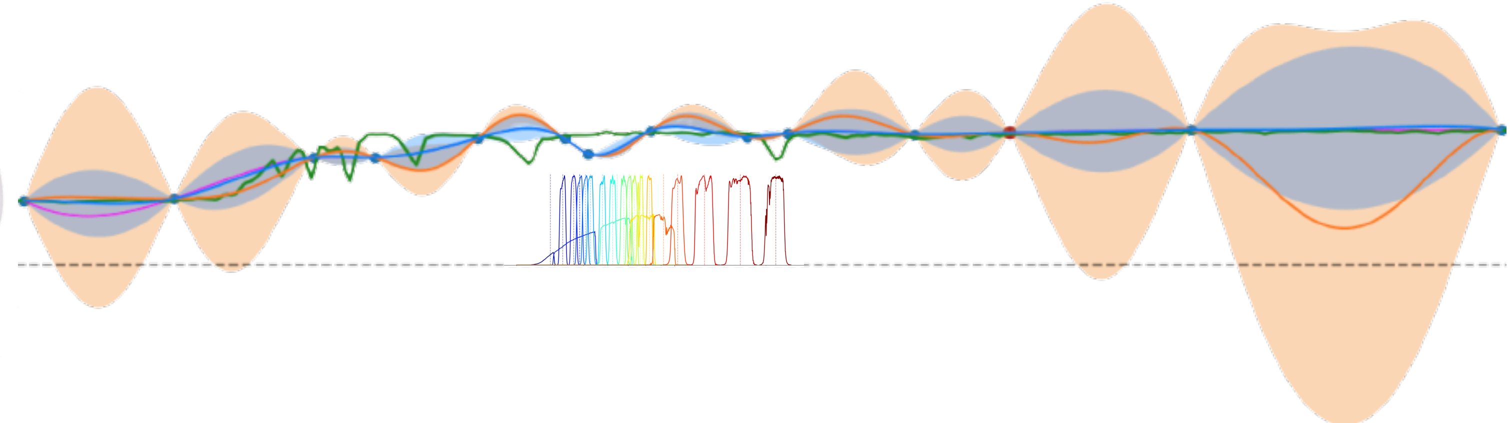


Exploring Methods for Accurate Flux Reconstruction in Galaxies SEDs:

Balancing Physics and Data-Driven Approaches

Ecogia Science Meeting, January 30th 2023

Federica Tarsitano, Stéphane Paltani, William Hartley
University of Geneva

Motivation

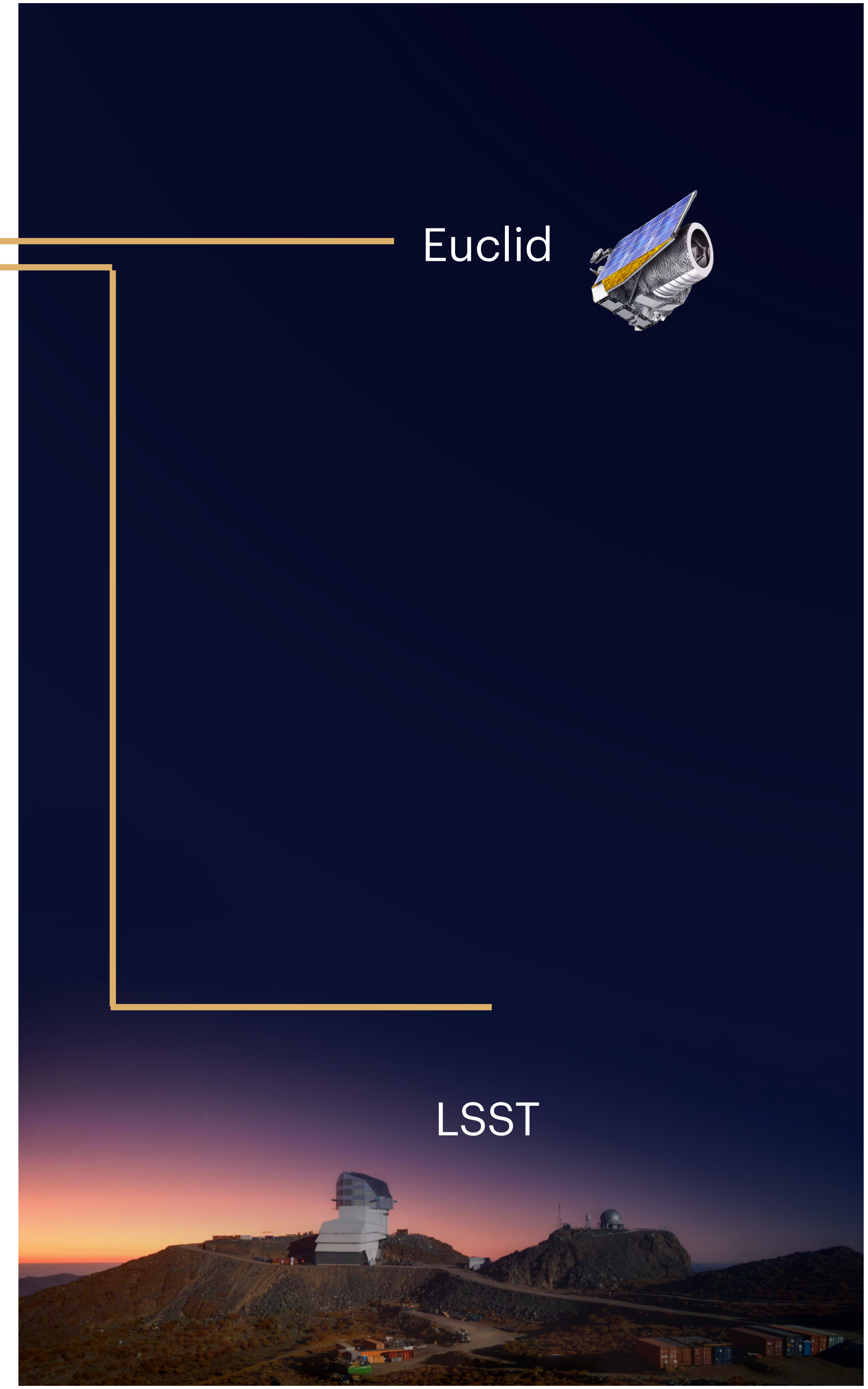
Shape measurement for billions of galaxies

Reduction of statistical uncertainty

↓

Reduction in residual systematics

Accurate estimation of cosmological parameters



Credits: Rubin Obs/NSF/AURA, Euclid Consortium

Motivation

Shape measurement for billions of galaxies

Challenges

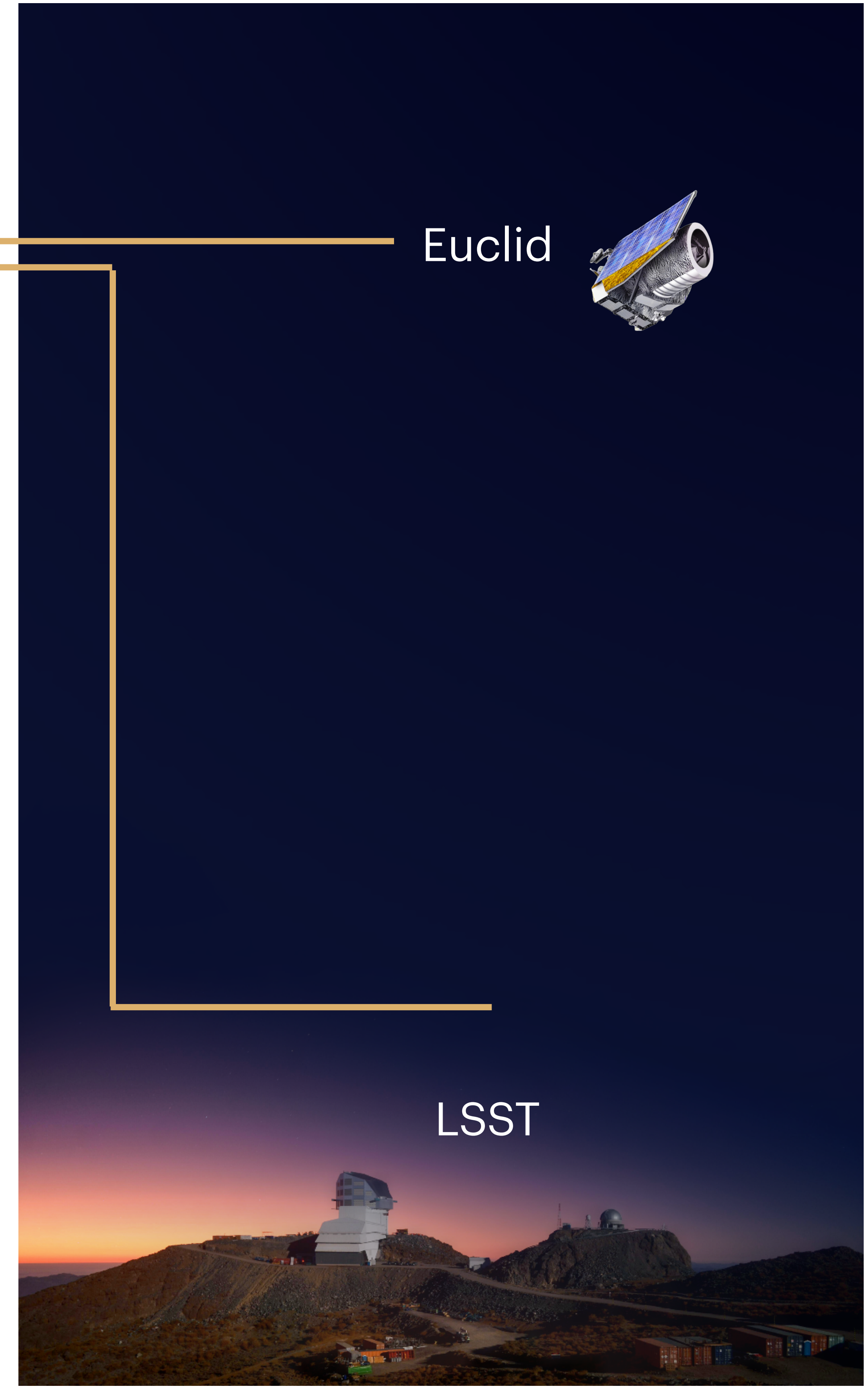
Convolution of the galaxy image with the wavelength-dependent PSF

(*)

$$I^{\text{obs}}(\boldsymbol{x}) = \int I^0(\boldsymbol{x};\lambda) \otimes P(\boldsymbol{x};\lambda)\mathrm{d}\lambda, \quad (**)$$

$$Q_{ij}^0 = Q_{ij}^{\text{obs}} - \frac{1}{F}\int F(\lambda)P_{ij}(\lambda)\mathrm{d}\lambda,$$

$$F(\lambda) \equiv \lambda S(\lambda)T(\lambda)$$



(*) Credits: Wikipedia

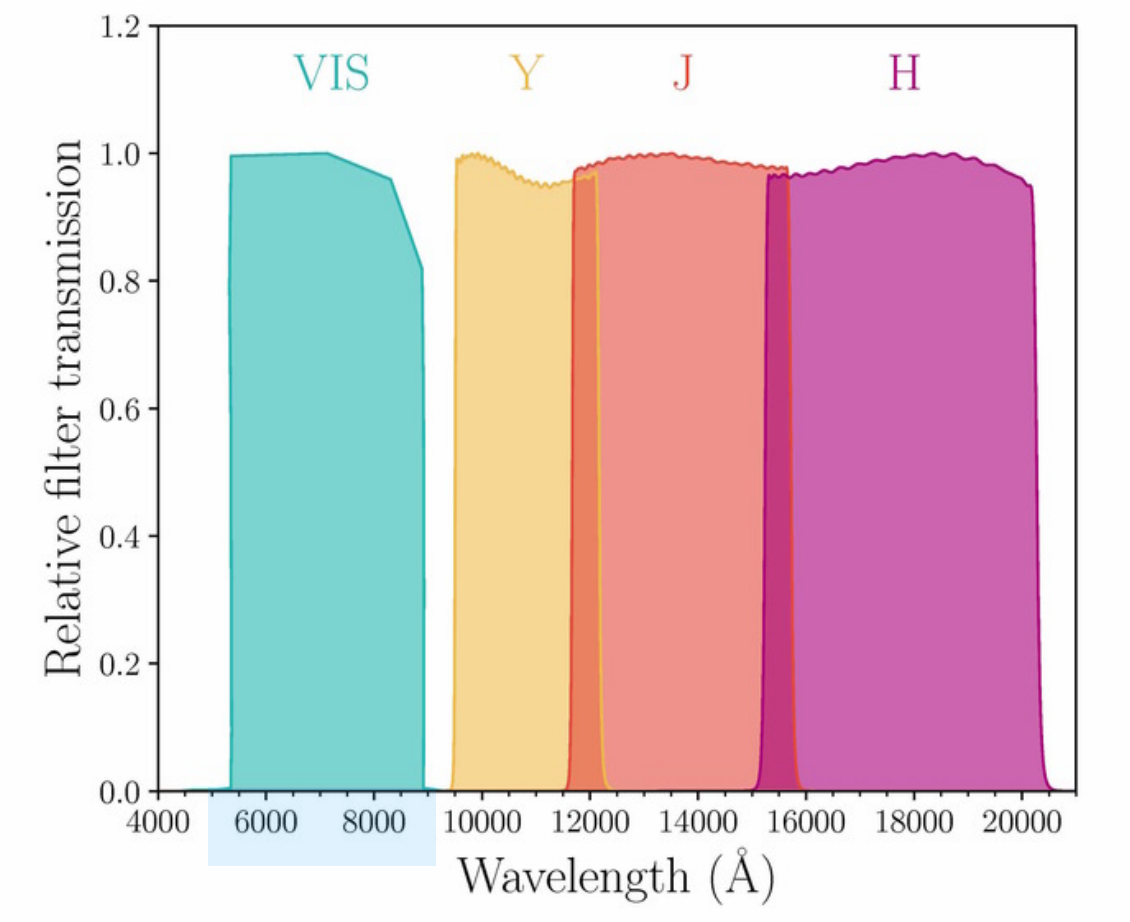
(**) M. Eriksen and H. Hoekstra, 2018

Motivation

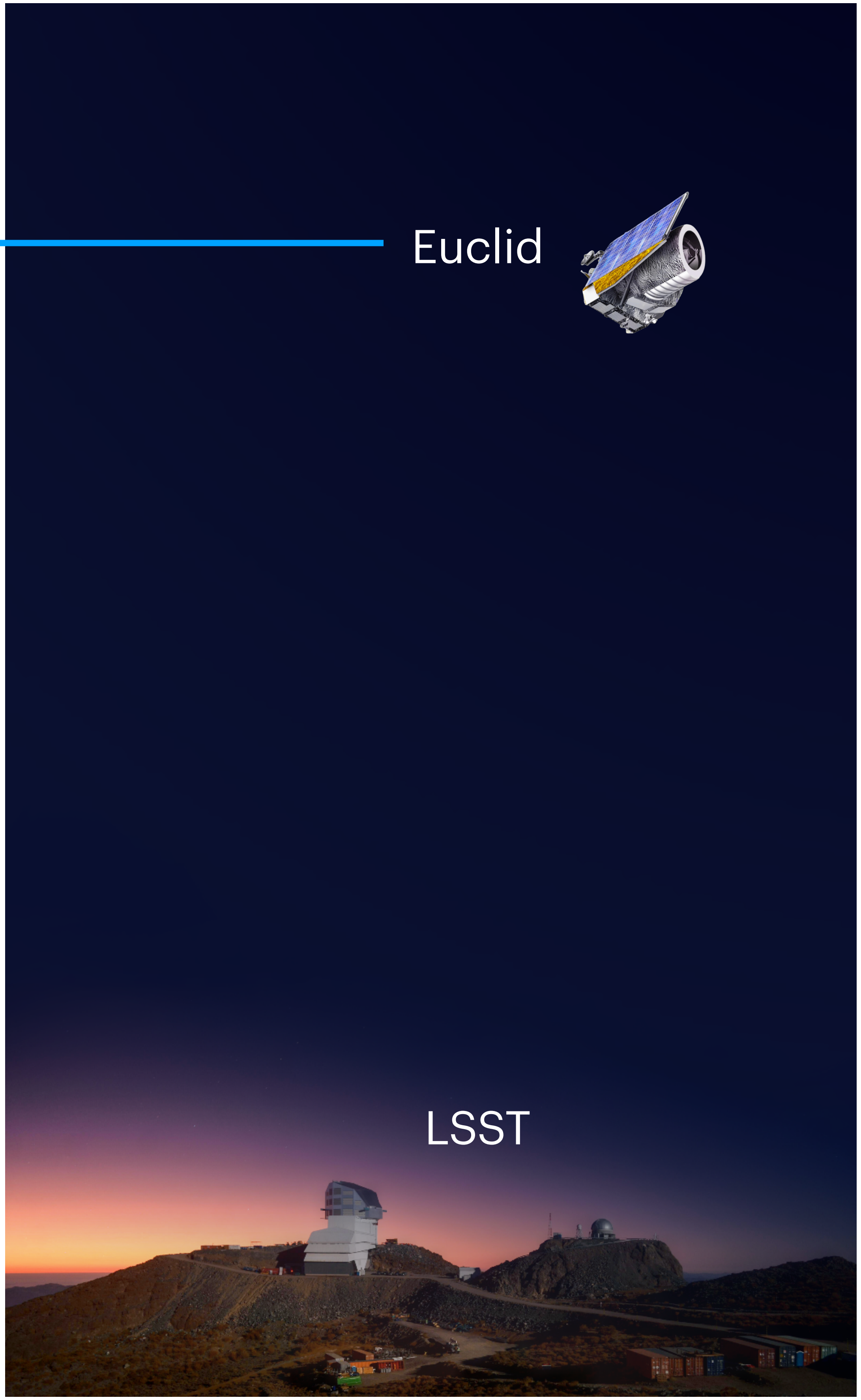
Euclid VIS filter

5500 - 9200 Å

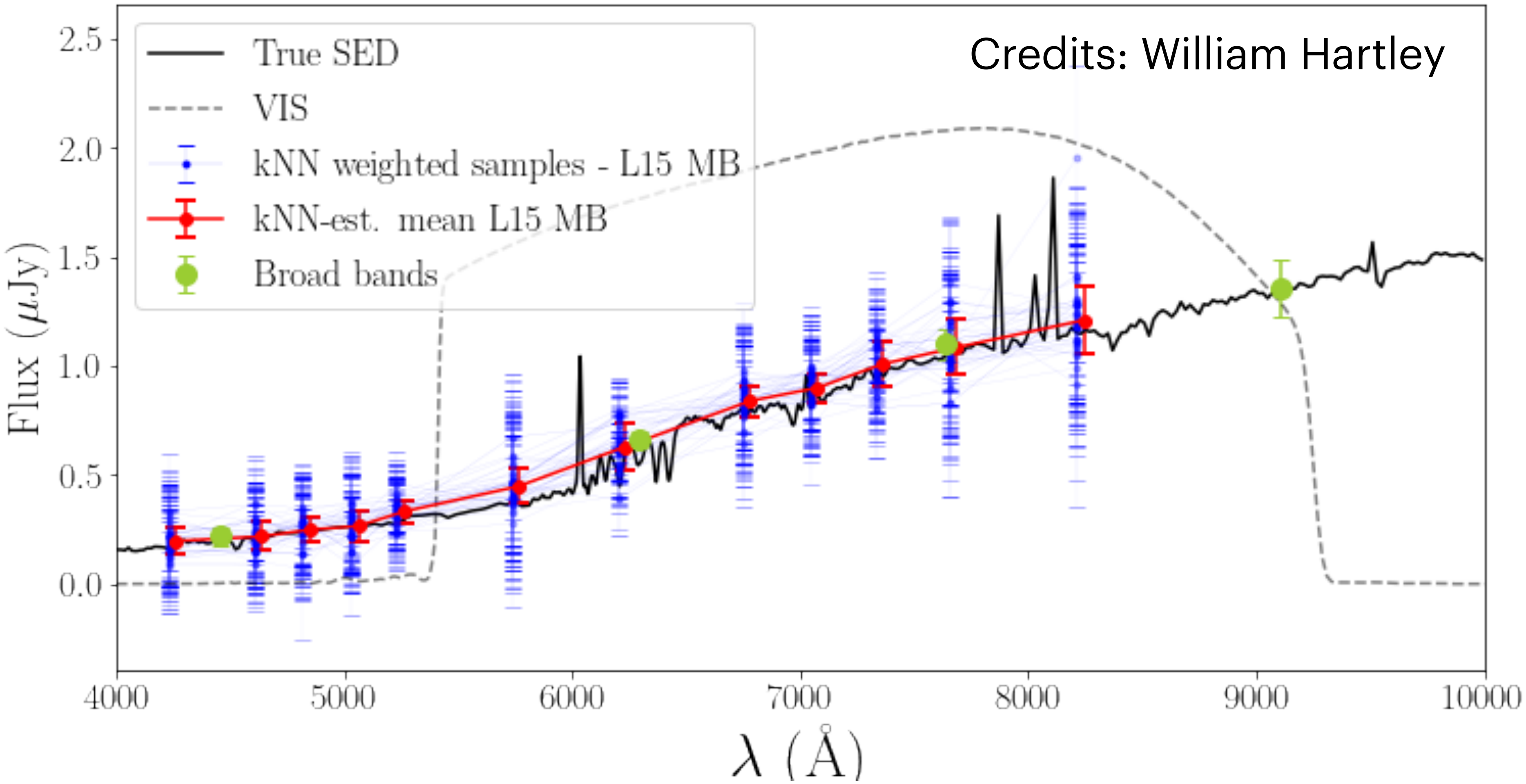
Challenges



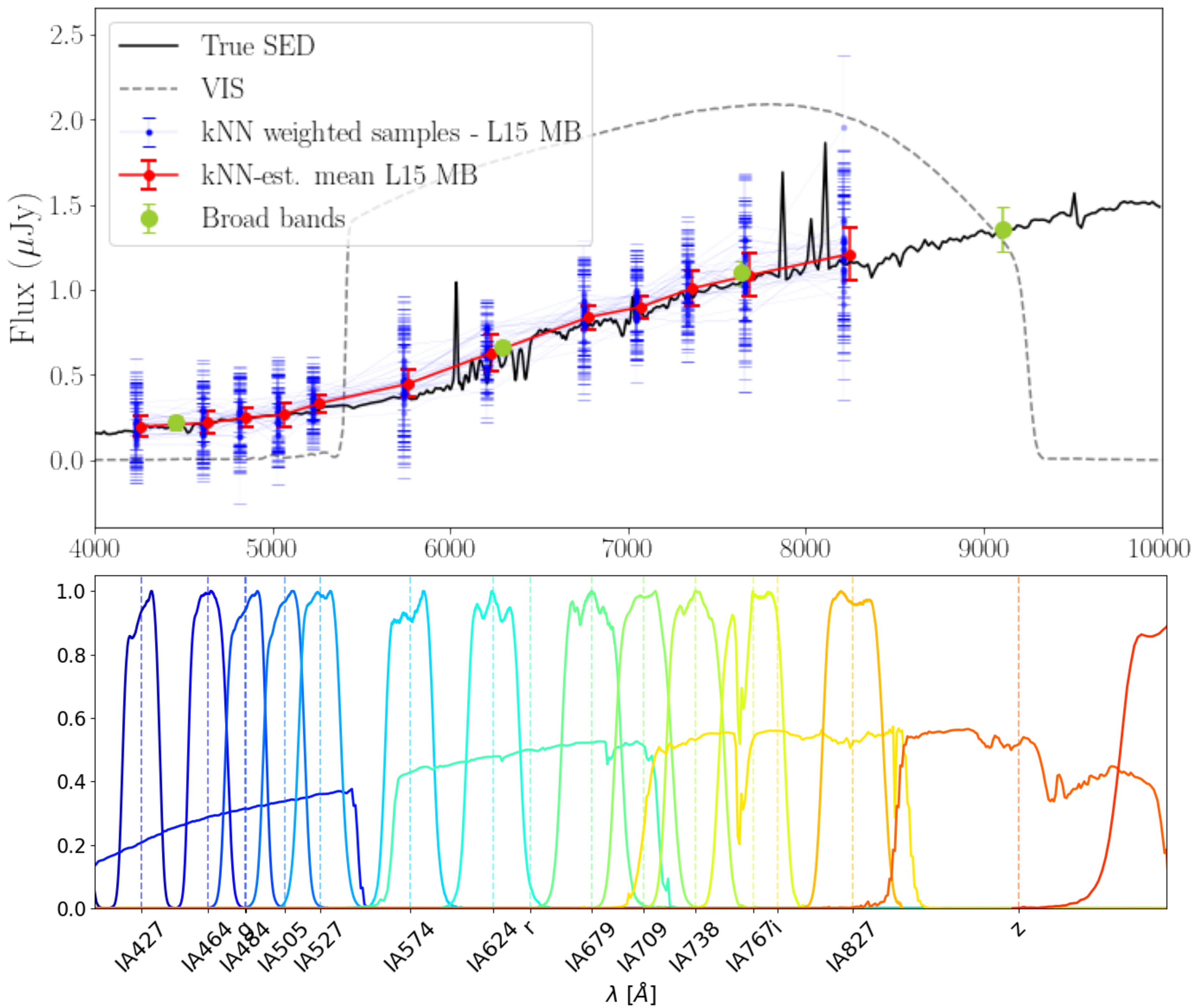
$$F(\lambda) \equiv \lambda S(\lambda) T(\lambda)$$



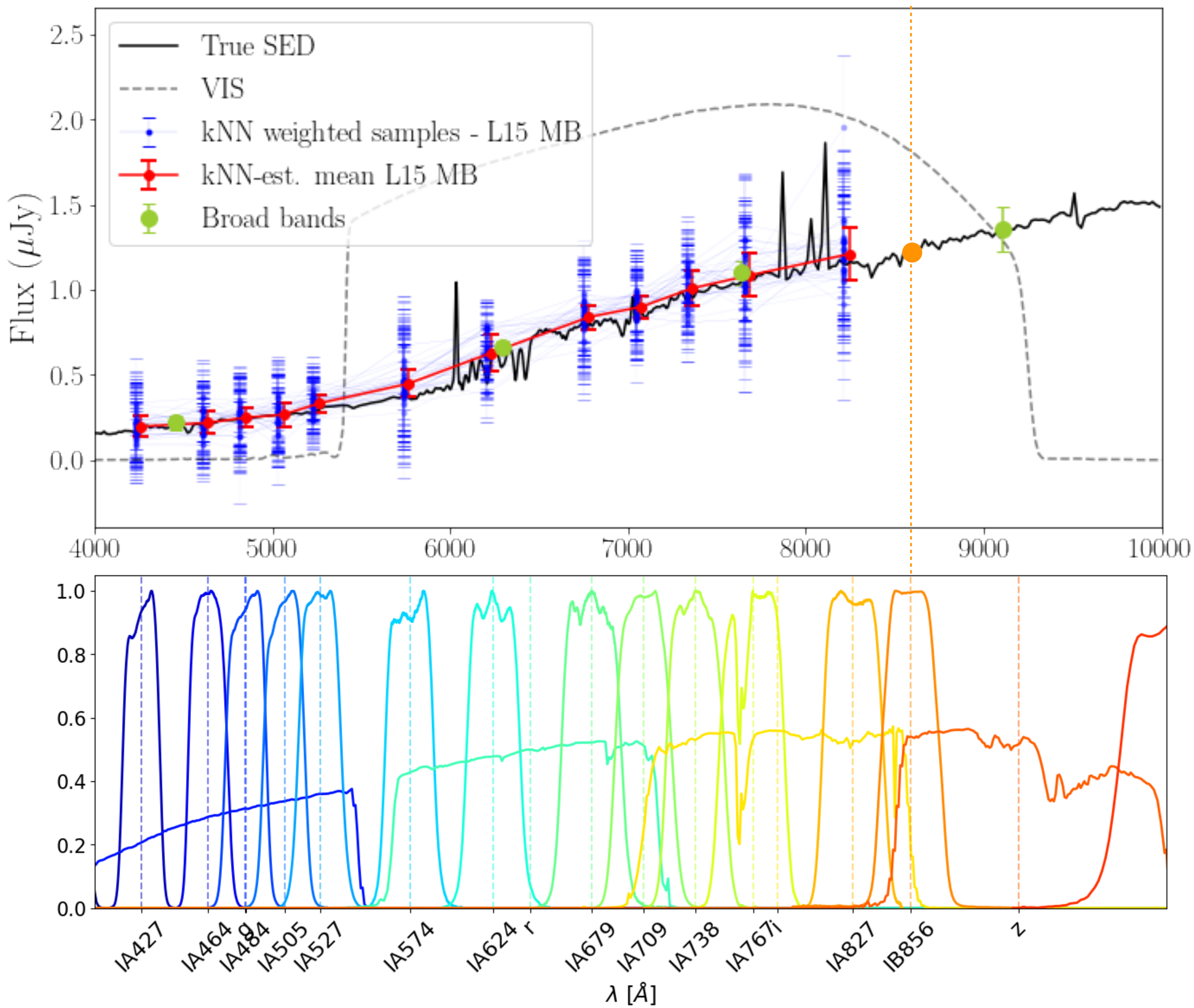
Sampling the Galaxy SED



Sampling the Galaxy SED



Sampling the Galaxy SED



Flux Reconstruction - Methods

Physics - driven approach

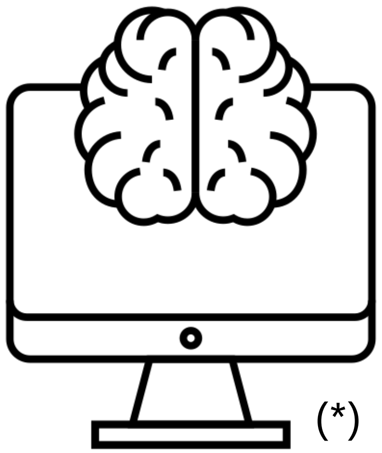
Template fitting

Data - driven approaches

Interpolation of the known data points:

Cubic Spline
Gaussian Processes

Random Forest algorithms



(*)

Hybrid method

Template Fitting plus relations in Color Space

Flux Reconstruction | single band (IB856)
multiband

(*) computer brain by ibrandify from the Noun Project

Template fitting for IB856 reconstruction

Gal-Pop simulations (William Hartley)

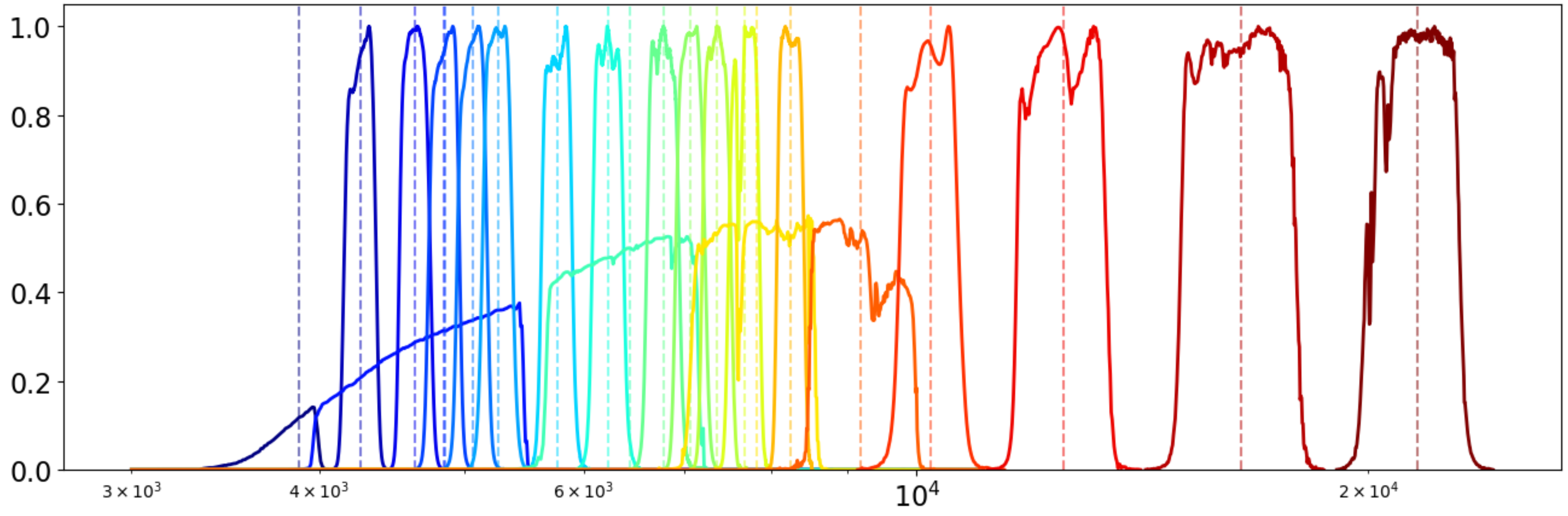
Set of simulated galaxies for which we have the SEDs

31 Cosmos SEDs templates saved in Phosphoros are used to simulate galaxies

Paltani et al., in prep.

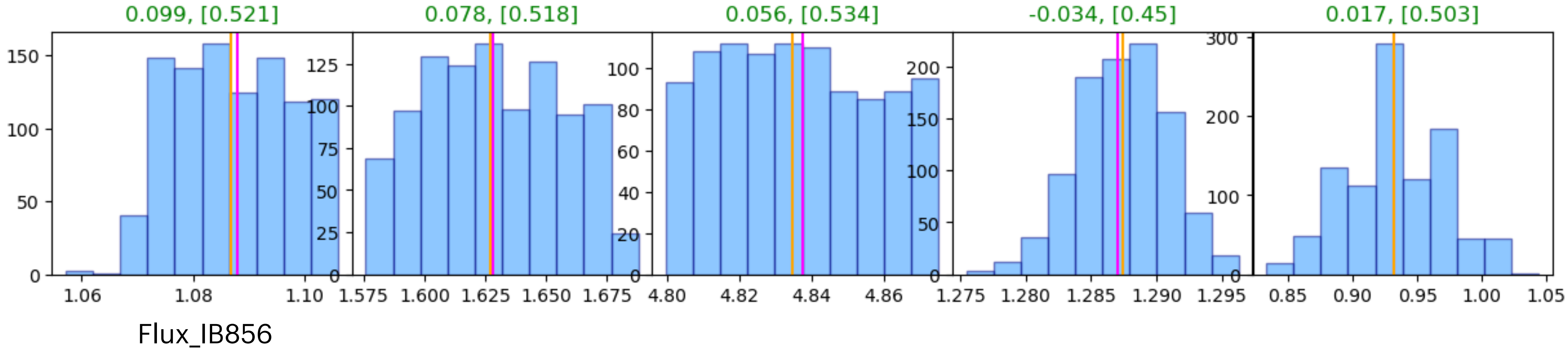
Output in the form of posterior distributions

Advantage for matching with the model grid in SED number - redshift parameter space

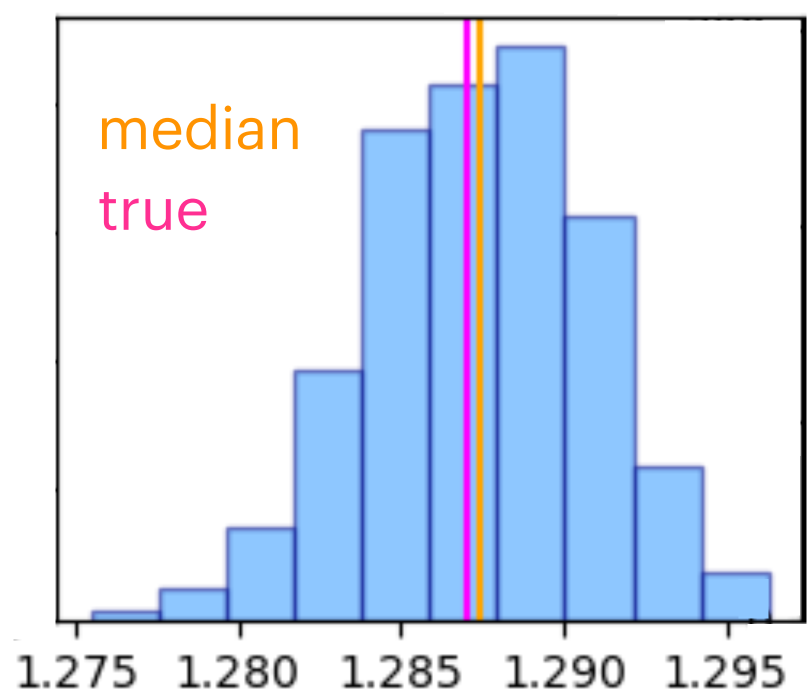


Results: bias on the flux posteriors (II)

Flux Posterior Distributions per object



-0.034, [0.45] PIT



$$b_i = \frac{m_i - t_i}{t_i},$$

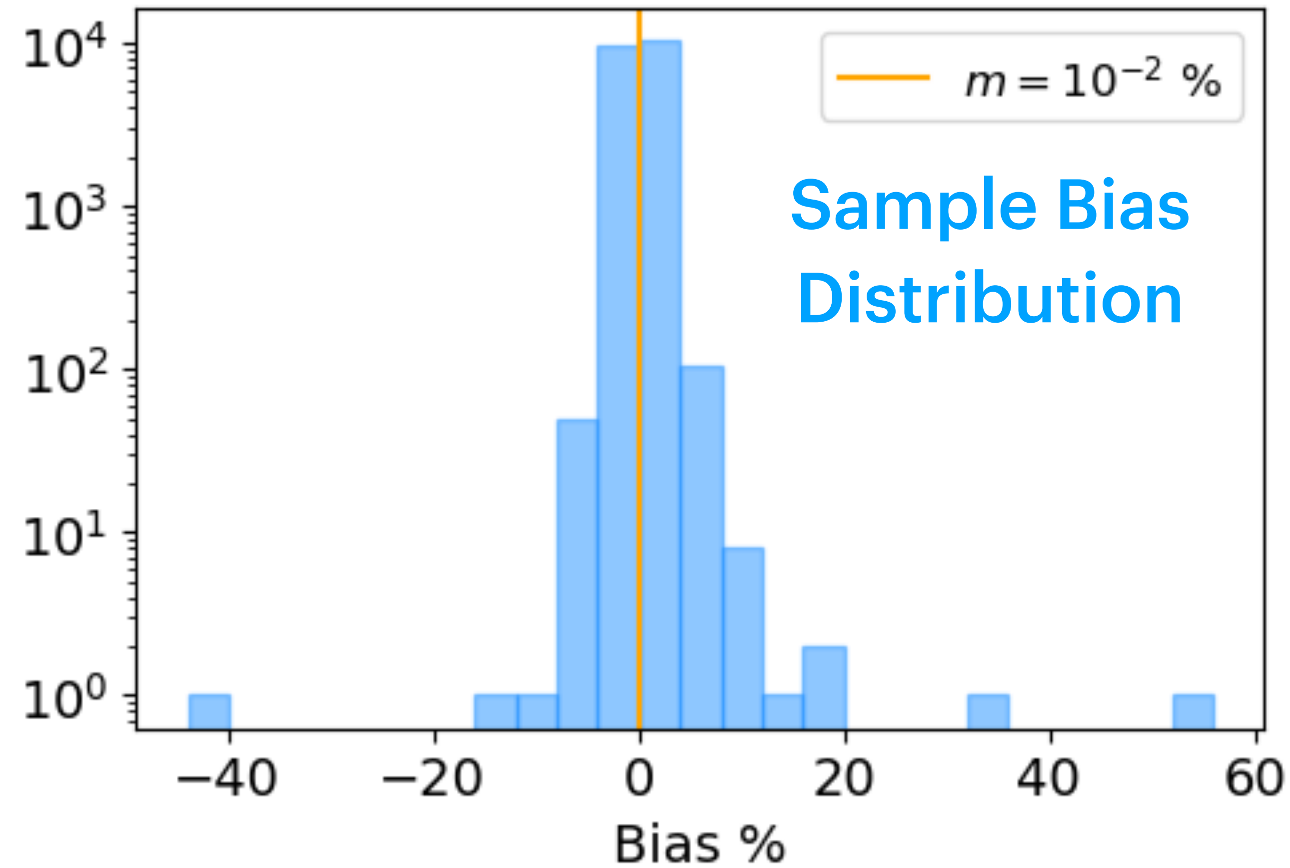
Bias per object

$$i = 0, \ldots, N_{sample}$$

Results: bias on the flux posteriors (II)

Accuracy in reconstructing each filter

PIT Plot



$$b_i = \frac{m_i - t_i}{t_i},$$

Bias per object

$$i = 0, \ldots, N_{sample}$$

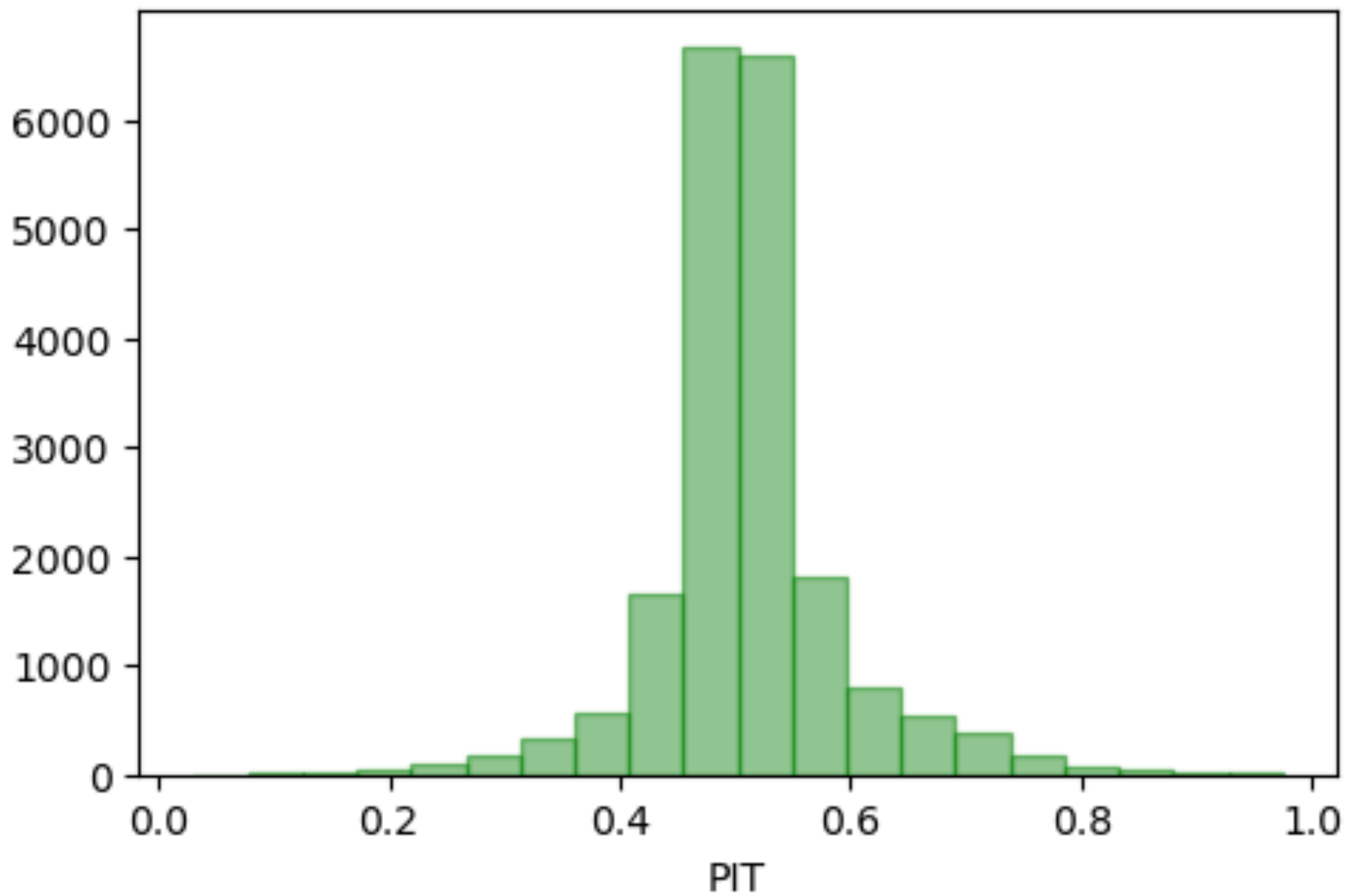
Tomographic analysis

Bias in redshift bins

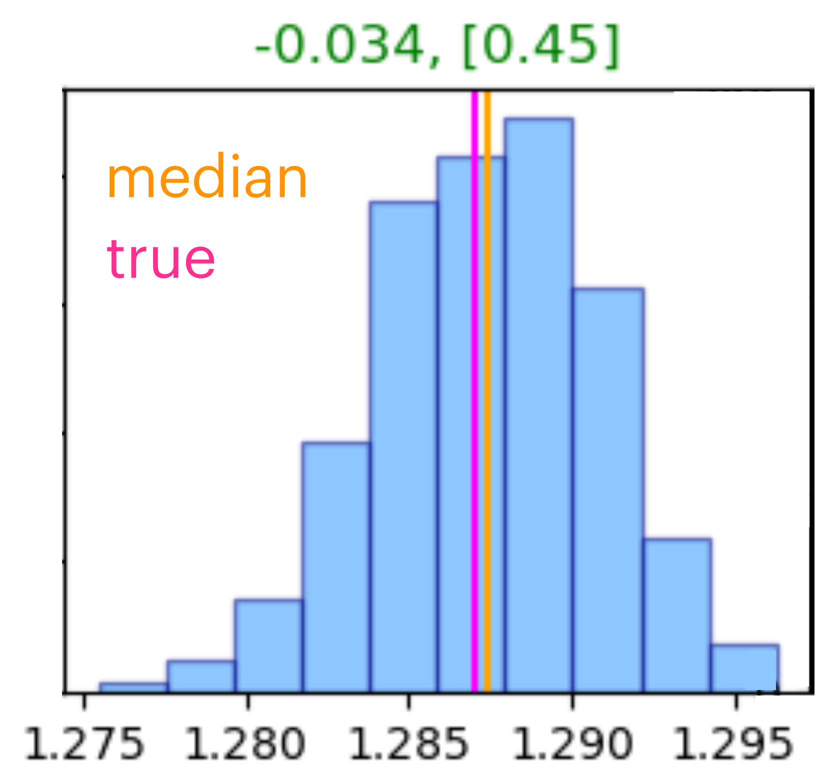
Bias on the SED metric

PIT plot and tomographic analysis (I)

Peaked PIT distribution



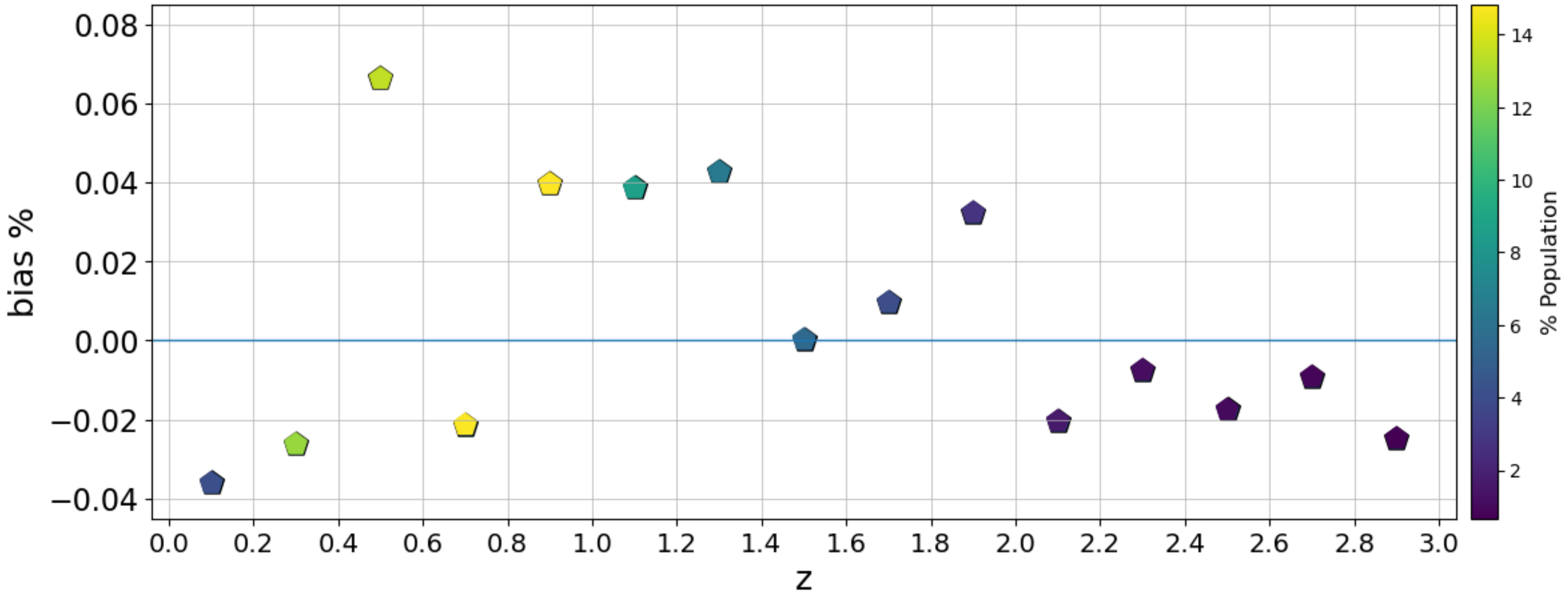
The flux posterior distributions tend to be broad



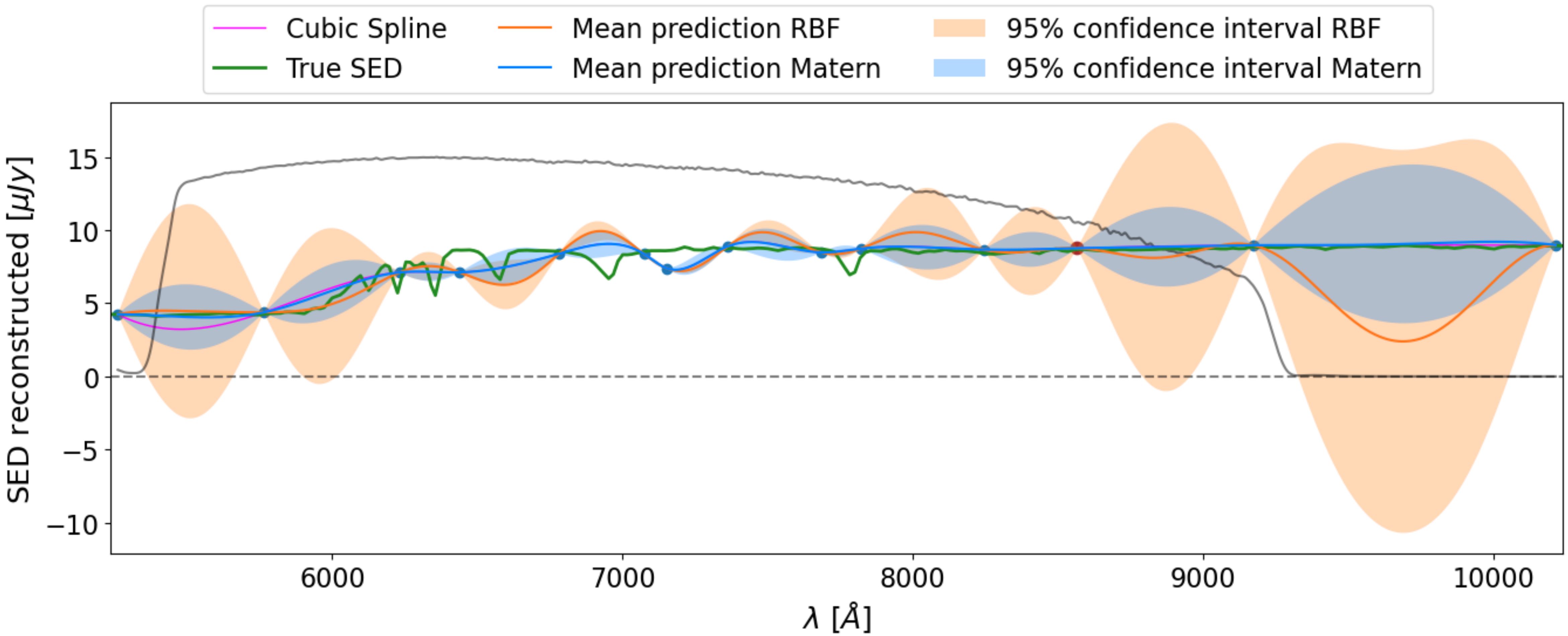
The median is used to calculate the bias

The effect is washed out in the tomographic analysis.

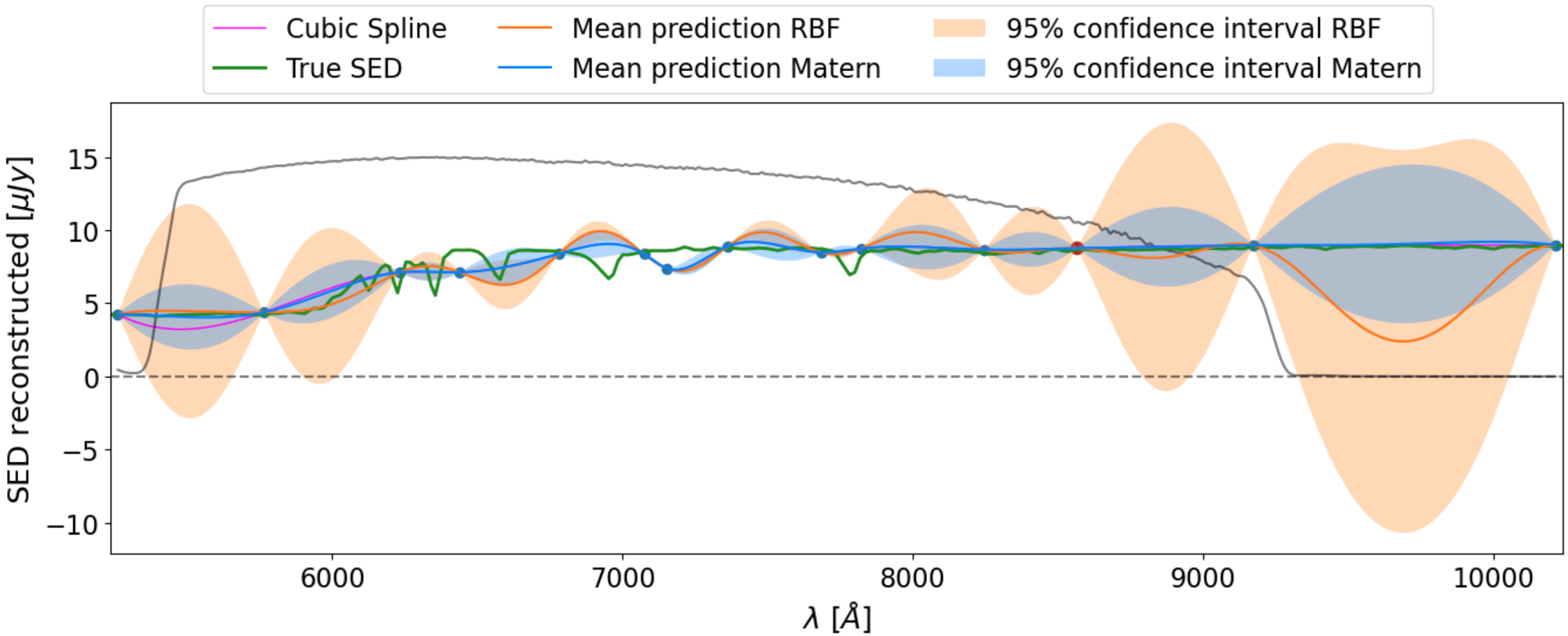
PIT plot and tomographic analysis (II)



SED Reconstruction: example



SED Reconstruction: interpolation schemes



Kernels

From scikit learn

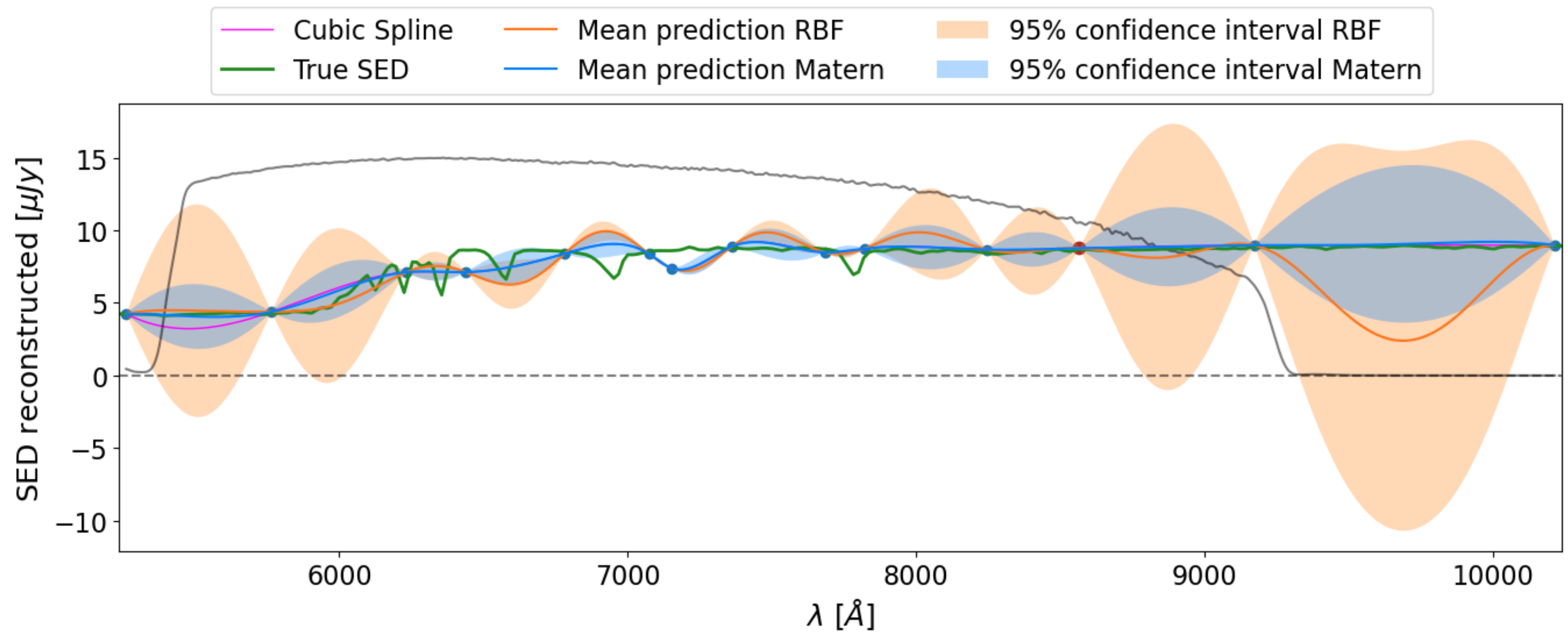
RBF (Radial Basis Function)

$$k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right)$$

Matérn

$$k(x_i, x_j) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}}{l}d(x_i, x_j)\right)$$

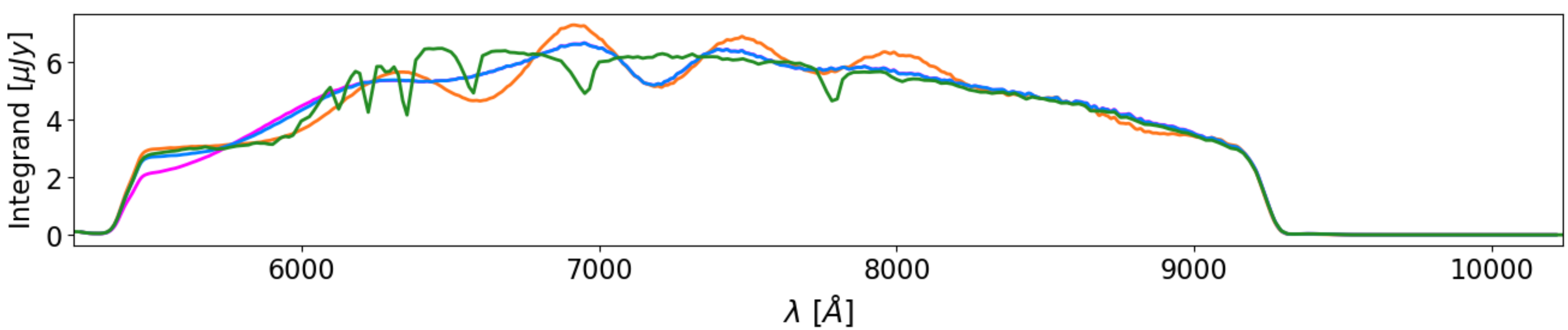
SED metric: integrating over the VIS filter



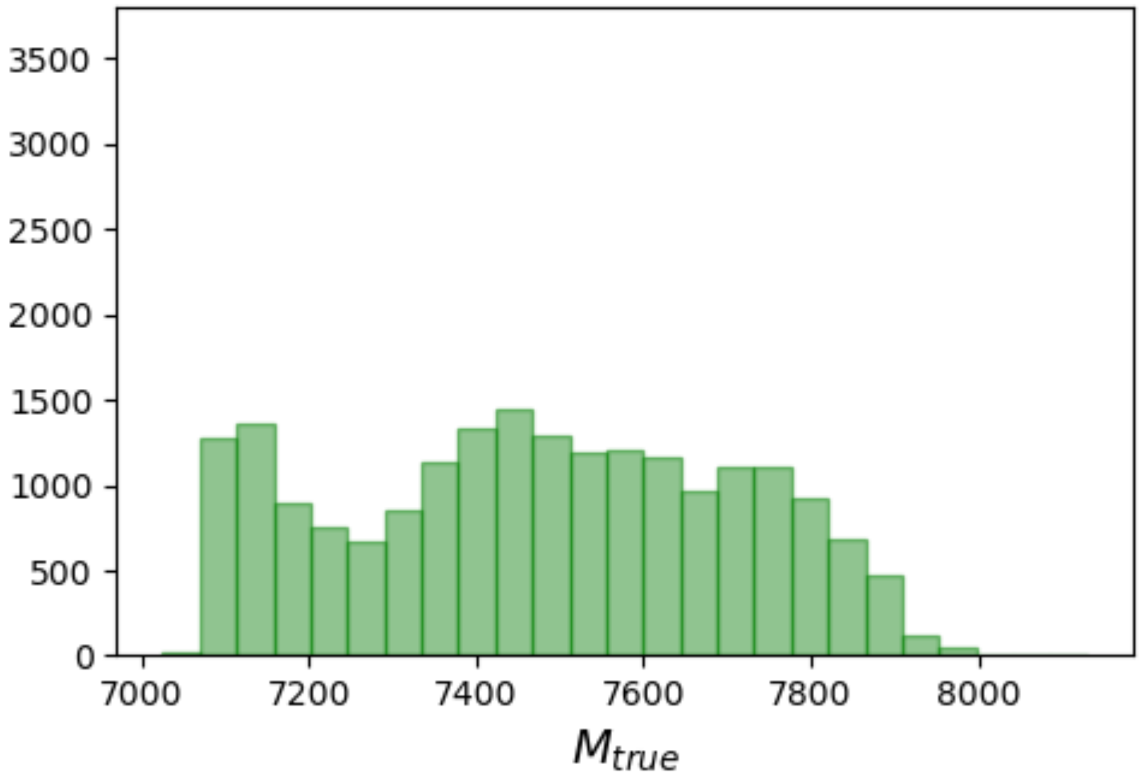
Goal: calculate the bias on the SED metric

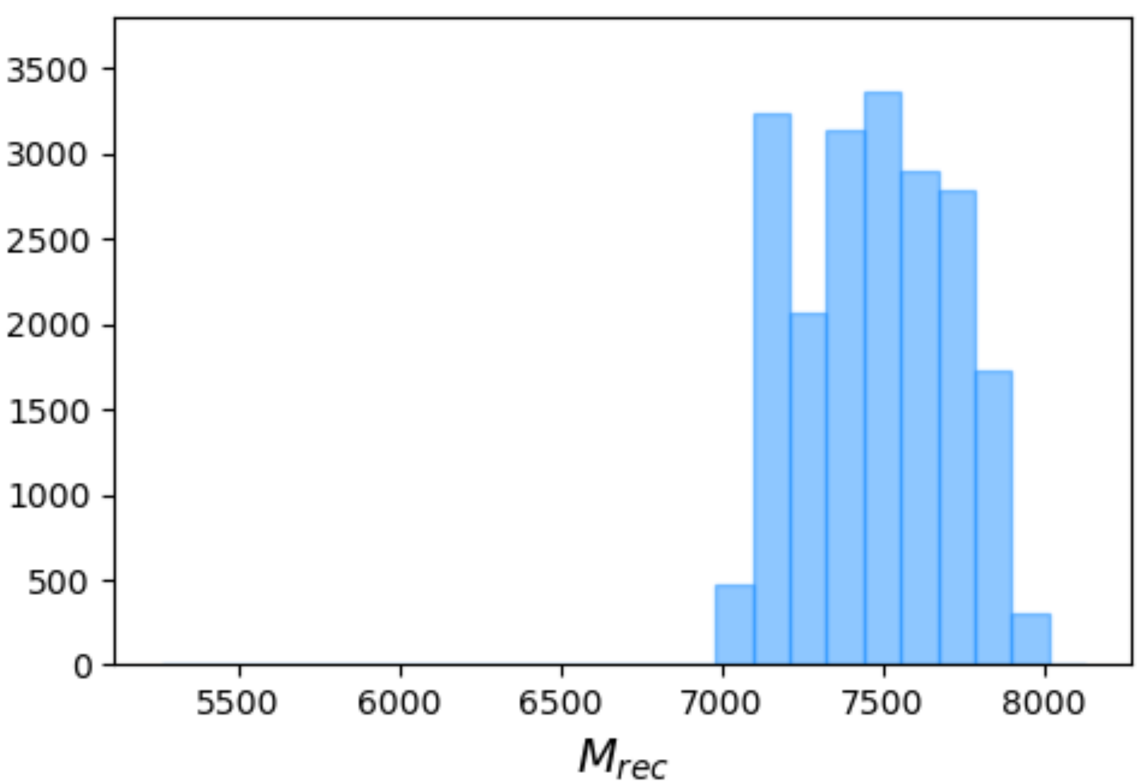
Schreiber et al., in prep.

$$\mathcal{M}_{\lambda}^{i} = \frac{\int d\lambda T(\lambda) S^{m,i}(\lambda)\lambda}{\int d\lambda T(\lambda) S^{m,i}(\lambda)}$$

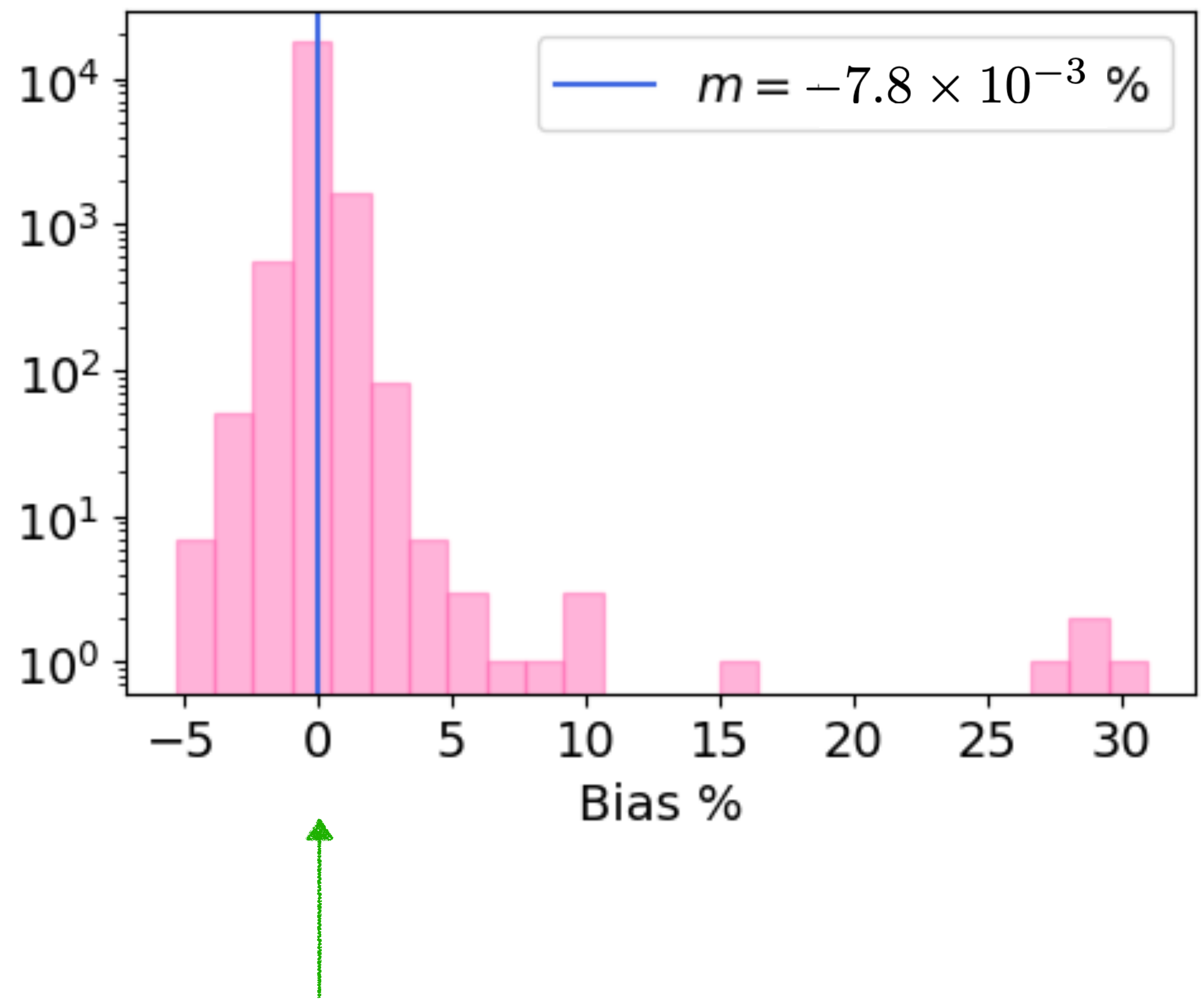


SED metric: bias - Results



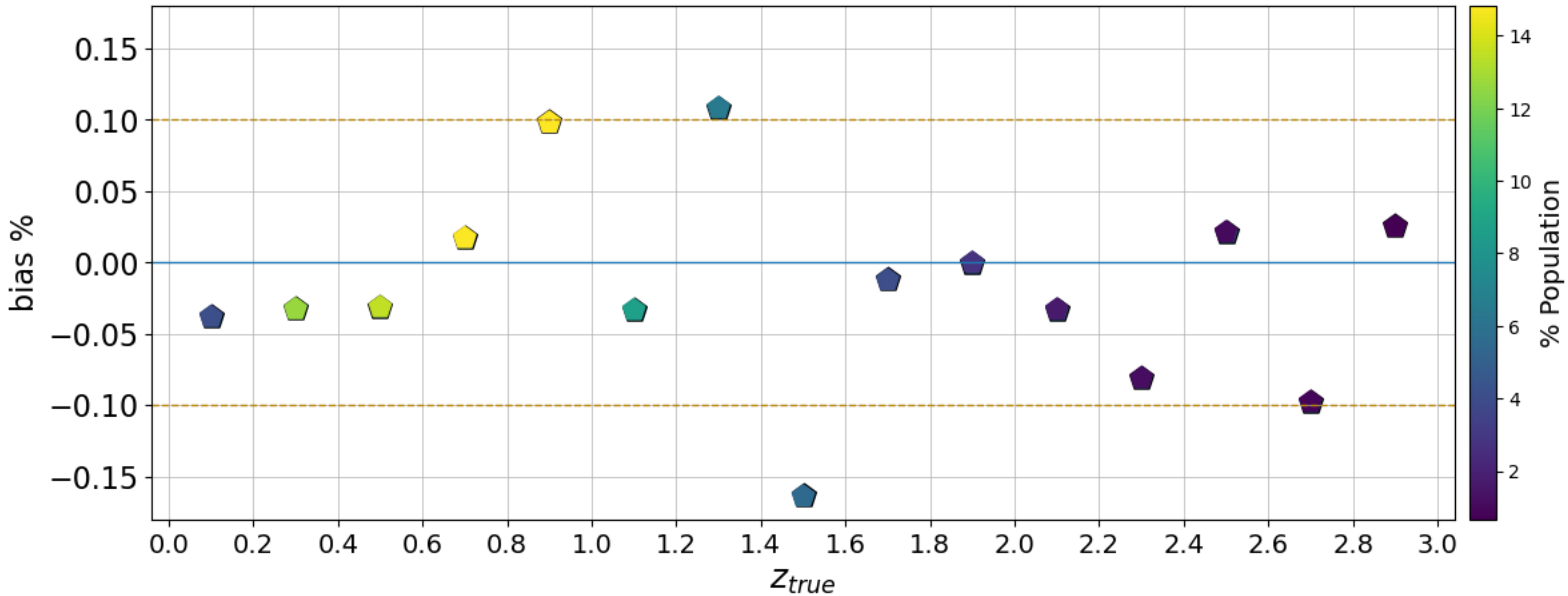


$$b^i = \frac{\mathcal{M}_\lambda^{*,i} - \mathcal{M}_\lambda^{m,i}}{\mathcal{M}_\lambda^{*,i}},$$



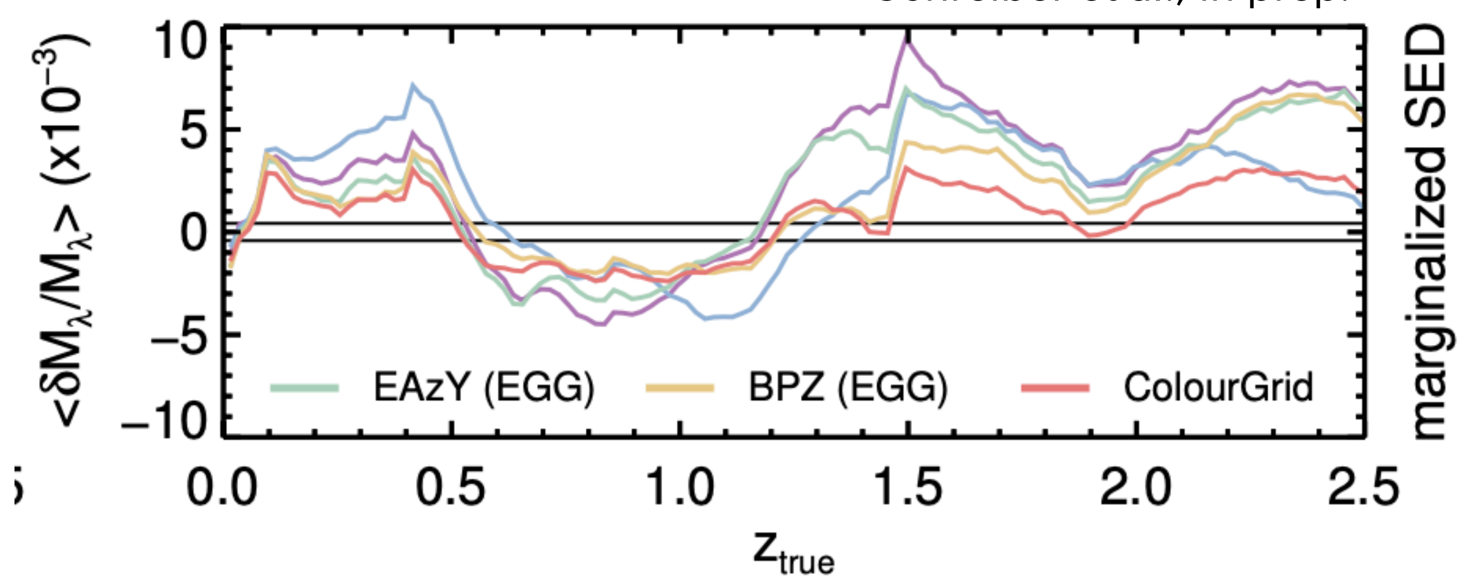
$\mathcal{M}_\lambda^i$ distributions

SED metric: bias - Tomographic bins

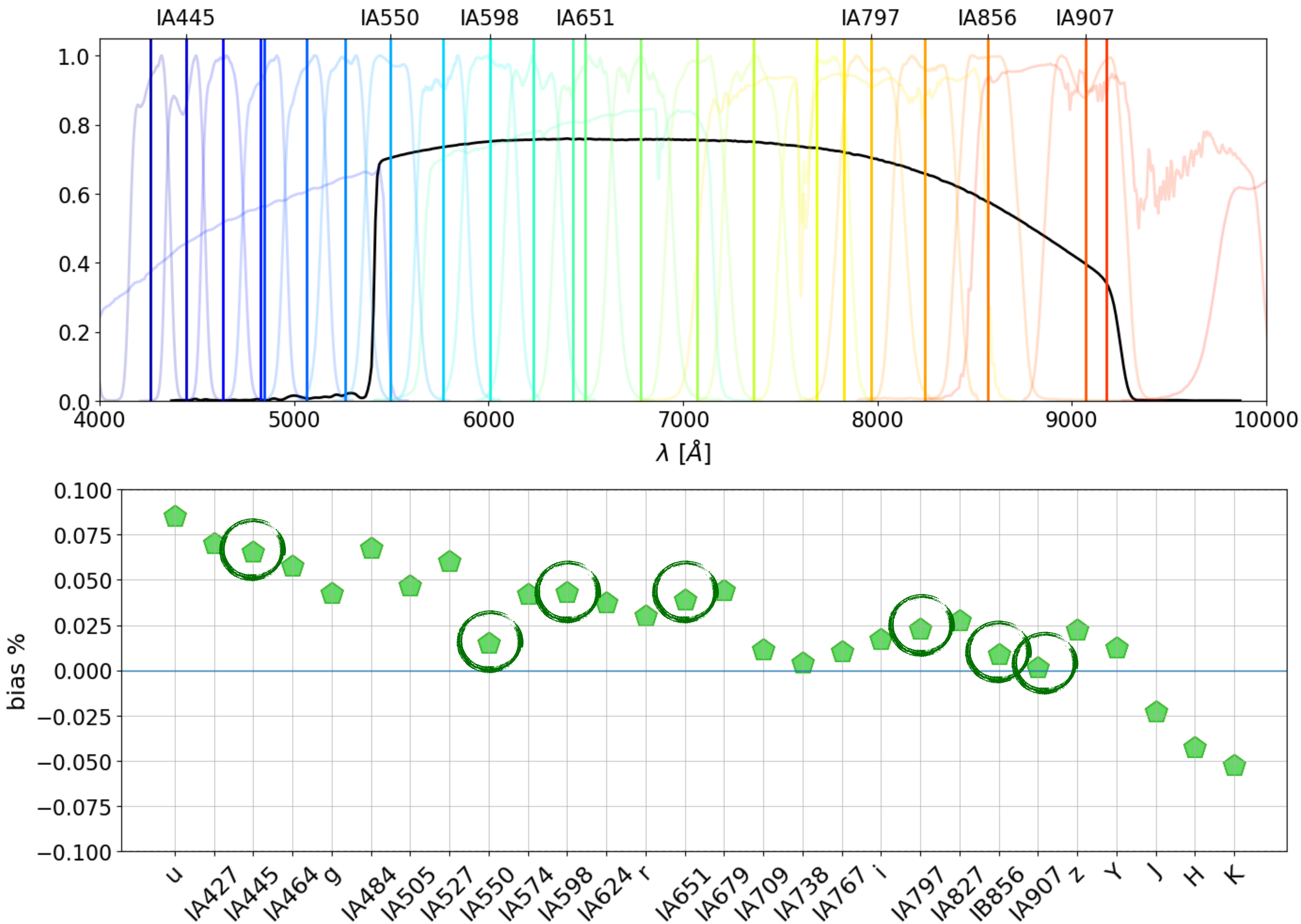


Schreiber et al., in prep.

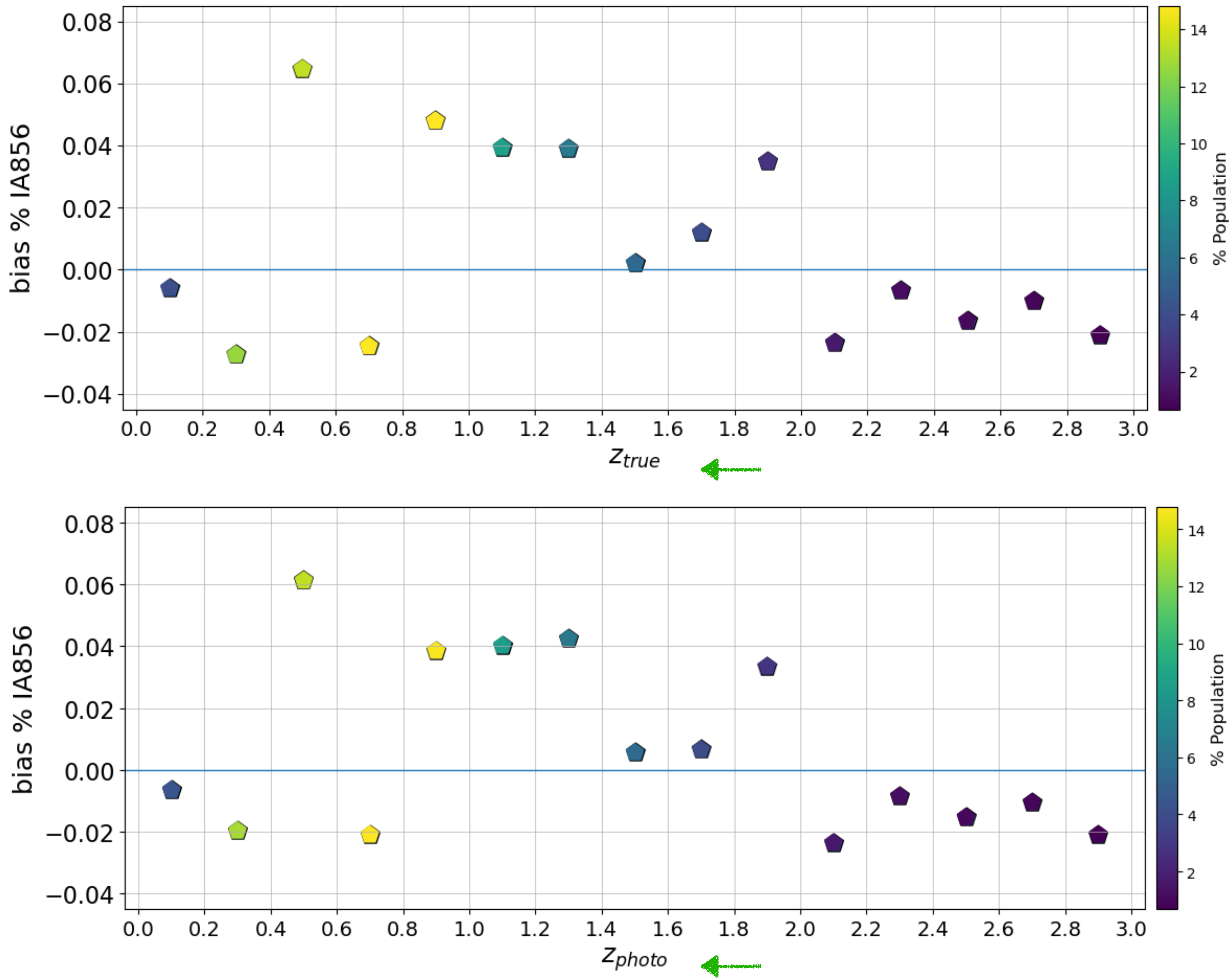
Overall passing the requirements



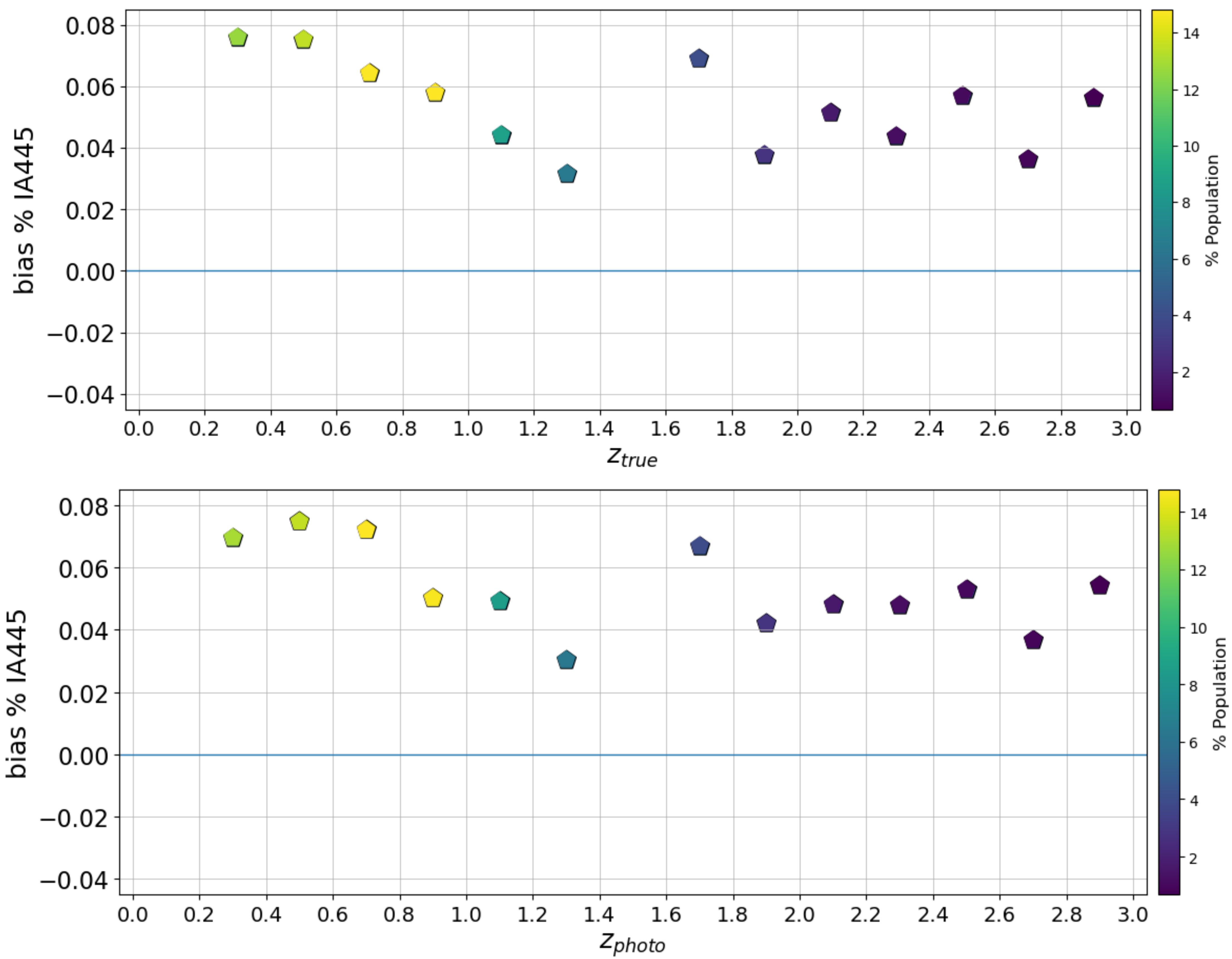
CDFS filters - Multi-band reconstruction



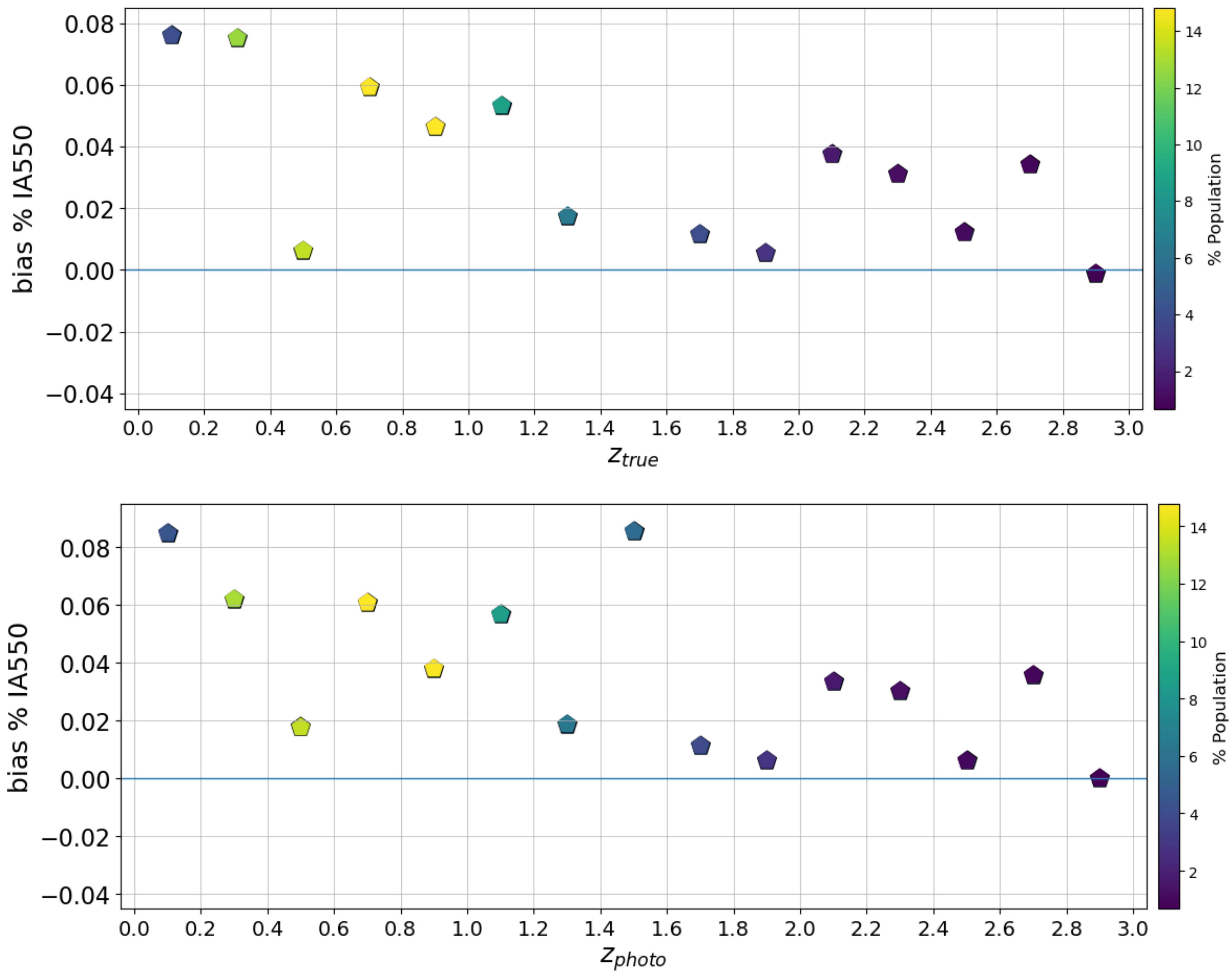
CDFS filters - tomographic bins - IA856



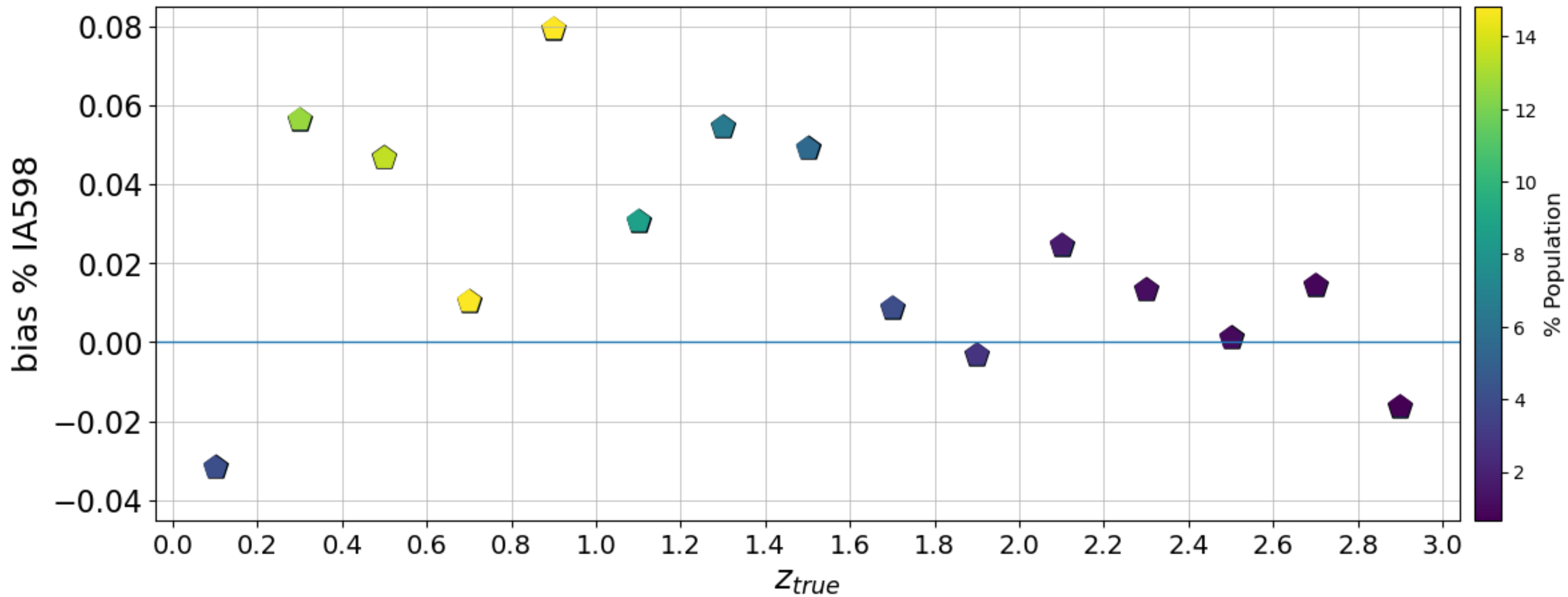
CDFS filters - tomographic bins - IA445

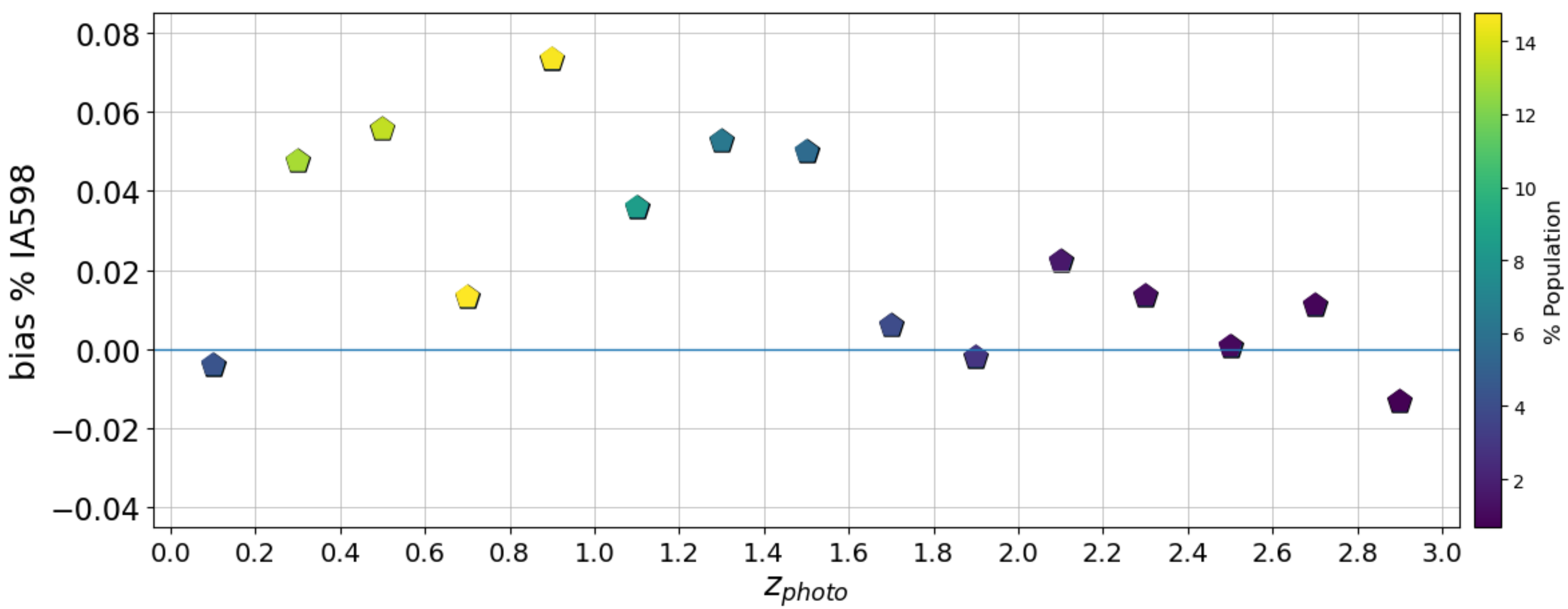


CDFS filters - tomographic bins - IA550

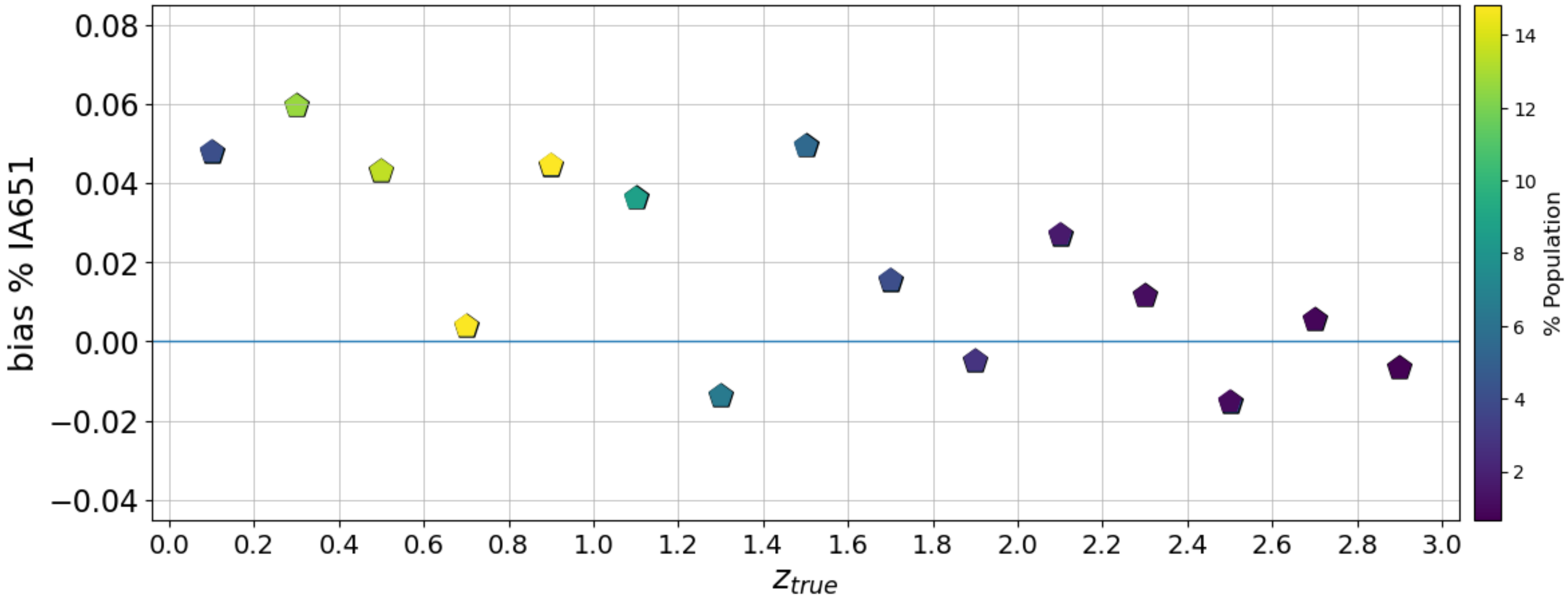


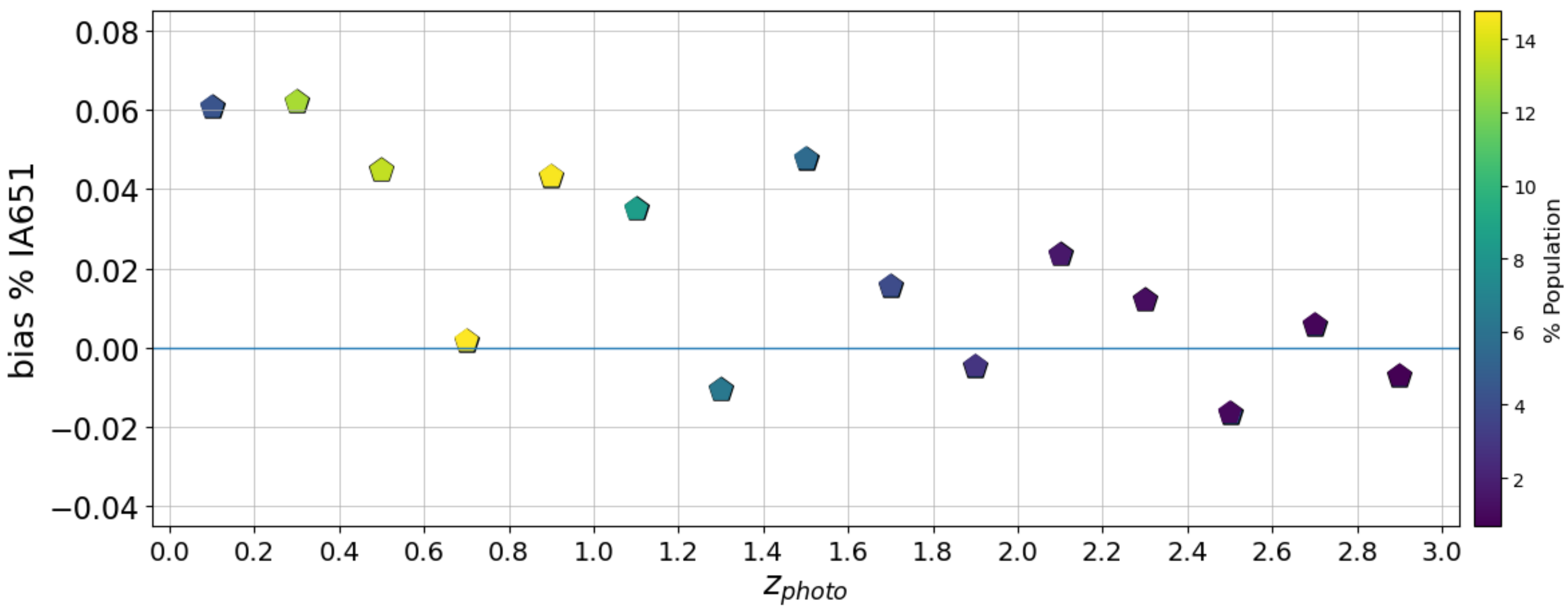
CDFS filters - tomographic bins - IA598



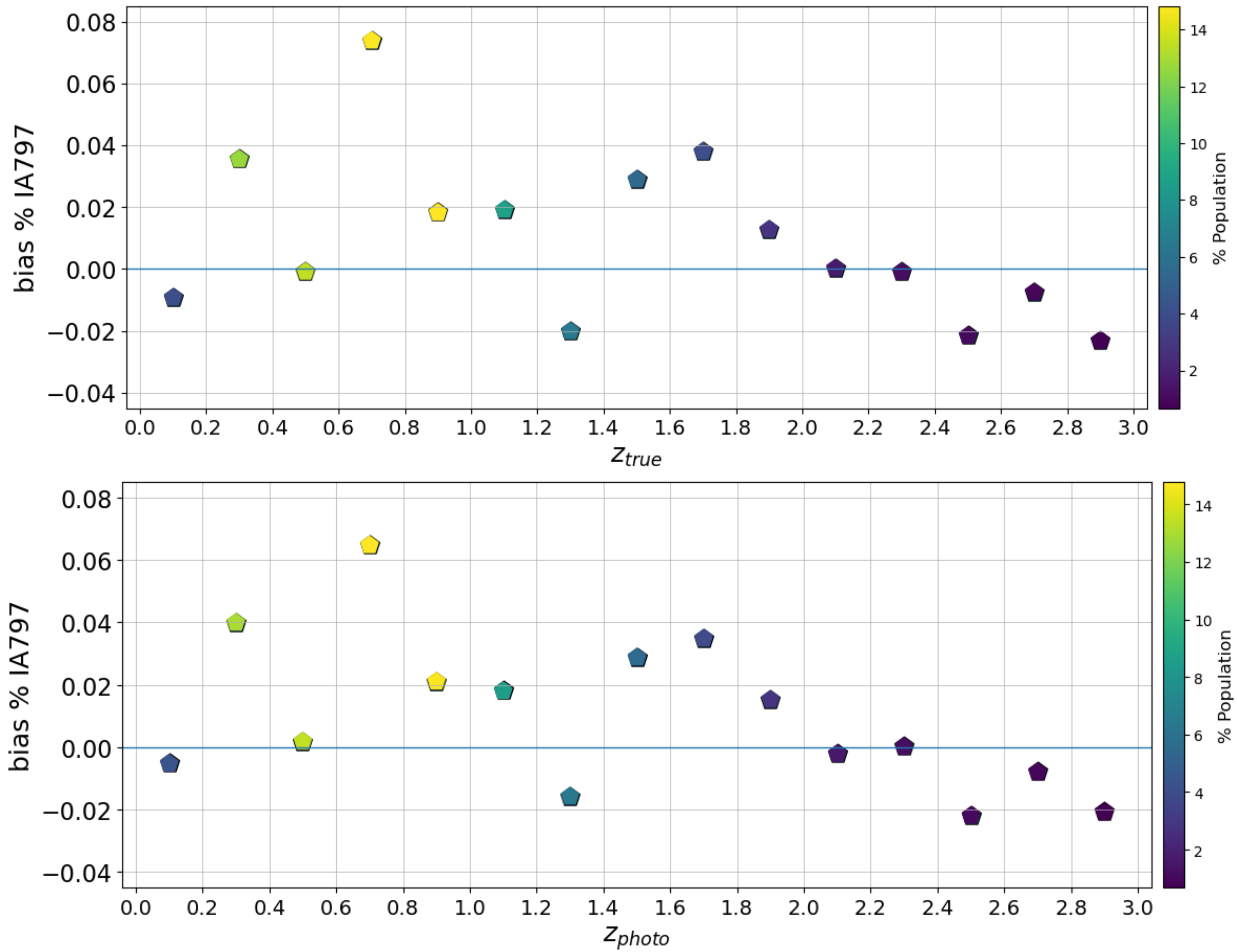


CDFS filters - tomographic bins - IA651

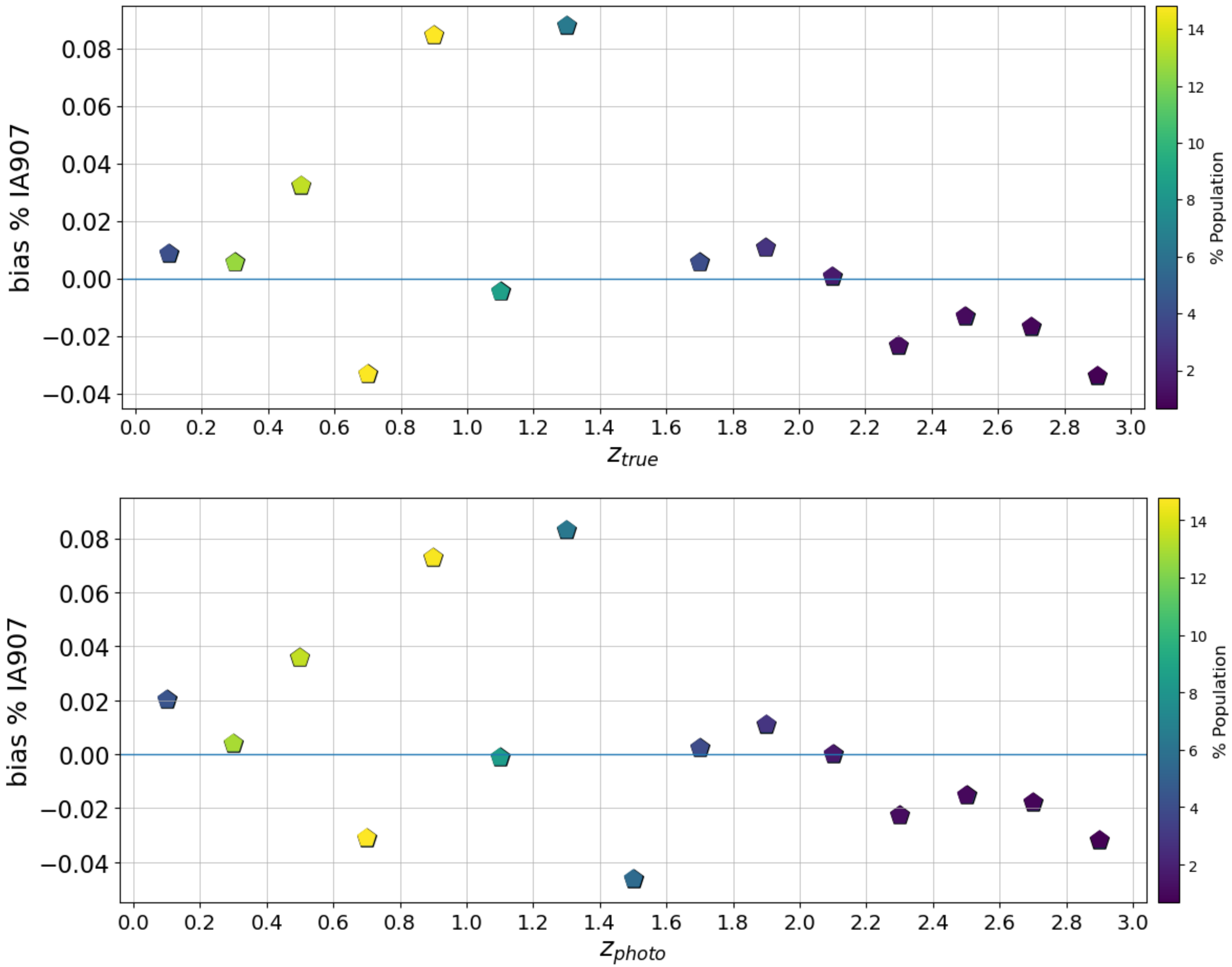




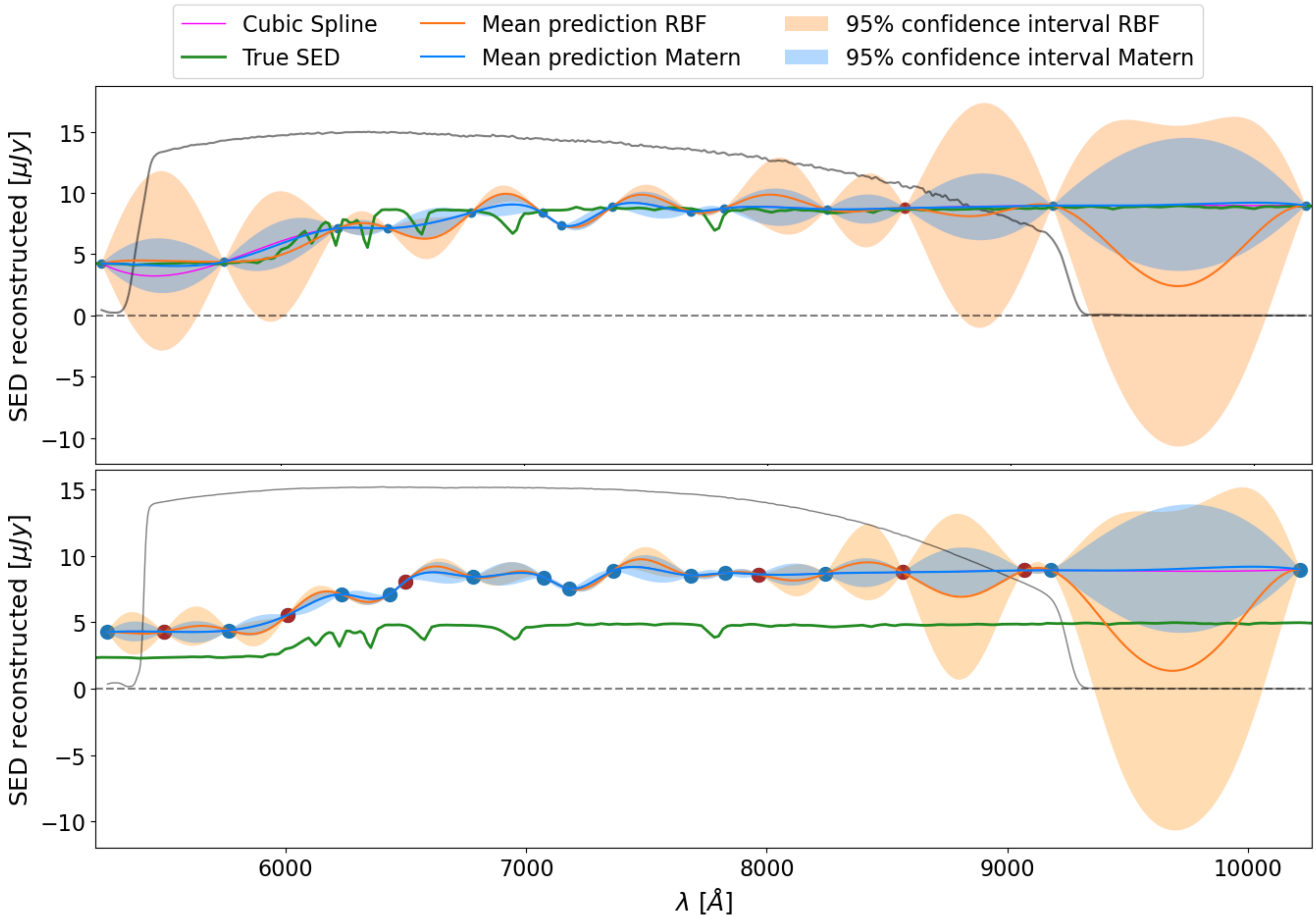
CDFS filters - tomographic bins - IA797

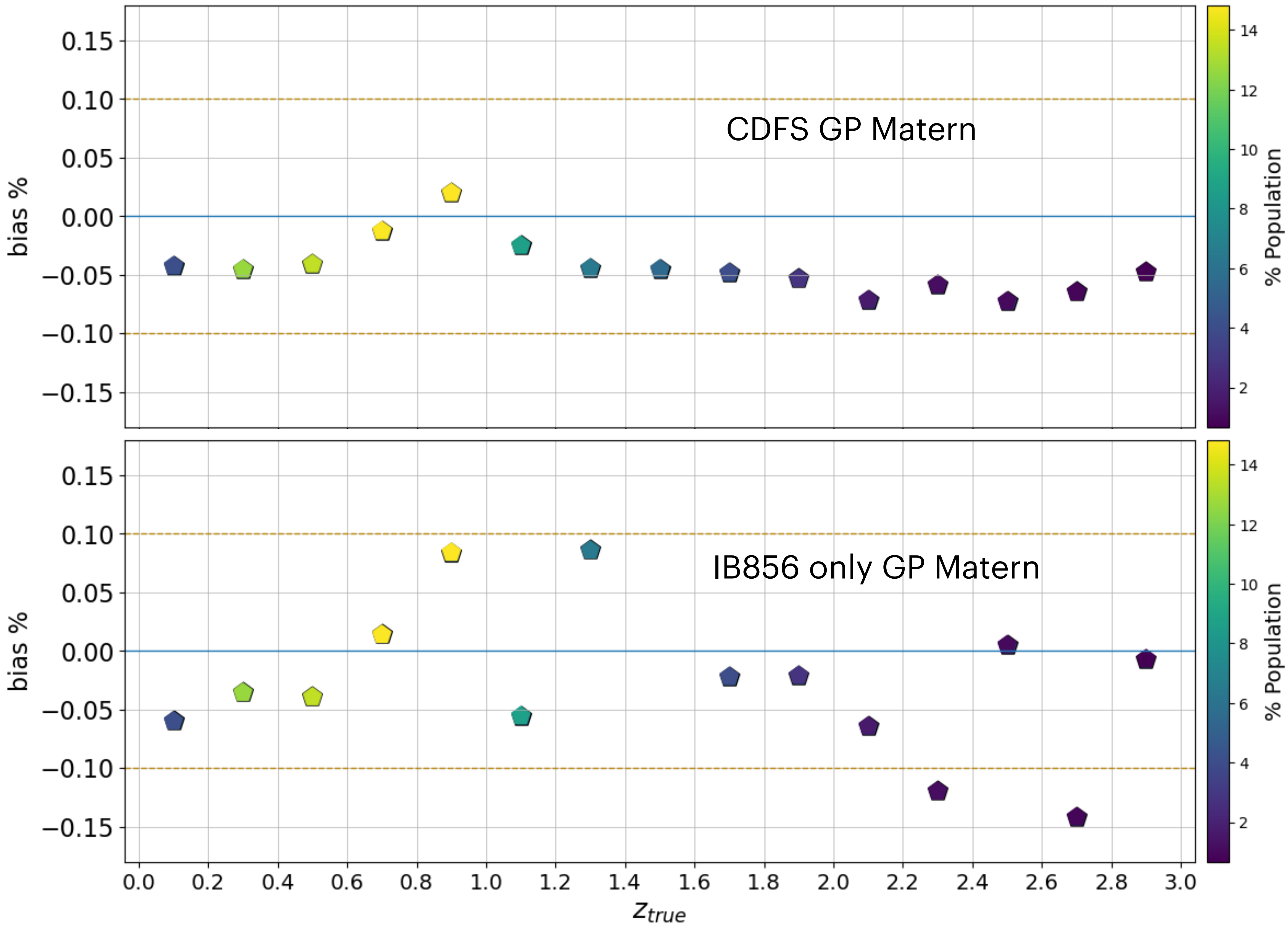


CDFS filters - tomographic bins - IA907



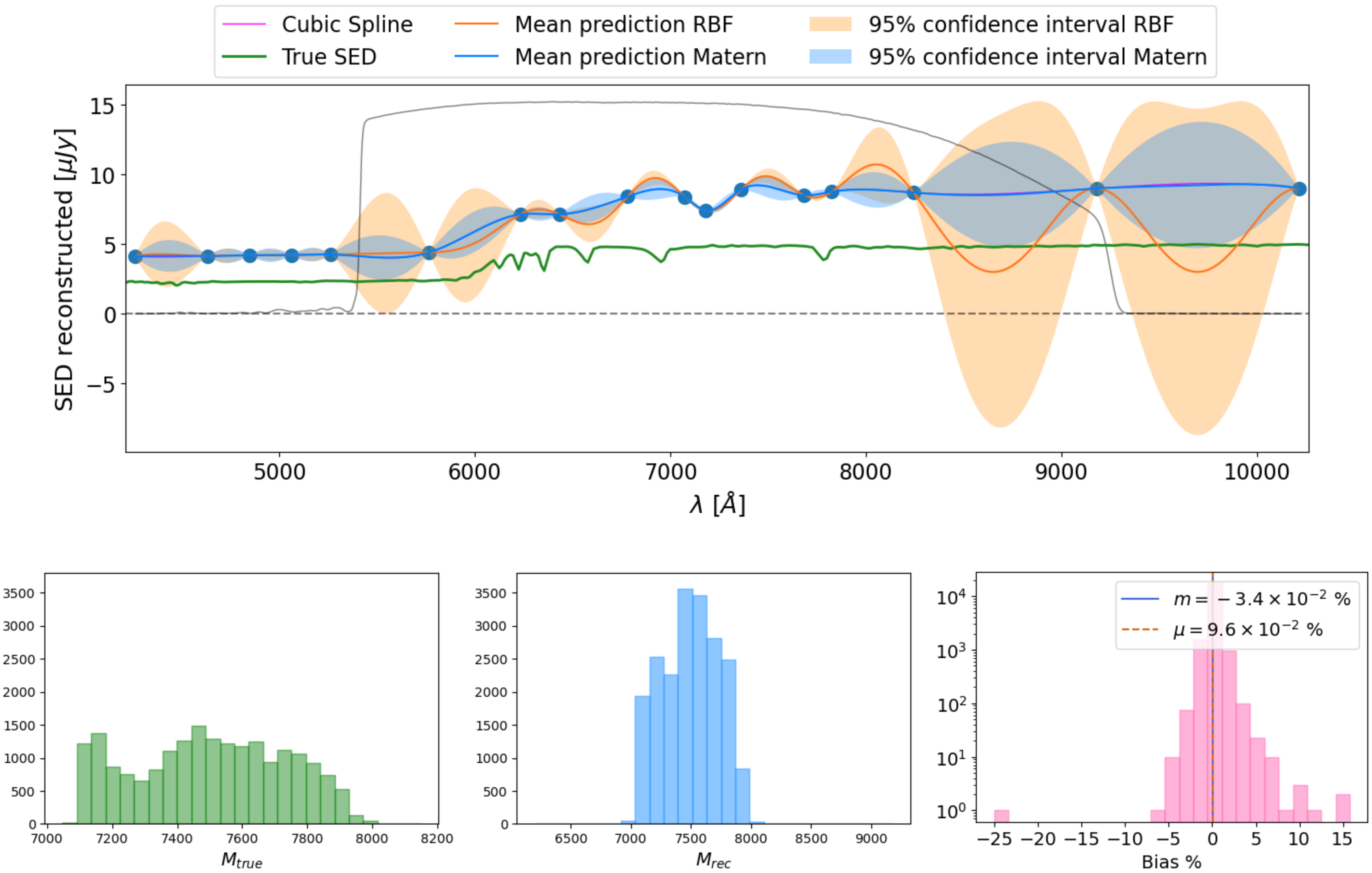
CDFS filters - SED Metric



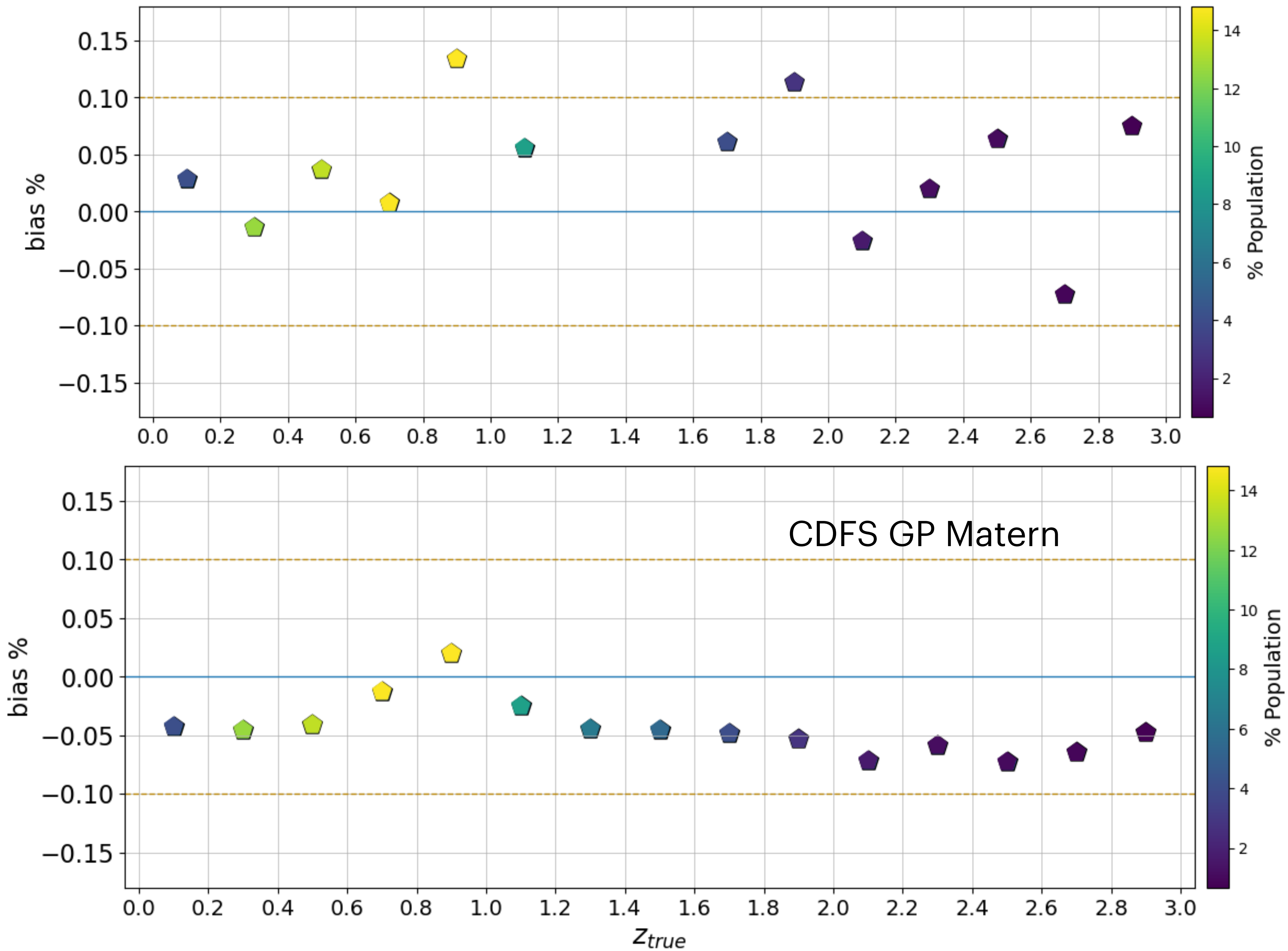


CDFS GP Matern
IB856 only GP Matern
bias %
% Population
z_{true}
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4 2.6 2.8 3.0
2 4 6 8 10 12 14

SED metric - no flux-reconstruction

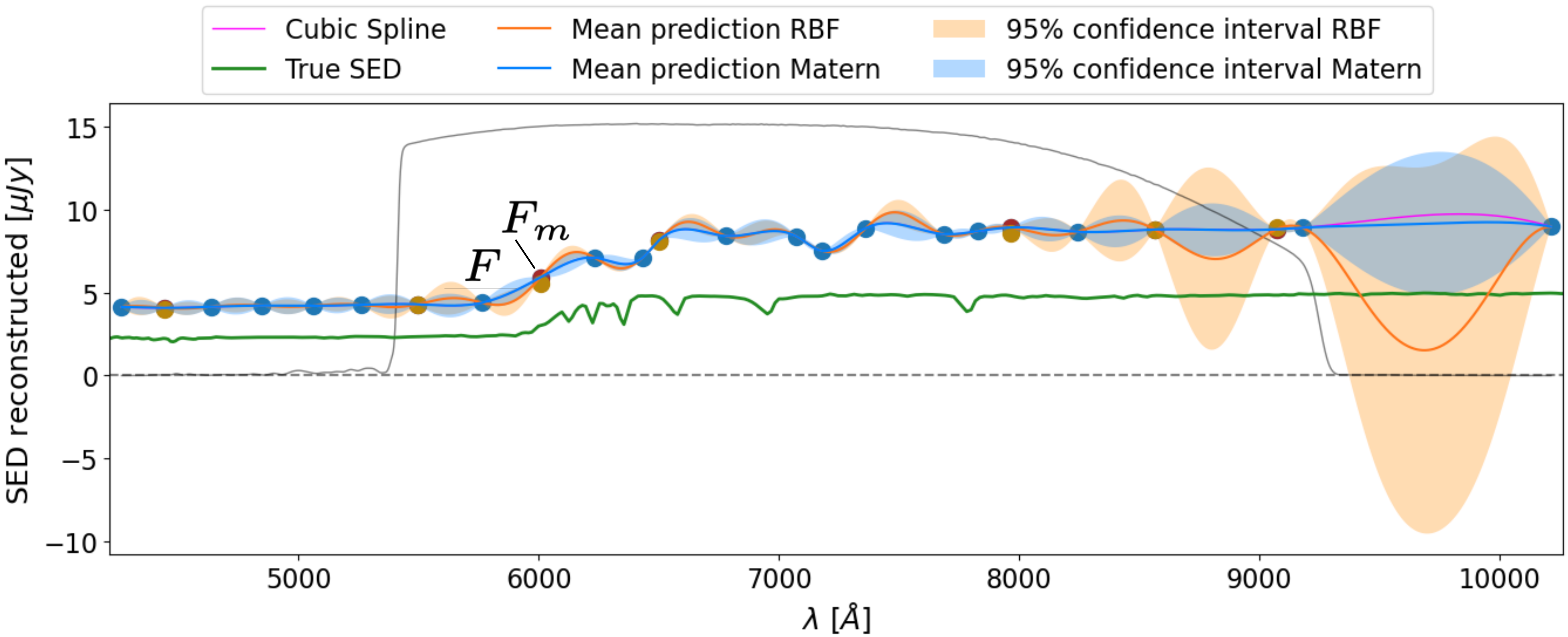


SED metric - no flux-reconstruction



CDFS reconstruction in color-space

Brown circles: fluxes reconstructed with color-space.
Golden space: fluxes reconstructed with template fitting.



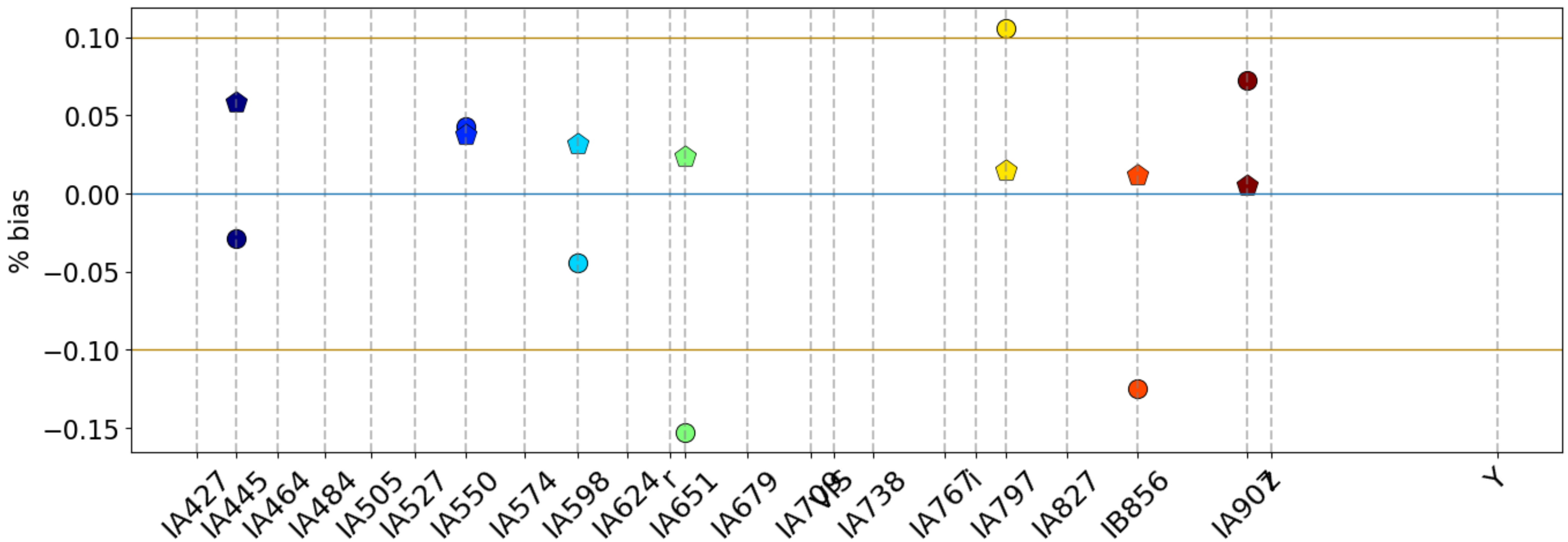
flux-ratio in color space

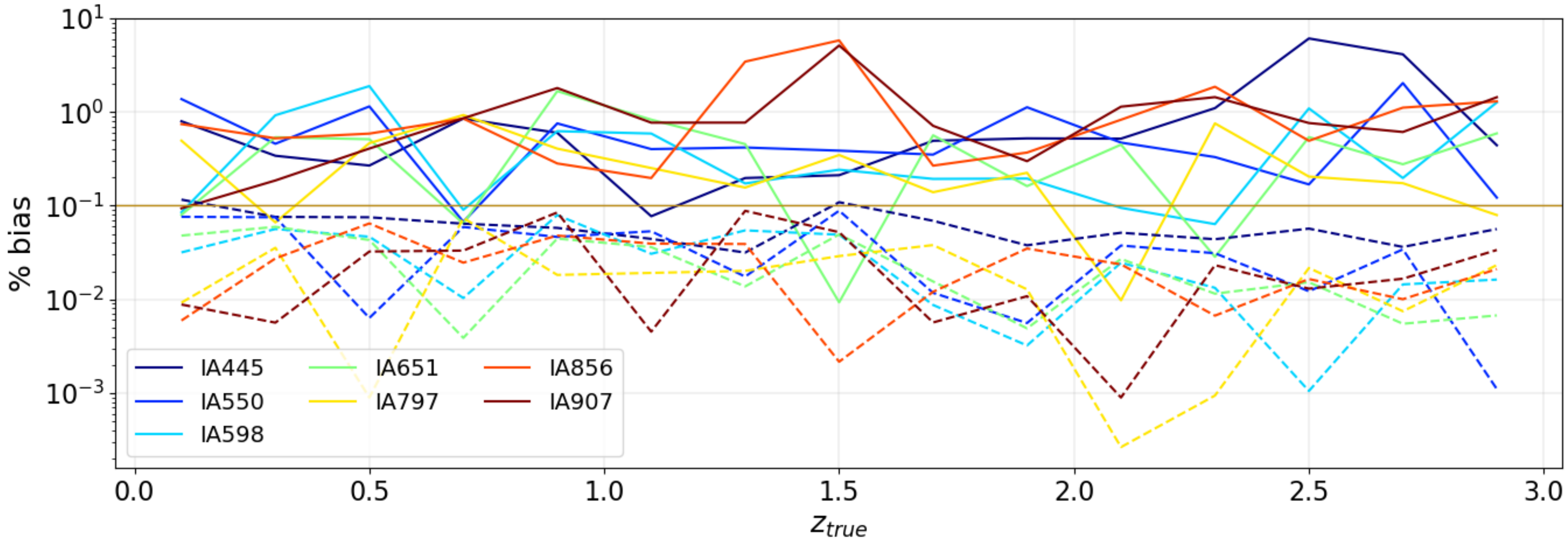
$$\frac{F}{F_m} = \frac{F^P}{F_m^P}$$

Posterior flux-ratio

CDFS reconstruction in color-space

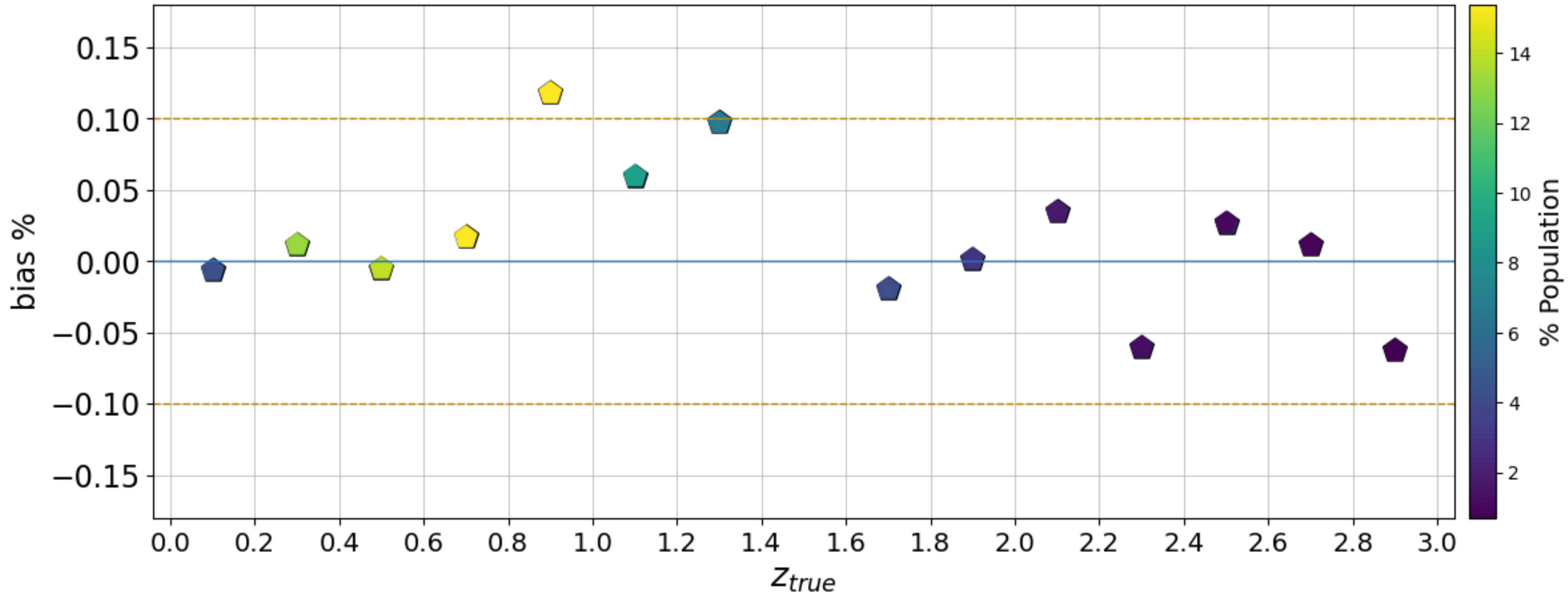
Scatter points are the median of the distributions previously showed.
Circles: reconstruction in color-space. Pentagons: reconstruction with template fitting.

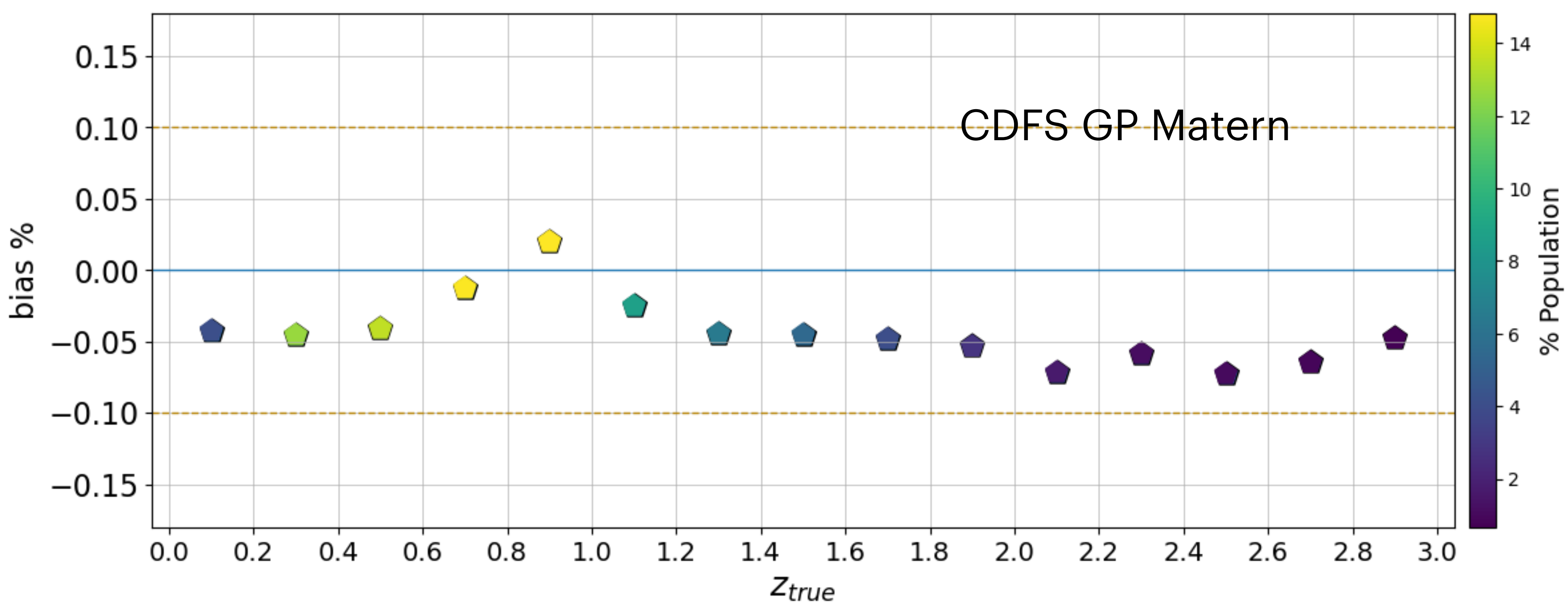




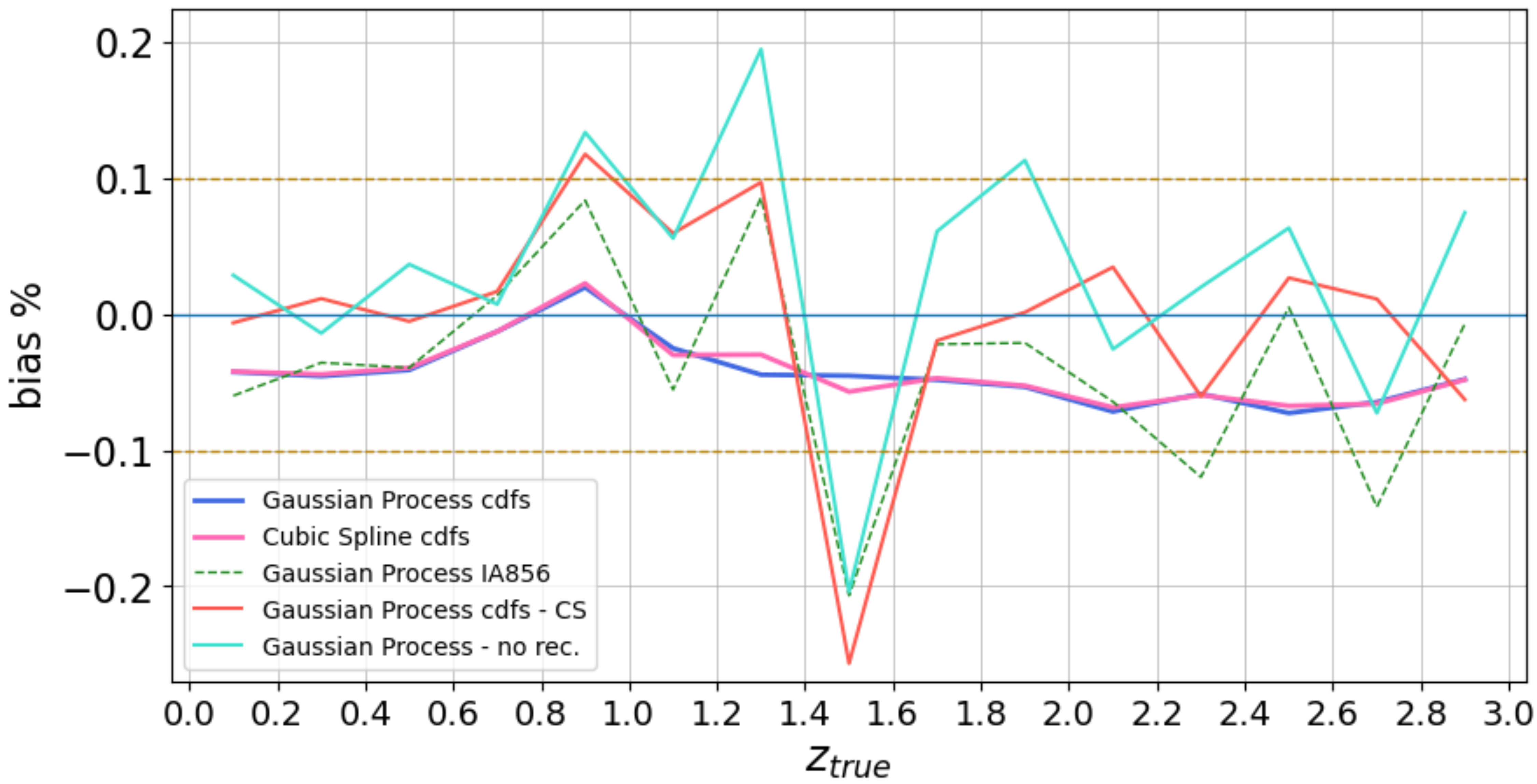
Solid lines: fluxes are reconstructed in color-space.
Dashed lines: fluxes are reconstructed with template fitting.

CDFS reconstruction in color-space - SED Metric - z



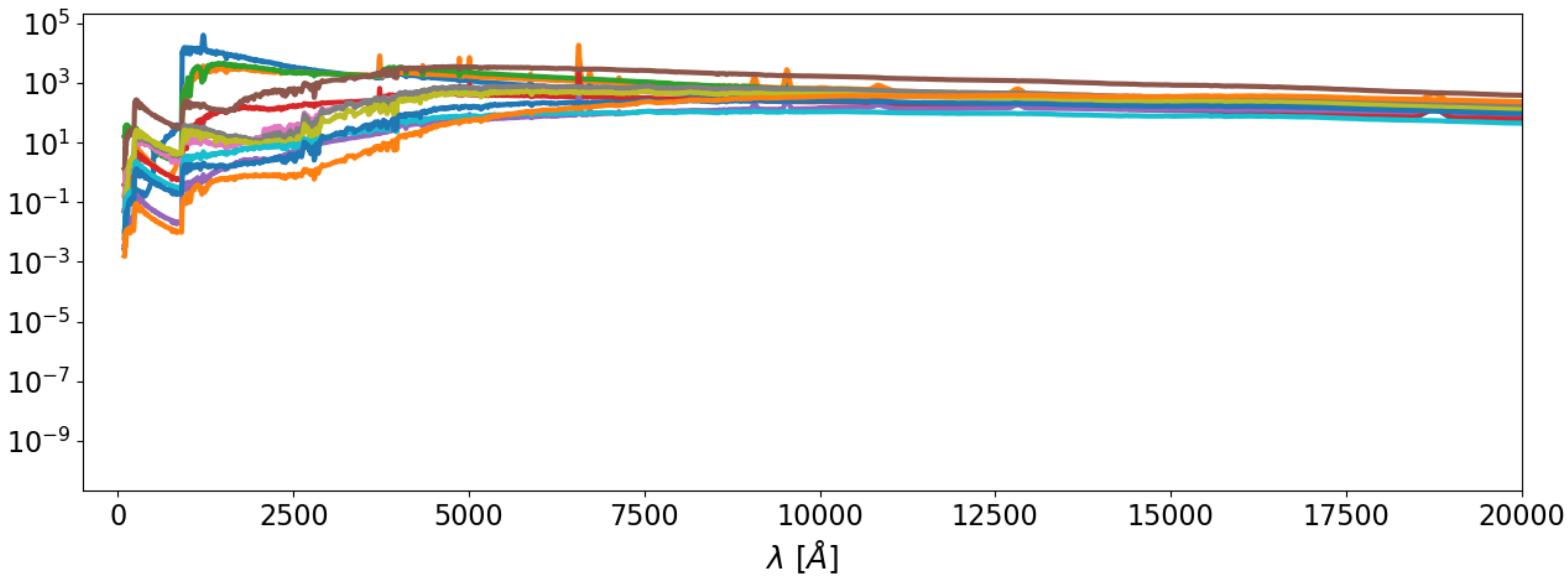


Tomographic SED Metric - comparison between reconstruction methods

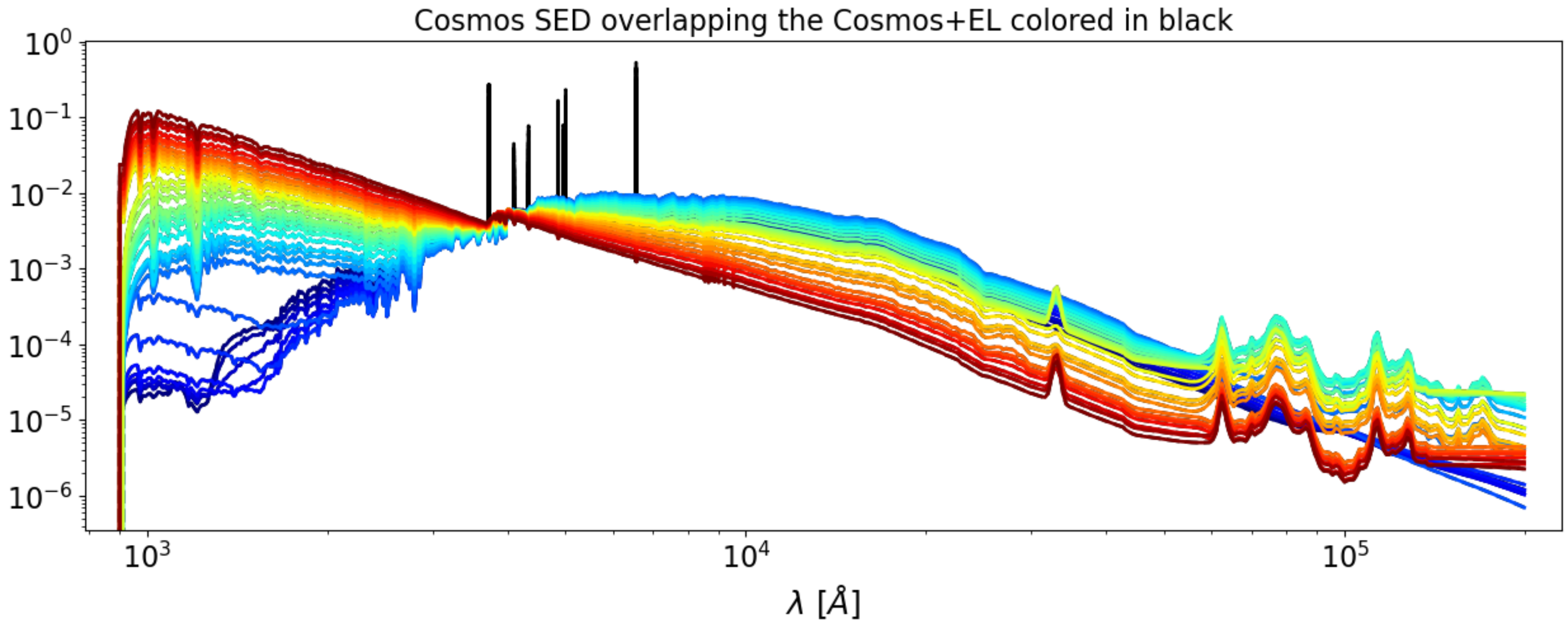


New catalogue generated with linearly combined SEDs

LC_P15_SIMS_cdfs_subsample - 31 Cosmos SEDs with Phosphoros-like emission lines
SEDs in the folder COSMOS_el

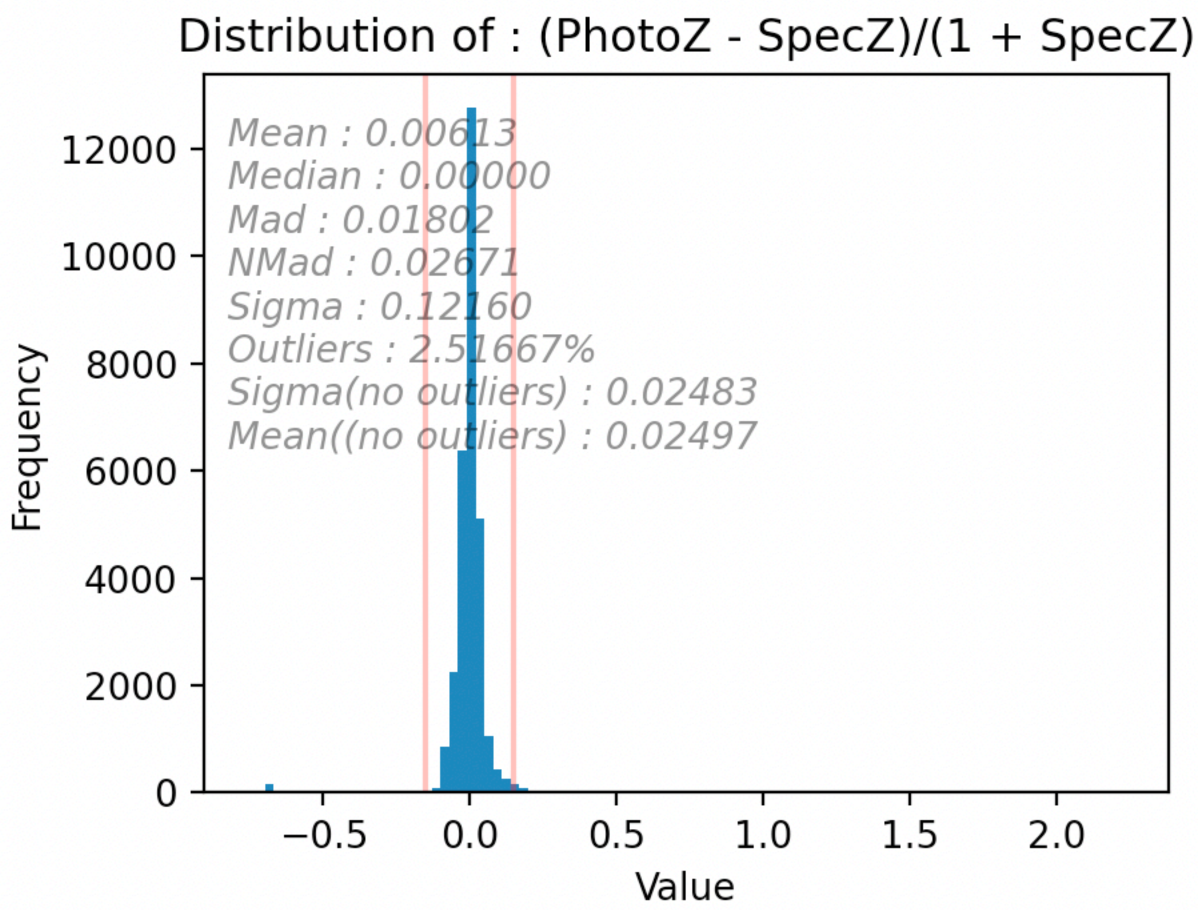


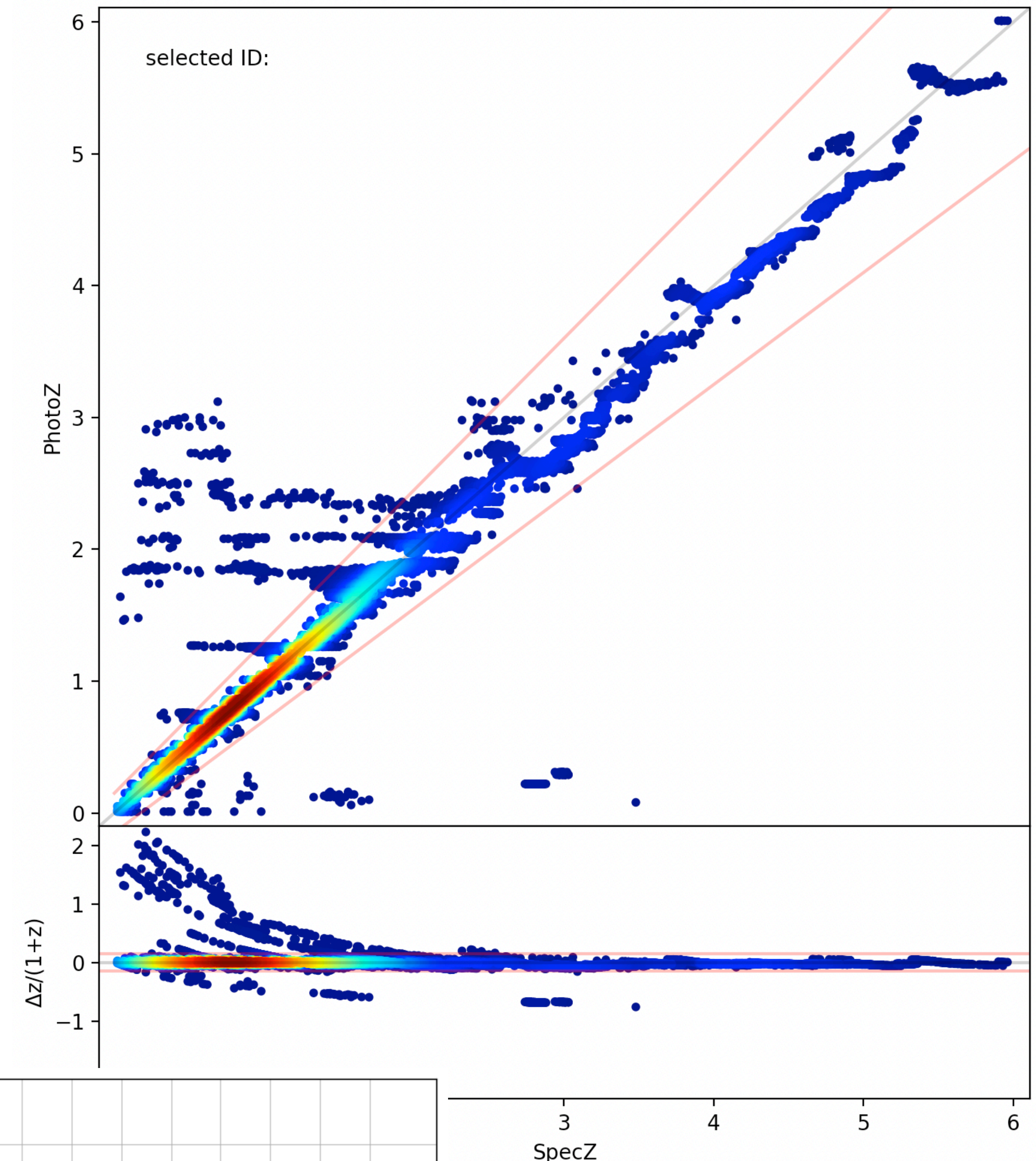
New simulations

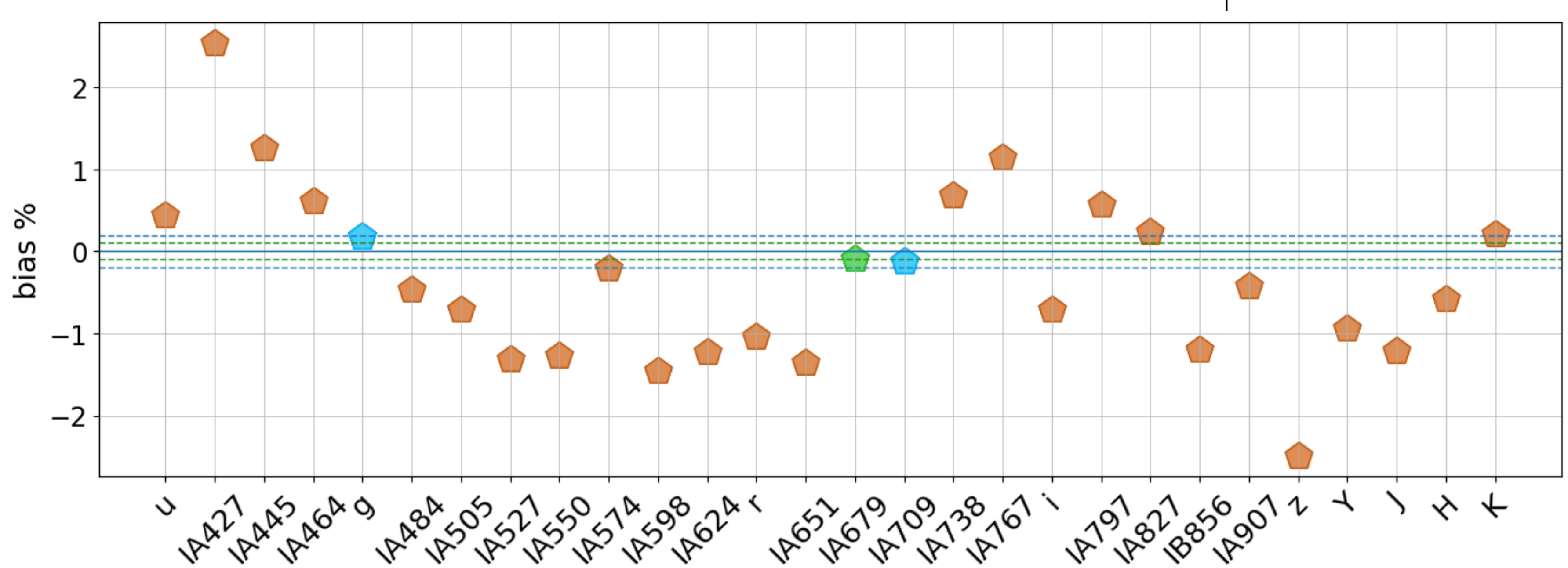


Cosmos SED + EL used for Phosphoros template fitting

Adding Reddening and using linearly interp. SEDs



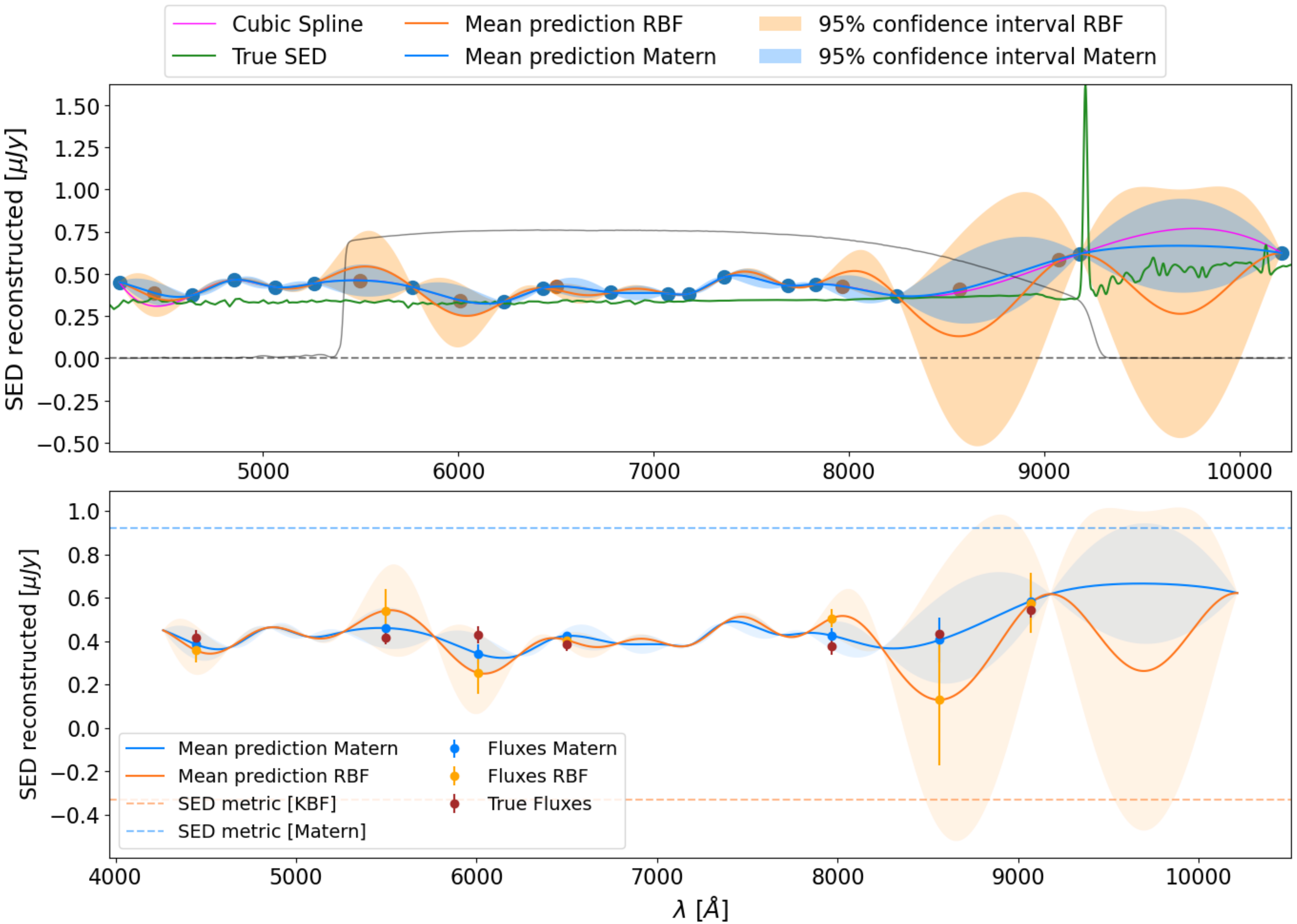




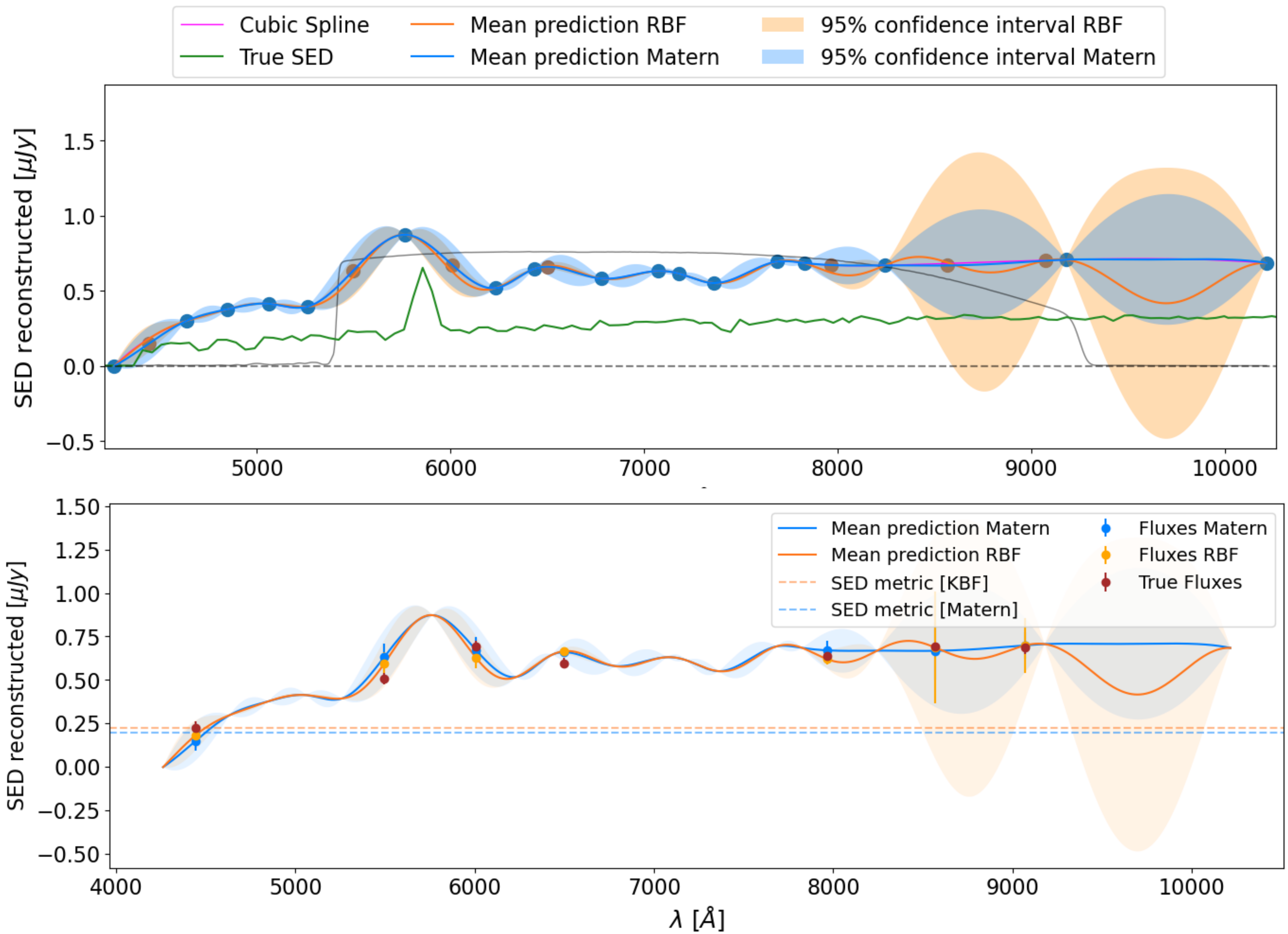
SED used to fit

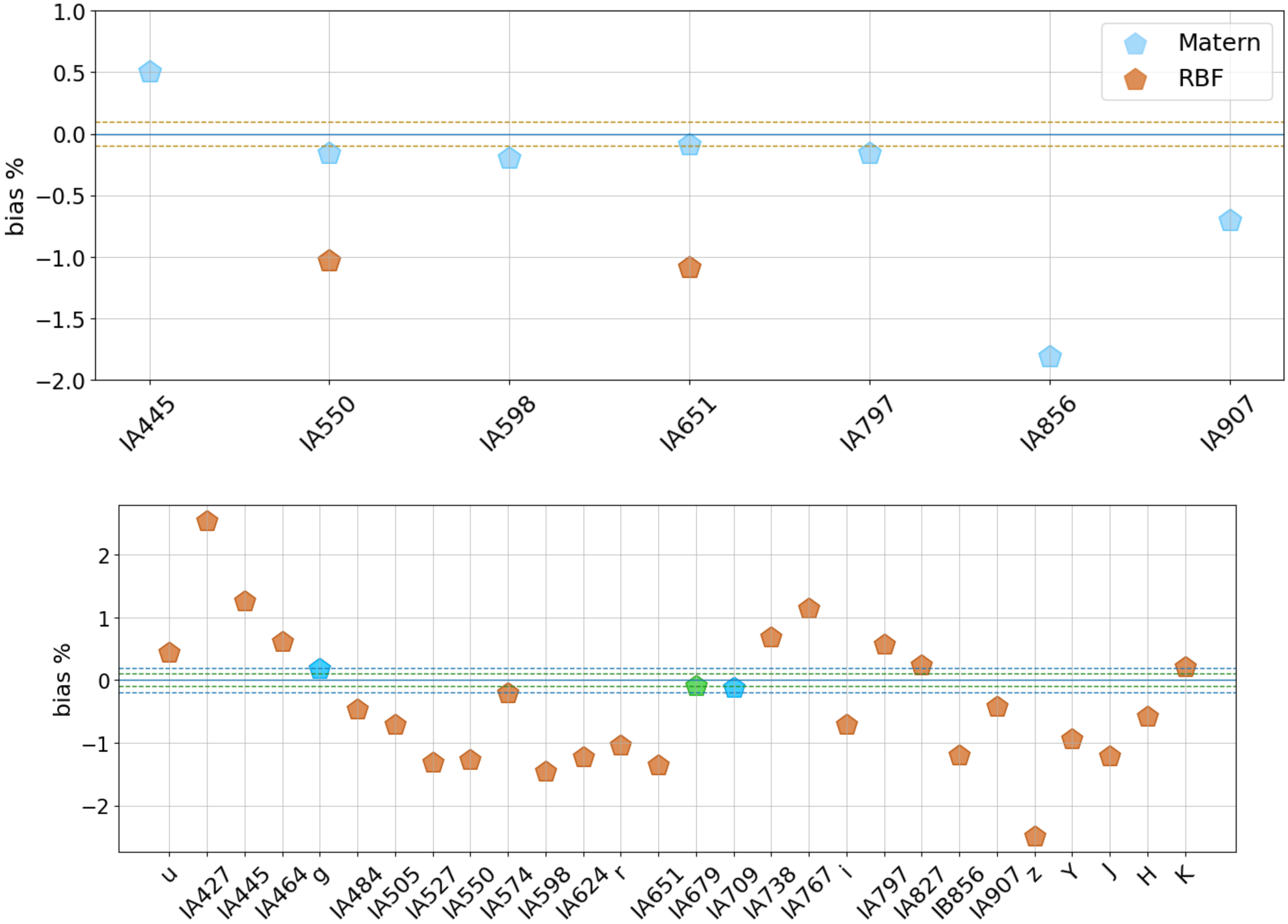
Template Fitting with Linearly Interp. SED COSMOS + Emission Lines

Reconstruction solely with GP - Matern nu=1.5 vs RBF



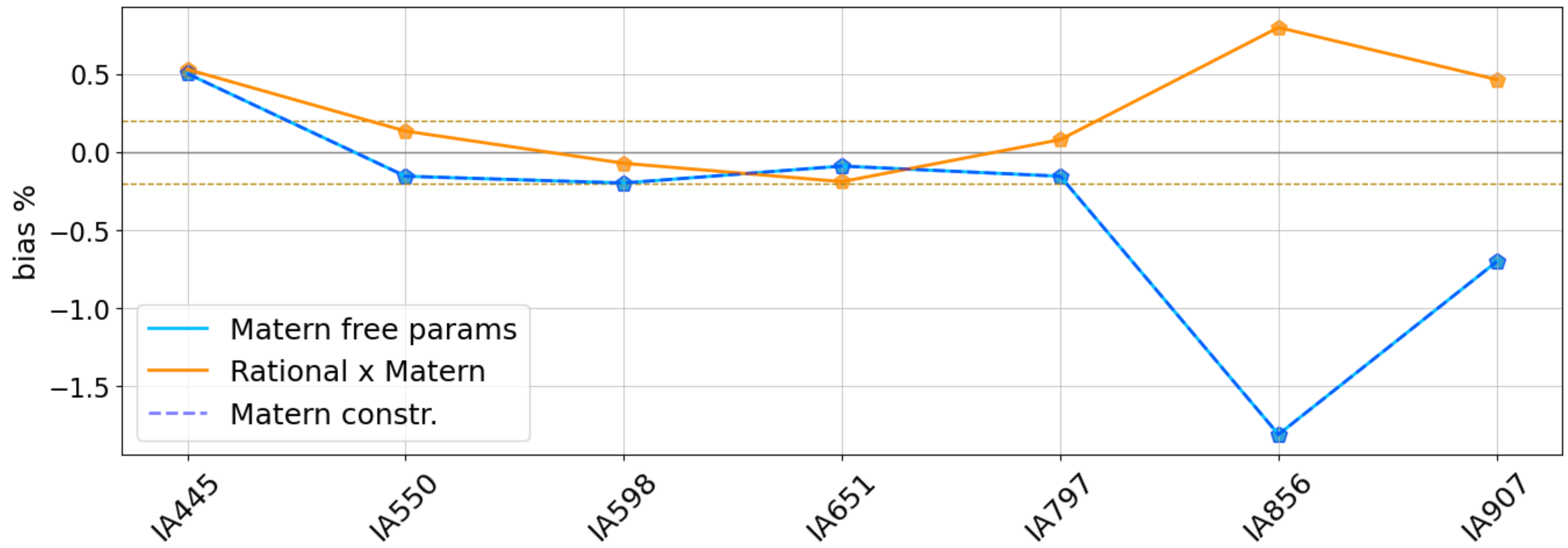
Reconstruction solely with GP - Matern nu=1.5 vs RBF

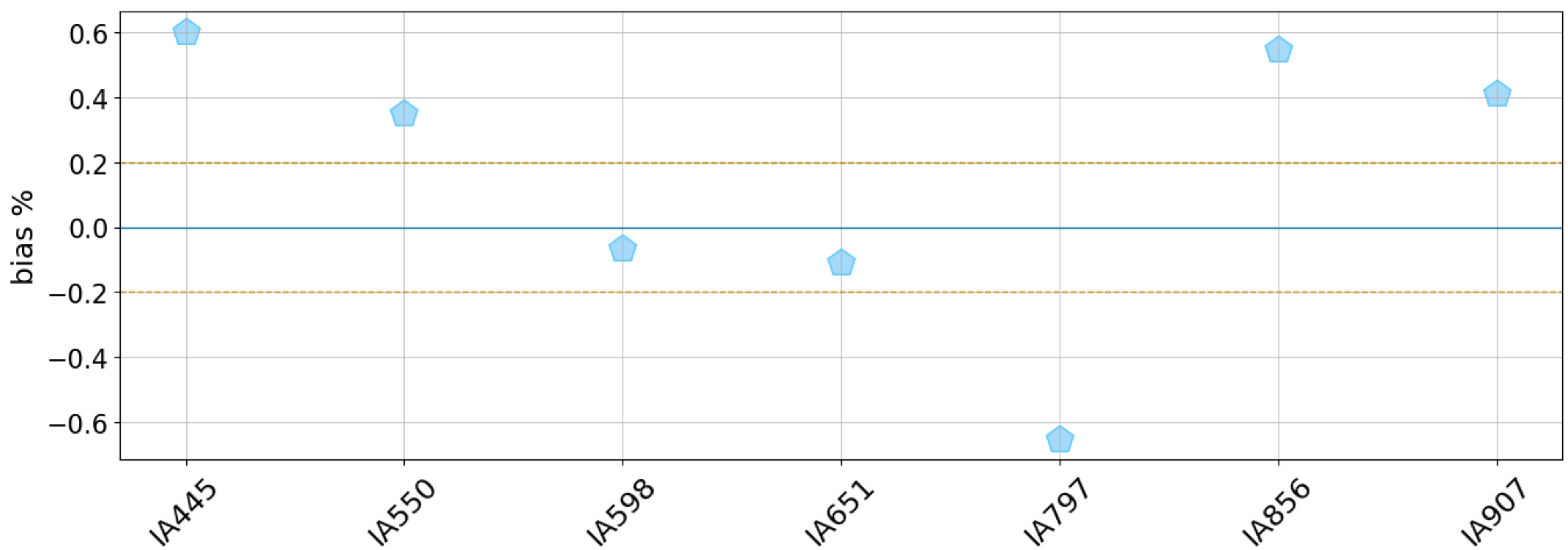




Matern
RBF
bias %
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
IA445
IA550
IA598
IA651
IA797
IA856
IA907
bias %
2
1
0
−1
−2
u
IA427
IA445
IA464
g
IA484
IA505
IA527
IA550
IA574
IA598
IA624
r
IA651
IA679
IA709
IA738
IA767
i
IA797
IA827
IB856
IA907
z
Y
J
H
K

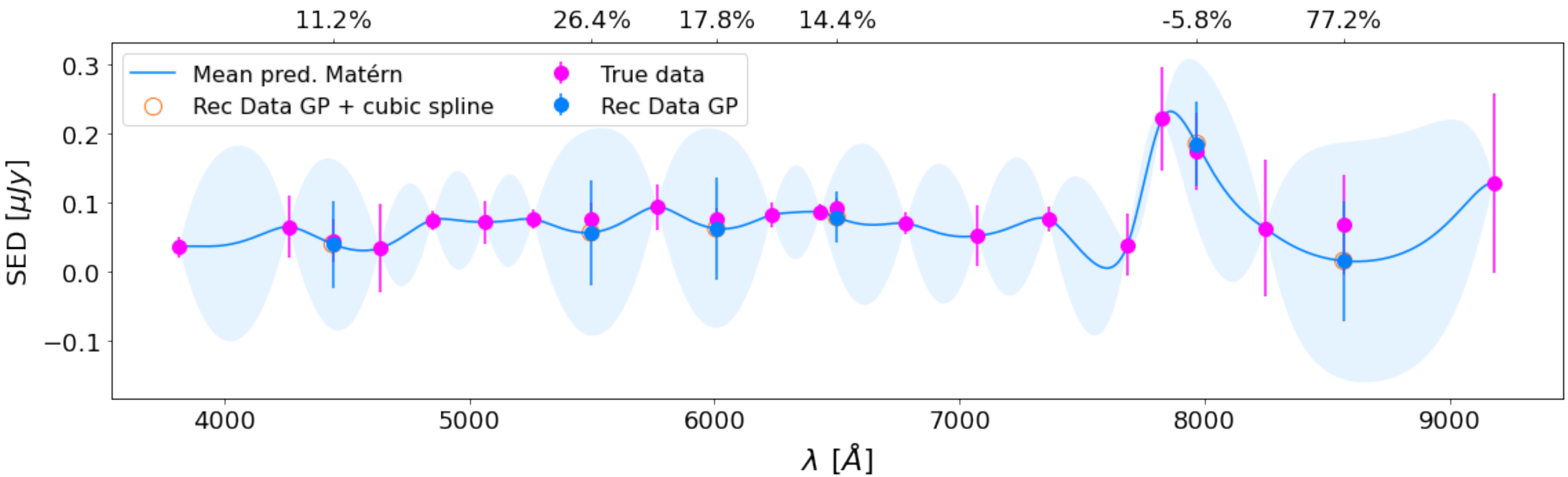
Reconstruction solely with GP - % bias on single fluxes
RationalQuadratic Kernel x Matern

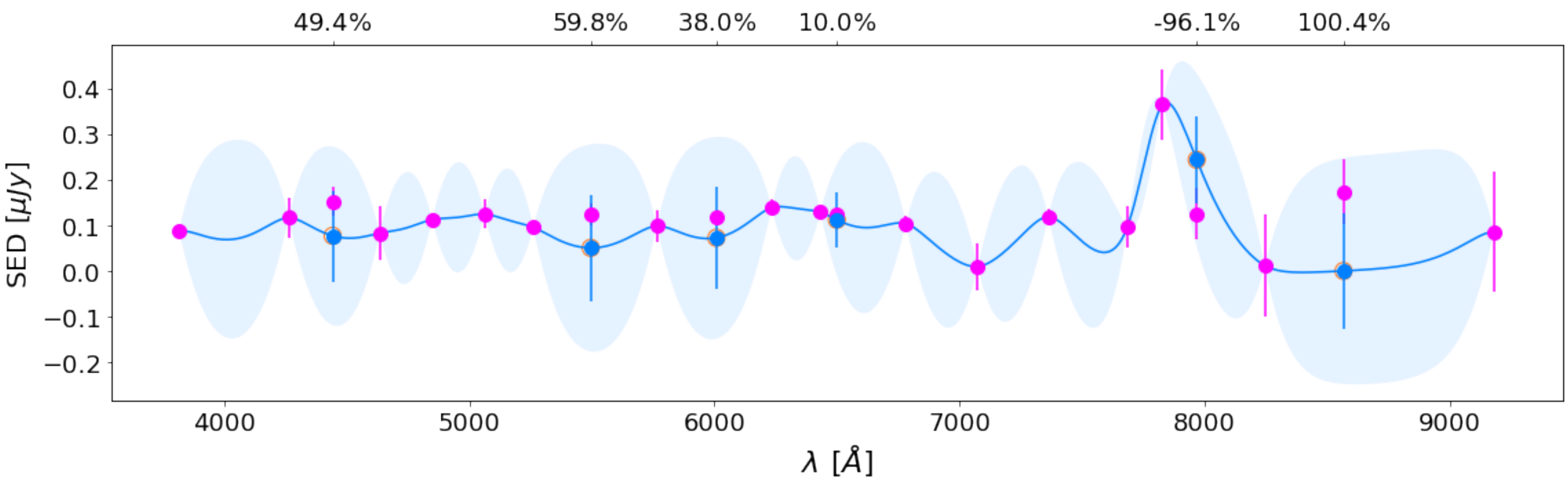




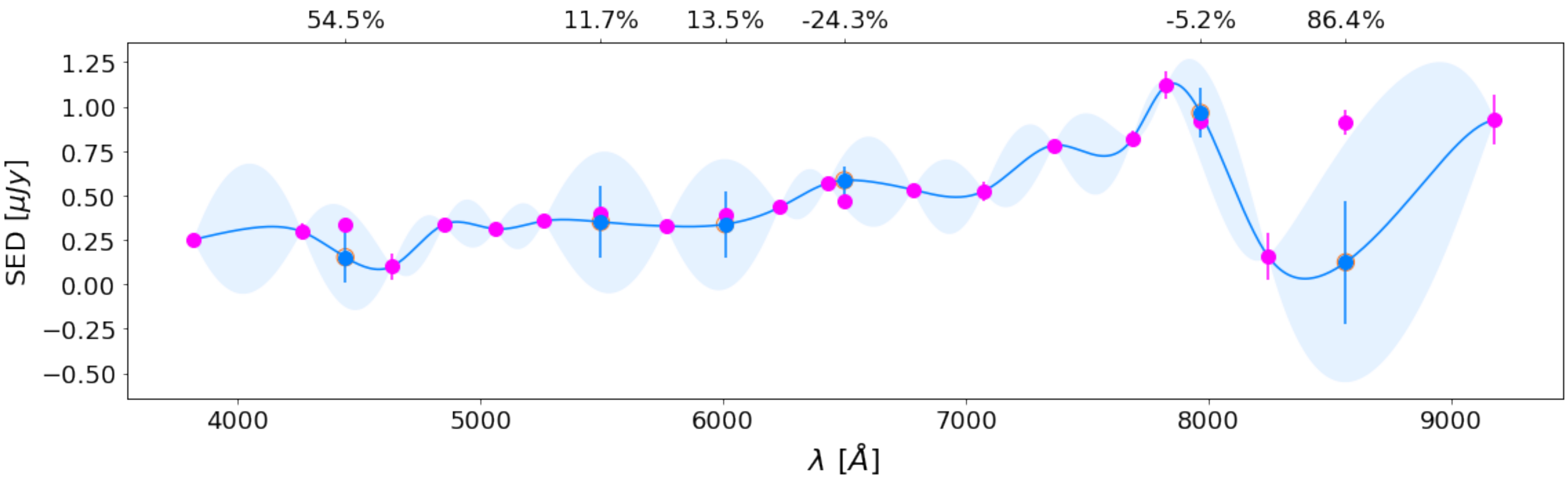
CDFS catalogue - GP without errors

MUSYC - Cardamone et al., 2010





CDFS catalogue - GP without errors

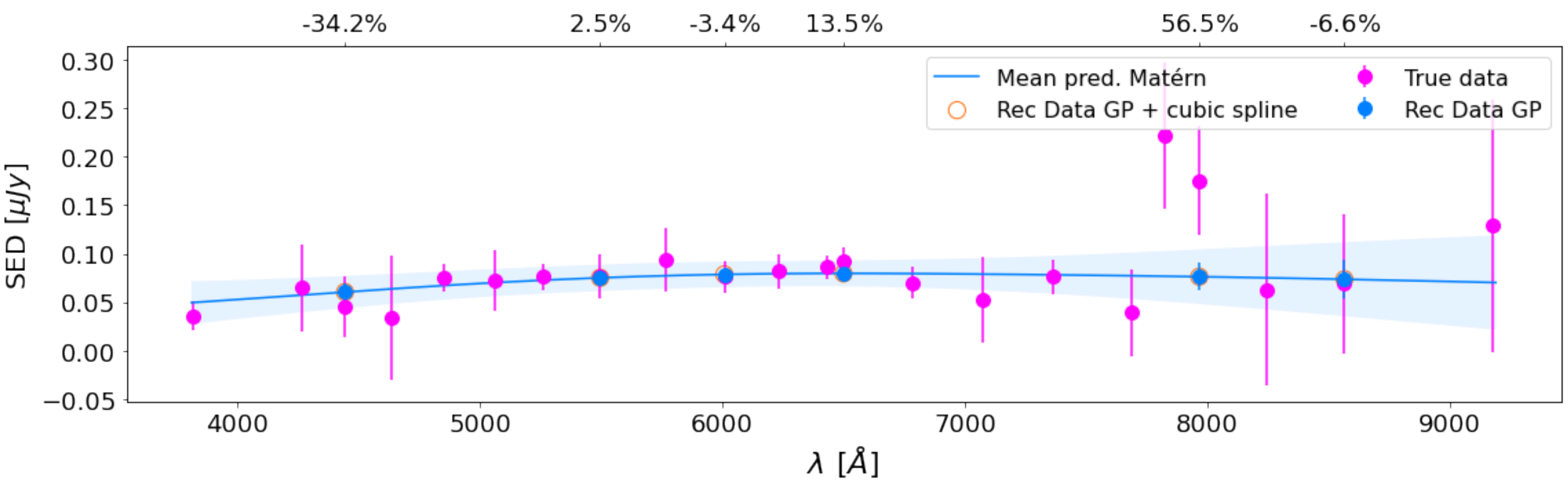


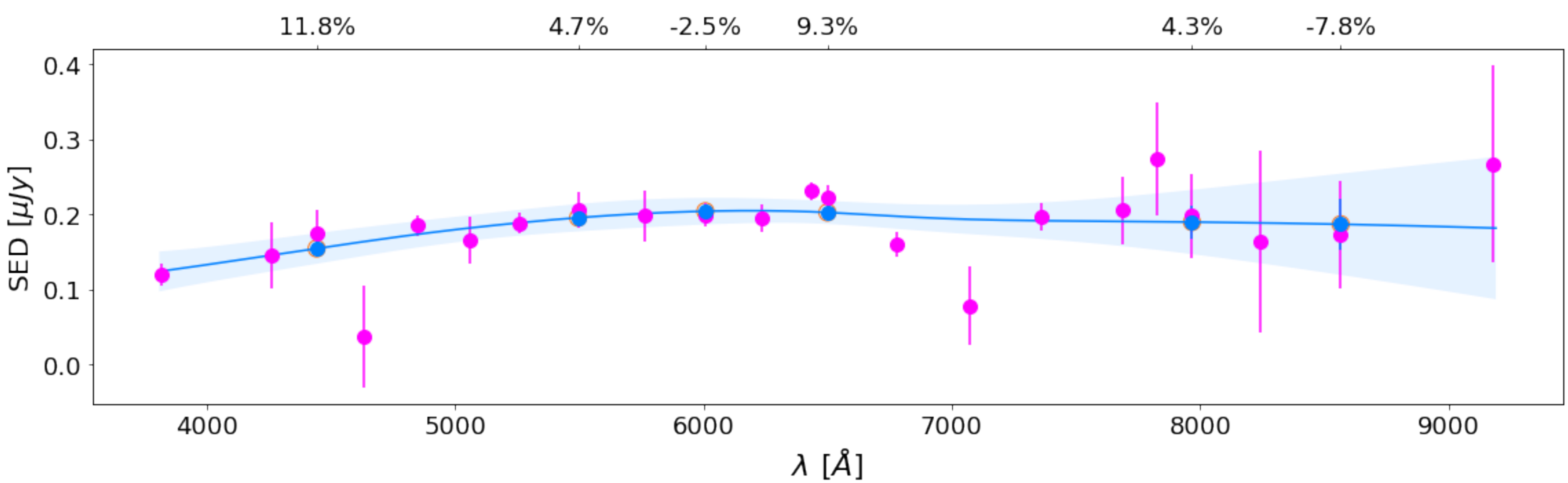
Median bias in each reconstructed filter

GP [%]

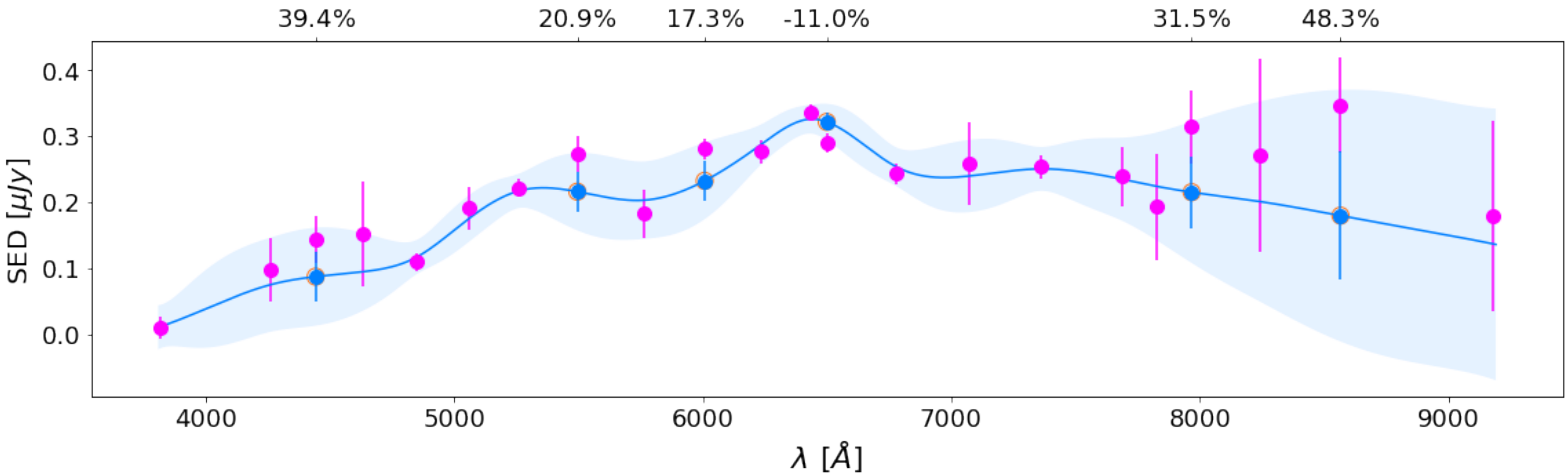
IA445	3.29
IA550	13.68
IA598	−0.93
IA651	−5.28
IA797	−8.95
IA856	32.24

CDFS catalogue - GP with errors





CDFS catalogue - GP with errors

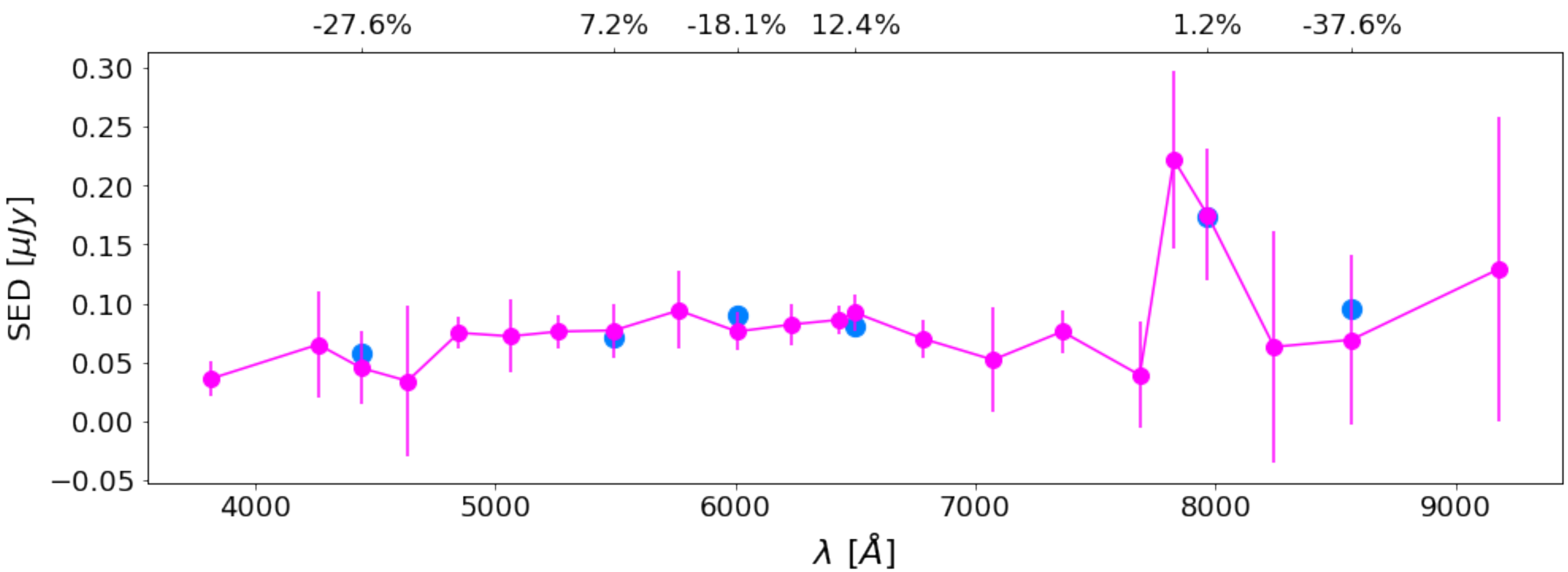


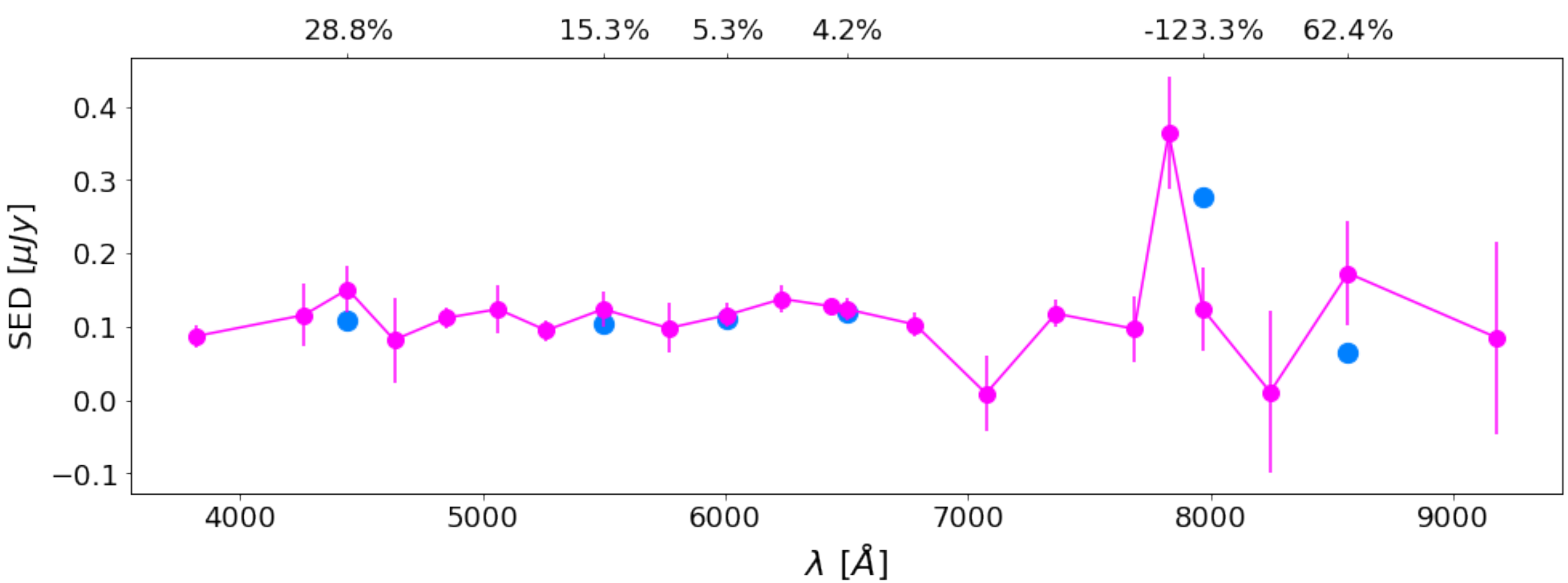
Median bias in each reconstructed filter

GP-err [%]	
IA445	1.16
IA550	8.05
IA598	−2.97
IA651	−3.7
IA797	0.97
IA856	15.81

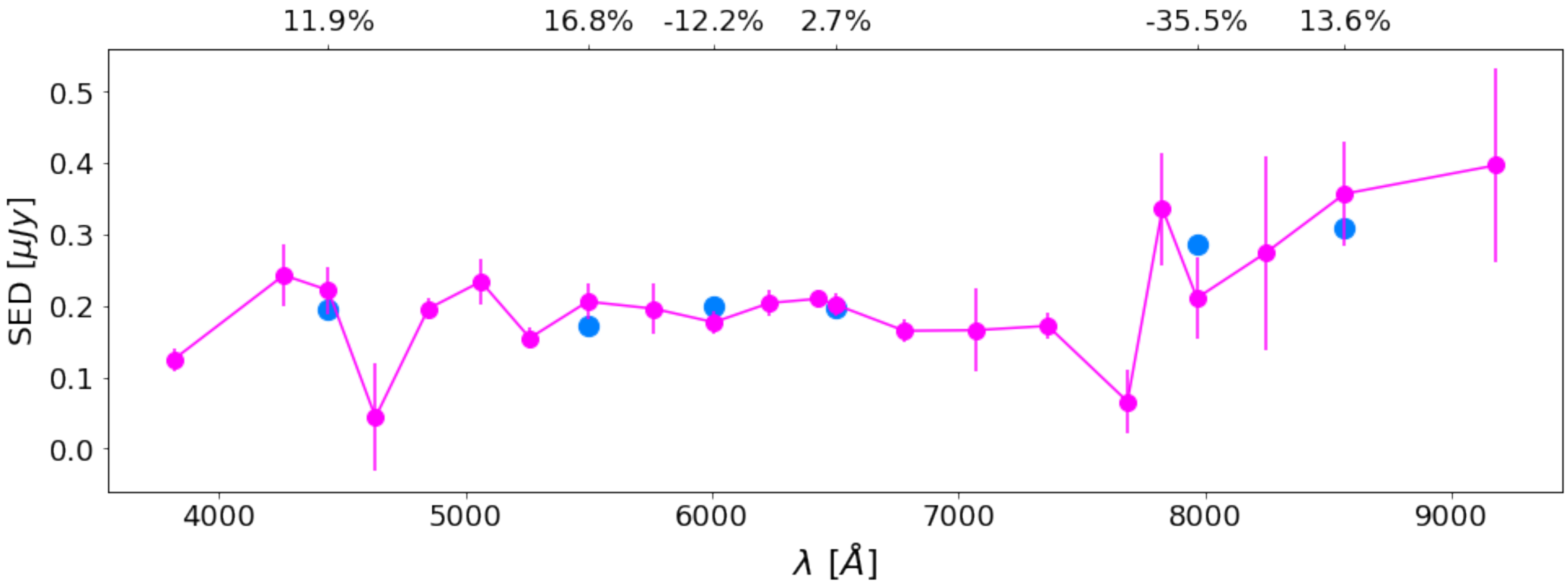
GP [%]	
IA445	3.29
IA550	13.68
IA598	−0.93
IA651	−5.28
IA797	−8.95
IA856	32.24

CDFS catalogue - Random Forest Regression





CDFS catalogue - Random Forest Regression



Median bias in each reconstructed filter

GP [%]

IA445	3.29
IA550	13.68
IA598	−0.93
IA651	−5.28
IA797	−8.95
IA856	32.24

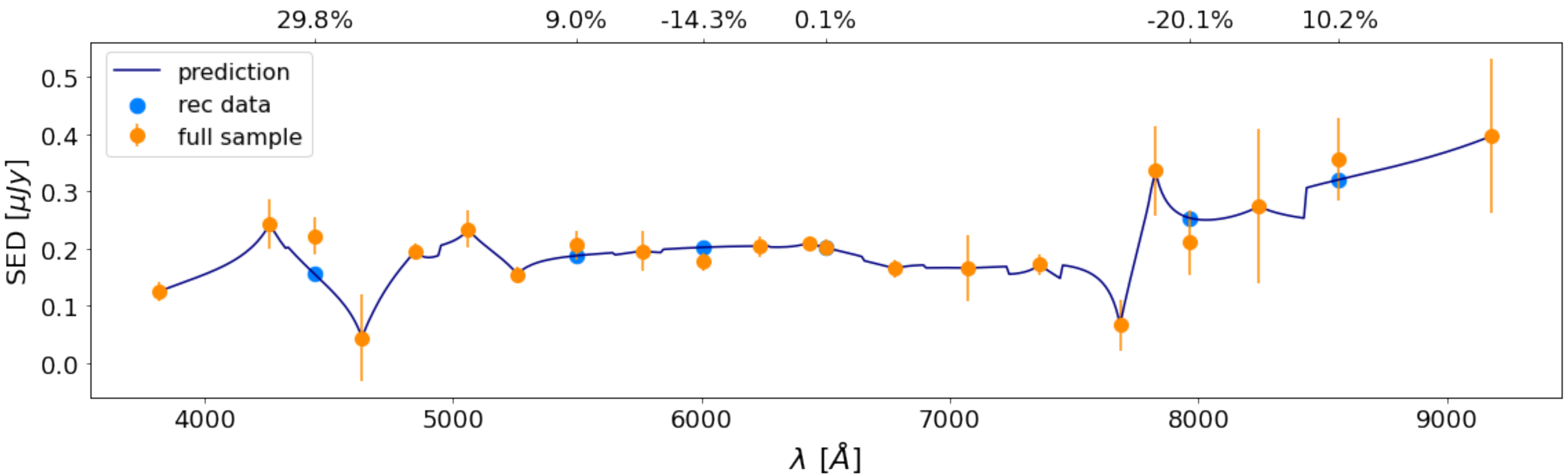
GP-err [%]

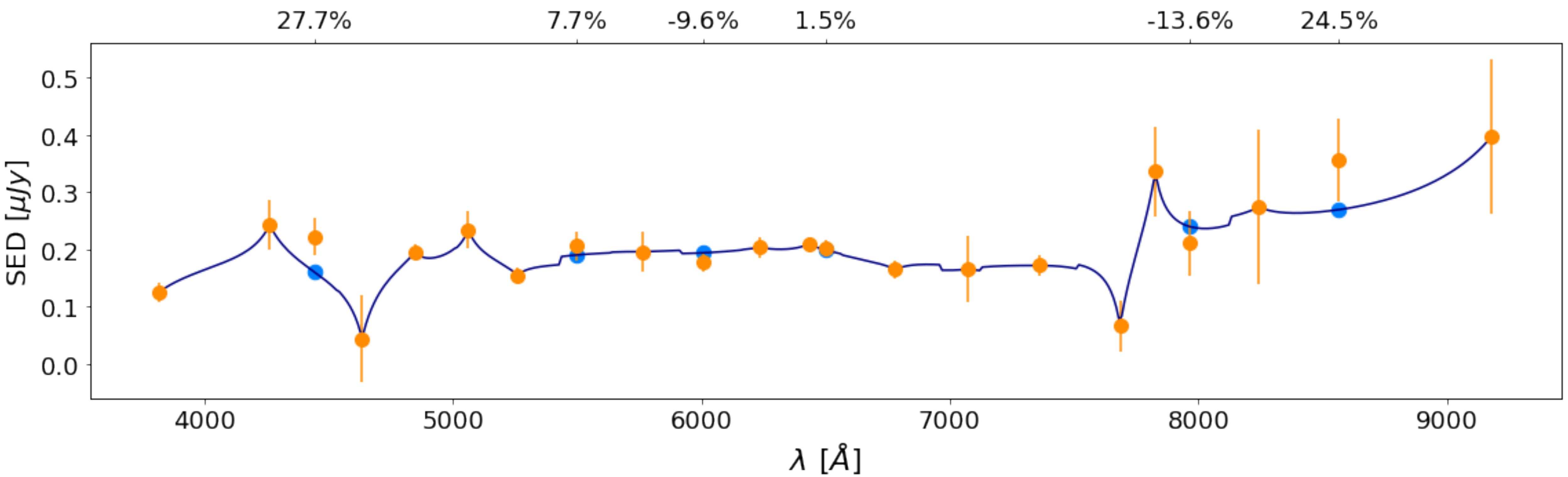
IA445	1.16
IA550	8.05
IA598	−2.97
IA651	−3.7
IA797	0.97
IA856	15.81

RF [%]

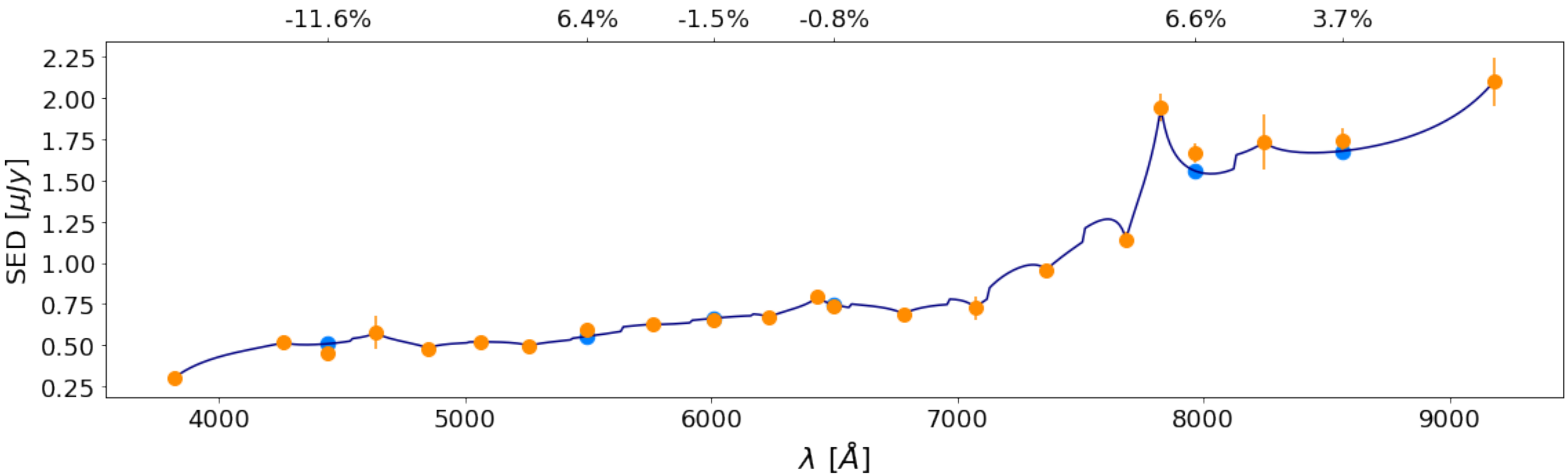
IA445	−0.53
IA550	4.69
IA598	−1.50
IA651	−1.10
IA797	−1.11
IA856	−2.61

CDFS catalogue - Nearest Neighbours





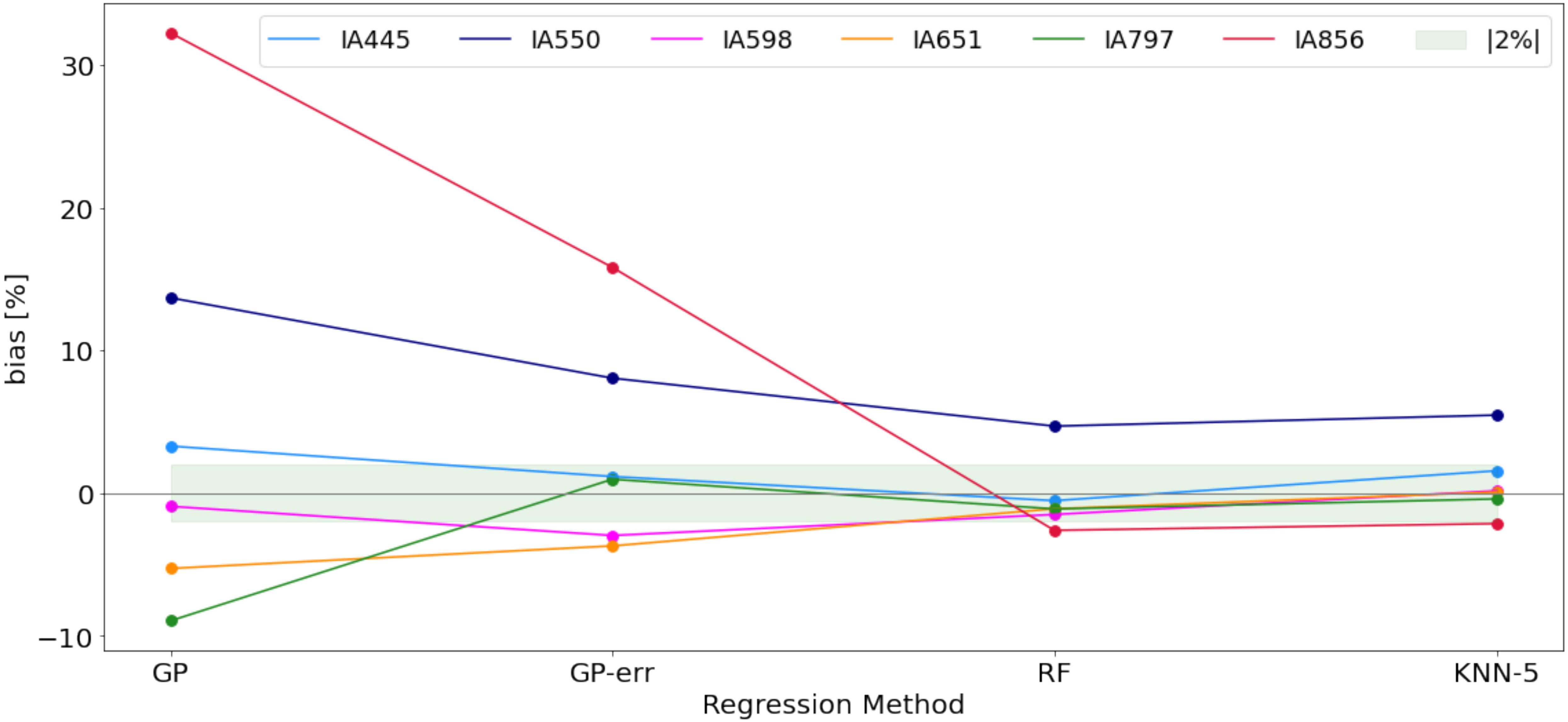
CDFS catalogue - Nearest Neighbours



Median bias in each reconstructed filter

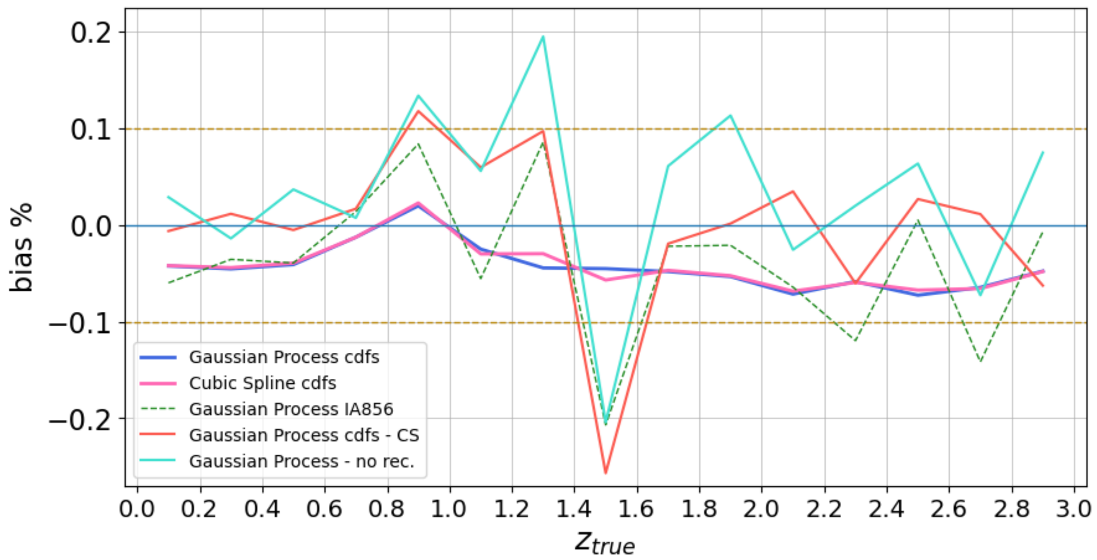
	GP [%]	GP-err [%]	RF [%]	NN - 5 [%]
IA445	3.29	1.16	−0.53	1.56
IA550	13.68	8.05	4.69	5.46
IA598	−0.93	−2.97	−1.50	0.16
IA651	−5.28	−3.7	−1.10	0.07
IA797	−8.95	0.97	−1.11	−0.41
IA856	32.24	15.81	−2.61	−2.14

CDFS catalogue - bias - summary



Summary

- Accuracy passing the requirements for weak lending analyses in all bands and tomographic bins with all choices of reconstruction of the missing fluxes.



RF [%]

IA445	−0.53
IA550	4.69
IA598	−1.50
IA651	−1.10
IA797	−1.11
IA856	−2.61

NN - 5 [%]

IA445	1.56
IA550	5.46
IA598	0.16
IA651	0.07
IA797	−0.41
IA856	−2.14

- Bias increases due to the noise in photometric data.

Future prospects

Optimisation of the data-driven regression methods (Random Forest, K-NN).

Thank you!