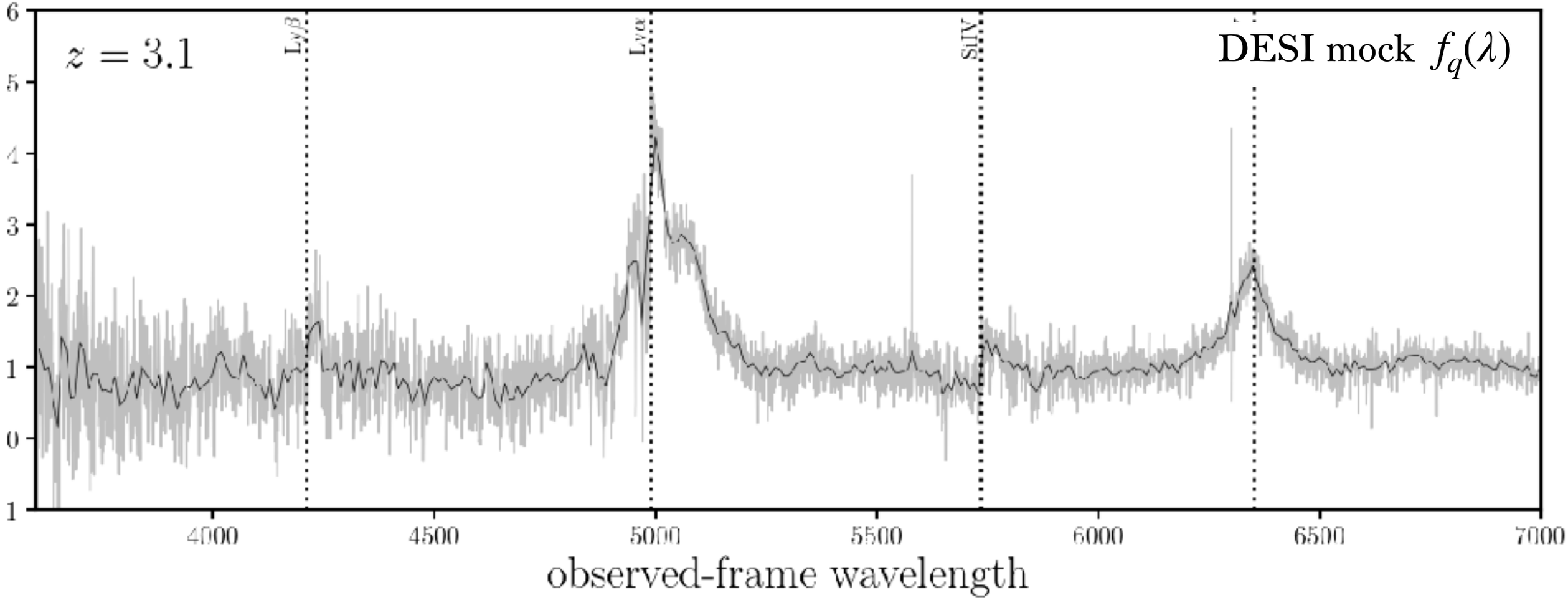

# Reconstructing Quasar Spectra and measuring the Ly$\alpha$ Forest

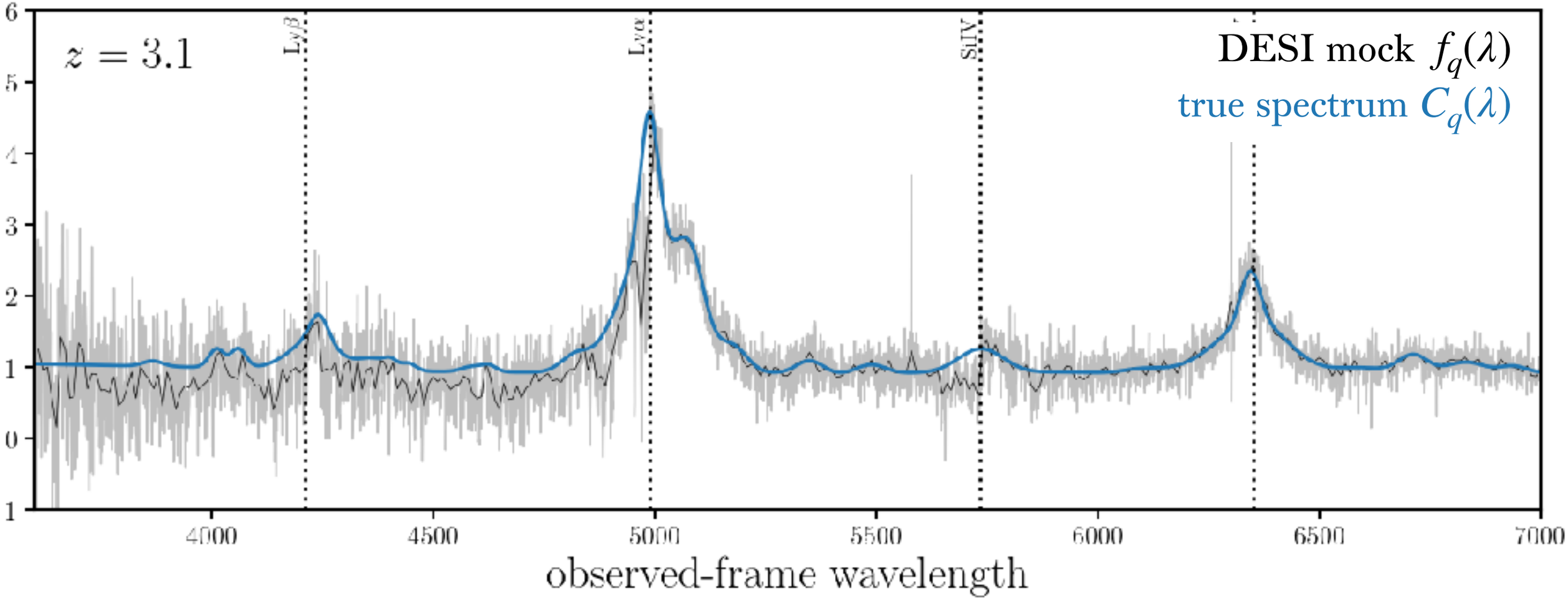
Changhoon Hahn

*with* Satya Gontcho A Gontcho, Peter Melchior, Abby Bault, Hiram Herrera

*GQP WG telecon — Nov 8, 2024*

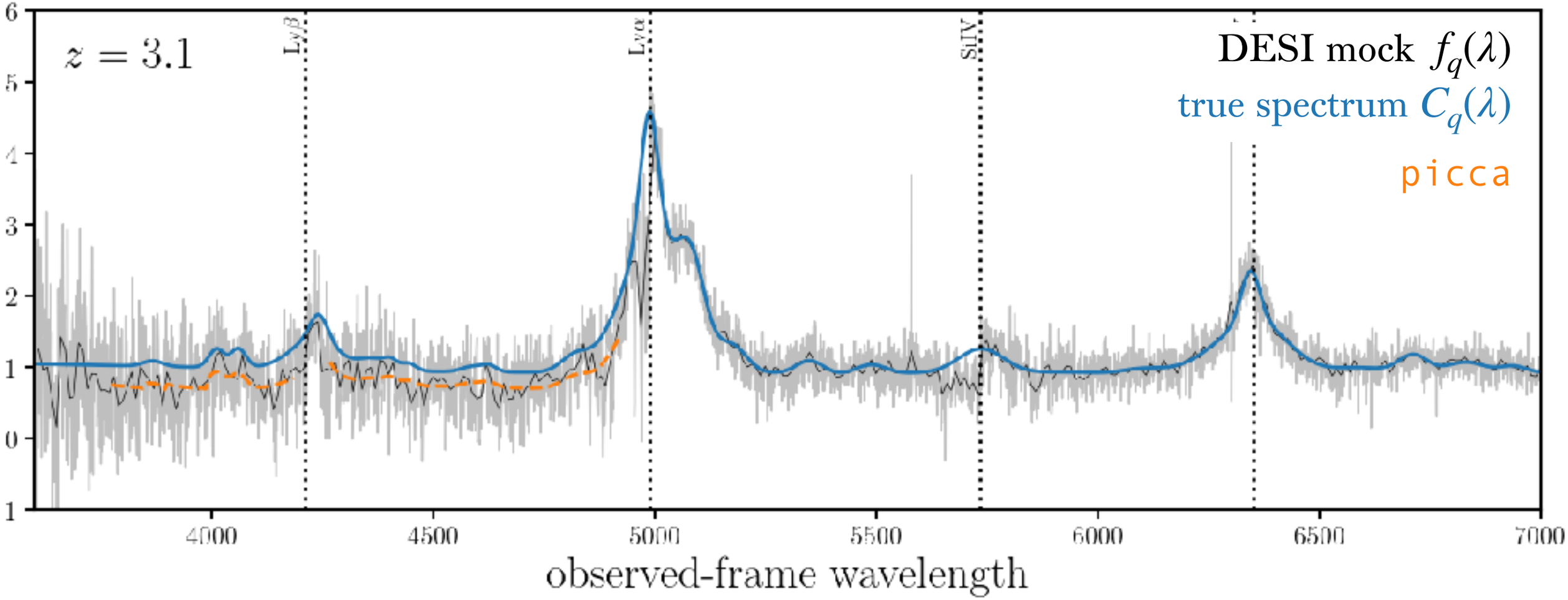



quasar spectra carry the imprint of foreground IGM (*e.g.* Ly$\alpha$ forest)



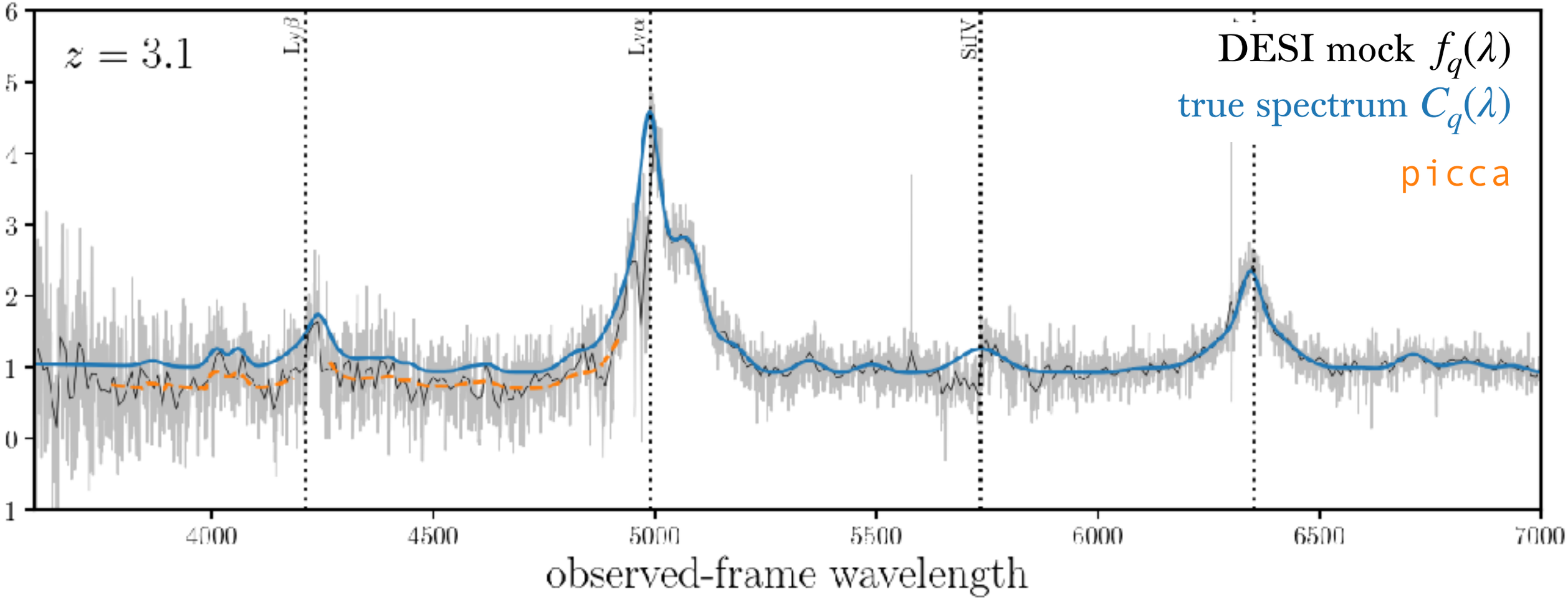

quasar spectra carry the imprint of foreground IGM (*e.g.* Ly$\alpha$ forest)

but requires estimating the true continuum $C_q(\lambda)$

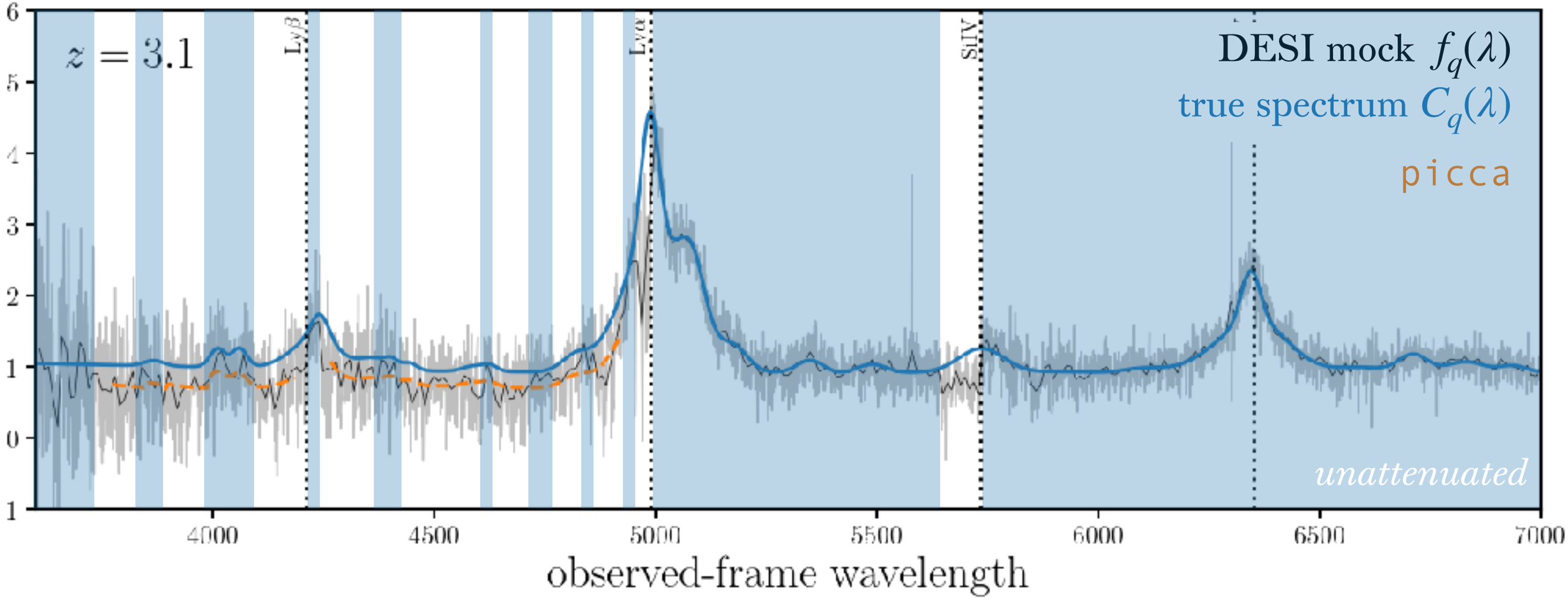



current state-of-the-art `picca` is *biased*

*du Mas des Bourboux et al. (2020)*
*https://github.com/igmhub/picca/*

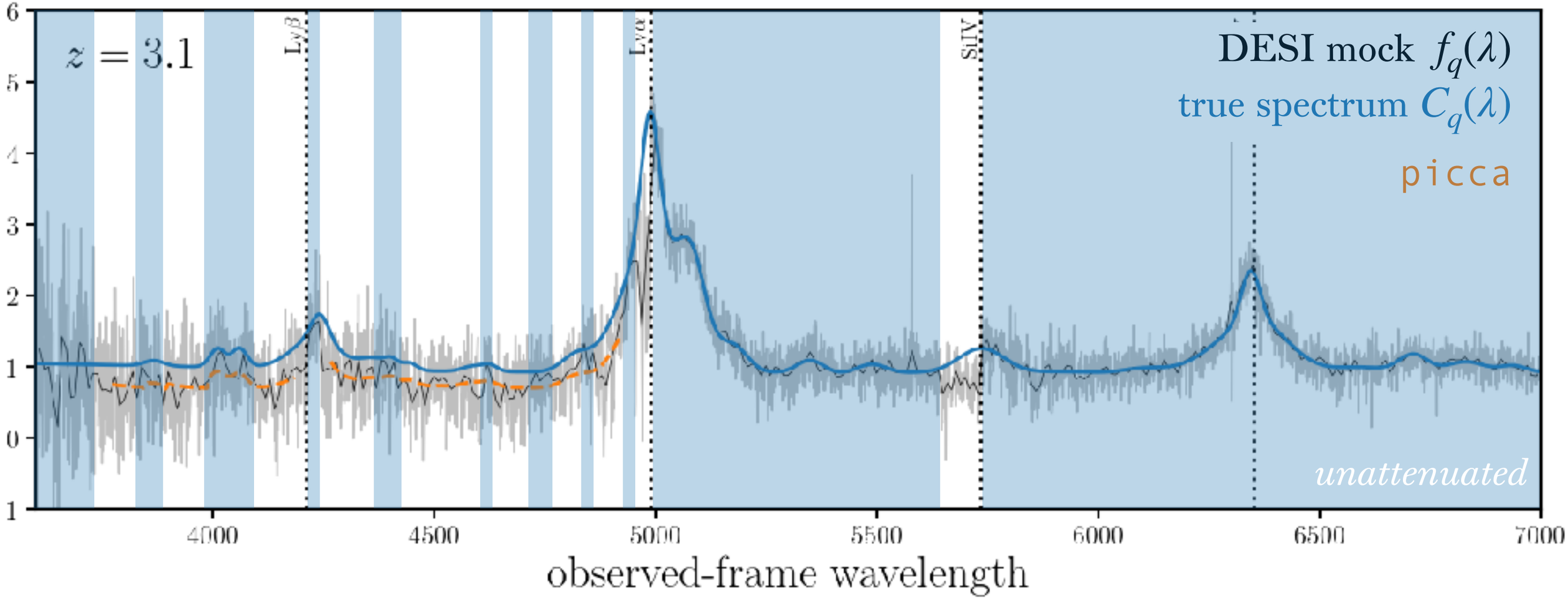



could we *reconstruct* the quasar continuum?




could we *reconstruct* the quasar continuum?

*parts of the spectra are unattenuated*

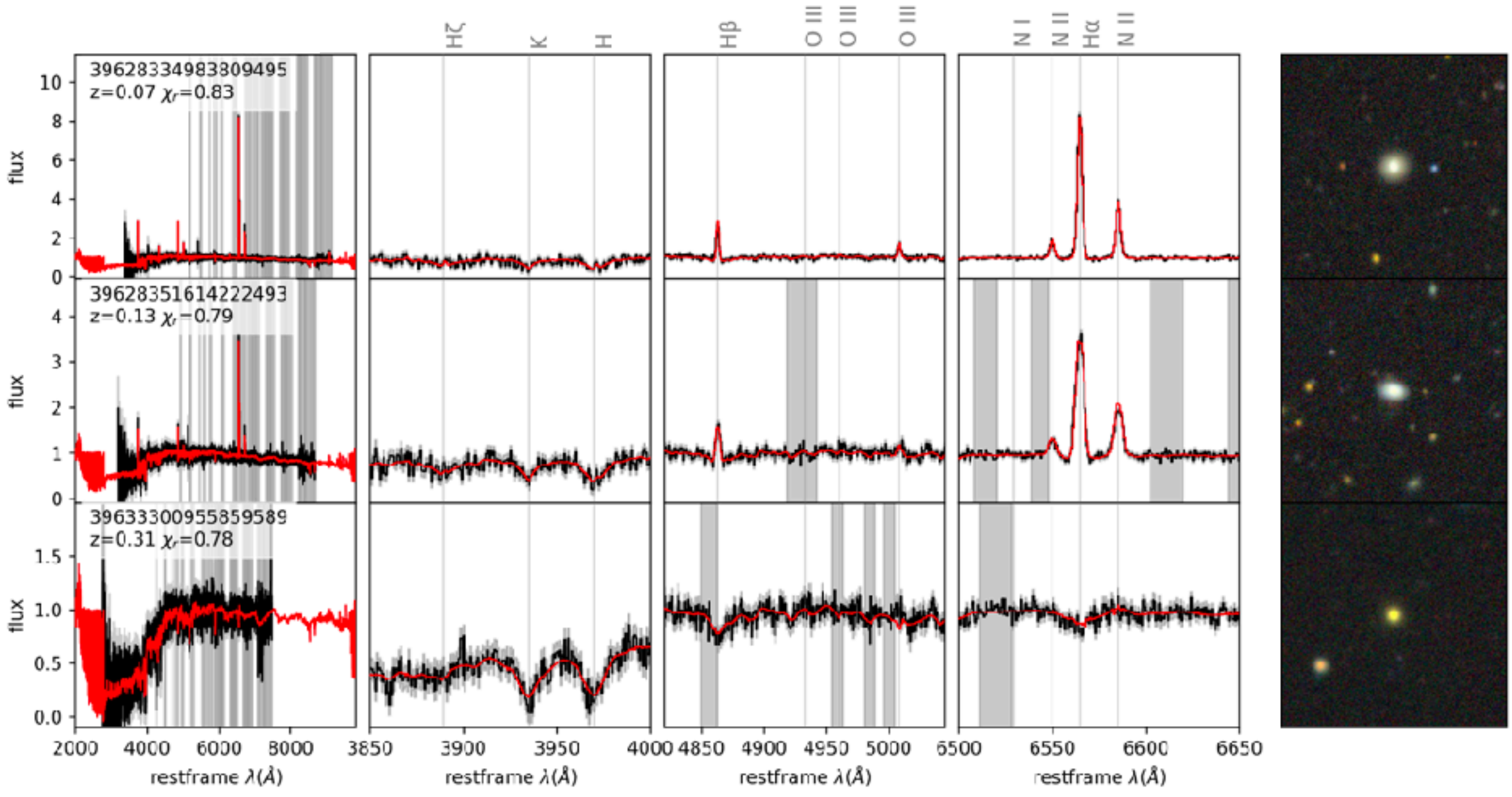



could we *reconstruct* the quasar continuum?

*parts of the spectra are unattenuated*

*foreground absorption is uncorrelated with the quasar*

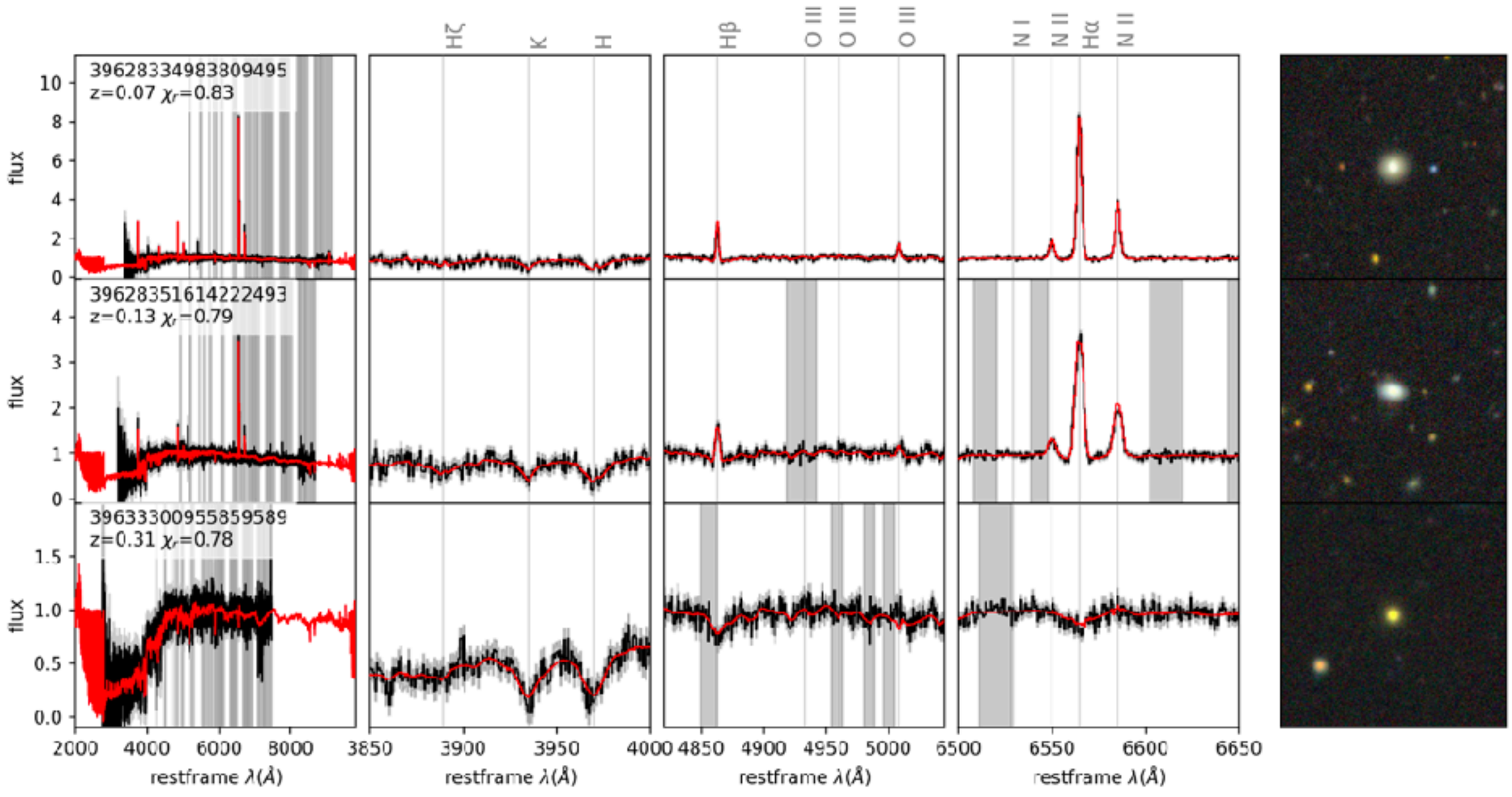
*spectral autoencoders* have been very successful in reconstructing galaxy spectra — *e.g.* **`spender`**




*DESI BGS galaxies*

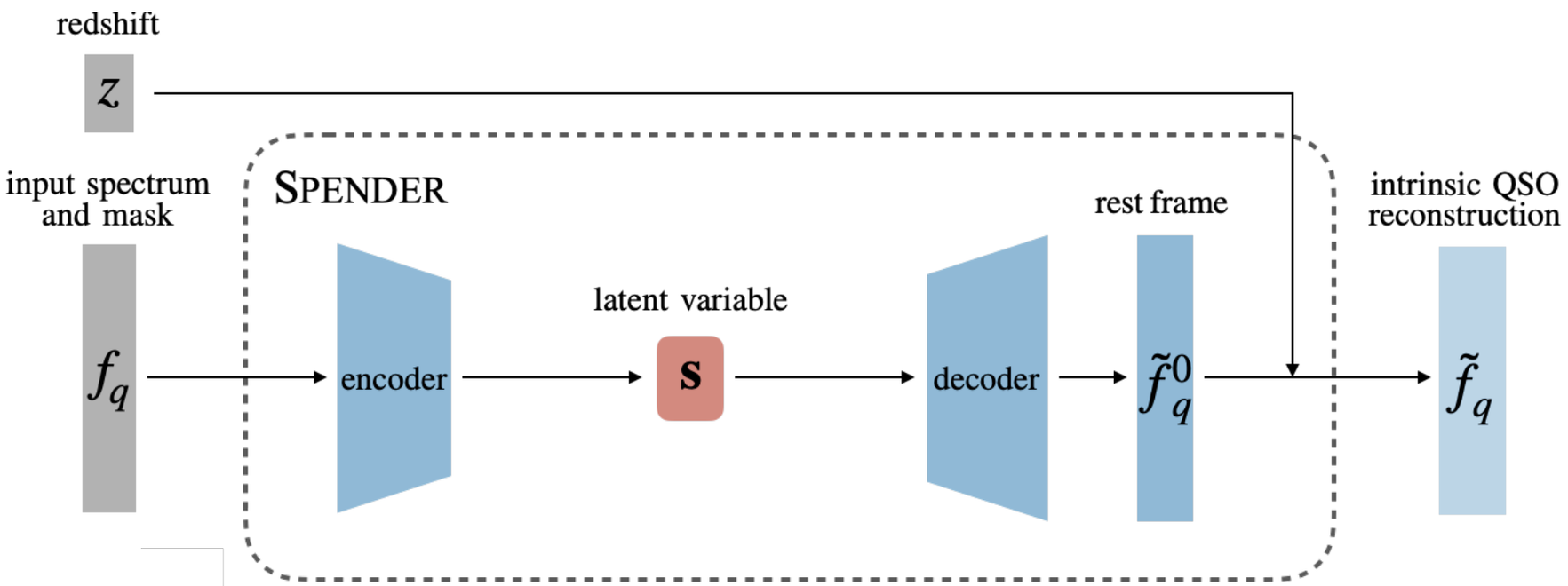
*Melchior+CH+(2022), Liang+CH+(2023)*
*see also Portillo et al.(2020), Teimoorina et al.(2022)*

*spectral autoencoders* have been very successful in reconstructing galaxy spectra — *e.g.* **`spender`**



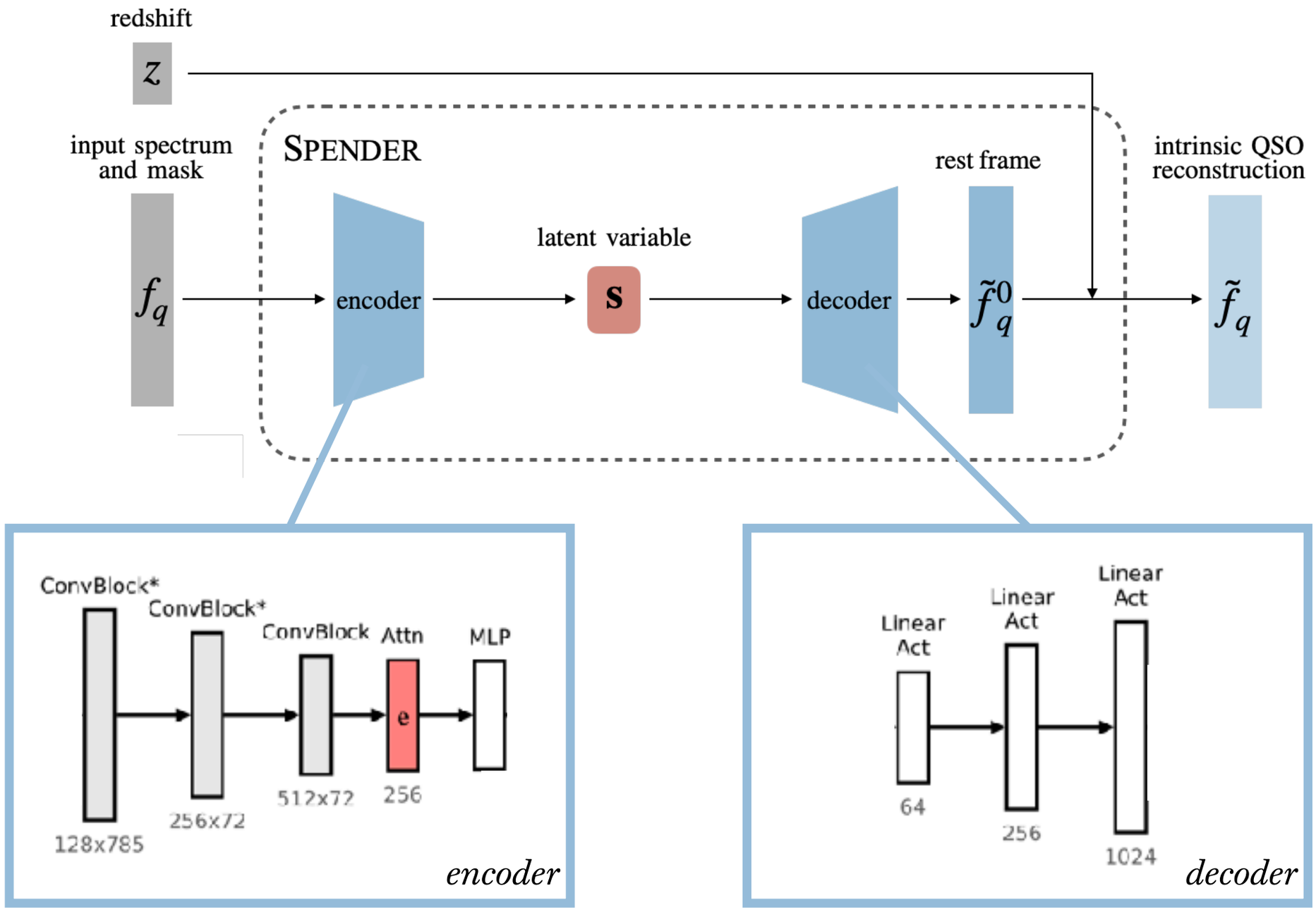

fully *data-driven* (*i.e. no training required on simulations)*

**spender**: spectral autoencoders learns low-dim representation of spectra in the latent space



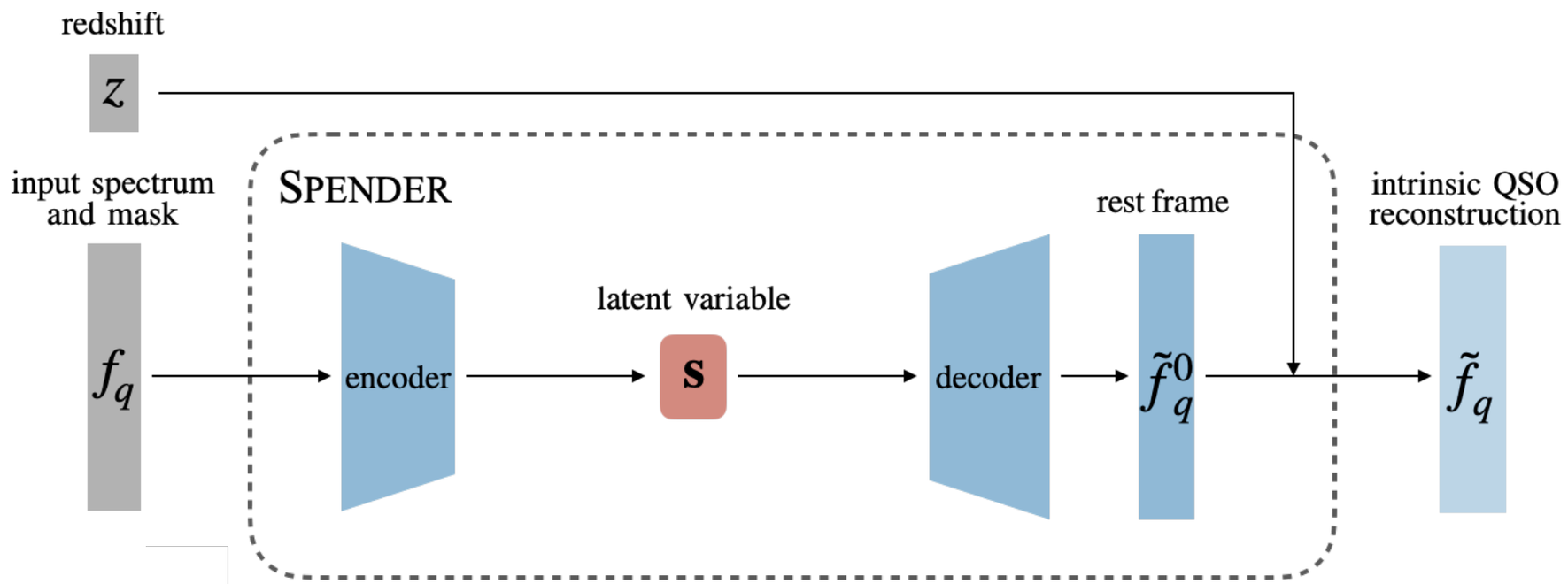

*Hahn, Gontcho A Gontcho, Melchior (2024) NeurIPS ML4PS*

**`spender`**: spectral autoencoder with attentive convolutional encoder and an explicit redshifting




*Hahn, Gontcho A Gontcho, Melchior (2024) NeurIPS ML4PS*

**spender**: spectral autoencoder with fidelity loss + similarity and consistency losses

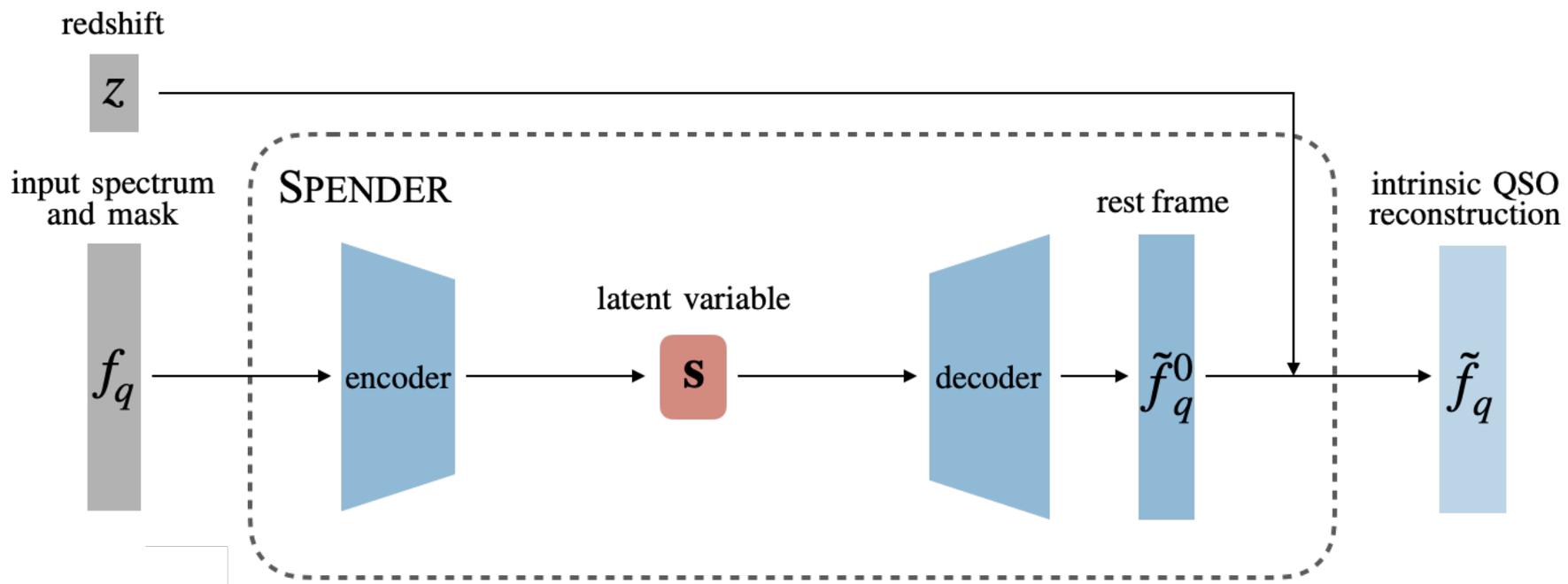



$$L_{\text{total}} = L_{\text{fid}} + L_{\text{sim}} + L_{\text{con}}$$

*latent distance* $\propto$ *spectral distance*

*Melchior+CH+(2022), Liang+CH+(2023)*

**spender**: spectral autoencoder with fidelity loss + similarity and consistency losses



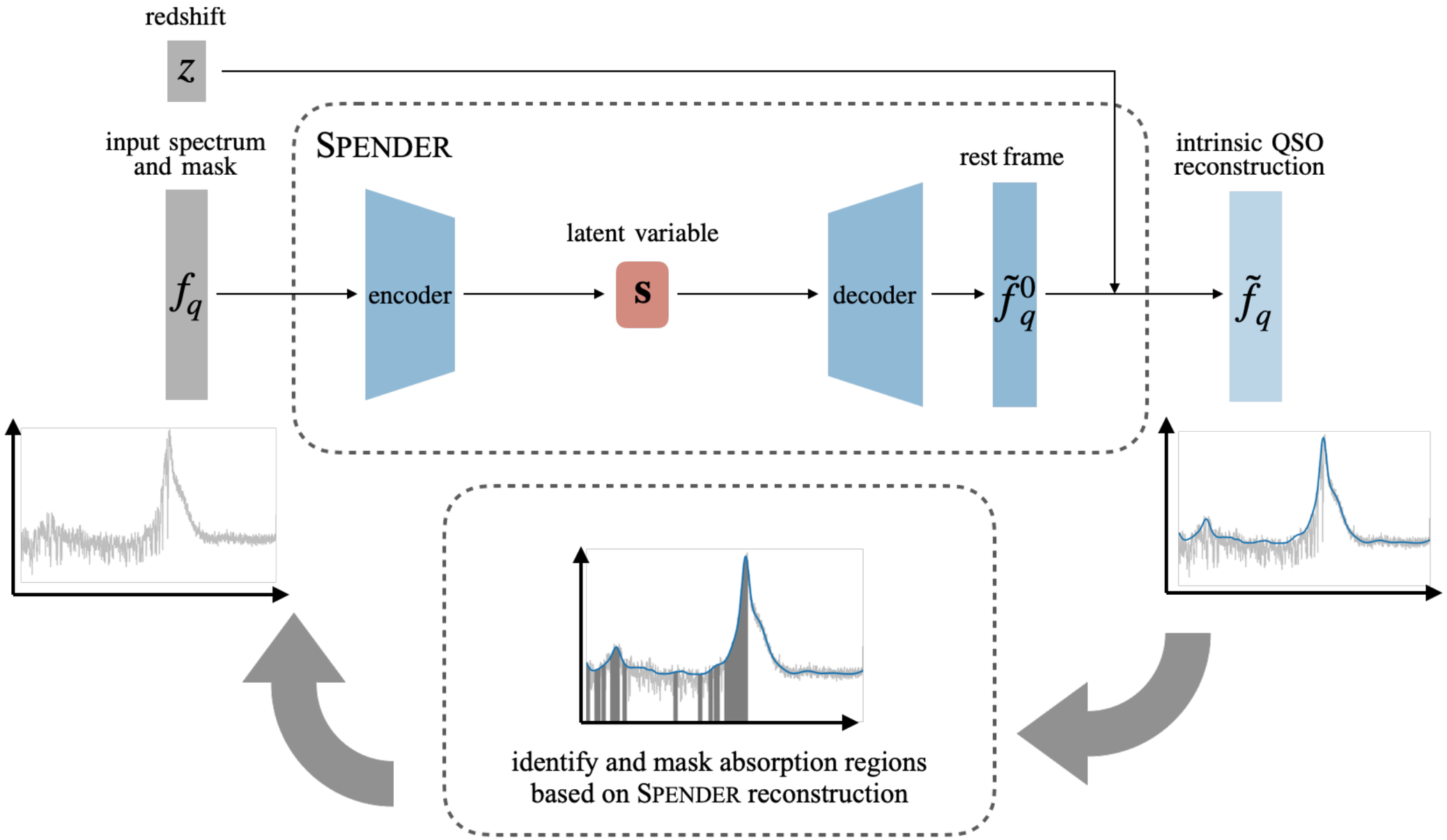

$$L_{\text{total}} = L_{\text{fid}} + L_{\text{sim}} + L_{\text{con}}$$

*redshift invariance*

*Melchior+CH+(2022), Liang+CH+(2023)*

SpenderQ

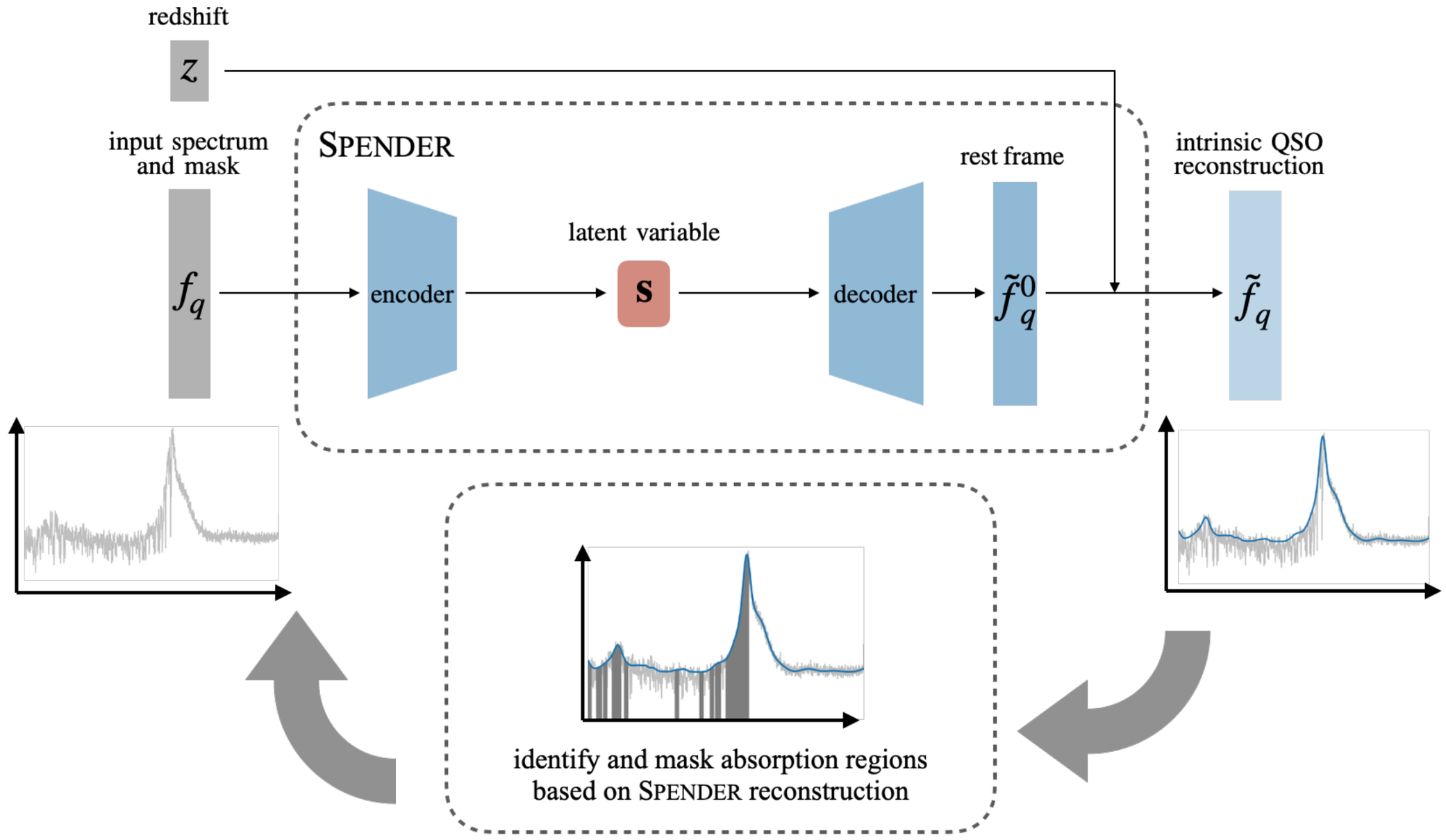
**spender** + iterative procedure for identifying and masking absorption



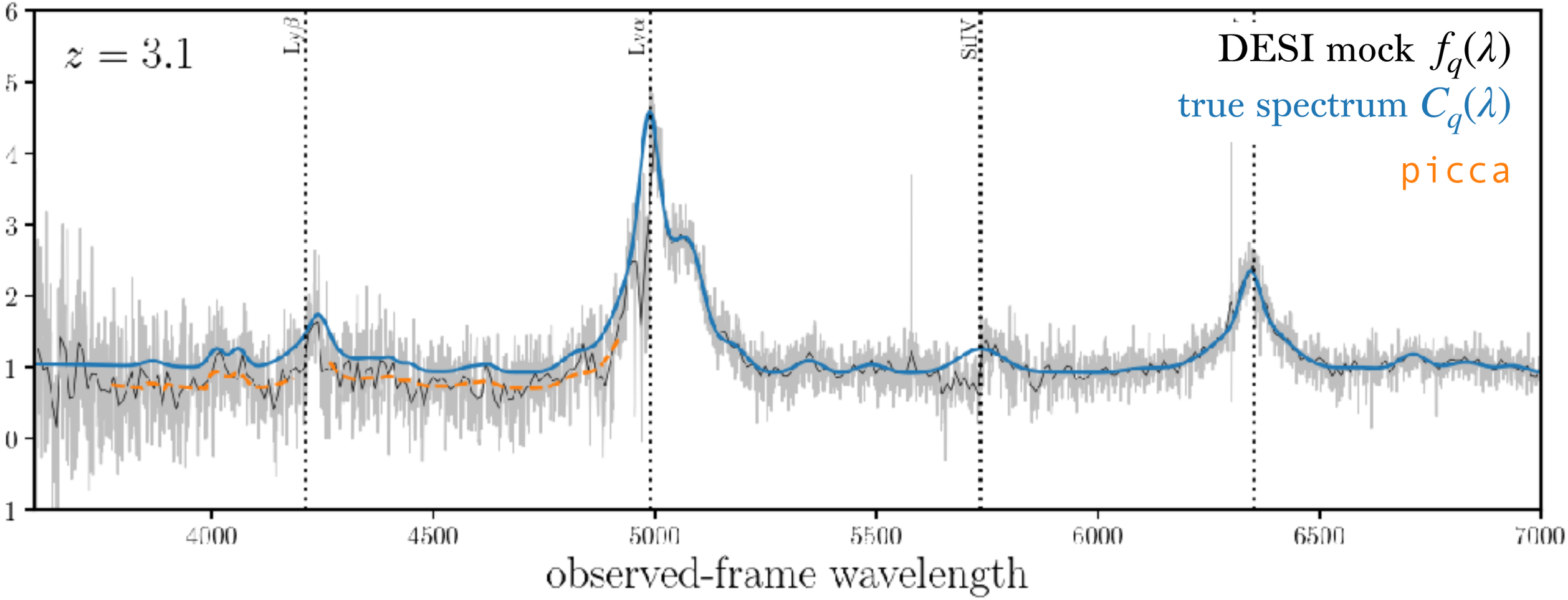

*Hahn, Gontcho A Gontcho, Melchior (2024) NeurIPS ML4PS*

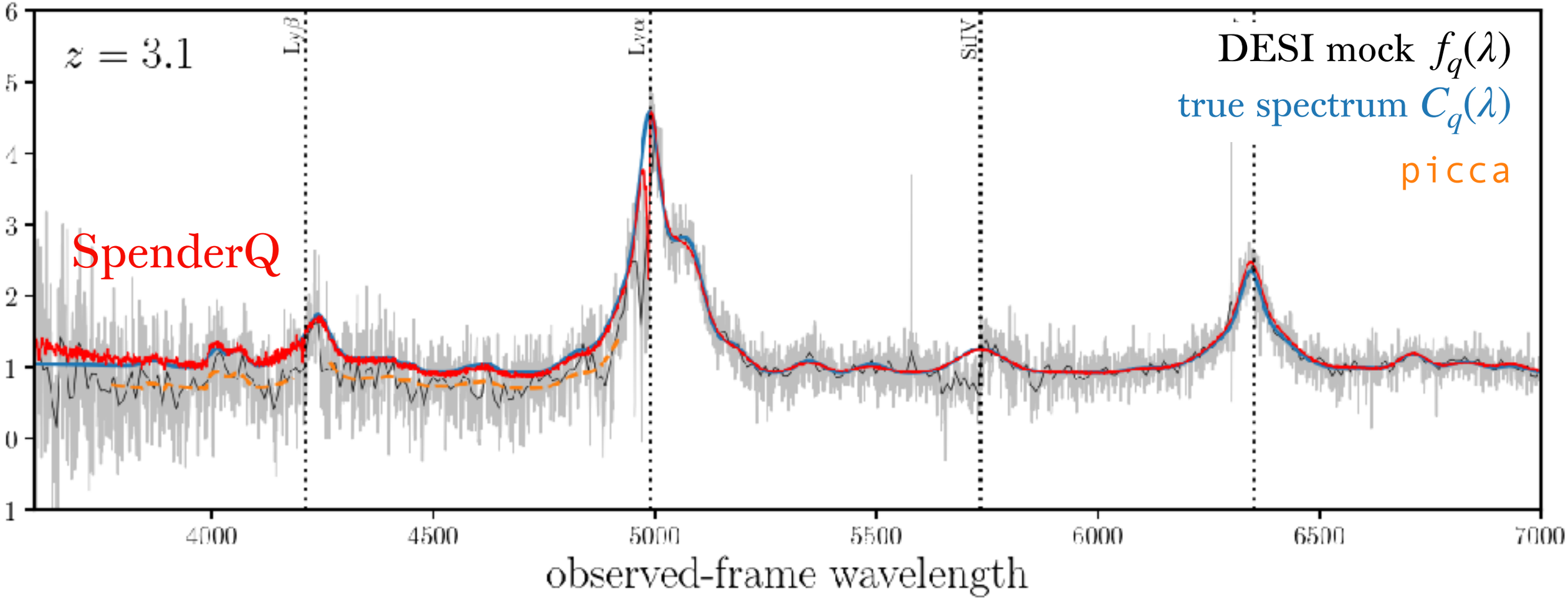
SpenderQ

**spender** + iterative procedure for identifying and masking absorption




SpenderQ *will not learn the absorption features since they are uncorrelated with the quasar spectra*


$z = 3.1$
Ly$\beta$
Ly$\alpha$
SiIV
DESI mock $f_q(\lambda)$
true spectrum $C_q(\lambda)$
picca
6
5
4
3
2
1
0
1
4000
4500
5000
5500
6000
6500
7000
observed-frame wavelength

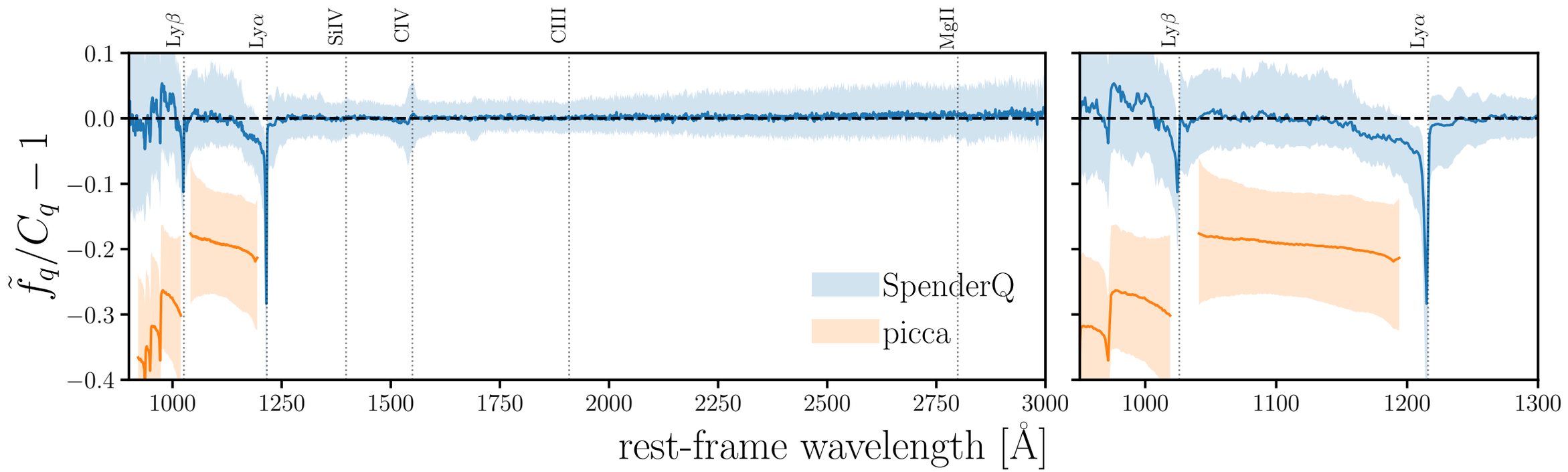

$z = 3.1$
Ly$\beta$
Ly$\alpha$
SiIV
DESI mock $f_q(\lambda)$
true spectrum $C_q(\lambda)$
picca
SpenderQ
6
5
4
3
2
1
0
1
4000
4500
5000
5500
6000
6500
7000
observed-frame wavelength

$z = 3.10$
$z = 3.04$
$z = 3.04$
$z = 3.13$
DESI Y1 Mock
(coarse binned)
True Continuum
picca
SpenderQ
Ly$\beta$
Ly$\alpha$
10.0
7.5
5.0
2.5
0.0
4000
5000
6000
7000
8000
9000
1000
1100
1200
1300
observed-frame wavelength
rest-frame wavelength

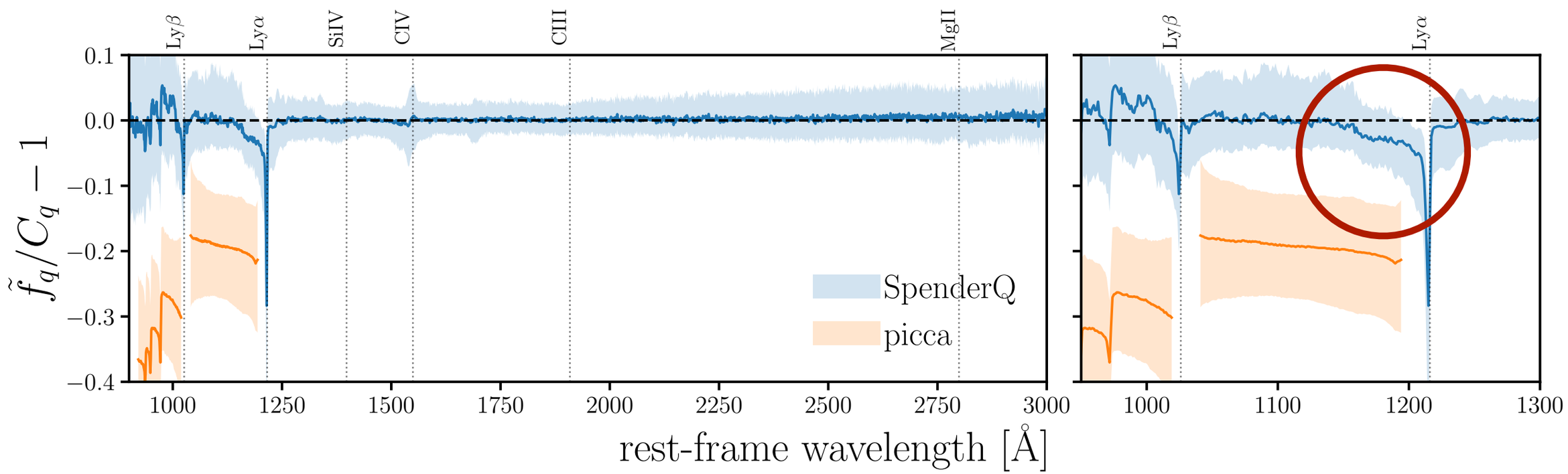
SpenderQ validation on DESI Year 1 "*London*" mocks, 400,000 spectra constructed using CoLoRe, Ly$\alpha$CoLoRe, and quickquasar

*Ramírez-Pérez et al. (2022); Farr et al. (2020); Herrera-Alcantar et al. (2023); Cuceu et al. (2024*

SpenderQ provides *percent-level* reconstructions of the quasar spectra — picca residuals: ~20% in Ly$\alpha$ and >30% in Ly$\beta$

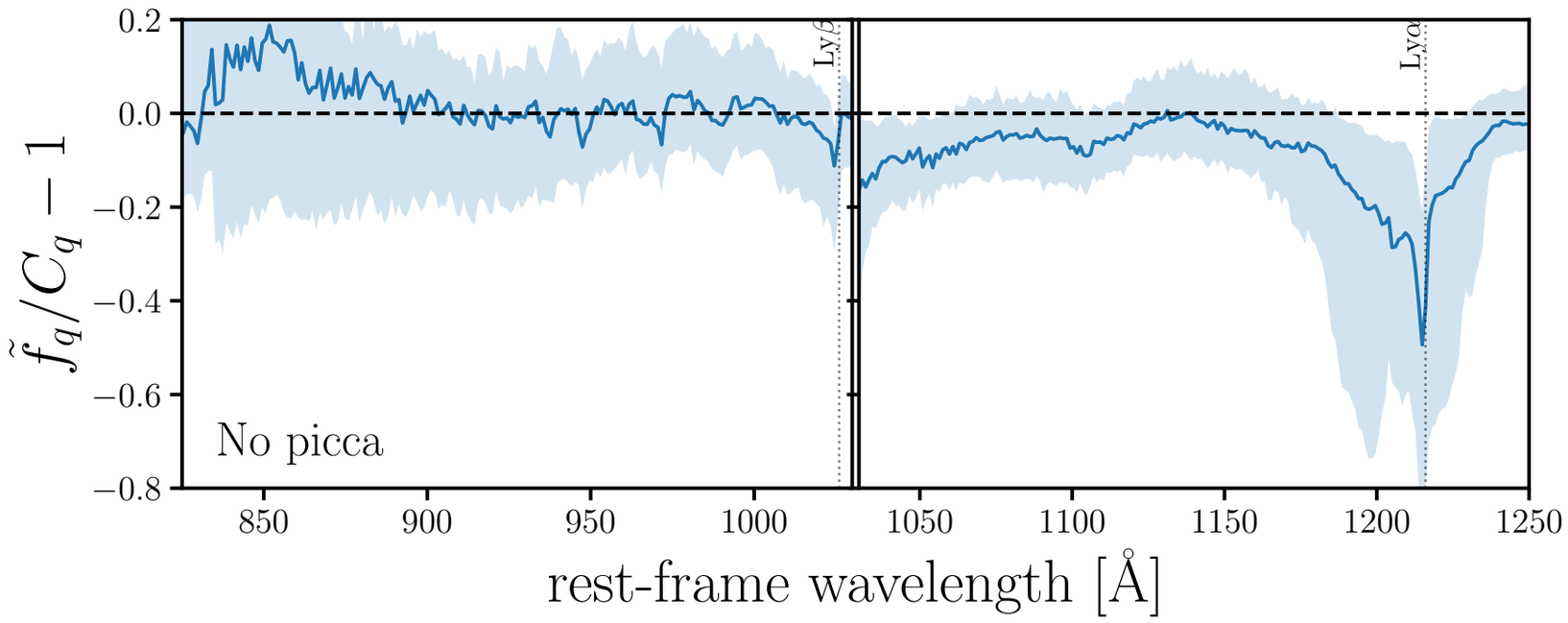


SpenderQ provides *percent-level* reconstructions of the quasar spectra — picca residuals: ~20% in Ly$\alpha$ and >30% in Ly$\beta$



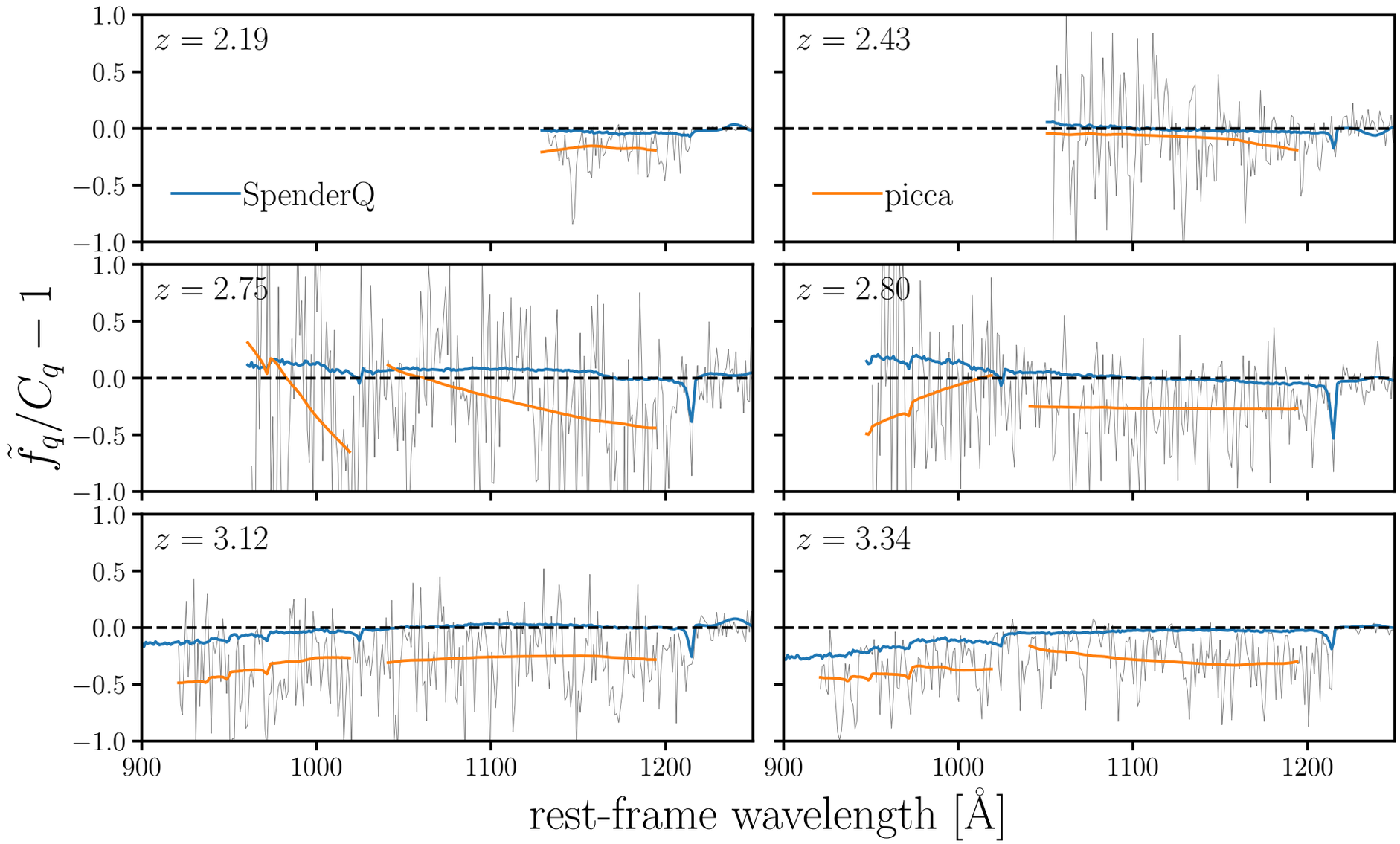

*we do not remove spectra flagged with BAL or DLAs in* SpenderQ

# SpenderQ provides reconstructions for *all quasars*

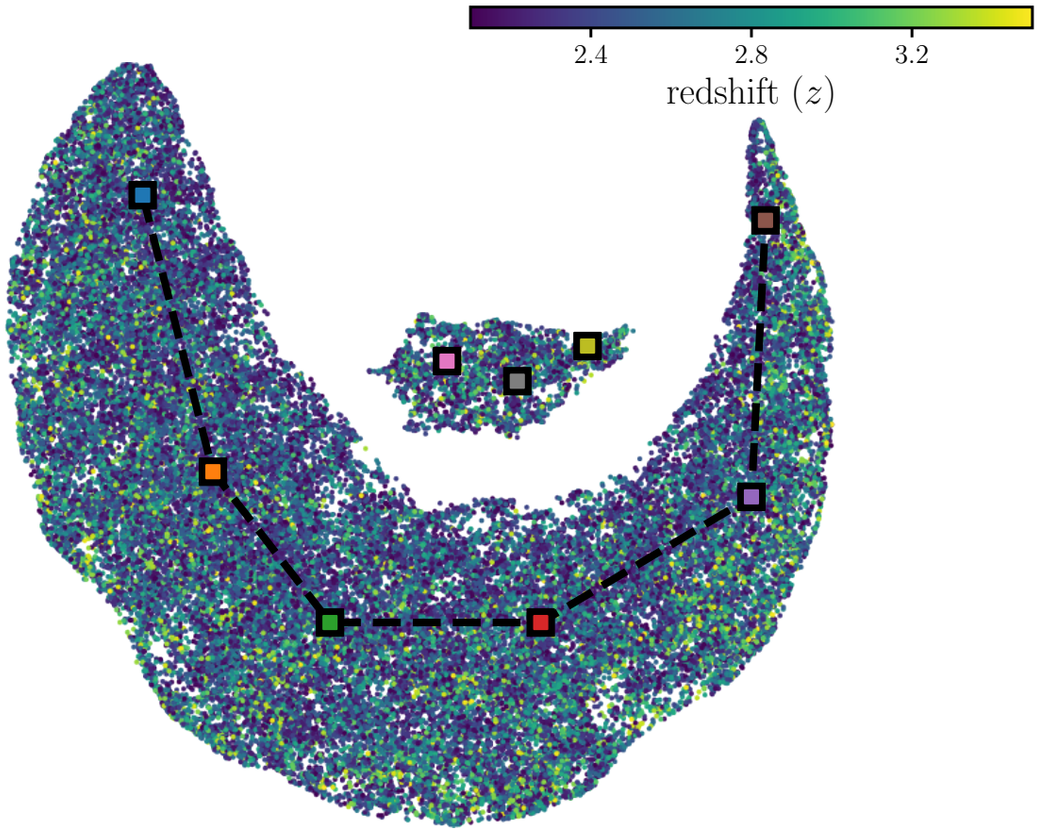



*picca requires >150 spectral elements in the Lyα/Lyβ region*

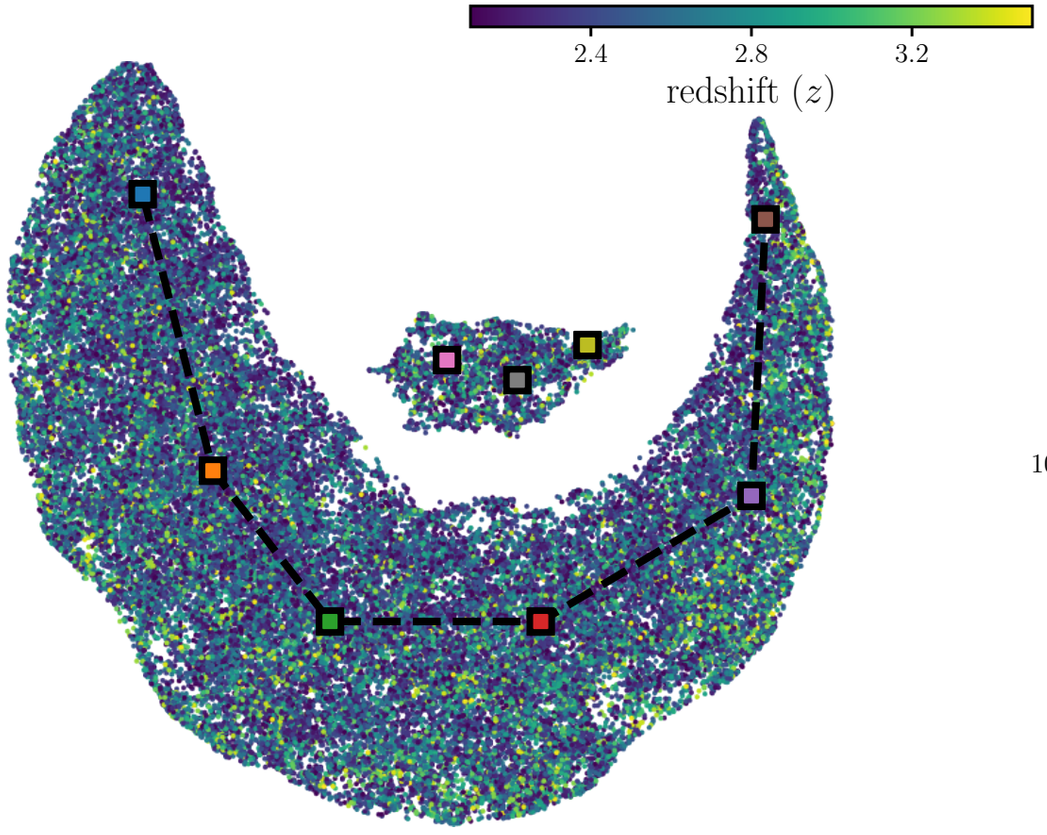
# SpenderQ accurately reconstructs the spectrum of *individual quasars*

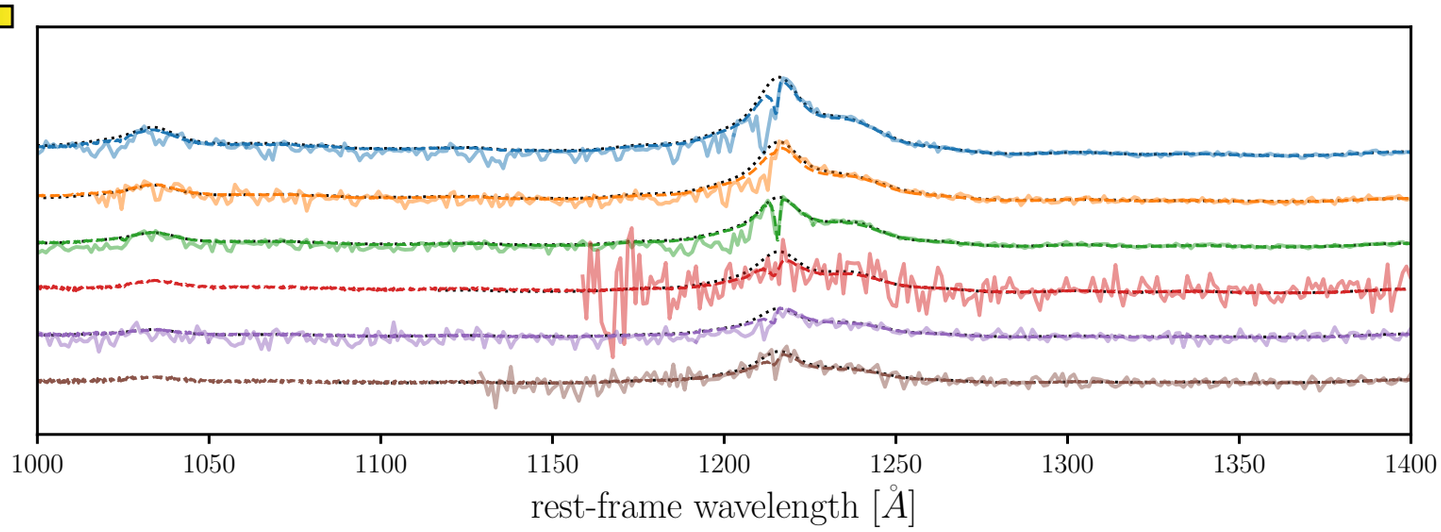


# SpenderQ provides a *redshift invariant* and physically informative latent encoding

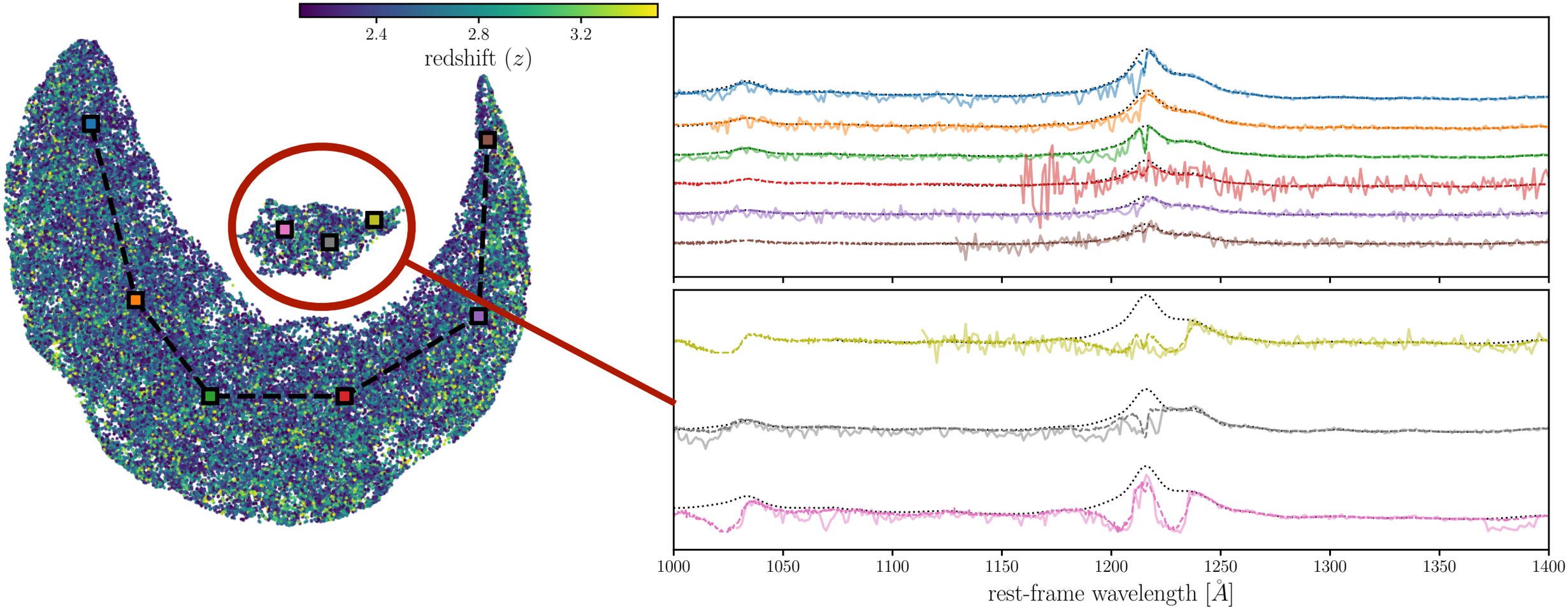


# SpenderQ provides a *redshift invariant* and *physically informative* latent encoding



# SpenderQ provides a *redshift invariant* and *physically informative* latent encoding — identifies Broad Absorption Lines (BALs)



# SpenderQ applied to DESI Year 1 "*London*" Ly$\alpha$ mocks

*Hahn, Gontcho A Gontcho et al (in prep.)*

# SpenderQ applied to DESI Year 1 "*London*" Ly$\alpha$ mocks

*Hahn, Gontcho A Gontcho et al (in prep.)*

# SpenderQ applied to DESI Early Data Release

*Ly$\alpha$ forest "delta" catalog*

*IGM absorber catalog*

*reconstructed quasar catalog*

*latent embedding catalog*

*Gontcho A Gontcho, Hahn et al (in prep.)*

SpenderQ applied to DESI Year 1 "*London*" Ly$\alpha$ mocks

*Hahn, Gontcho A Gontcho et al (in prep.)*

SpenderQ applied to DESI Early Data Release

*Gontcho A Gontcho, Hahn et al (in prep.)*

SpenderQ for Ly$\alpha$ clustering analyses — BAO and full-shape analyses

SpenderQ applied to DESI Year 1+ Data Release