magnitude, second/third most probable redshift, …