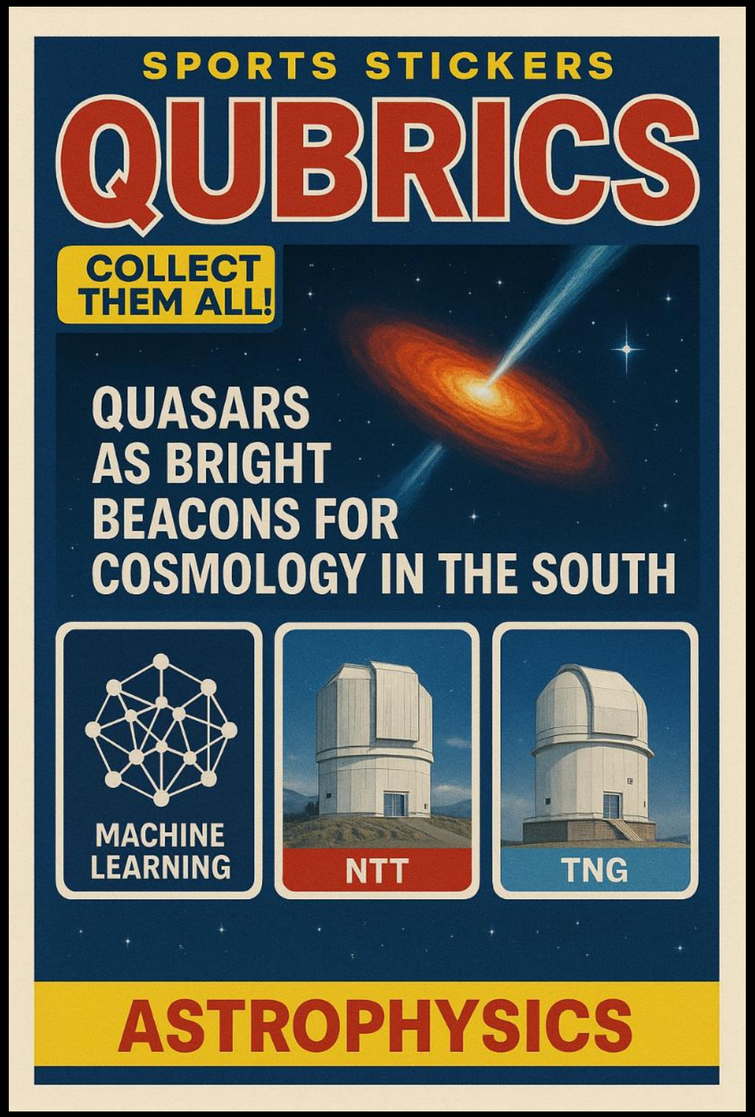
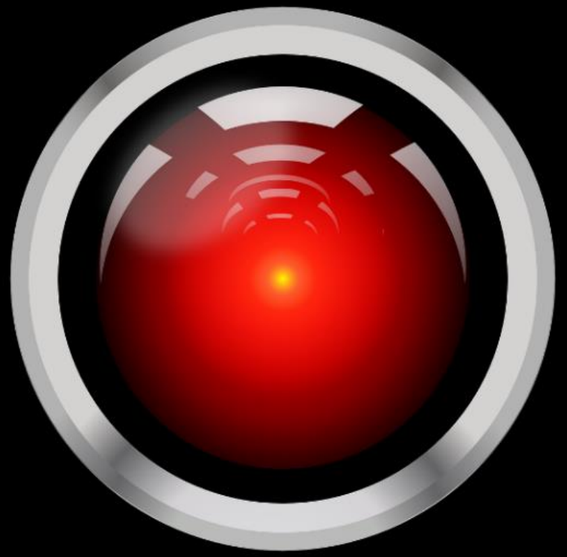
QUBRICS – finding the beacons





QUasars as **BRI**ght beacons for **C**osmology in the **S**outhern hemisphere

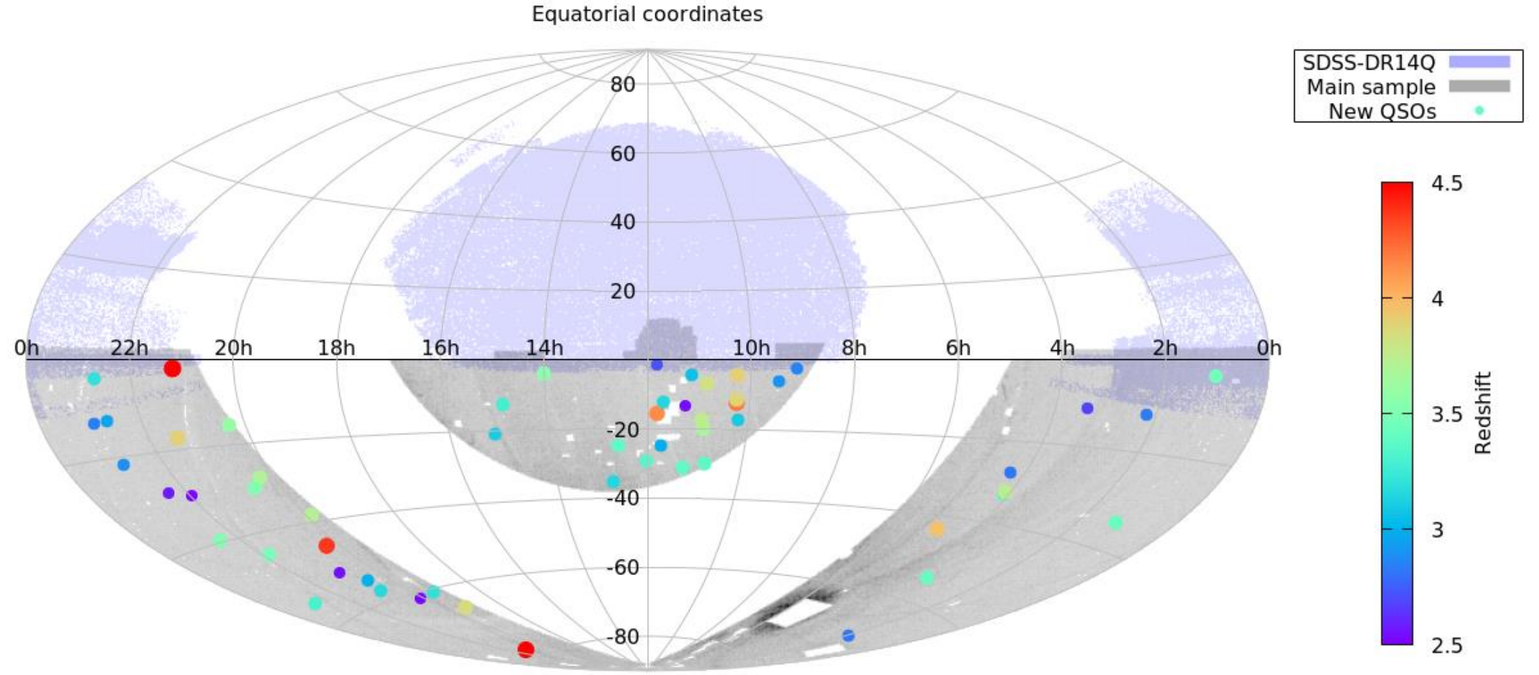
QUBRICS-Team & References

- Konstantina Boutsia, Giorgio Calderone, Stefano Cristiani, Andrea Grazian
- Guido Cupani,Valentina D'Odorico, Fabio Fontanot, Francesco Guarneri, Matteo Porru...

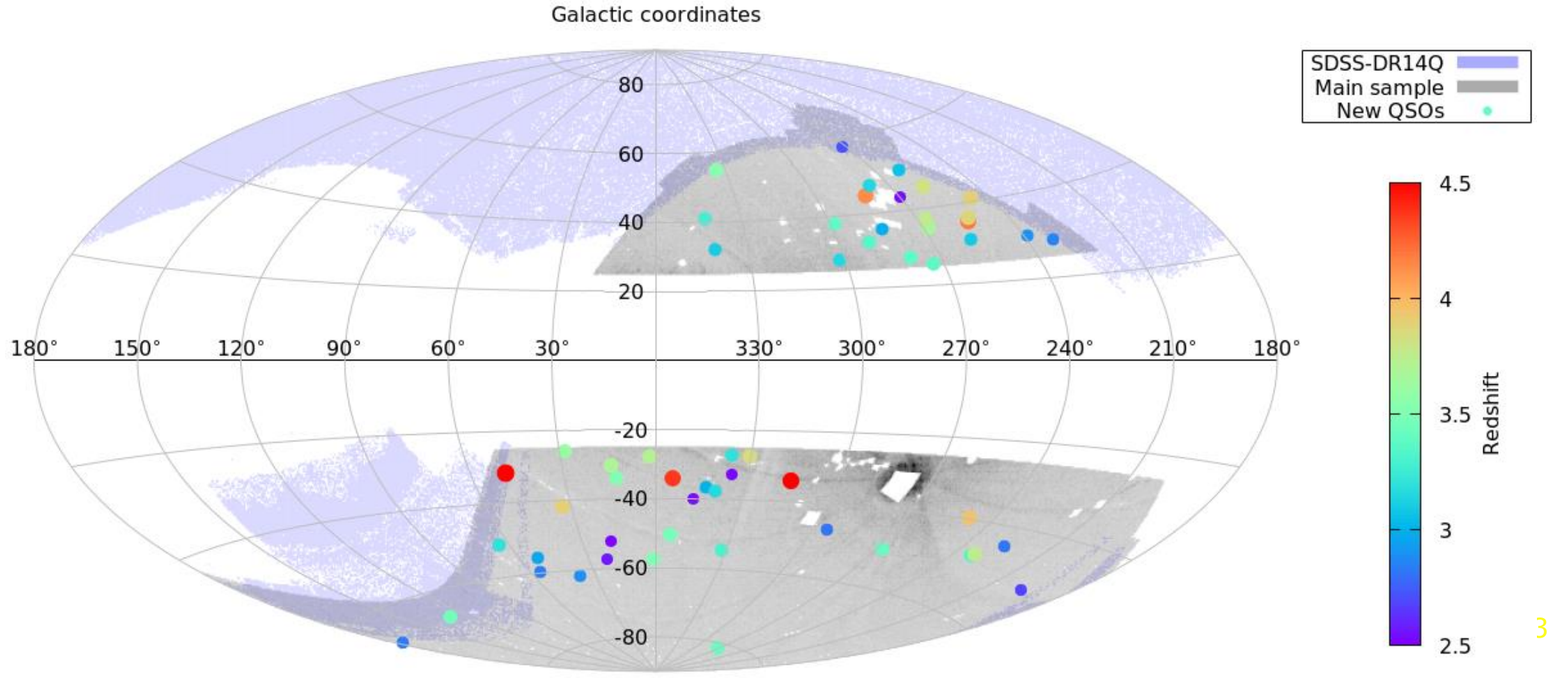
Calderone G., et al., 2019, ApJ, 887, 268… … …Cristiani, S., et al., 2023, MNRAS, 522, 2019

Main goal:

- **Identify bright, high-redshift QSOs** using data from publicly available photometric survey:
  - Gaia
  - Skymapper
  - PAN-STARRS
  - 2MASS
  - WISE
  - DES
  - eROSITA
  - Rubin Observatory
  - Euclid …
- Two-fold problem: first identify QSOs, then remove low-redshift (z<2.5) QSO/ (Classification + Interpolation)



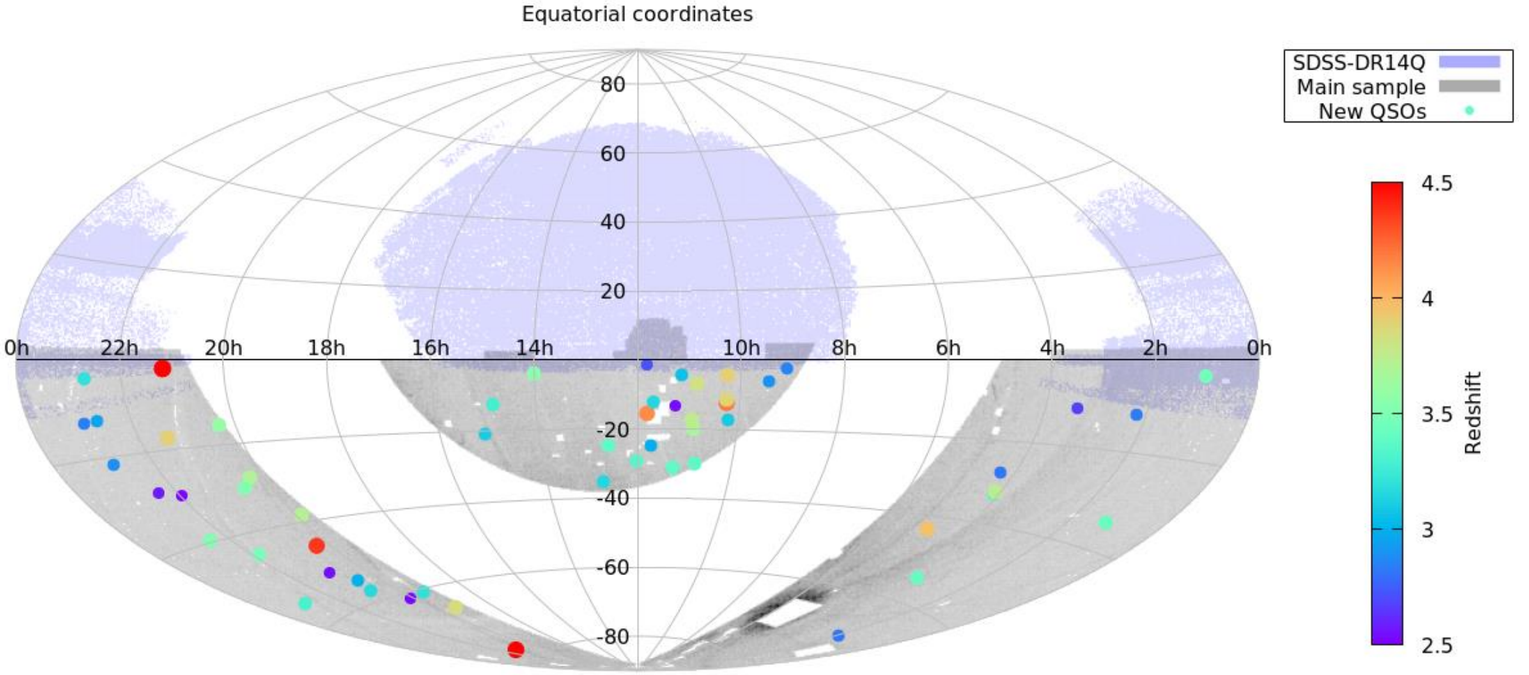
$|b| > 25$




My God
It's full of stars
quickmeme.com

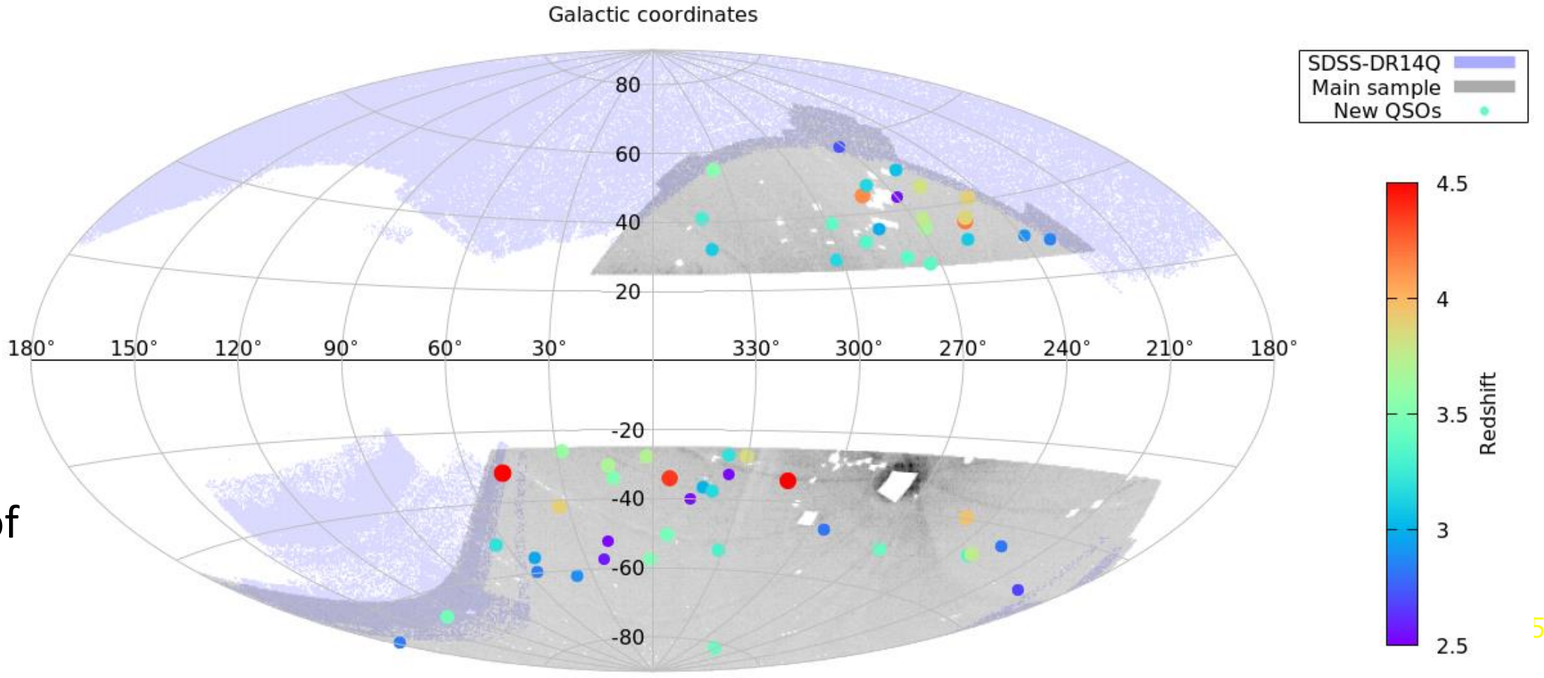
Main goal:

- Identify bright, high-redshift QSOs using data from publicly available photometric survey:
  - SkyMapper
  - Gaia
  - PAN-STARRS
  - 2MASS
  - WISE
  - DES …
- Two-fold problem: first identify QSOs, then remove low-redshift objects



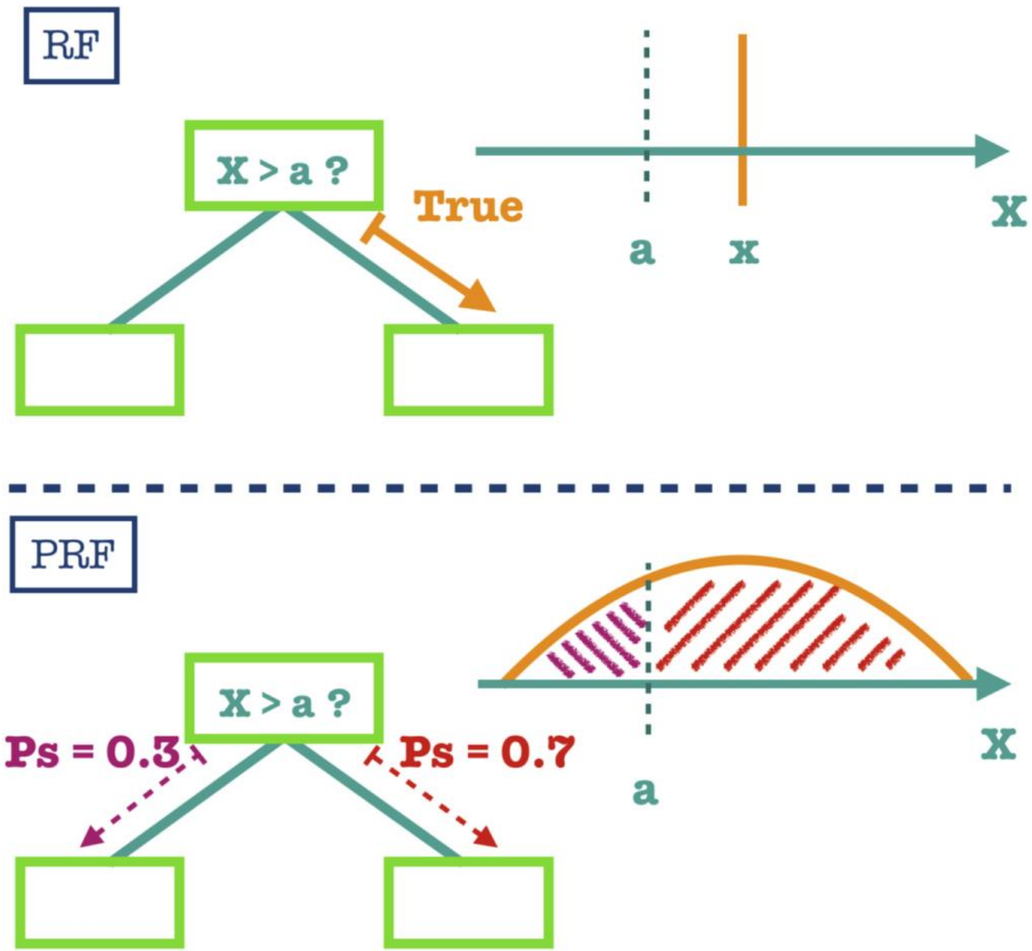
Method:

- Apply **ML techniques** on photometric datasets:
  - Canonical Correlation Analysis (**CCA**)
    - Calderone et al 2019 *ApJ* 887 268
    - Boutsia et al 2020 *ApJS* 250 26
  - Probabilistic Random Forest (**PRF**)
    - Guarneri et al 2021 MNRAS 506 2
    - Guarneri et al 2023 MNRAS
  - **XGB**: Calderone et al. 2024
- Spectroscopic follow-up to confirm the nature of high-redshift candidates



The QUBRICS survey: Probabilistic Random Forest (Reis et al. 2019)

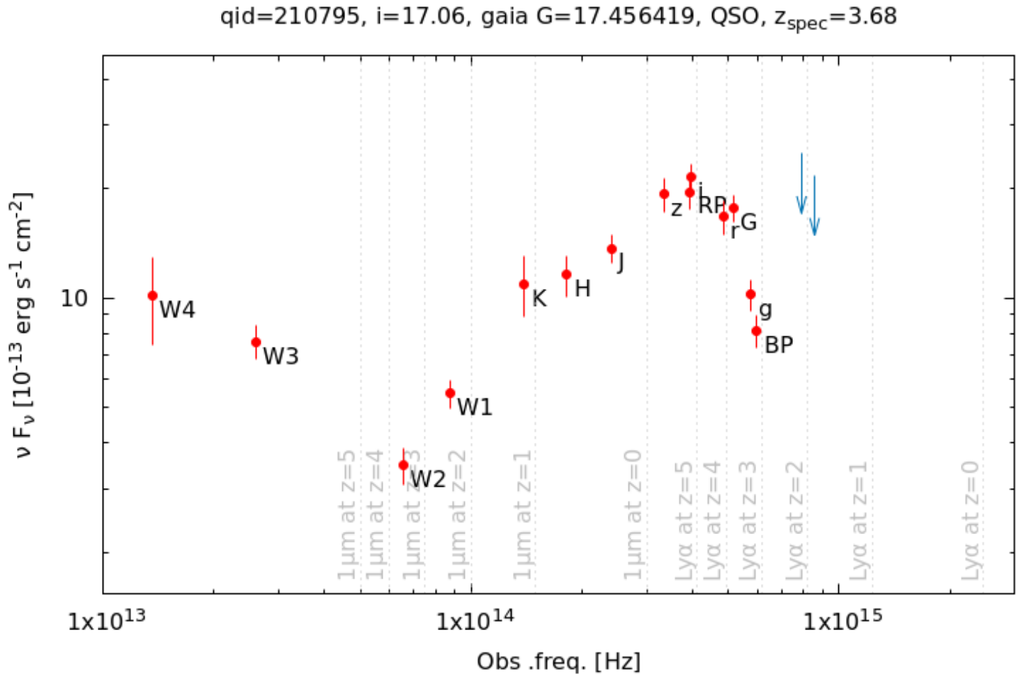
- Generalization of the original Random Forest (RF) to account for measurement uncertainties
- In the PRF each feature is a probability distribution function: this improves performances and takes into account errors as variance of the distribution
- Naturally handles missing data (upper limits and lack of observations)!

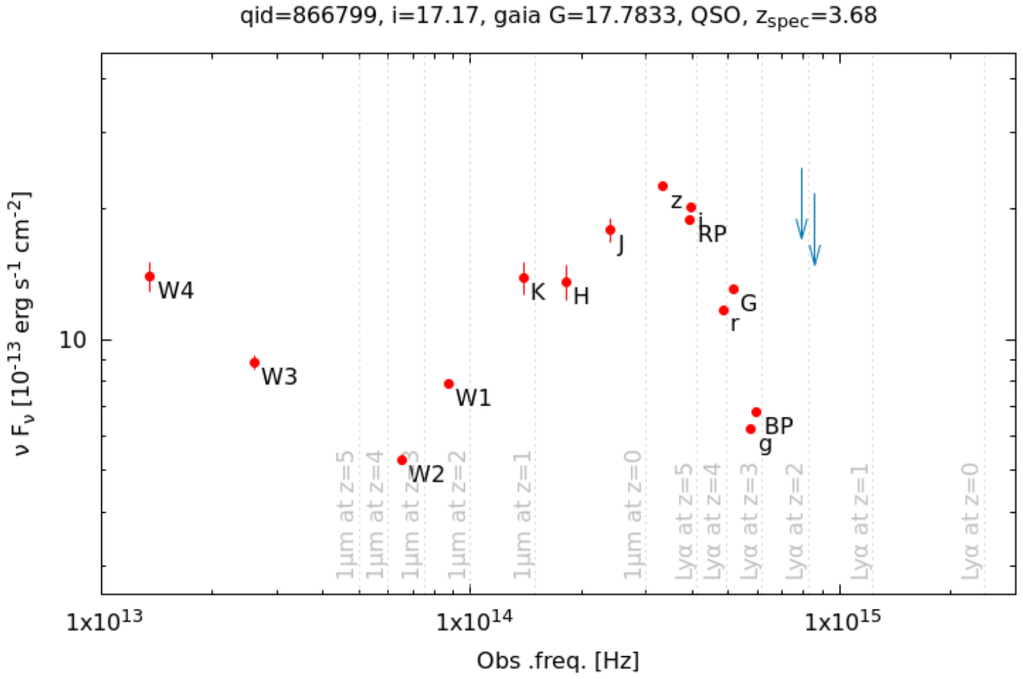


Reis et al. 2019 - arxiv:1811.05994

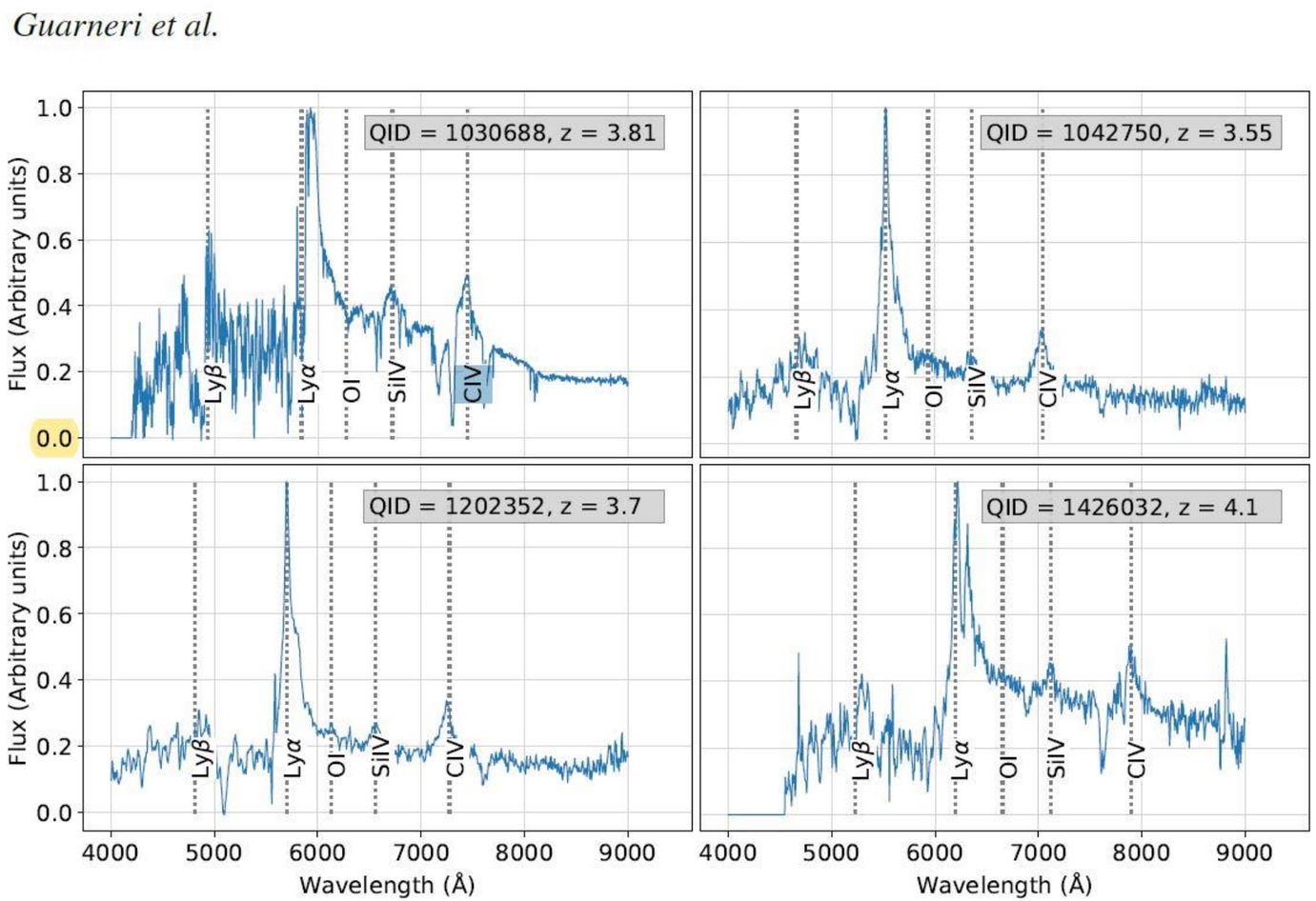
Guarneri et al 2021, MNRAS 506 2

Guarneri et al 2022, MNRAS 517 2436





Example SEDs and follow-up spectra



Learning (many) lessons from ML

Beware of:

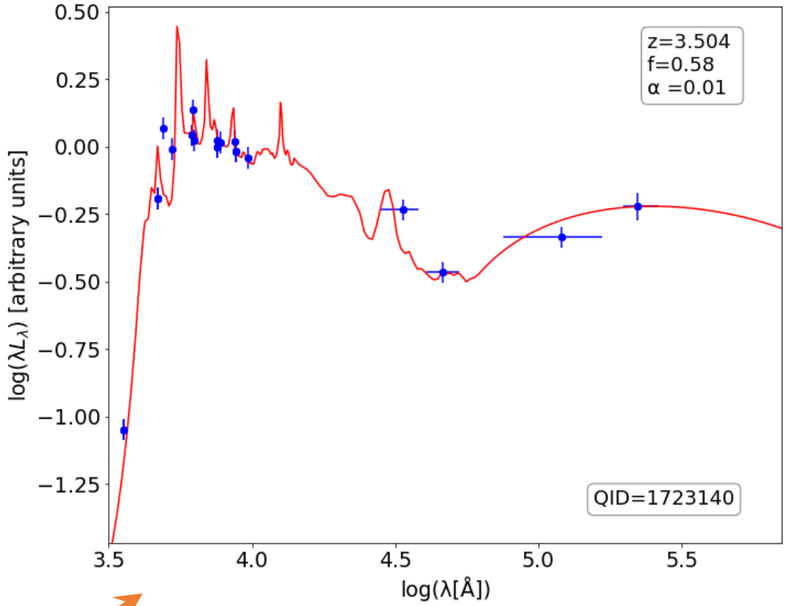
- Black box syndrome (Petch et al. 2022, e.g.)
- Overfitting (complementary methods)
- Fancy interpretation of unphysical features
- Amazing success rates and completeness

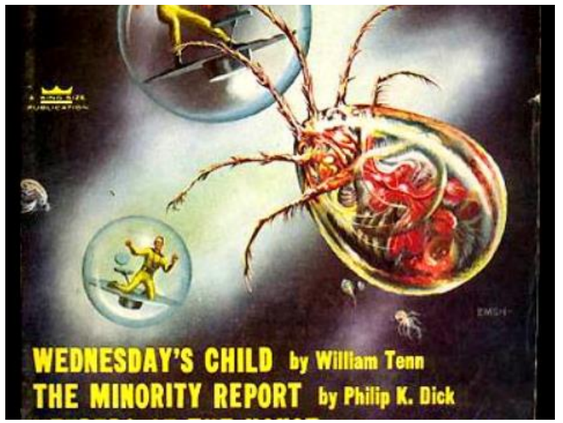
Consider that:

- Good for classification may be less good for regression
- On a well defined class, fitting a model may be fine

Need for:

- Large and balanced training sets (synthetic data)
- Proper error treatment
- Physical insight





Remember *Minority Report (Dick, 1956)*

Mervin *et al. J Cheminform* (2021) 13:62
https://doi.org/10.1186/s13321-021-00539-7

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

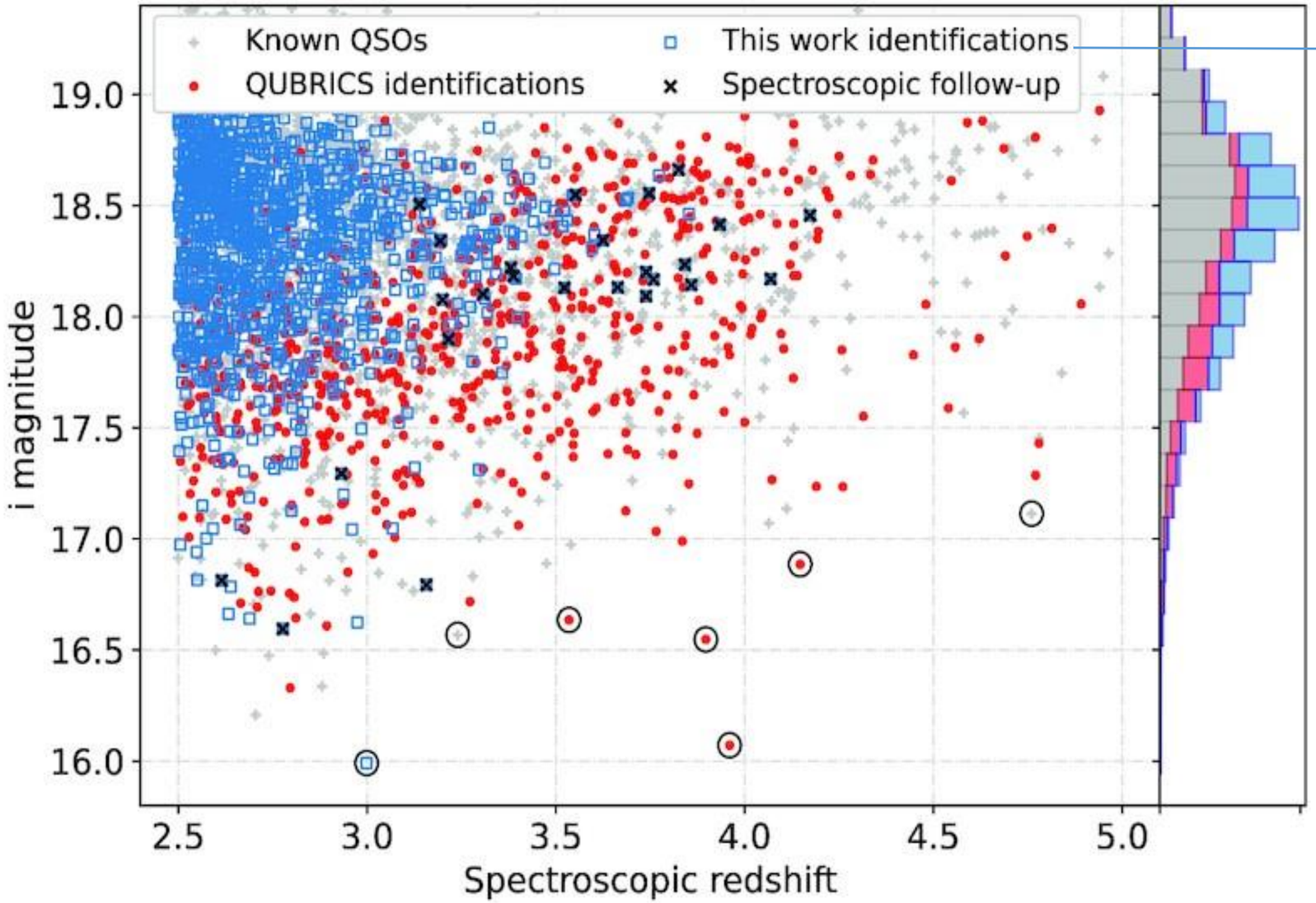
Probabilistic Random Forest improves bioactivity predictions close to the classification threshold by taking into account experimental uncertainty

Lewis H. Mervin[1*†], Maria-Anna Trapotsi[2†], Avid M. Afzal[3], Ian P. Barrett[3], Andreas Bender[2] and Ola Engkvist[4,5]

*Correspondence: lewis.mervin1@astrazeneca.com
†Lewis H. Mervin and Maria-Anna Trapotsi contributed equally to this work
[1] Molecular AI, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK
Full list of author information is available at the end of the article

"*It seems the nature & bias in our data shares much similarity to the ones in yours, in the search for quasars.*"

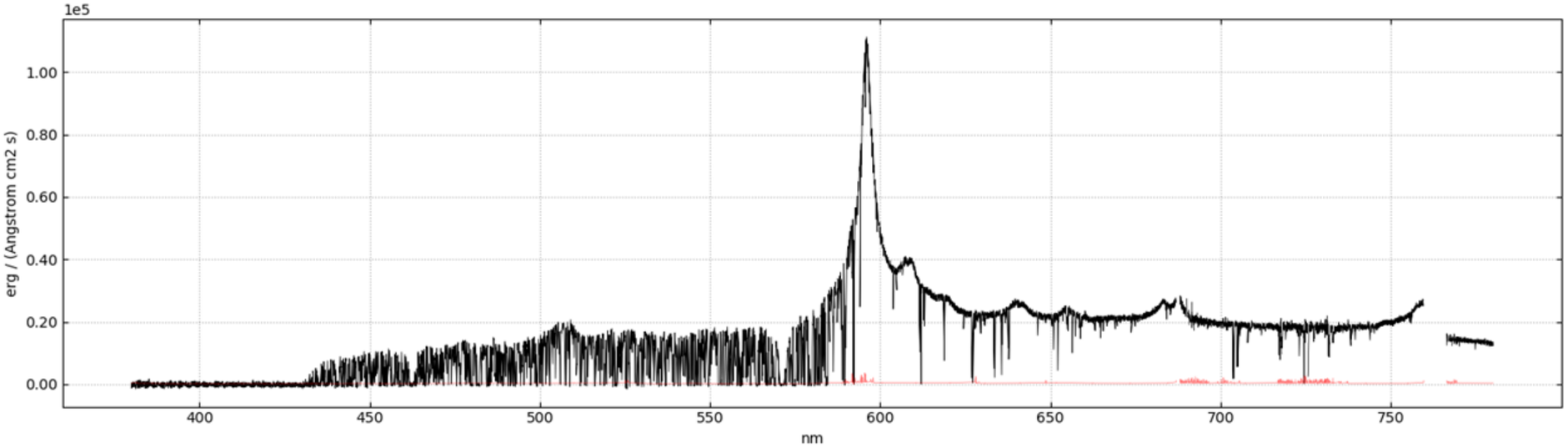
Observations – Finding the cosmic beacons



Cristiani et al. 2023

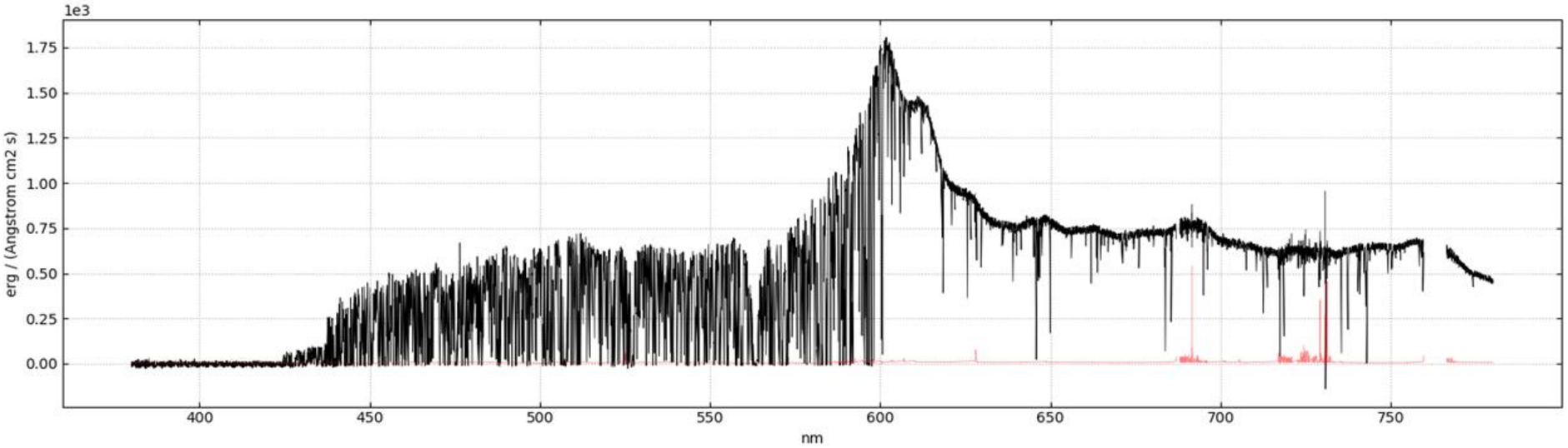
SB1

total int.: ~8.6 h
⟨SNR⟩: ~60
last obs.: 15/8/23

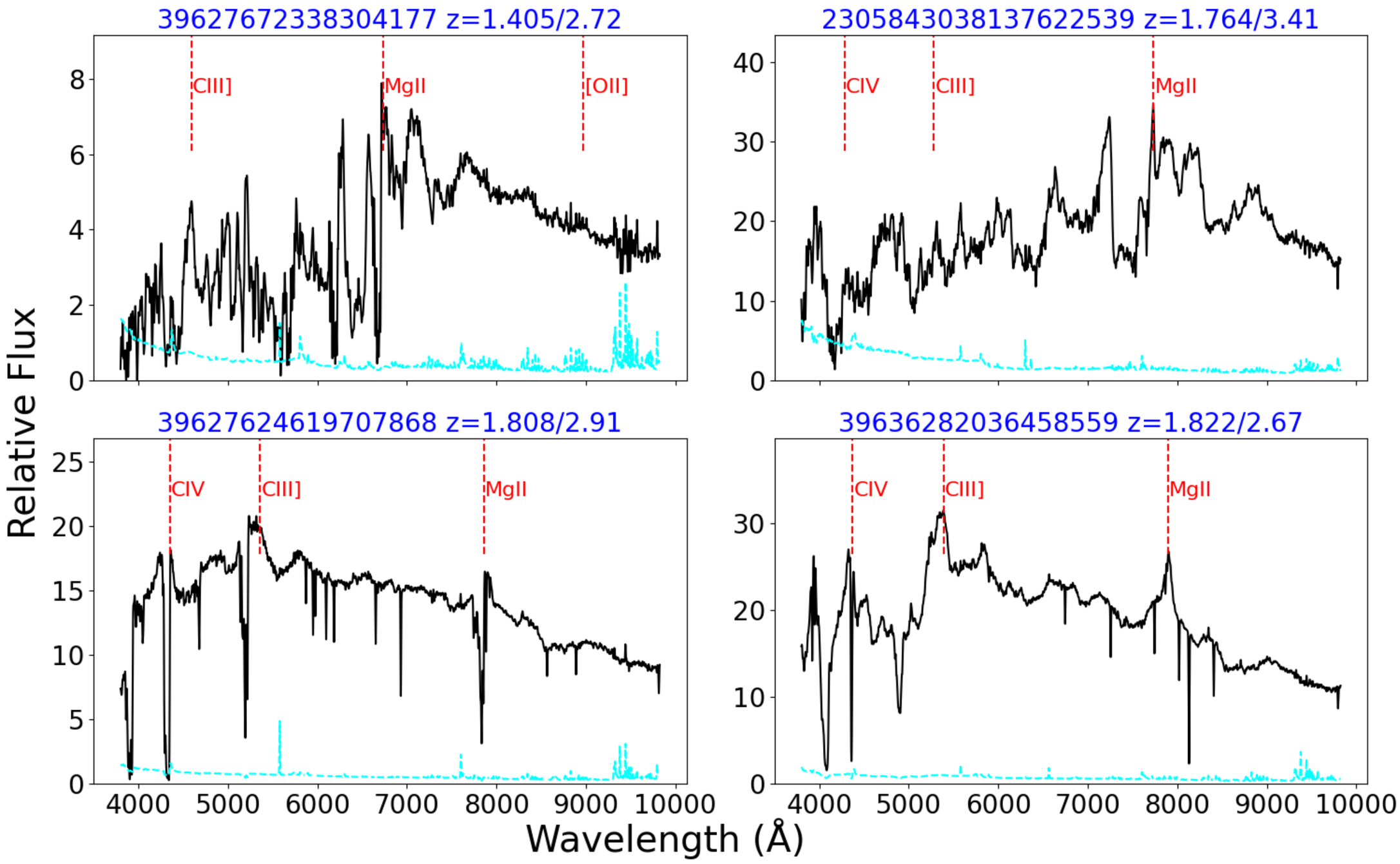


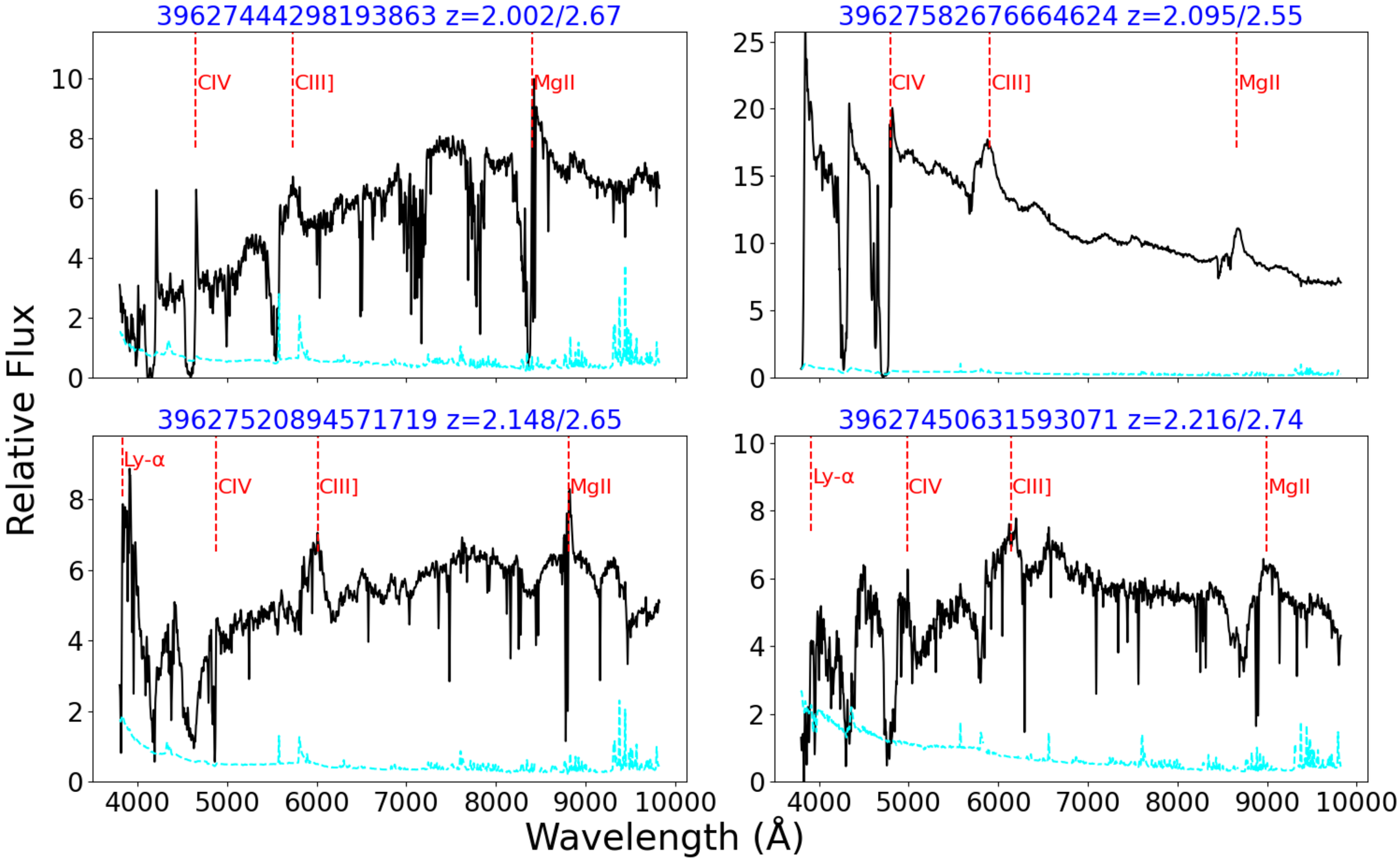
SB2

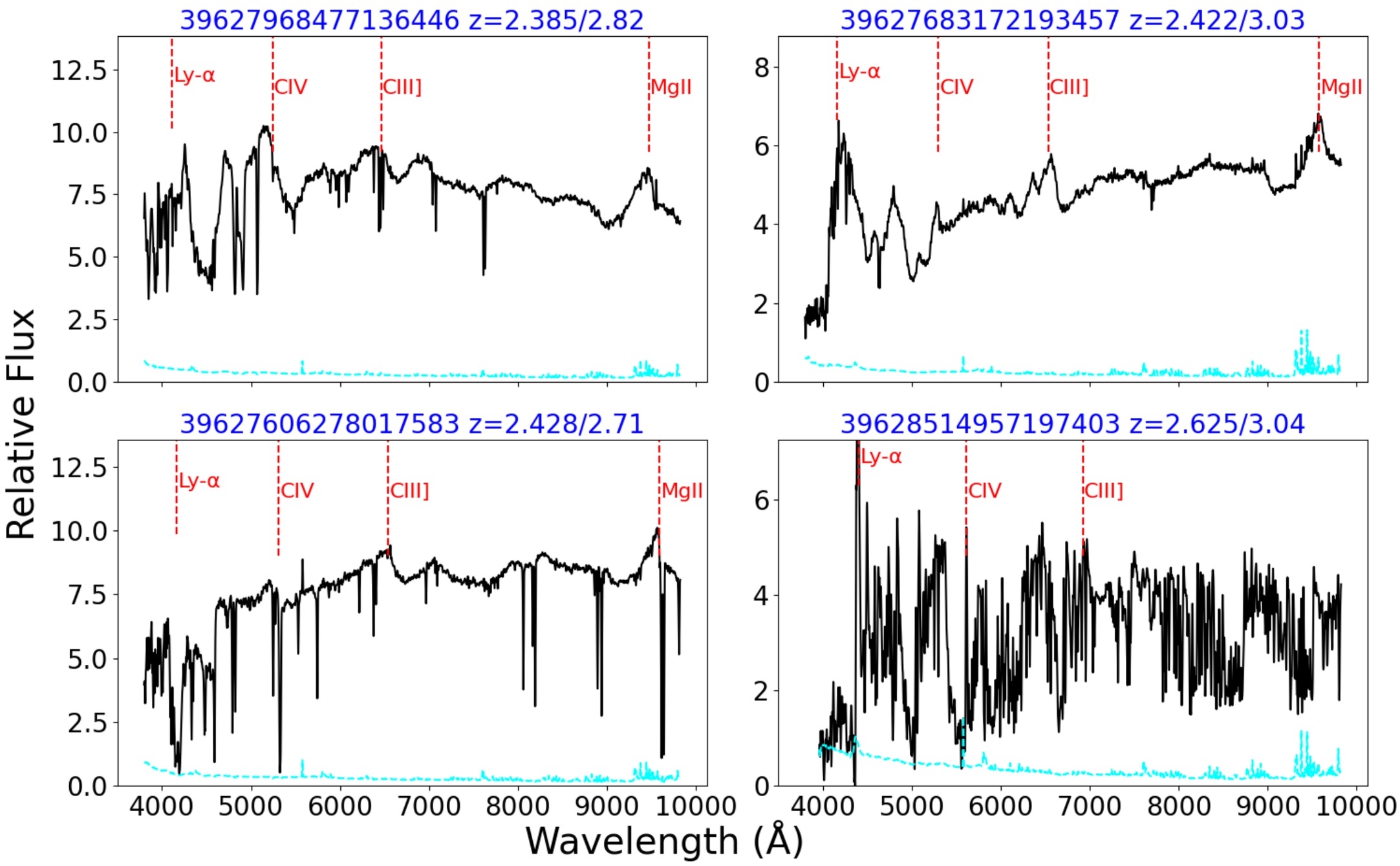
total int.: ~22 h
⟨SNR⟩: ~120
last obs.: 30/12/24



Rogues’ Gallery


3962767233830417 7 z=1.405/2.72
CIII]
MgII
[OII]
2305843038137622539 z=1.764/3.41
CIV
CIII]
MgII
3962762461970786 8 z=1.808/2.91
CIV
CIII]
MgII
3963628203645855 9 z=1.822/2.67
CIV
CIII]
MgII
Relative Flux
Wavelength (Å)

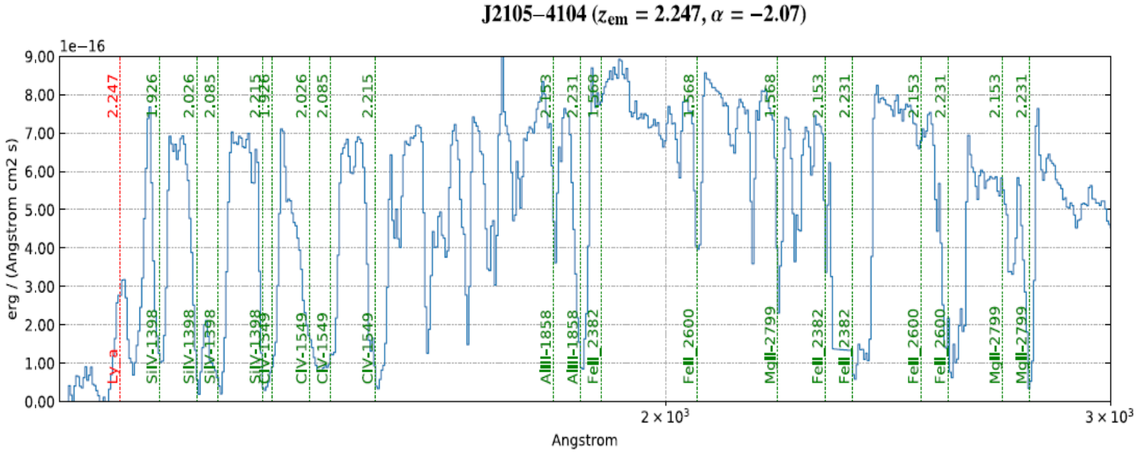

396274429819386 3 z=2.002/2.67
396275826766462 4 z=2.095/2.55
396275208945717 19 z=2.148/2.65
396274506315930 71 z=2.216/2.74
CIV
CIII]
MgII
Ly-α
Relative Flux
Wavelength (Å)

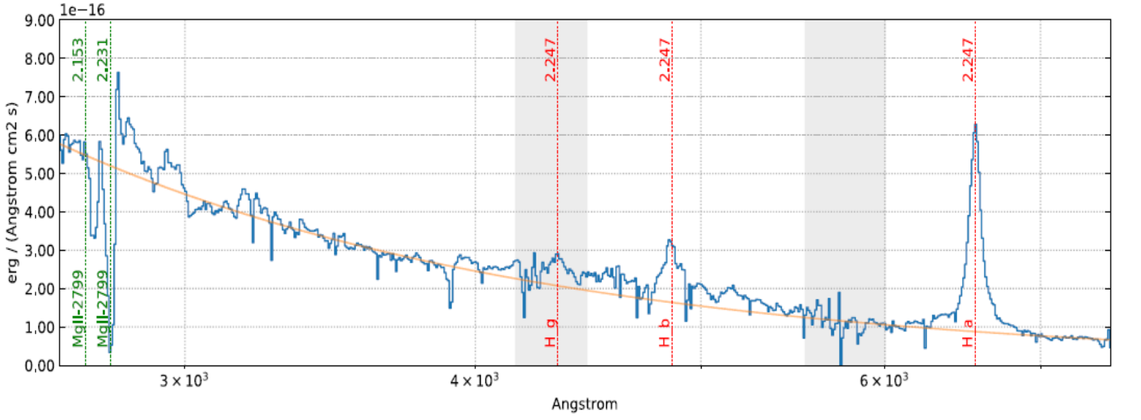

396279684771364 46 z=2.385/2.82
396276831721934 57 z=2.422/3.03
396276062780175 83 z=2.428/2.71
396285149571974 03 z=2.625/3.04
Ly-α
CIV
CIII]
MgII
Relative Flux
Wavelength (Å)

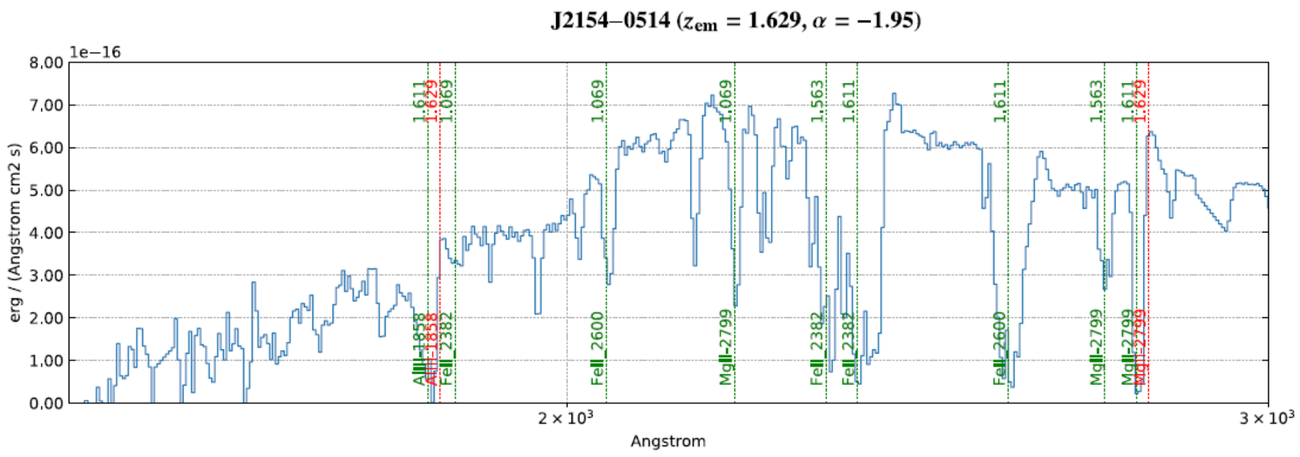
Near-infrared spectroscopy of extreme BAL QSOs from the QUBRICS bright quasar survey

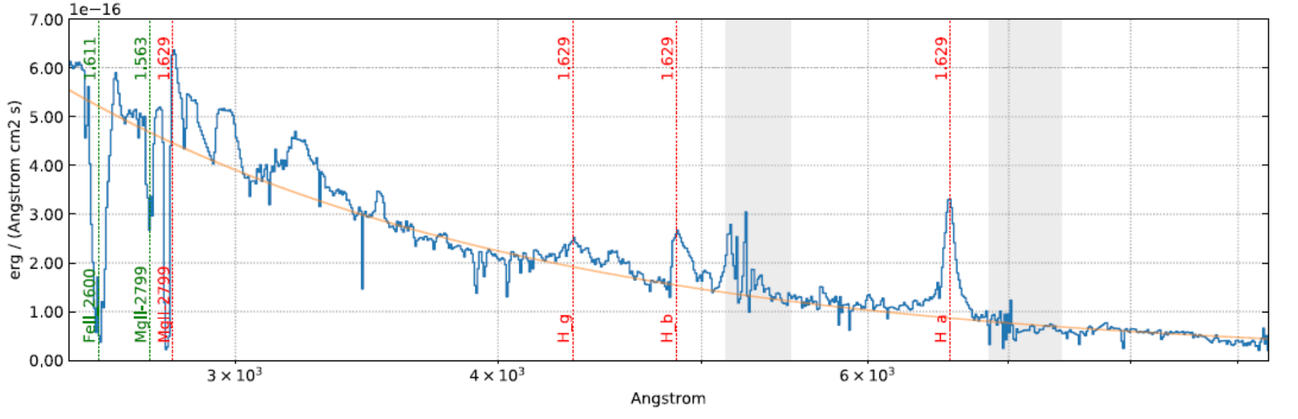
Guido Cupani [1,2★] Giorgio Calderone [1] Pierluigi Selvelli,[1] Stefano Cristiani,[1,2,3]
Konstantina Boutsia [4] Andrea Grazian,[5] Fabio Fontanot [1,2] Francesco Guarneri [1,6]
Valentina D'Odorico,[1,2,7] Emanuele Giallongo[8] and Nicola Menci[8]

2022









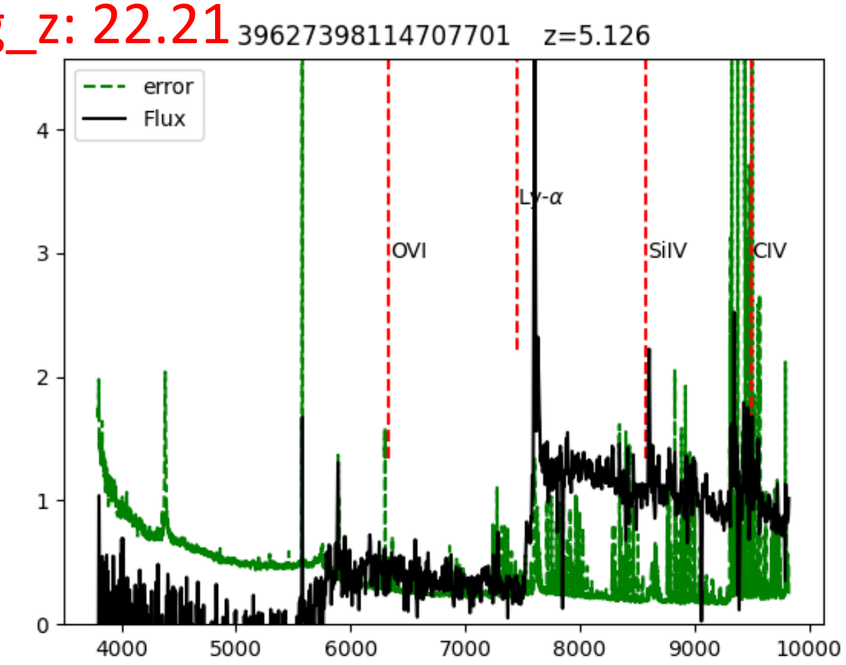
DESI DR1

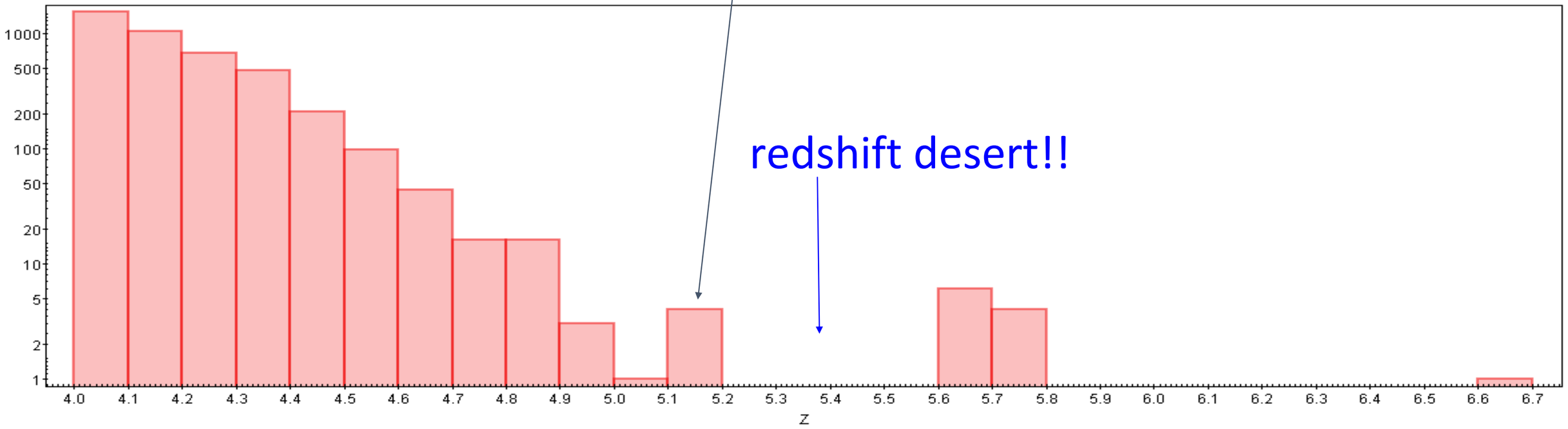
4171 QSOs

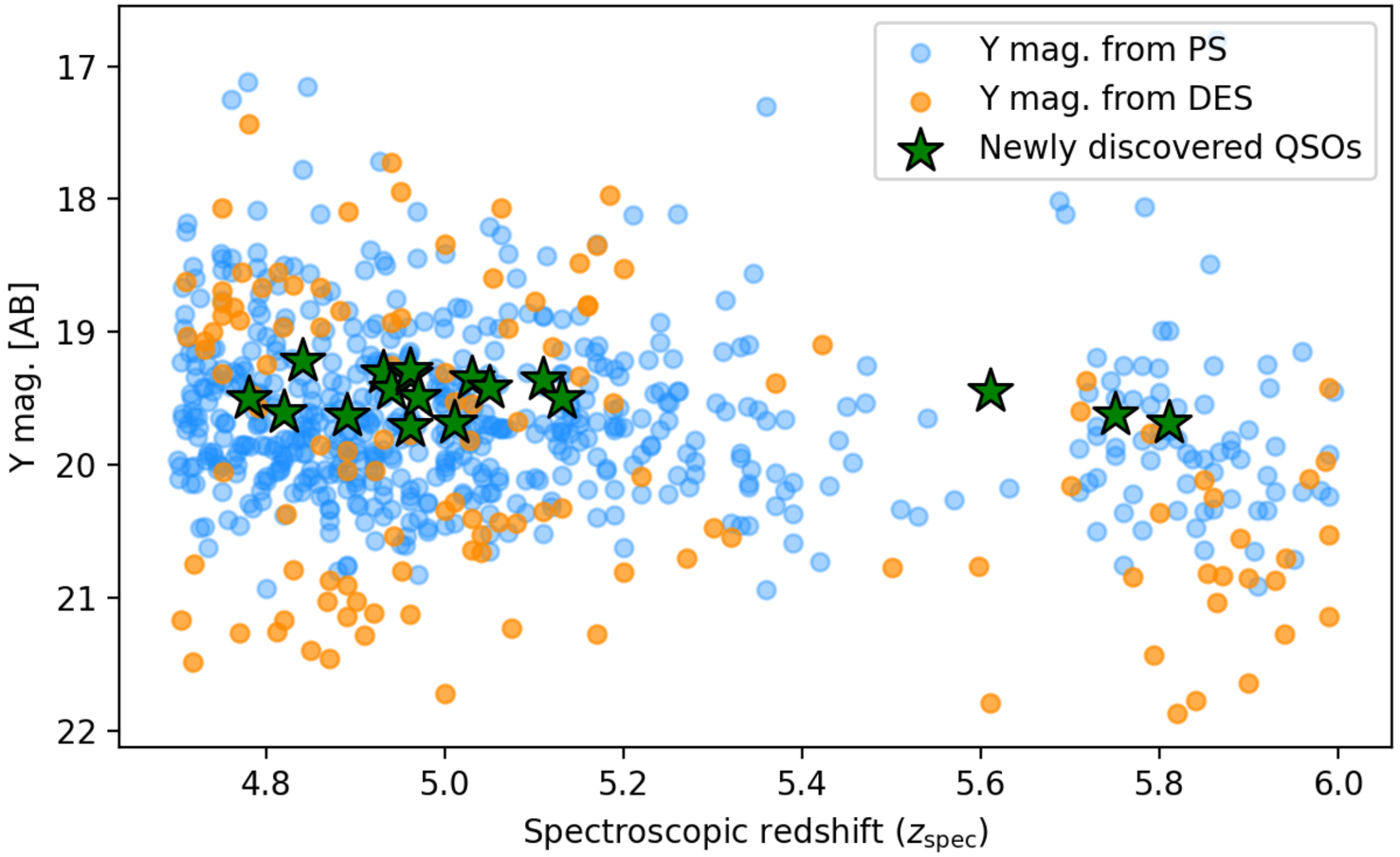
z>4

DEC<15

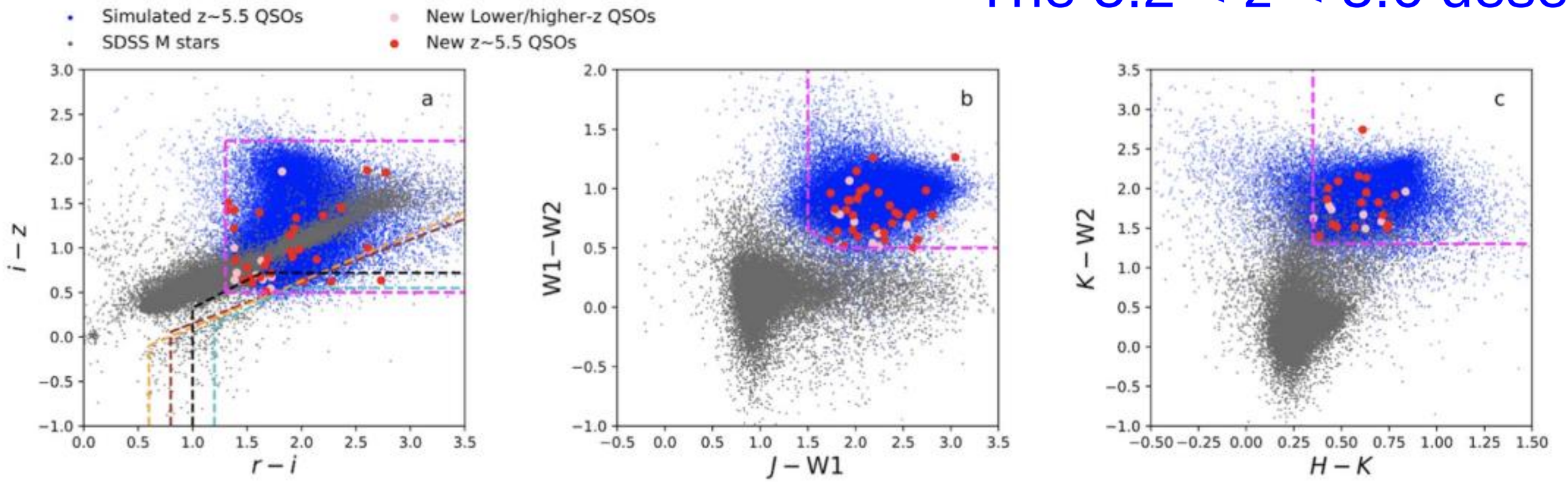
highest z mag_z: 22.21






Y mag. from PS
Y mag. from DES
Newly discovered QSOs
Y mag. [AB]
17
18
19
20
21
22
4.8
5.0
5.2
5.4
5.6
5.8
6.0
Spectroscopic redshift (z_{spec})

The 5.2 < z < 5.6 desert



THE ASTROPHYSICAL JOURNAL, 871:199 (17pp), 2019 February 1

https://doi.org/10.3847/1538-4357/aaf858

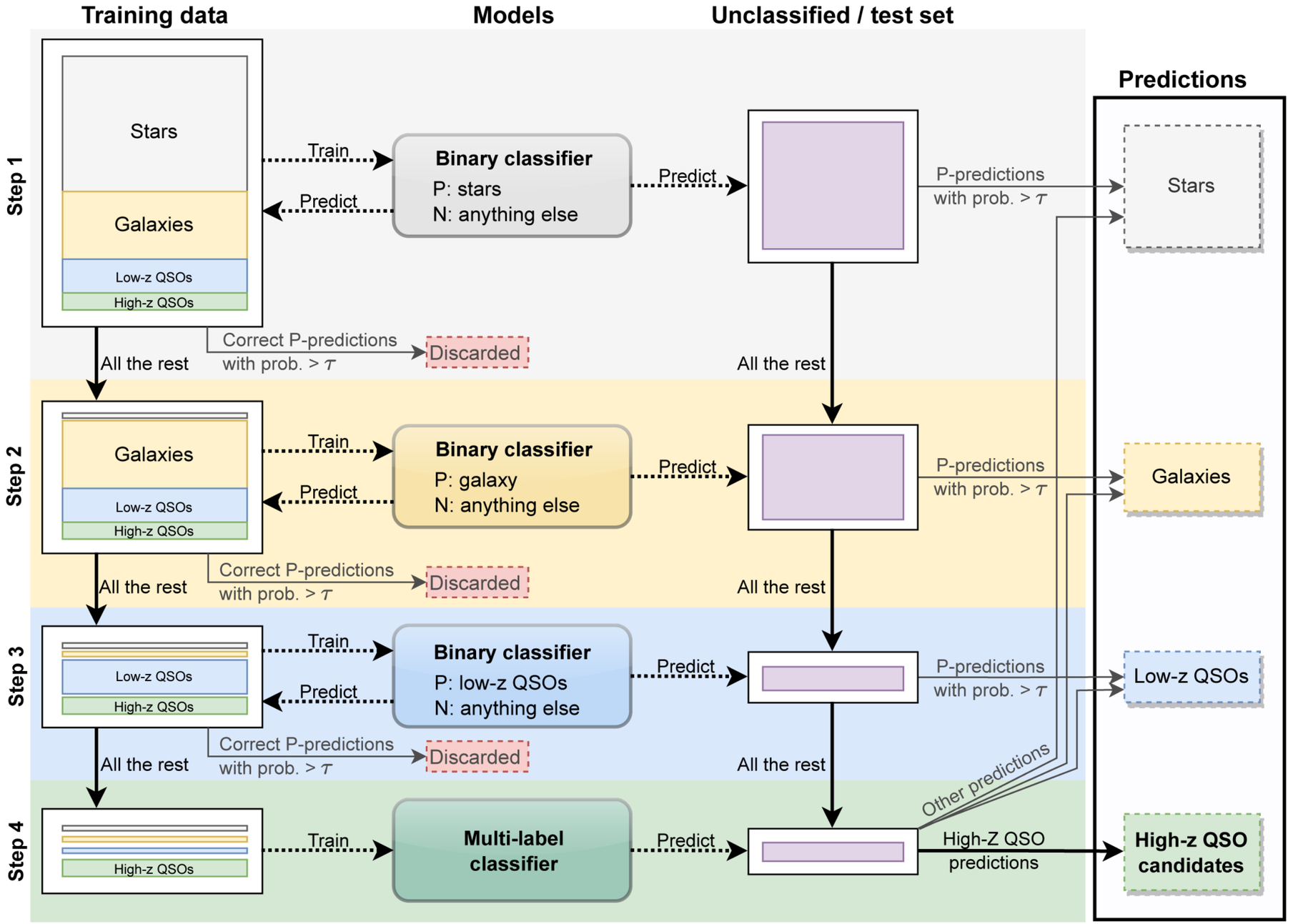
CrossMark

Filling in the Quasar Redshift Gap at $z \sim 5.5$. II. A Complete Survey of Luminous Quasars in the Post-reionization Universe

Jinyi Yang[1,2], Feige Wang[2,3], Xiaohui Fan[1], Xue-Bing Wu[2,4], Fuyan Bian[5], Eduardo Bañados[6], Minghao Yue[1], Jan-Torge Schindler[1], Qian Yang[2,4], Linhua Jiang[2], Ian D. McGreer[1], Richard Green[1], and Simon Dye[7]

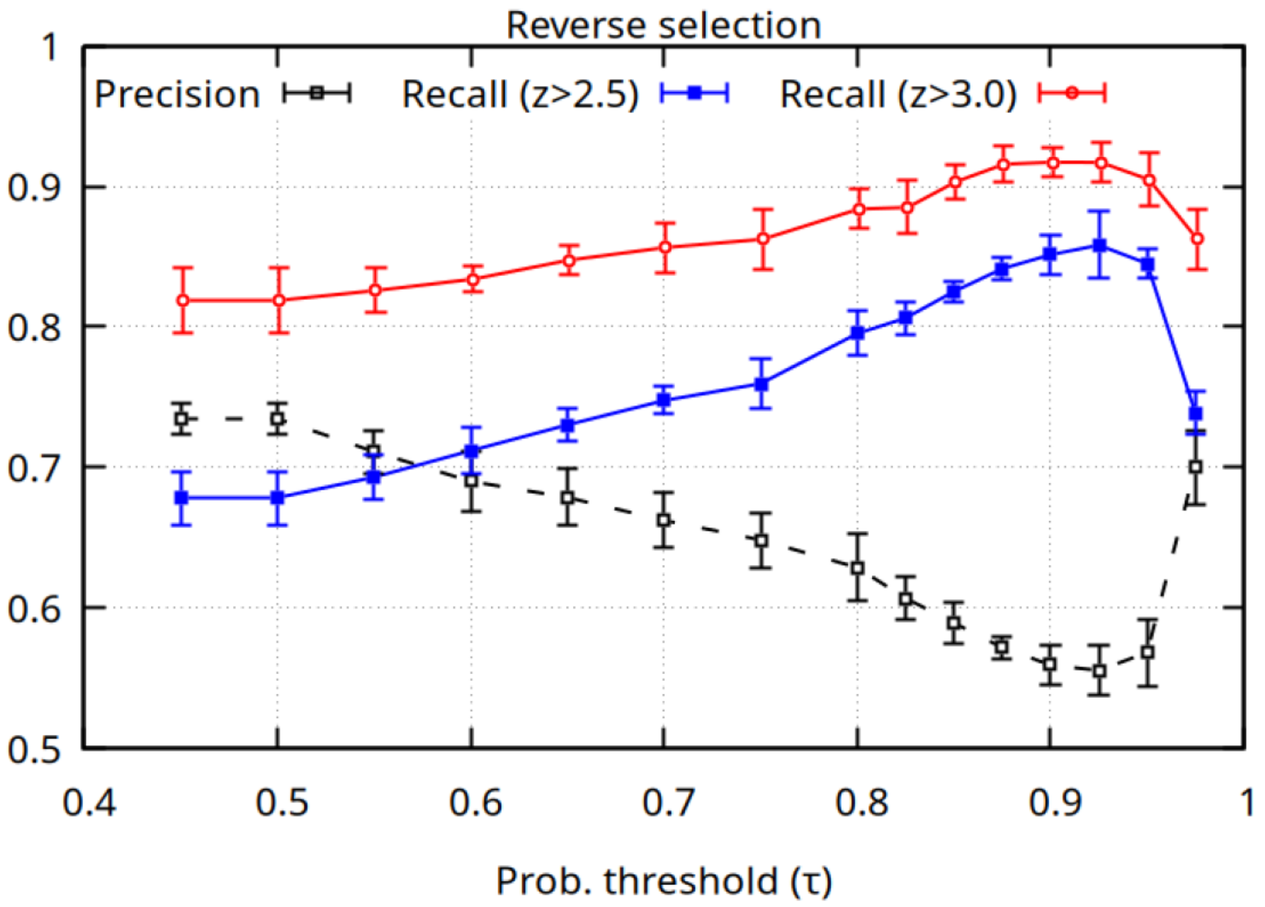
[1] Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ, USA; jinyiyang@as.arizona.edu
[2] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China
[3] Department of Physics, University of California, Santa Barbara, CA 93106-9530, USA

Reverse selection method XGB – Unbalanced data sets



Calderone+2024
A&A, 683, A34

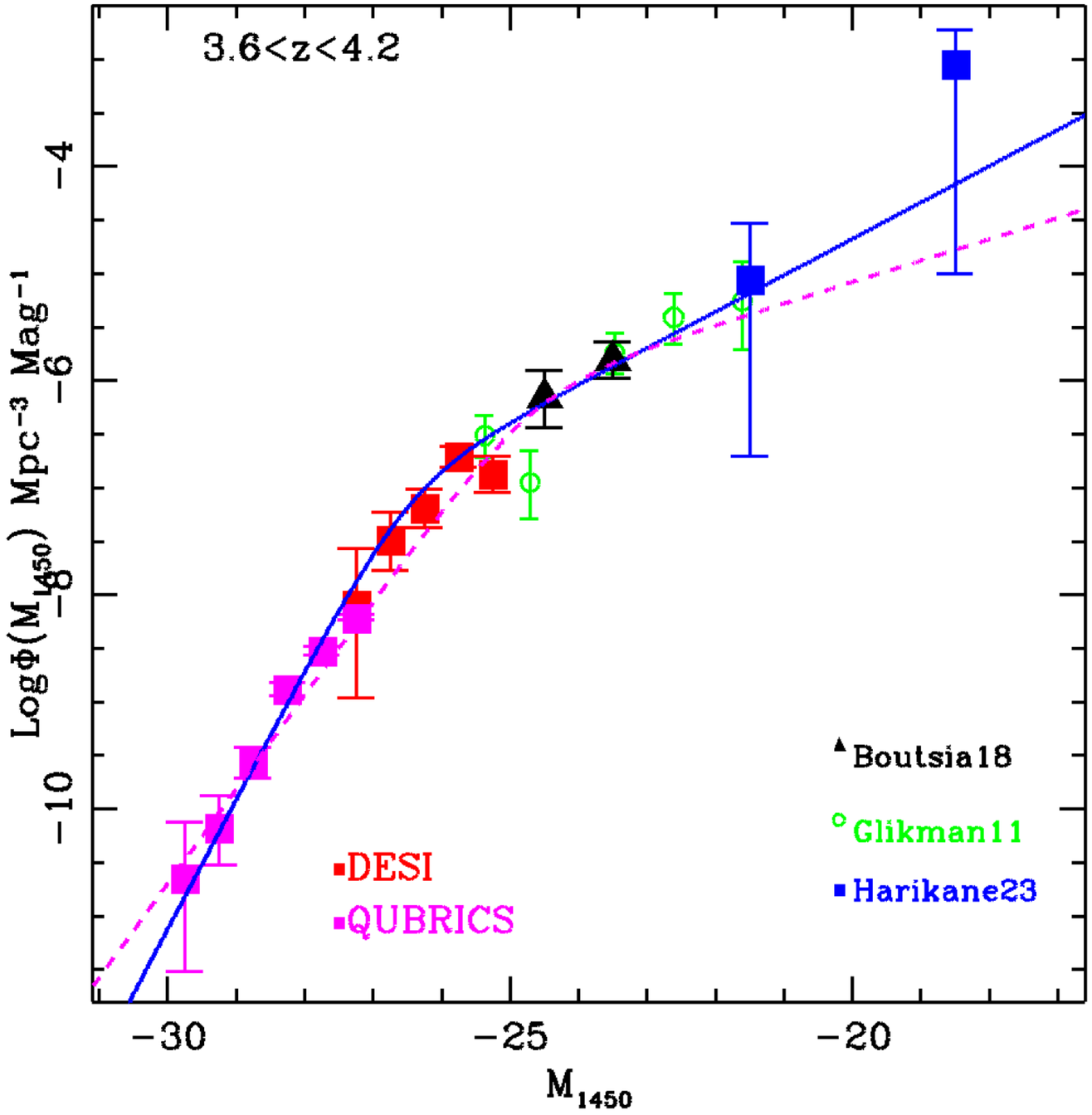
Calibrate probability threshold τ



Calderone+2024

- Increase recall, **slightly affect precision**;
- Optimal τ is the one maximizing recall;
- The better performances depend on the test set choice?
- ...run 100 analysis with random train/test splits, estimate recall at $z > 2.5, 3$.

SCIENTIFIC RESULTS



Boutsia+2025 in prep.

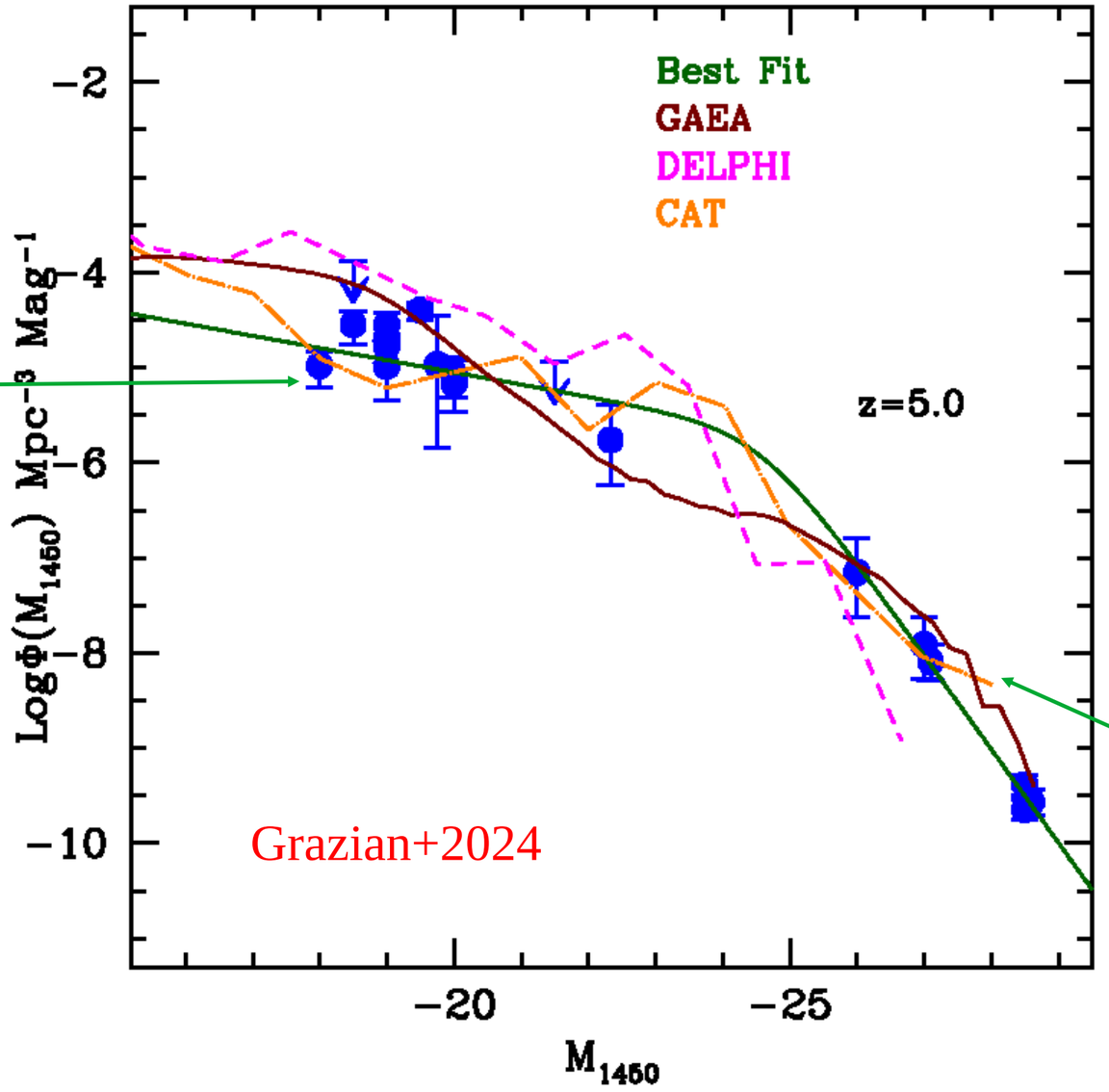
Table 3. The Best-Fit Parameters of the QSO LF at $3.6 \leq z \leq 4.2$ in the QUBRICS Footprint

α	β	M^*_{1450}	$Log\Phi^*$
$-1.850^{+0.150}_{-0.250}$	$-4.025^{+0.575}_{-0.425}$	$-26.50^{+0.85}_{-0.60}$	$-6.85^{+0.60}_{-0.45}$

Boutsia+2021

The z~5 AGN Luminosity Function

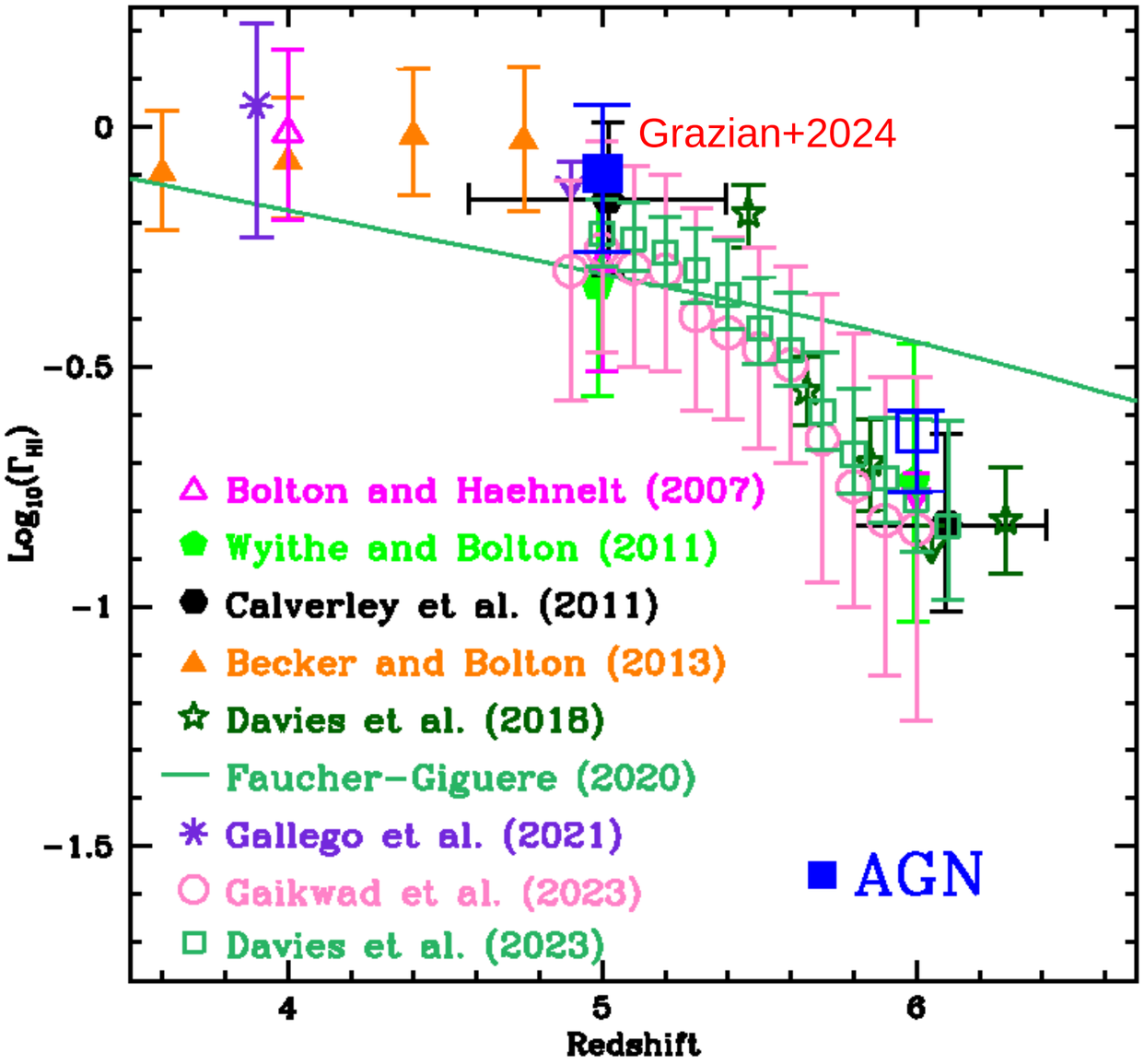
JWST:
Kocevski+2023
Harikane+2023
Matthee+2024
Greene+2024



RUBICON (Reionizing the Universe with BrIght COsmological Nuclei)
Grazian+2023

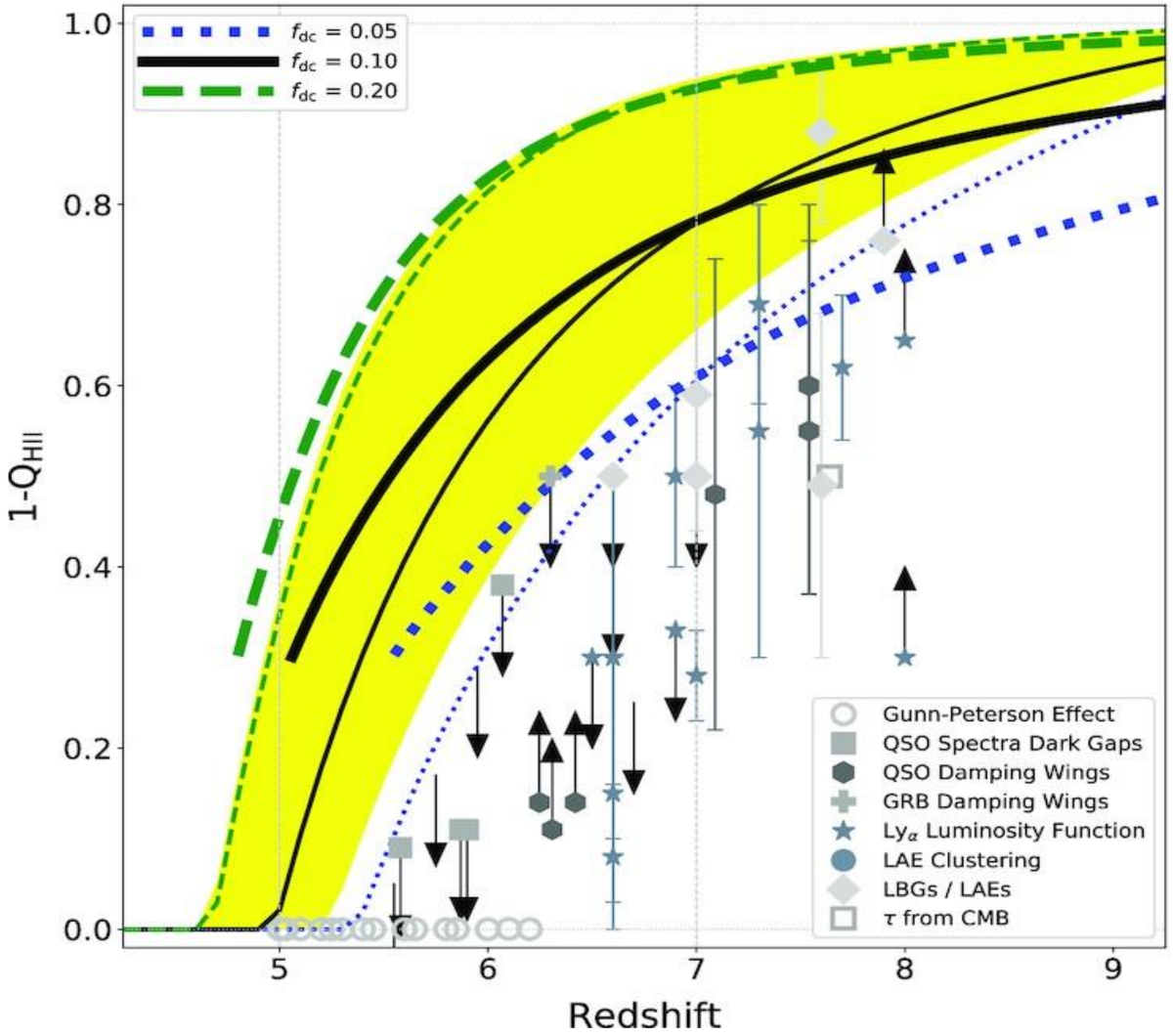
QUBRICS and RUBICON Surveys

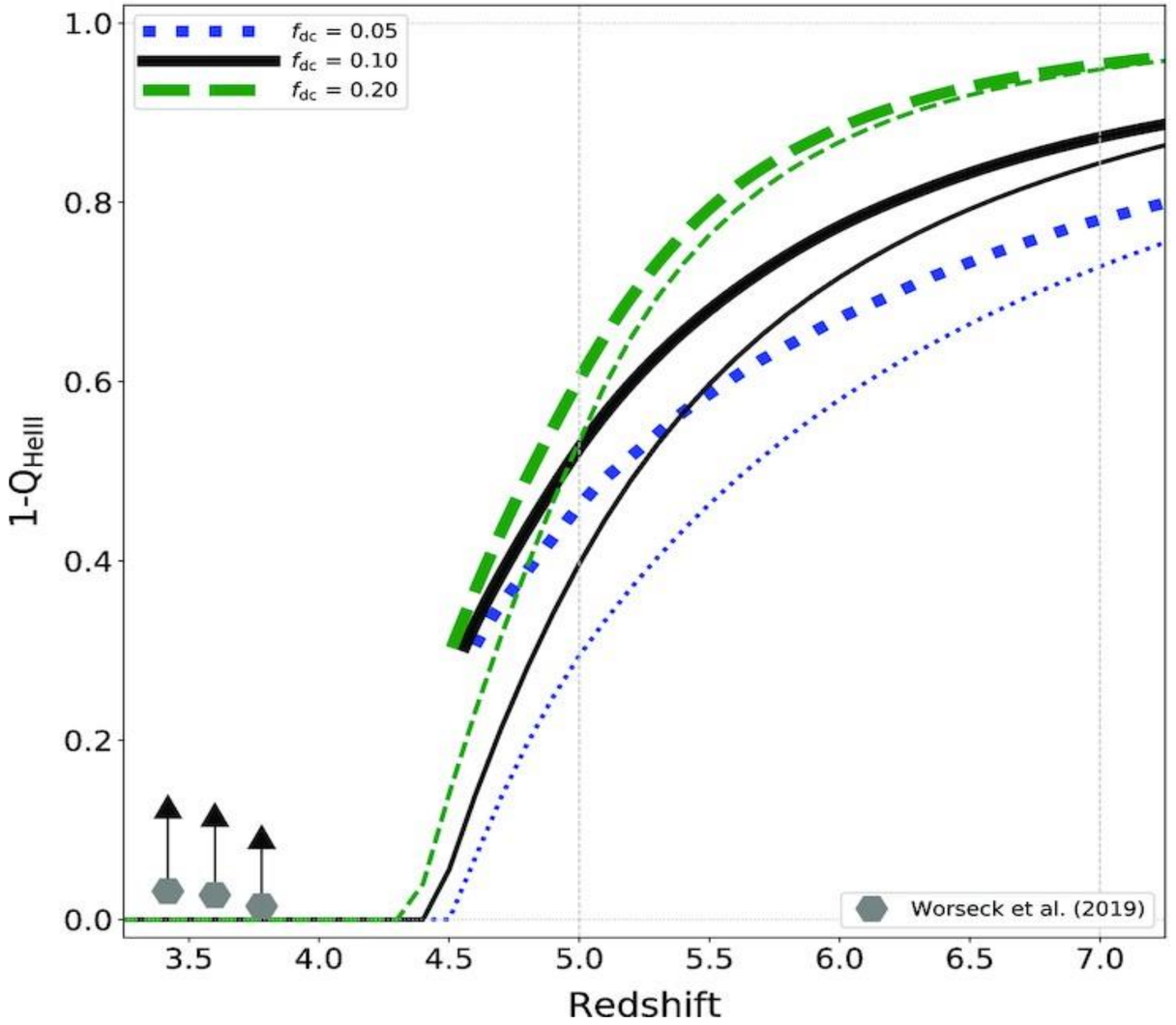
The AGN contribution to the Photoionization rate at z~5



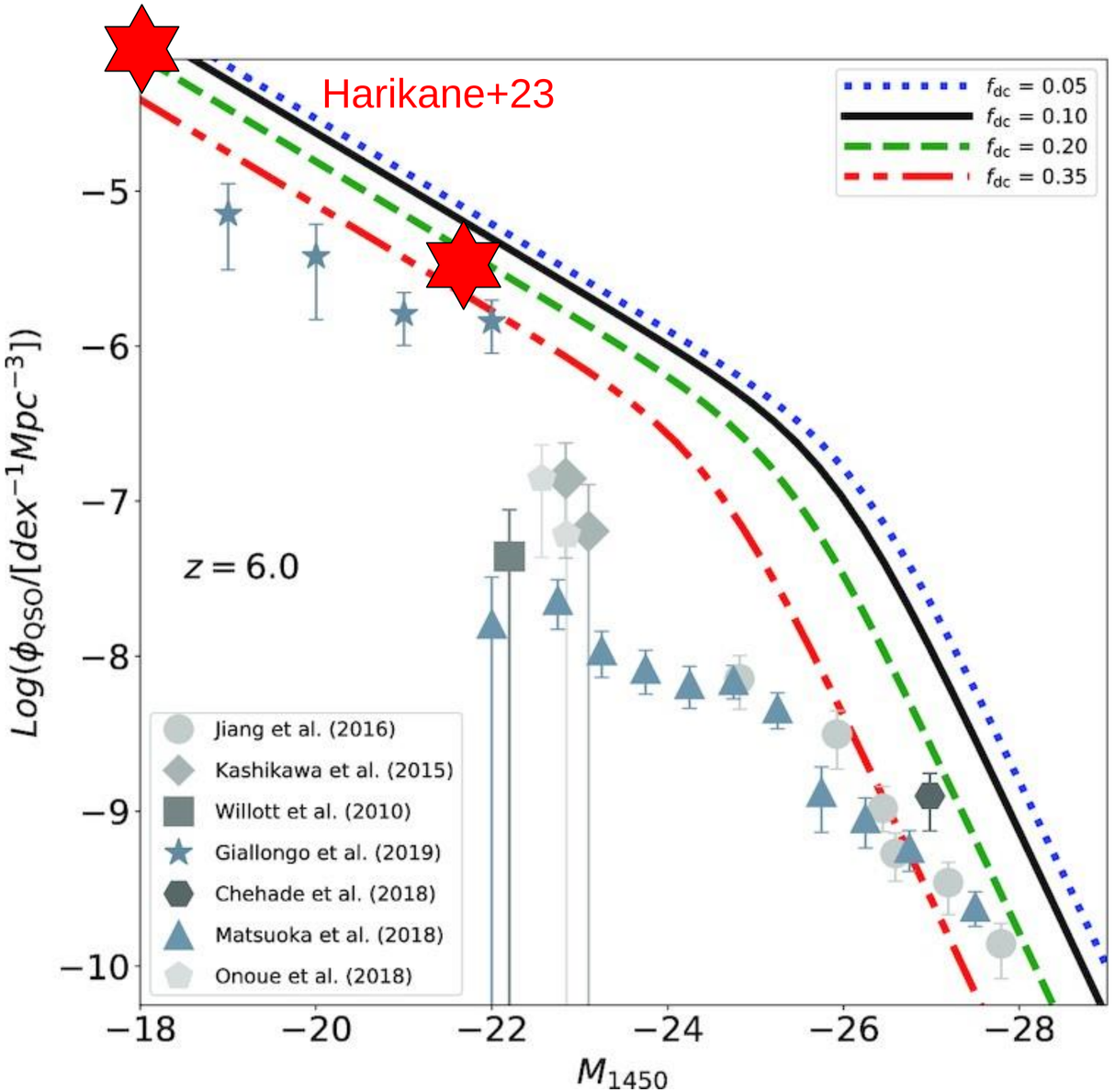
See also Madau+2024

Fontanot+2023: HI and HeII neutral fraction





Estimated z = 6 AGN LF. The predictions from Fontanot+23 models compared with observations



INTERGALACTIC MEDIUM

Trost et al. 2025 PhRvD, 111, 083034

2410.02858

The Intergalactic Medium as a particle detector

The IGM = all baryons not in galaxies

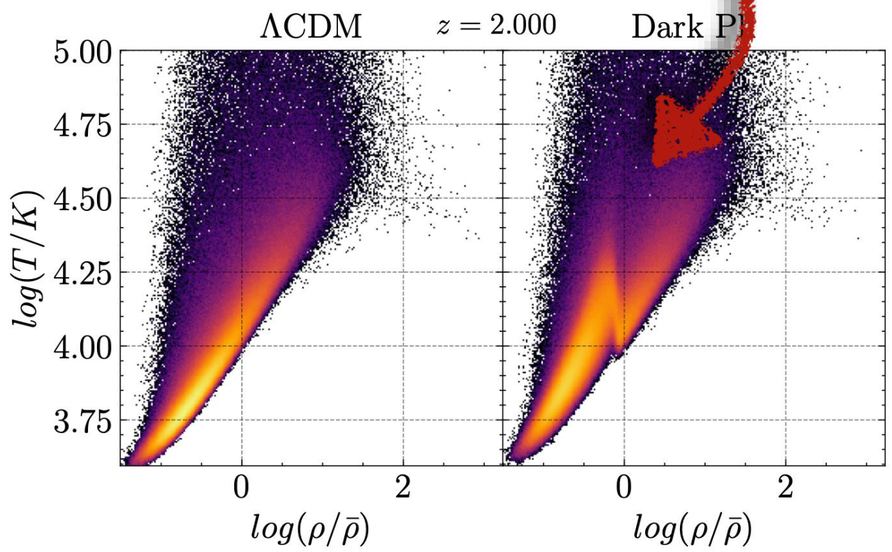
Mostly ionised hydrogen

Dark Photon

When resonance occurs $m_\gamma^2(n_e) = m_{A'}^2$

Production of γ free-free absorbed by gas

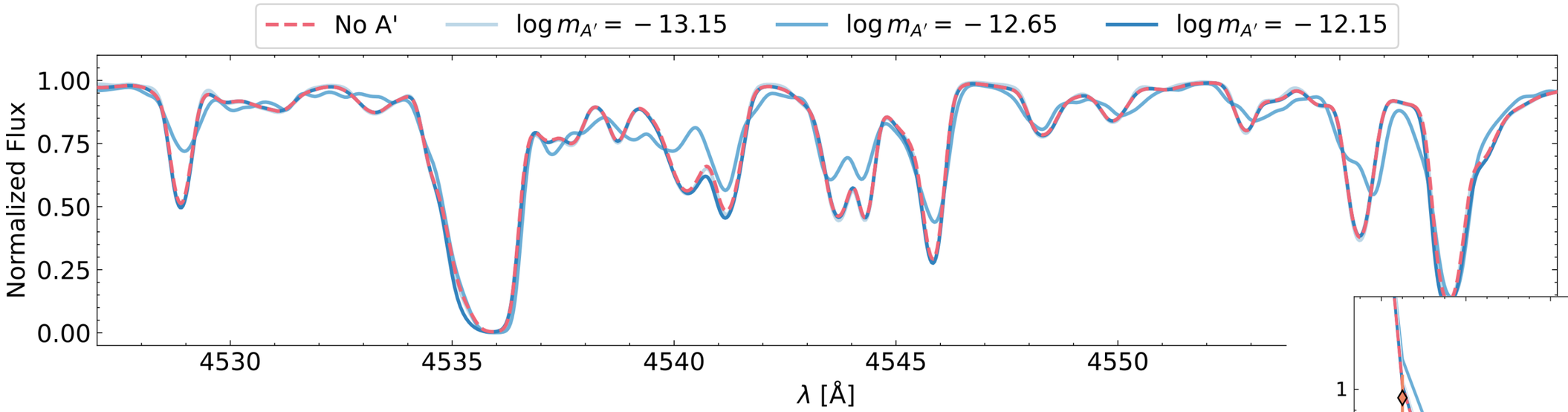
Net energy injection in IGM $E_{A' \to \gamma}$

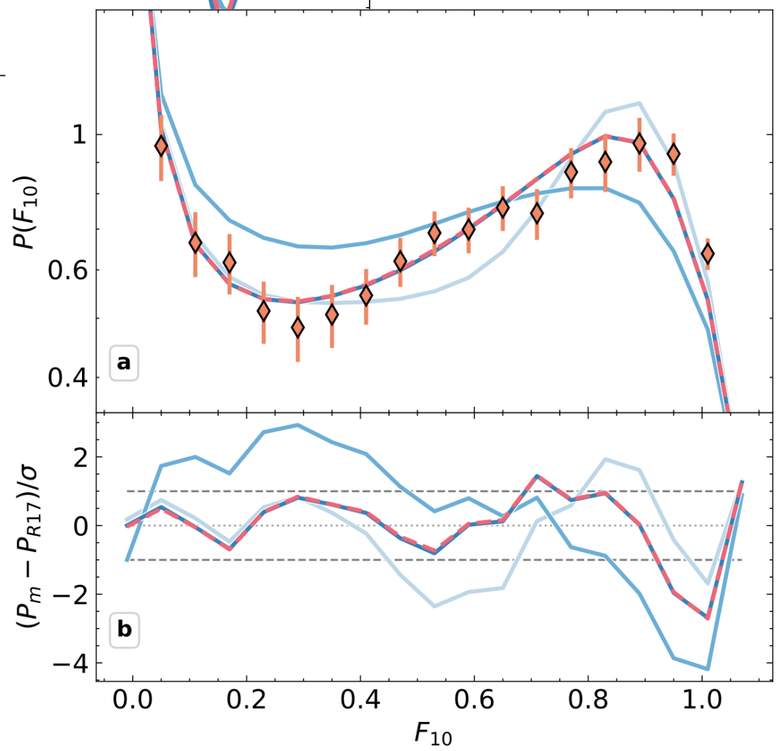


Andrea Trost — andrea.trost@inaf.it — OATs

A' in the Lyman Forest

2410.02858





Spectral lines are sensitive to gas temperature

A' deforms the
Transformed and regulated flux PDF

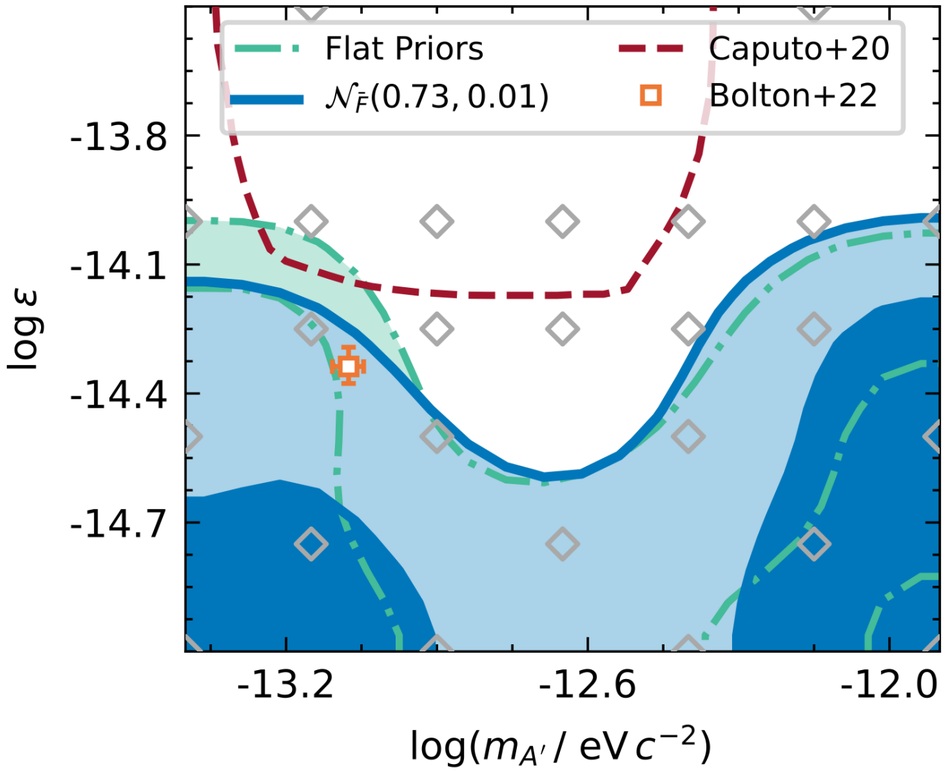
Comparing simulations to high-signal-to-noise data constraints A' parameters

Andrea Trost — andrea.trost@inaf.it — OATs

Constraining A'

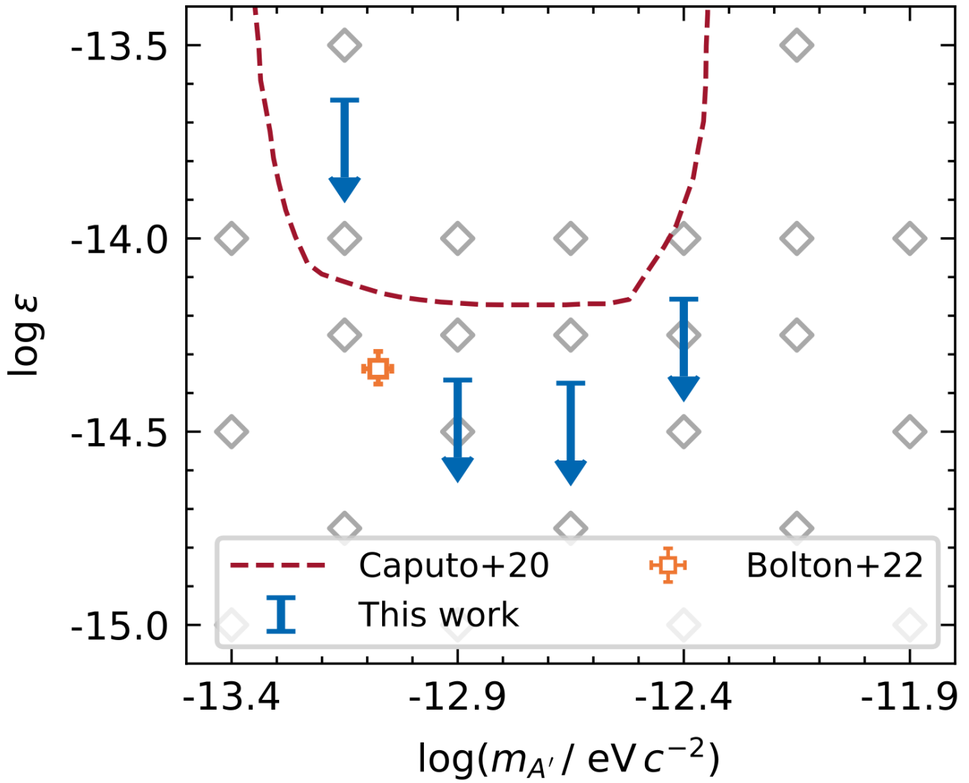
2410.02858

Bayesian — looking for preferred models



No extra heating required!!

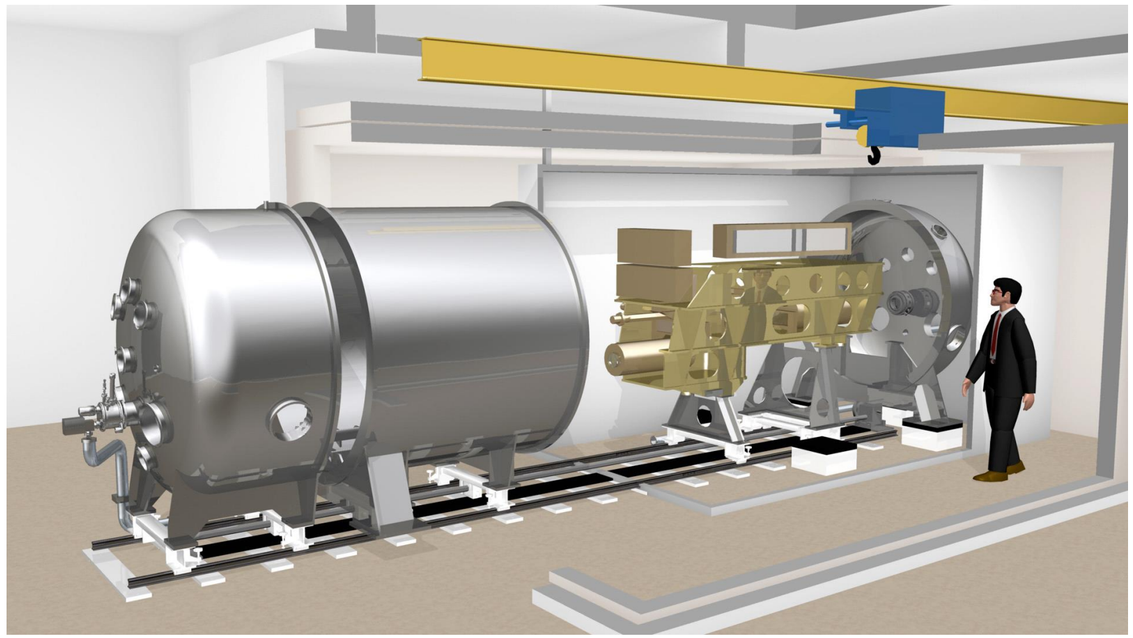
Frequentist — put constraints



New very solid constraints on mass and mixing

Andrea Trost — andrea.trost@inaf.it — OATs

ESPRESSO: designed for stability



ΔRV =1 m/s

ΔT =0.01 K

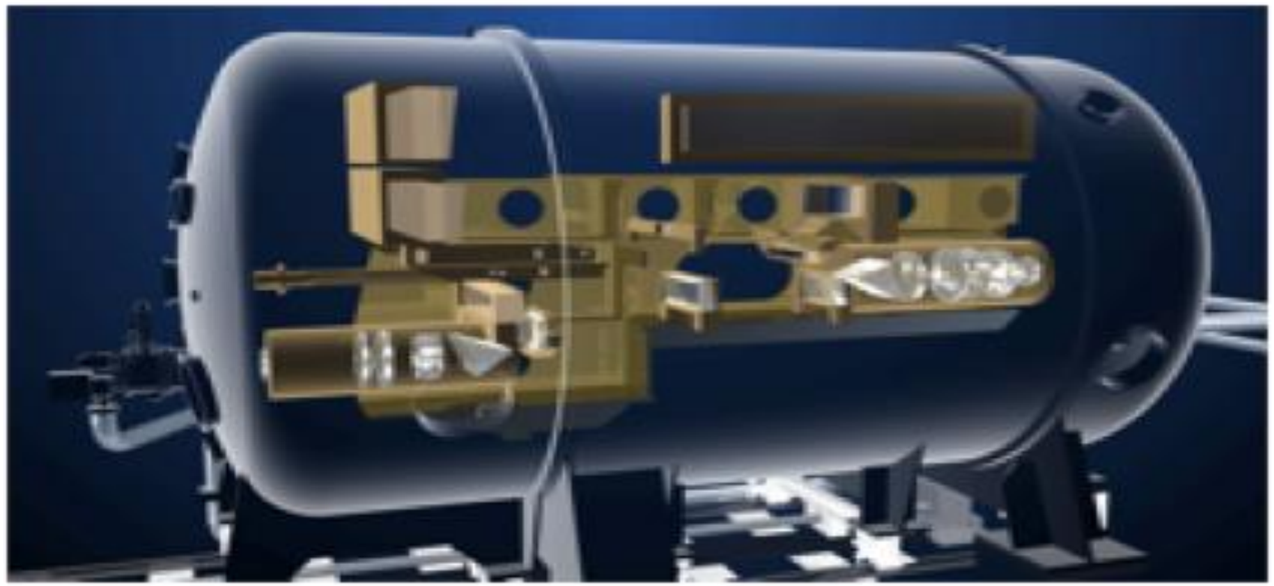
Δp=0.01 mBar

ΔRV =1 m/s

Δλ=0.00001 A

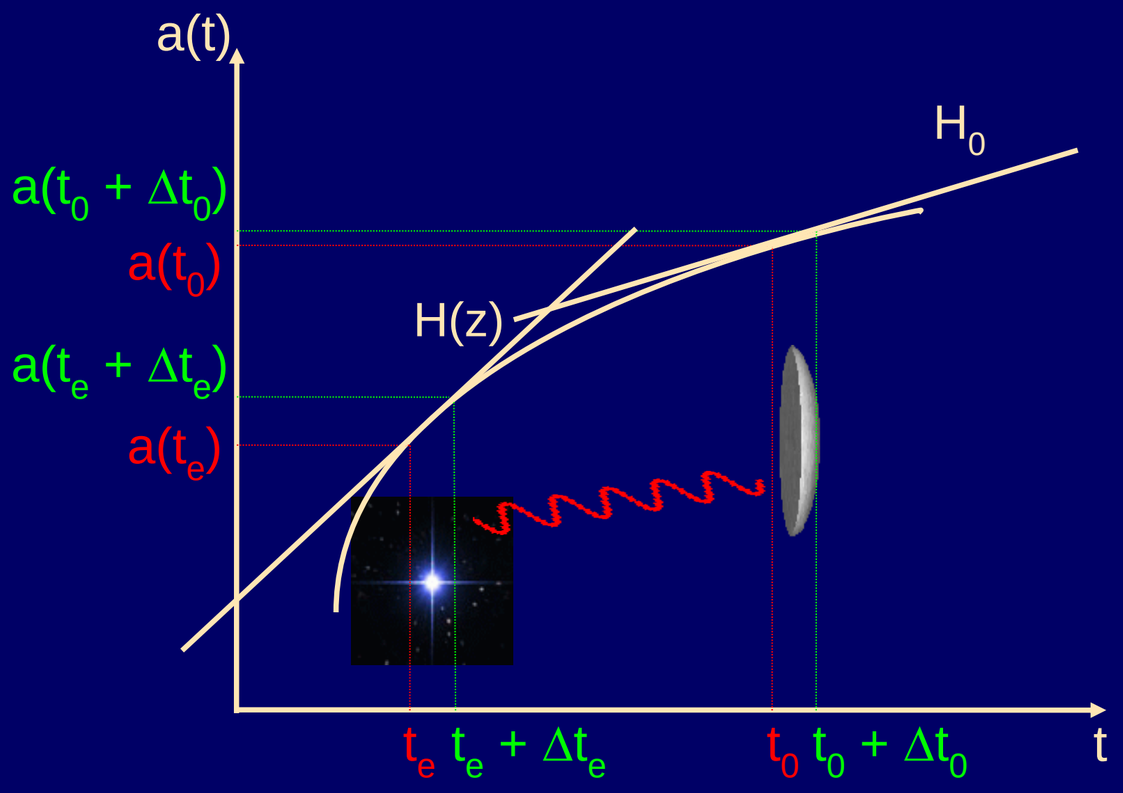
15 nm

1/1000 pixel



The Sandage Test of the Cosmic Redshift Drift

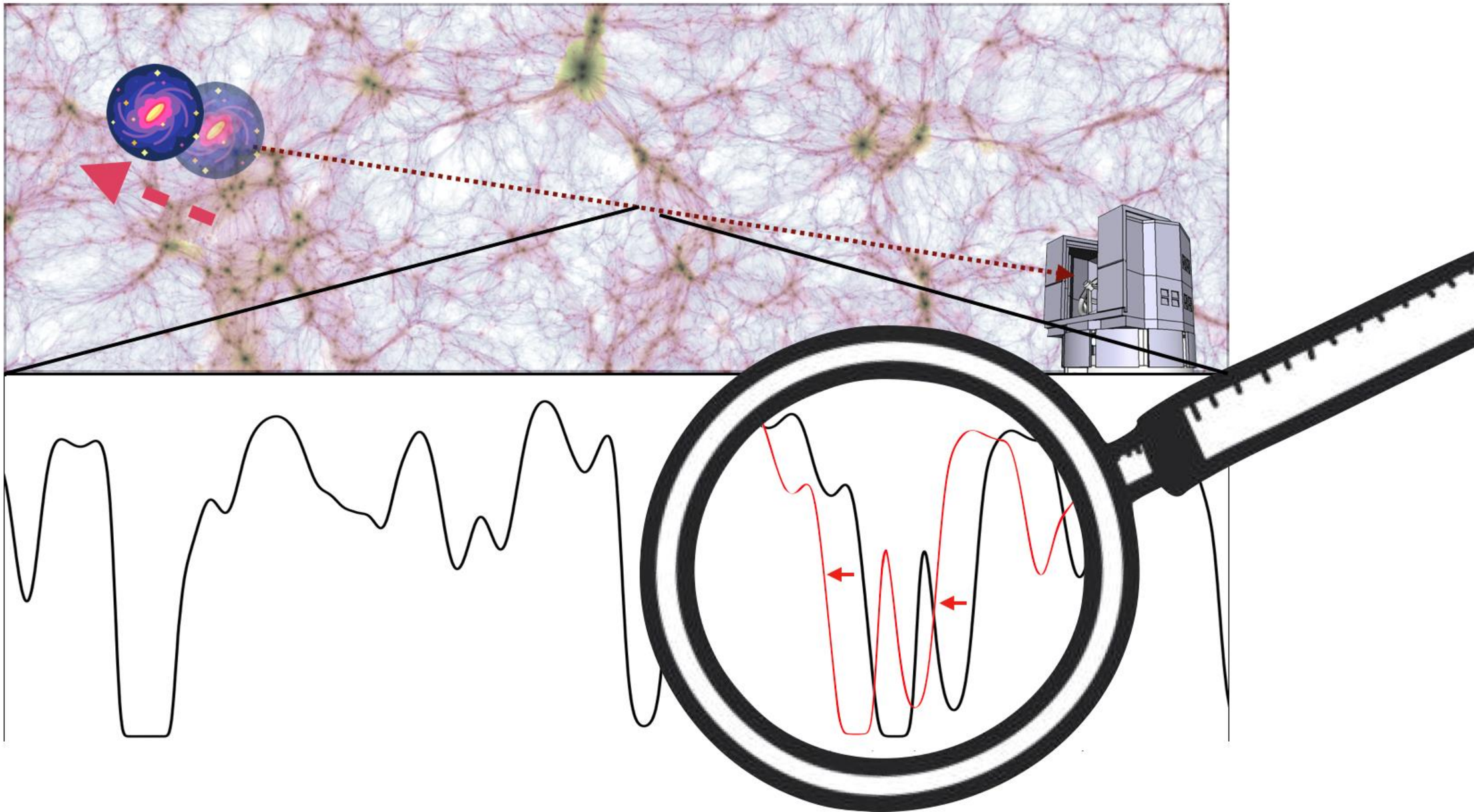
Measuring H(z)



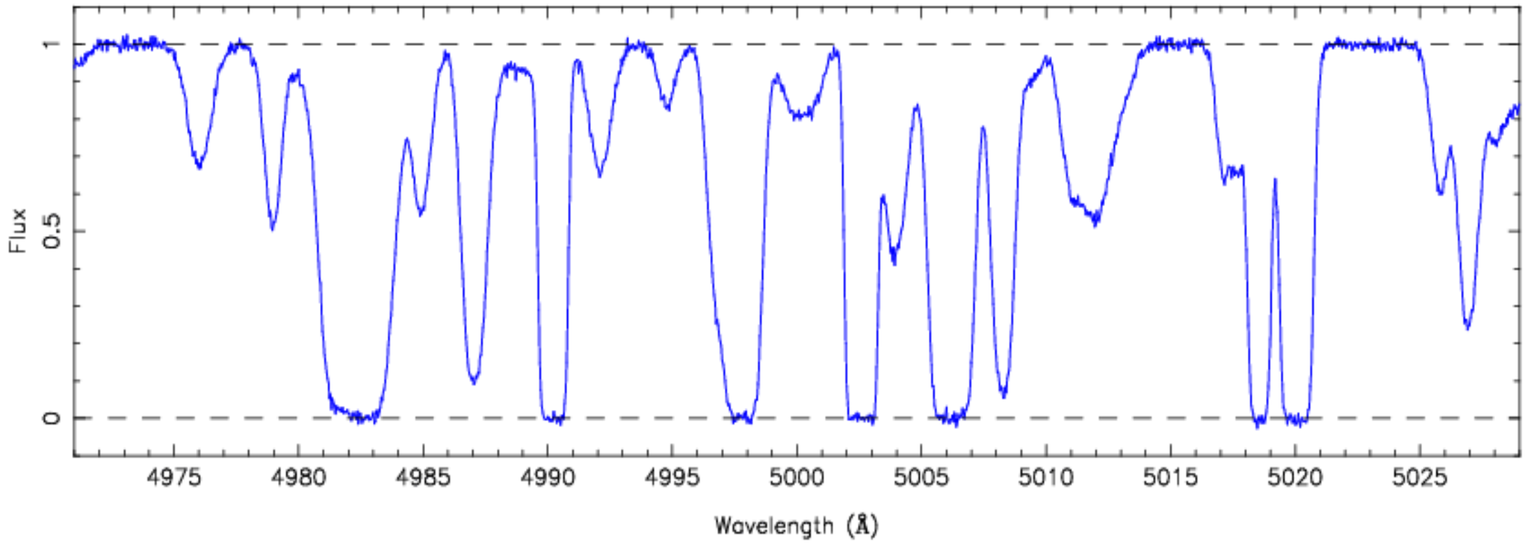
Courtesy J.Liske

$$\frac{z(t_0 + \Delta t_0) - z(t_0)}{\Delta t_0} = \frac{\Delta z}{\Delta t_0} \approx \frac{dz}{dt_0} = (1+z)H_0 - H(z)$$

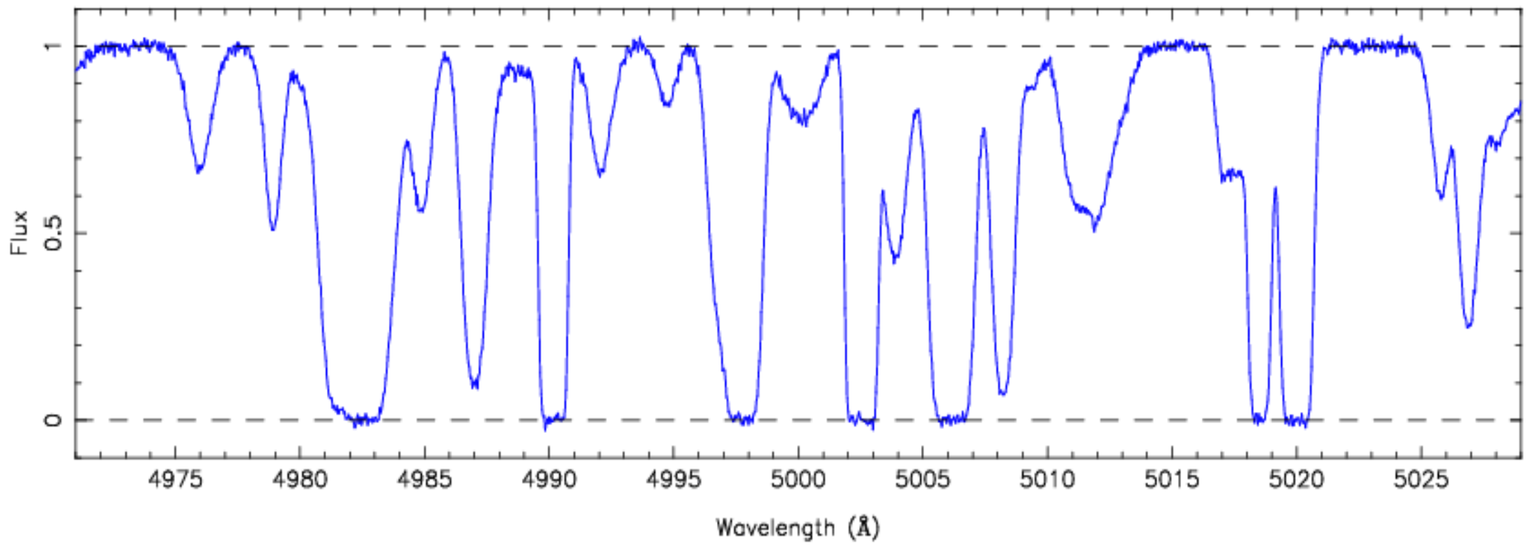
The Lyman Forest - today and … years after



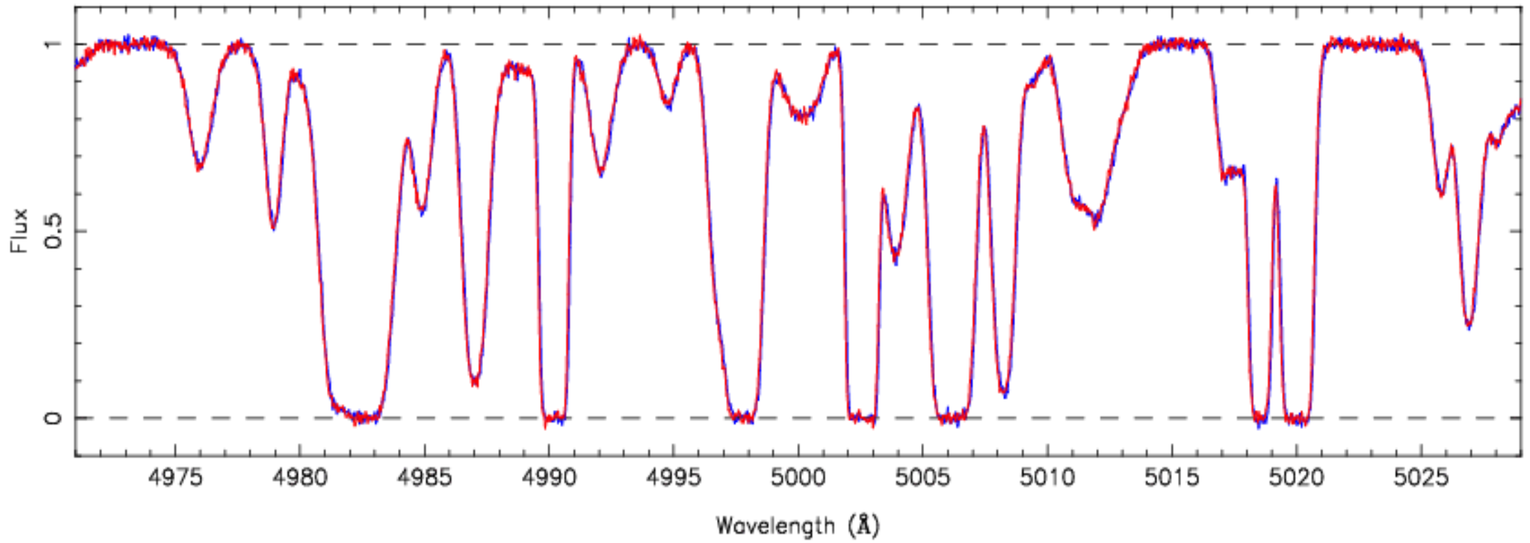
Observing dz/dt in the Ly-α Forest



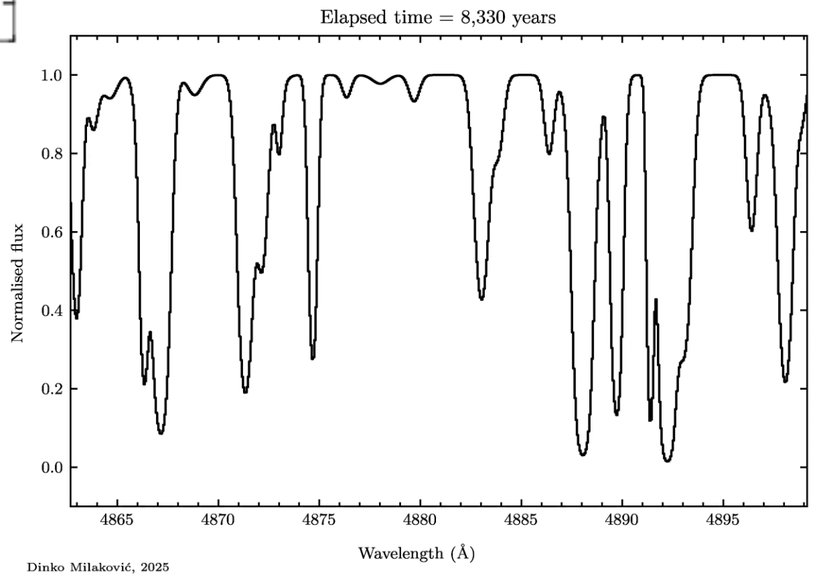
Observing dz/dt in the Ly-α Forest



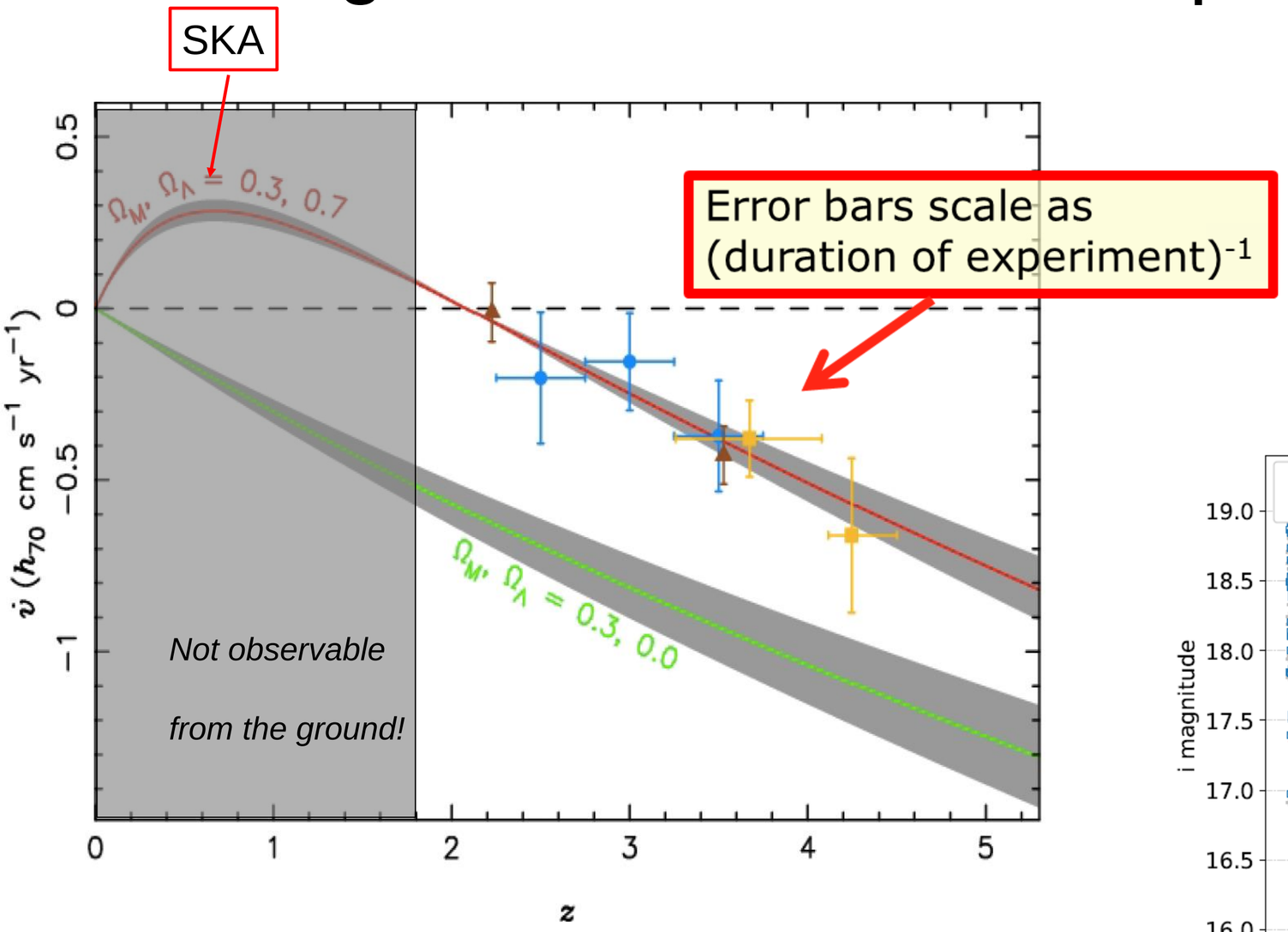
Observing dz/dt in the Ly-α Forest



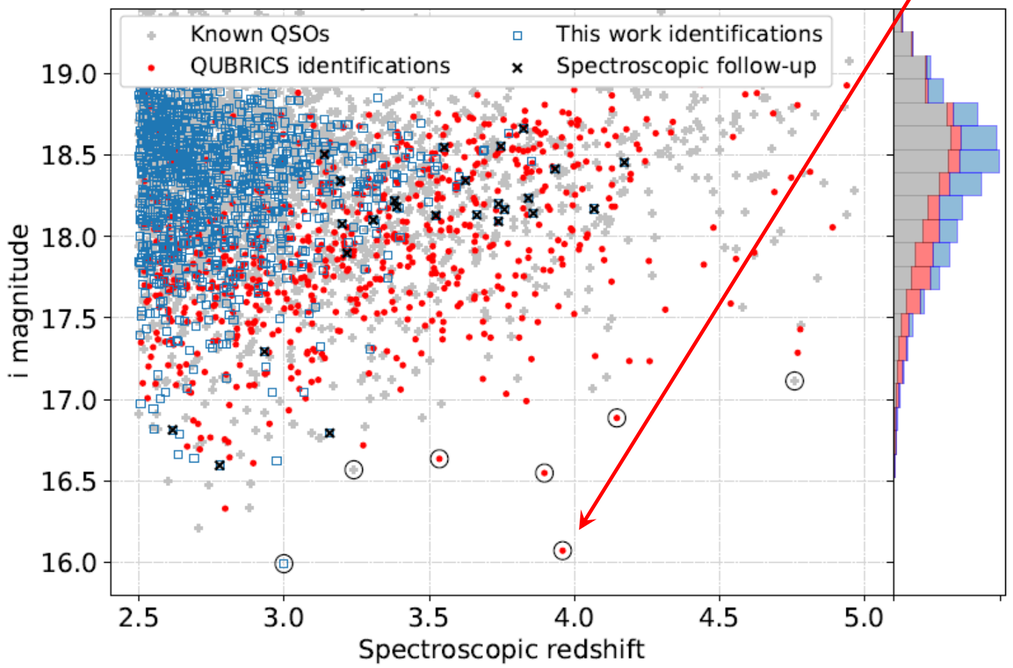
$\Delta t = 10^6$ years!



Sandage Test with a Rs~10^5 spectrograph at the E-ELT

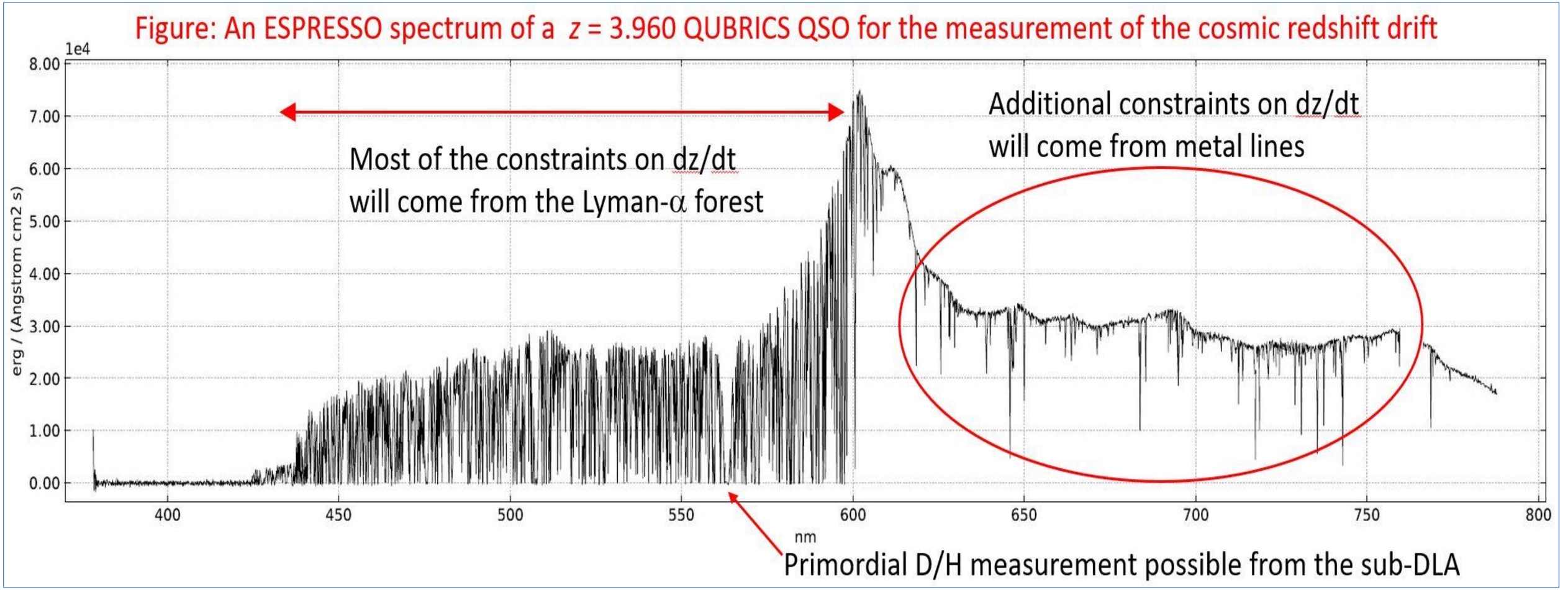


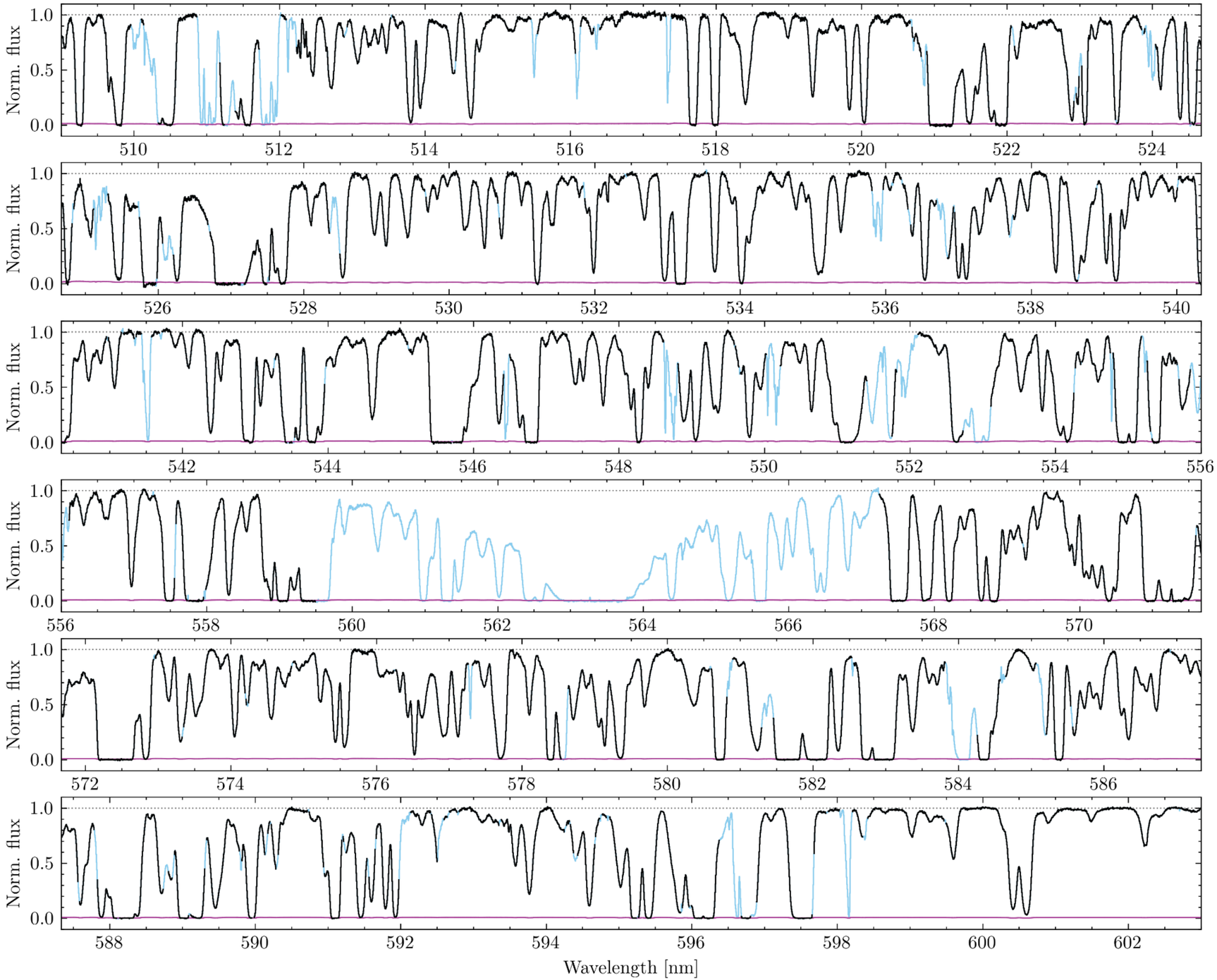
- Different coloured points reflect different targeting strategies
- "**Golden Sample**" of 7 QSOs (Cristiani+2023) from **QUBRICS** Survey (Calderone+2019)



Pasquini et al. 2005, Cristiani et al. 2007, Liske et al. 2008

Sandage test of the cosmic redshift drift

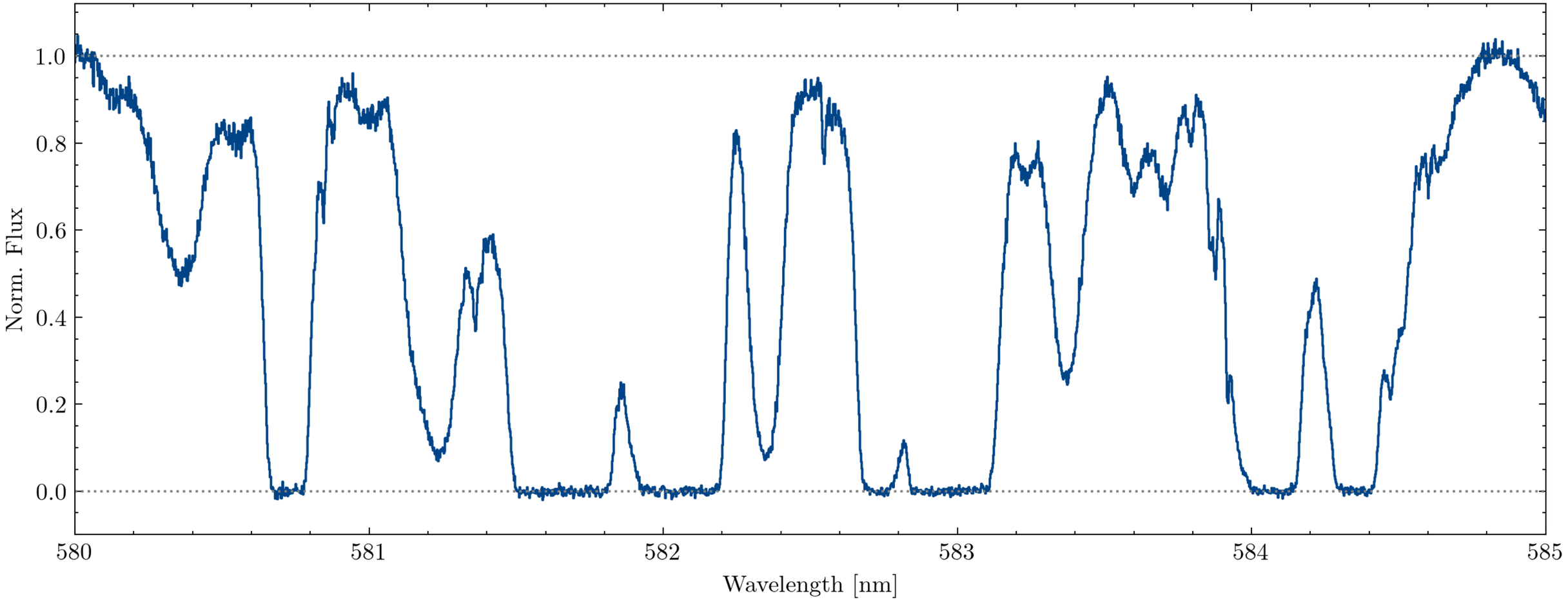




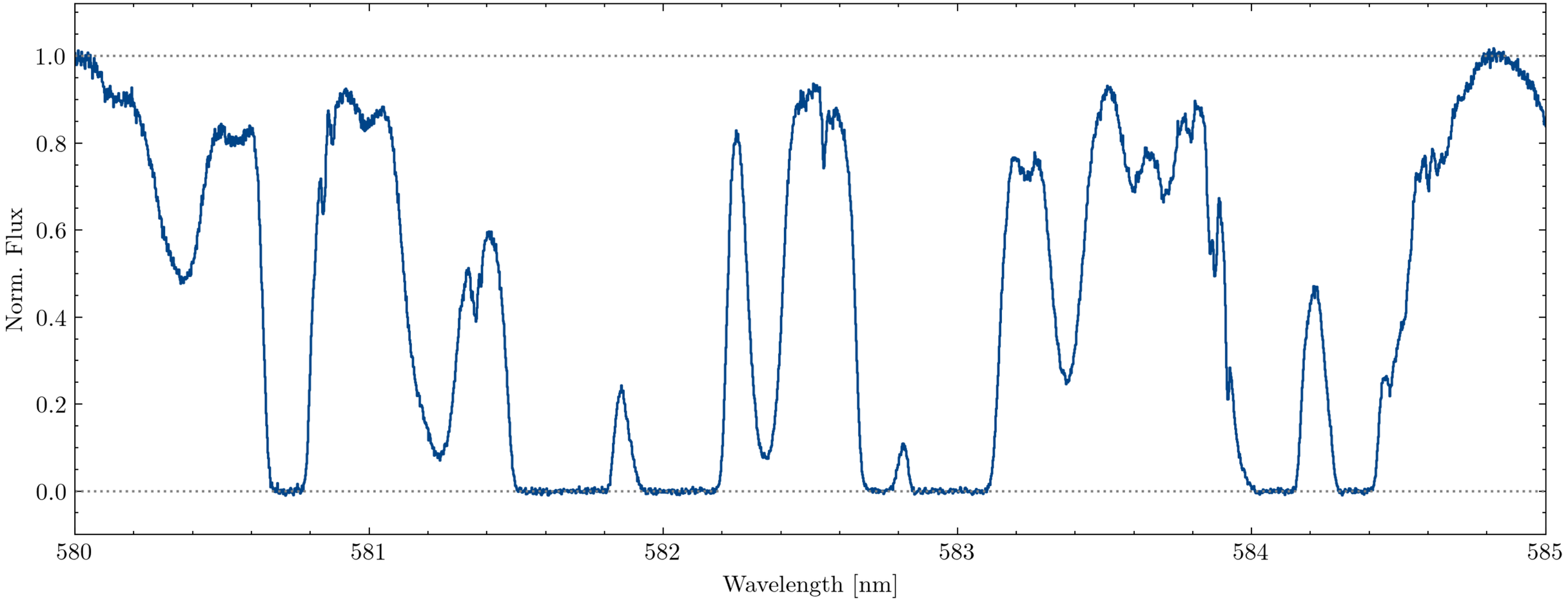
SB-2
Trost et al. 2025

$$\Delta v = -1.25^{+4.44}_{-4.46}\ \mathrm{m\,s^{-1}}$$
equivalent to
$$\dot{v} = -1.43^{+5.08}_{-5.10}\ \mathrm{m\,s^{-1}\,yr^{-1}}$$
or
$$\dot{z} = -2.19^{+7.75}_{-7.78} \times 10^{-8}\ \mathrm{yr^{-1}}$$

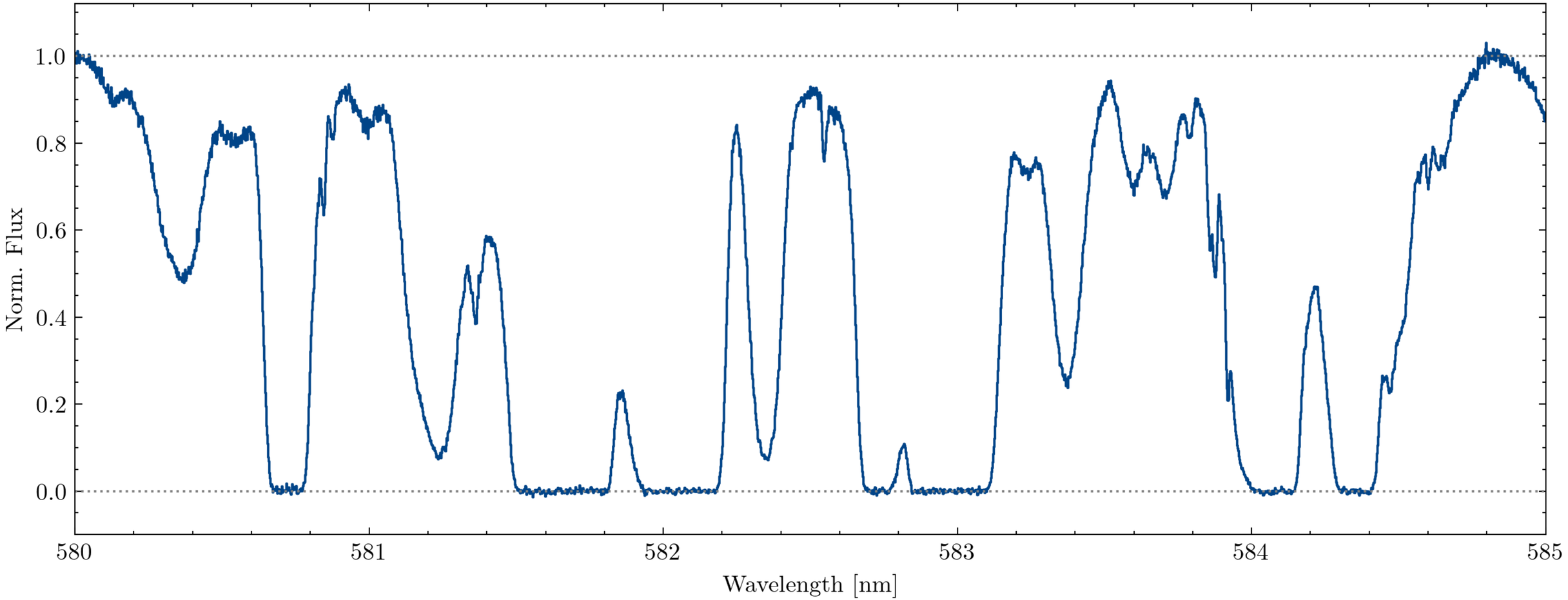
Epoch 1 - 2022


Norm. Flux
1.0
0.8
0.6
0.4
0.2
0.0
580
581
582
583
584
585
Wavelength [nm]

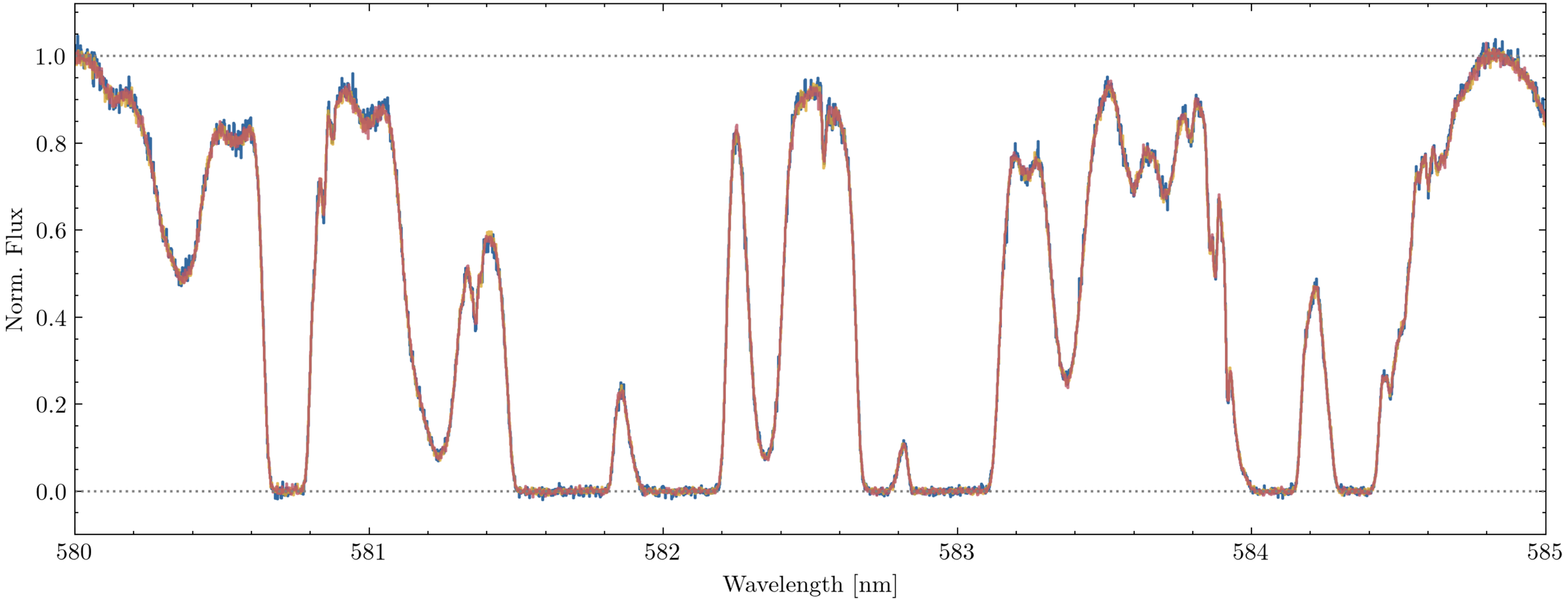
Epoch 2 - 2023


1.0
0.8
0.6
0.4
0.2
0.0
Norm. Flux
580
581
582
583
584
585
Wavelength [nm]

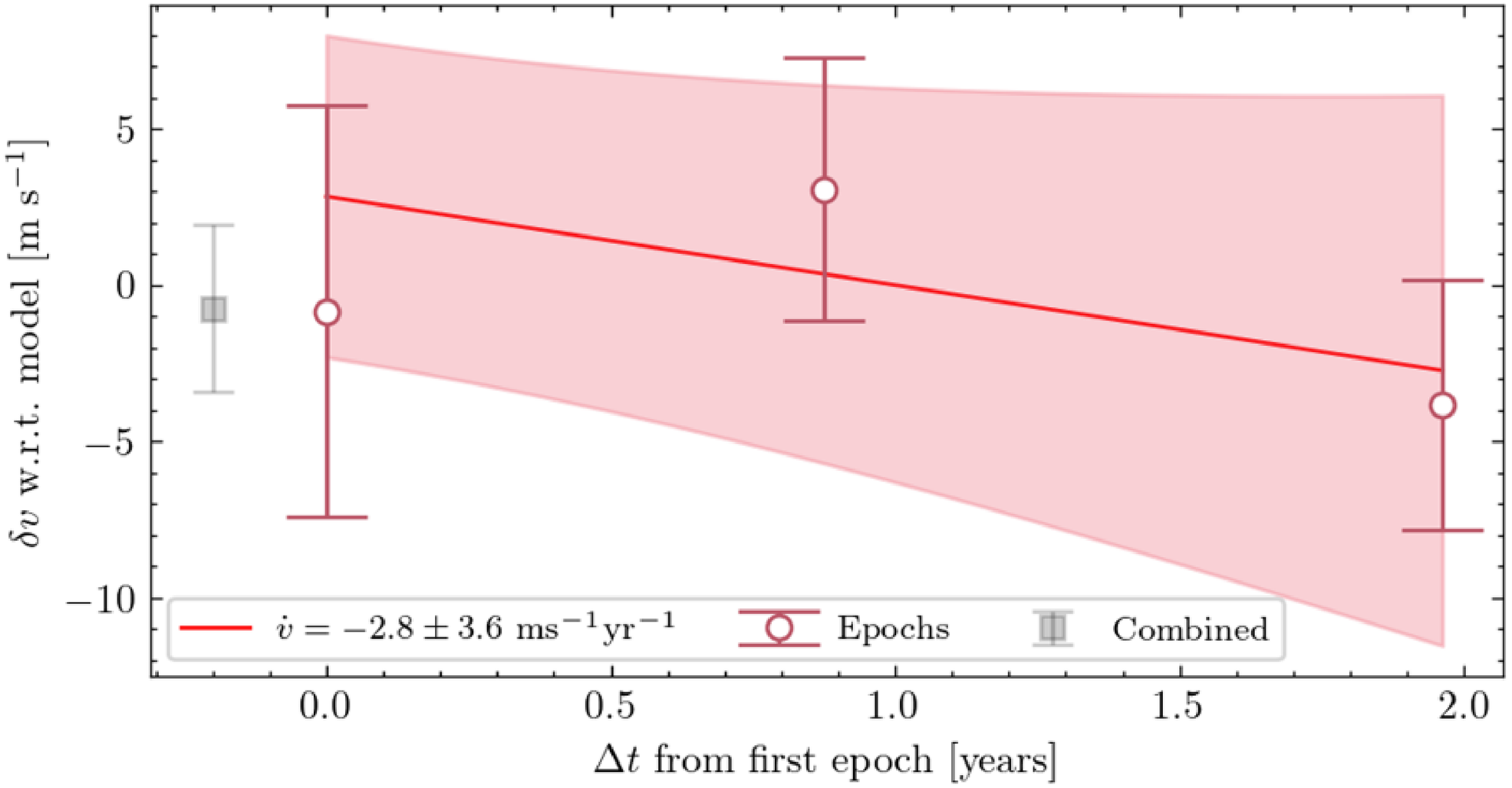
Epoch 3 - 2024


1.0
0.8
0.6
0.4
0.2
0.0
Norm. Flux
580
581
582
583
584
585
Wavelength [nm]

Epoch 1-2-3


1.0
0.8
0.6
0.4
0.2
0.0
Norm. Flux
580
581
582
583
584
585
Wavelength [nm]

SB-2 - Trost et al. in prep

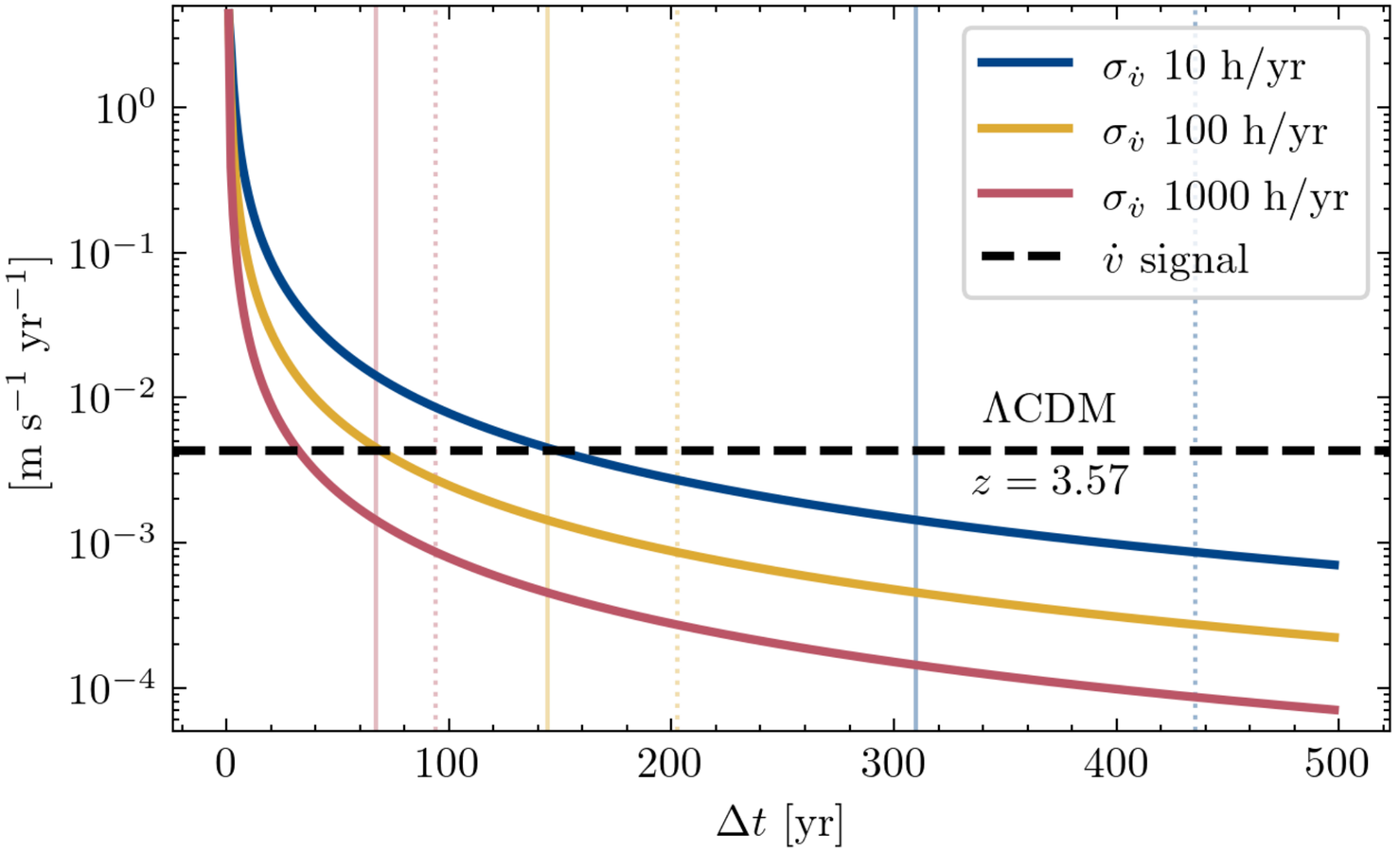


Liske et al. 2008

$$\sigma_v = 2 \left(\frac{\mathrm{S/N}}{2370} \right)^{-1} \left(\frac{N_{\mathrm{QSO}}}{30} \right)^{-1/2} \left(\frac{1 + z_{\mathrm{QSO}}}{5} \right)^{-1.7} \mathrm{cm\,s^{-1}}$$

$$\frac{\mathrm{S}}{\mathrm{N}} = 700 \left[\frac{Z_X}{Z_r} \, 10^{0.4(16 - m_X)} \left(\frac{D}{42\,\mathrm{m}} \right)^2 \frac{t_{\mathrm{int}}}{10\,\mathrm{h}} \frac{\epsilon}{0.25} \right]^{1/2}$$

Velocity shift uncertainty reached with ESPRESSO spectra of SB2, assuming three different observational strategies with an integration time of 10 hours per year (blue), 100 hours per year (yellow), and 1000 hours per year (red).



3σ detection of the redshift drift can be achieved with ESPRESSO(ANDES) after 145 (54) years, assuming 100 hours of observation per year for SB2

$$\Delta t_{3\sigma}^{\mathrm{ANDES}} \approx 54 \left(\frac{T}{100\ \mathrm{h\ yr^{-1}}}\right)^{-1/3} \left(\frac{\epsilon}{0.1}\right)^{-1/3} \mathrm{yr}$$

SB-2
Trost et al. 2025


QUBRICS
Gotta catch 'em all!
QUASARS
AS BRIGHT BEACONS IN THE SOUTHERN SKY
NTT
TNG
LBT
STICKER ALBUM

THE
END