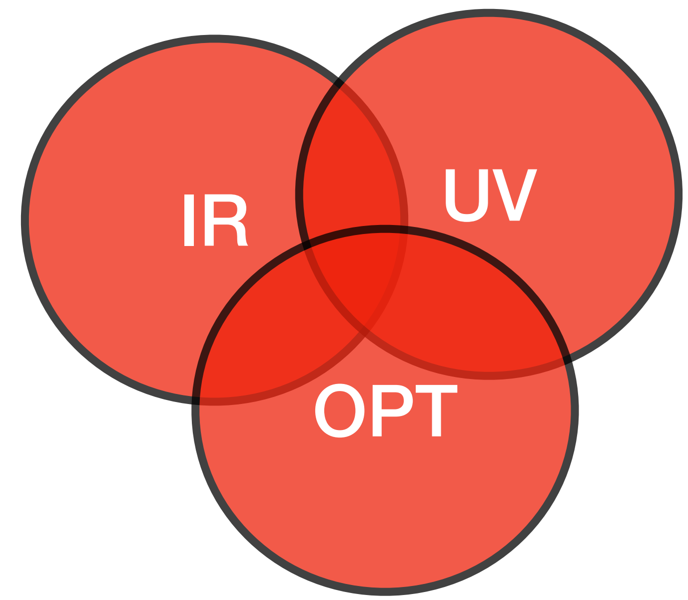
A proto-foundation model for galaxy SED: the J-PAS case.

Ginés Martínez Solaeche, R. M. González Delgado,
R. García-Benito, L. A. Díaz-García

Instituto de Astrofísica de Andalucía, IAA-CSIC

EXCELENCIA
SEVERO
OCHOA

Galaxy observations, looking for the cross-match



1. **Mulitwavelenght observation of galaxies are more the exception than the norm.**

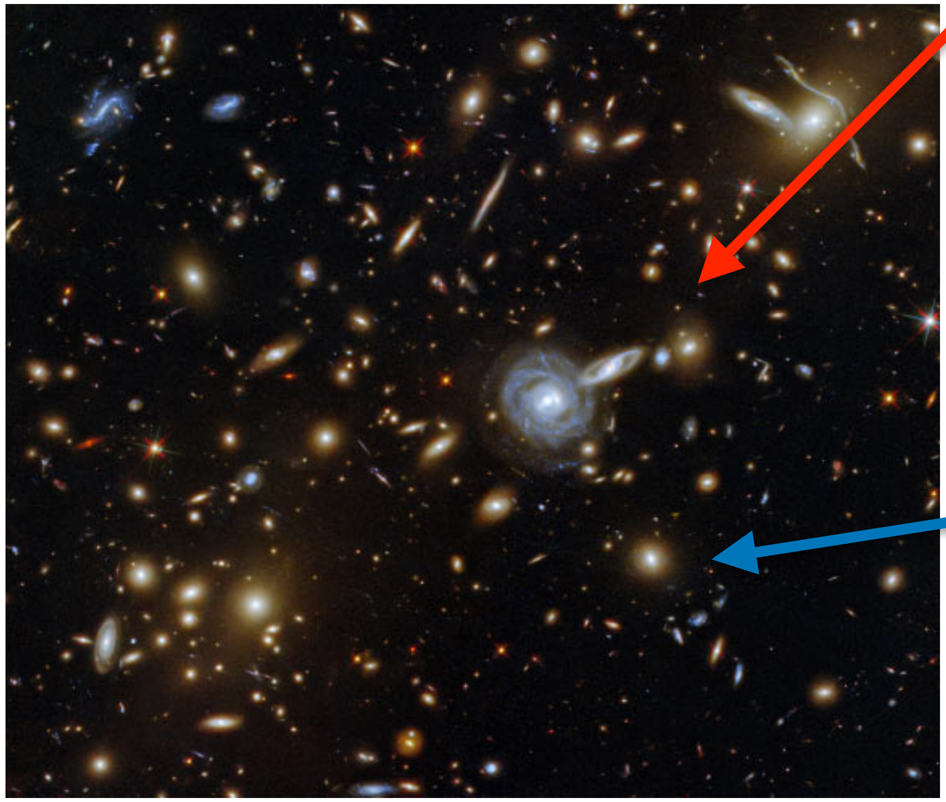
2. **DL algorithms can benefit from training on the union of available data rather than the intersection**

Galaxy observations, dealing with missing data

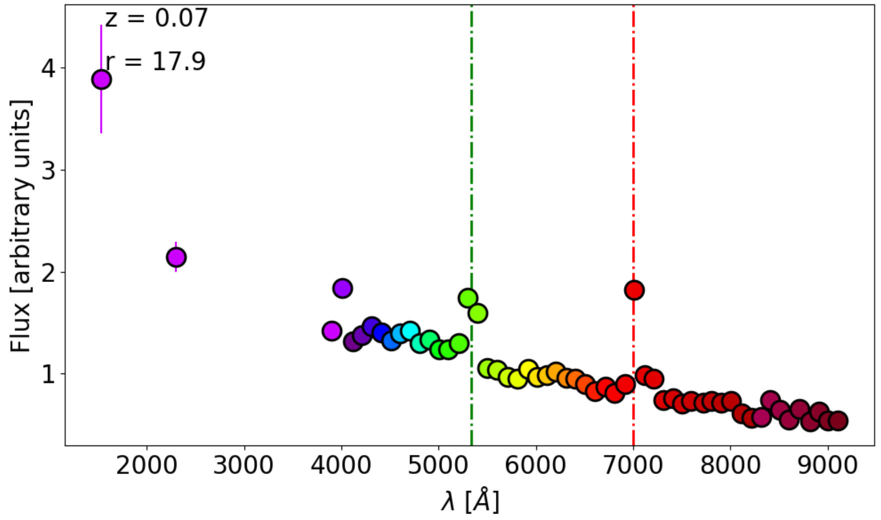
esa euclid

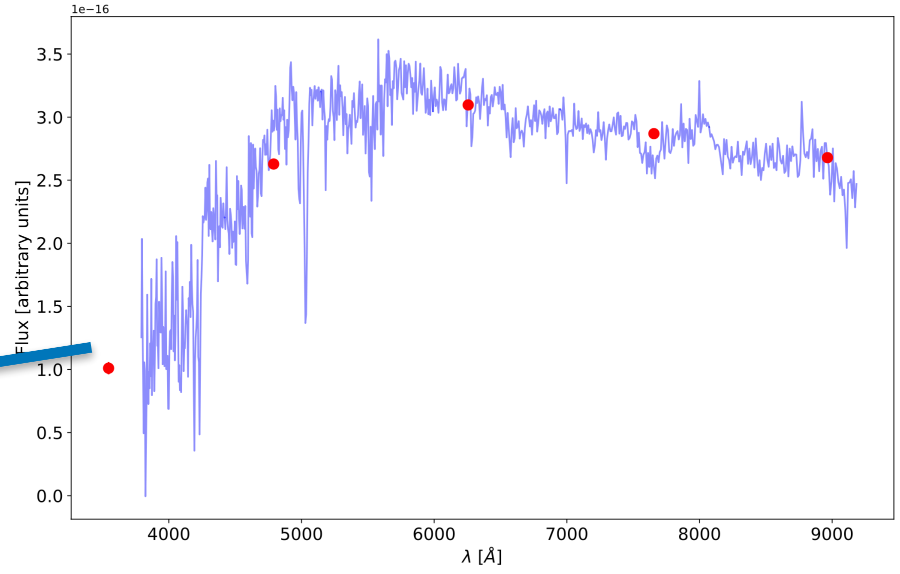
DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

SDSS



J-PAS
Javalambre Physics of the Accelerating Universe Astrophysical Survey





Galaxy Properties:

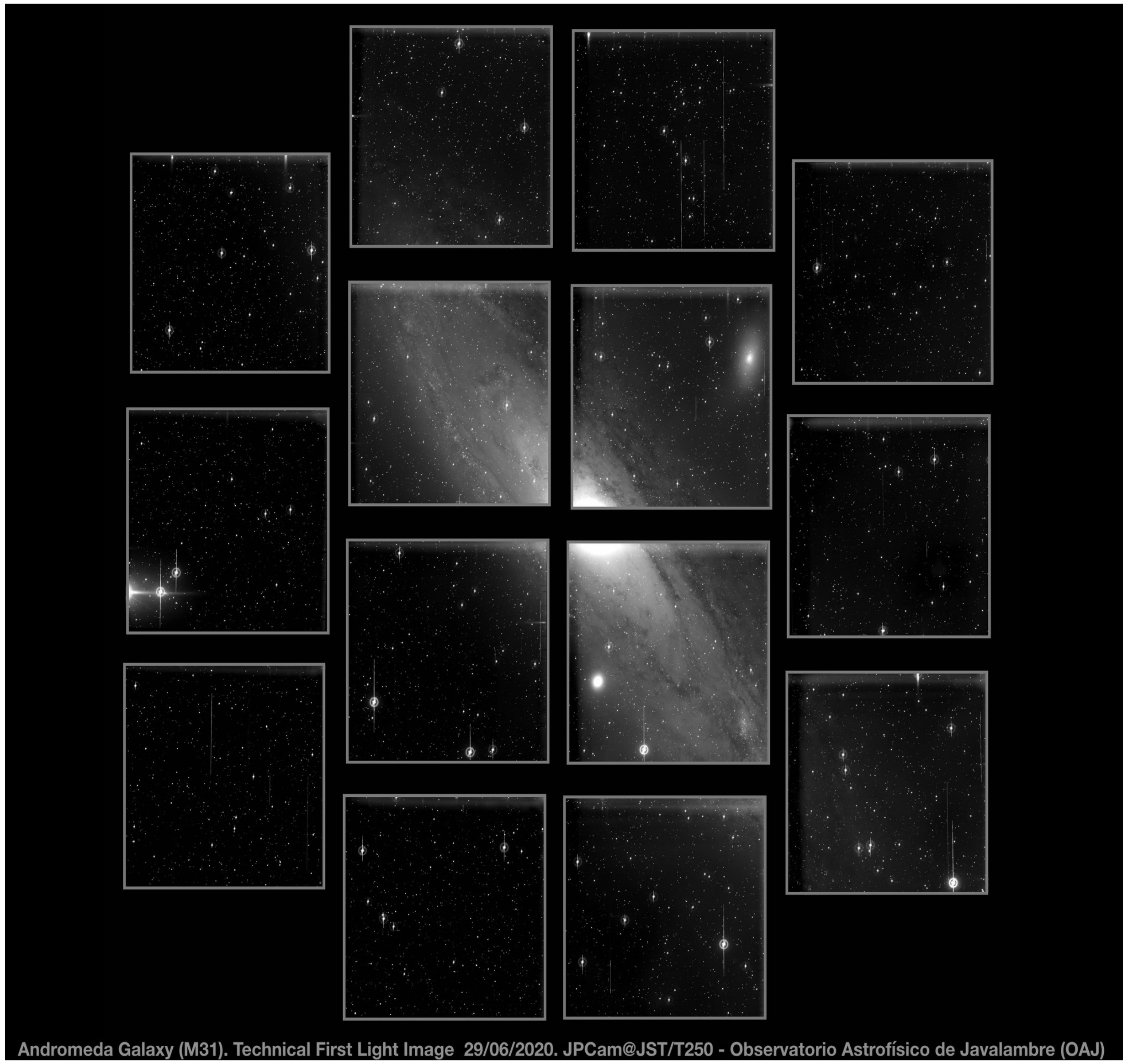
Redshift
Stellar Mass
Metallicity

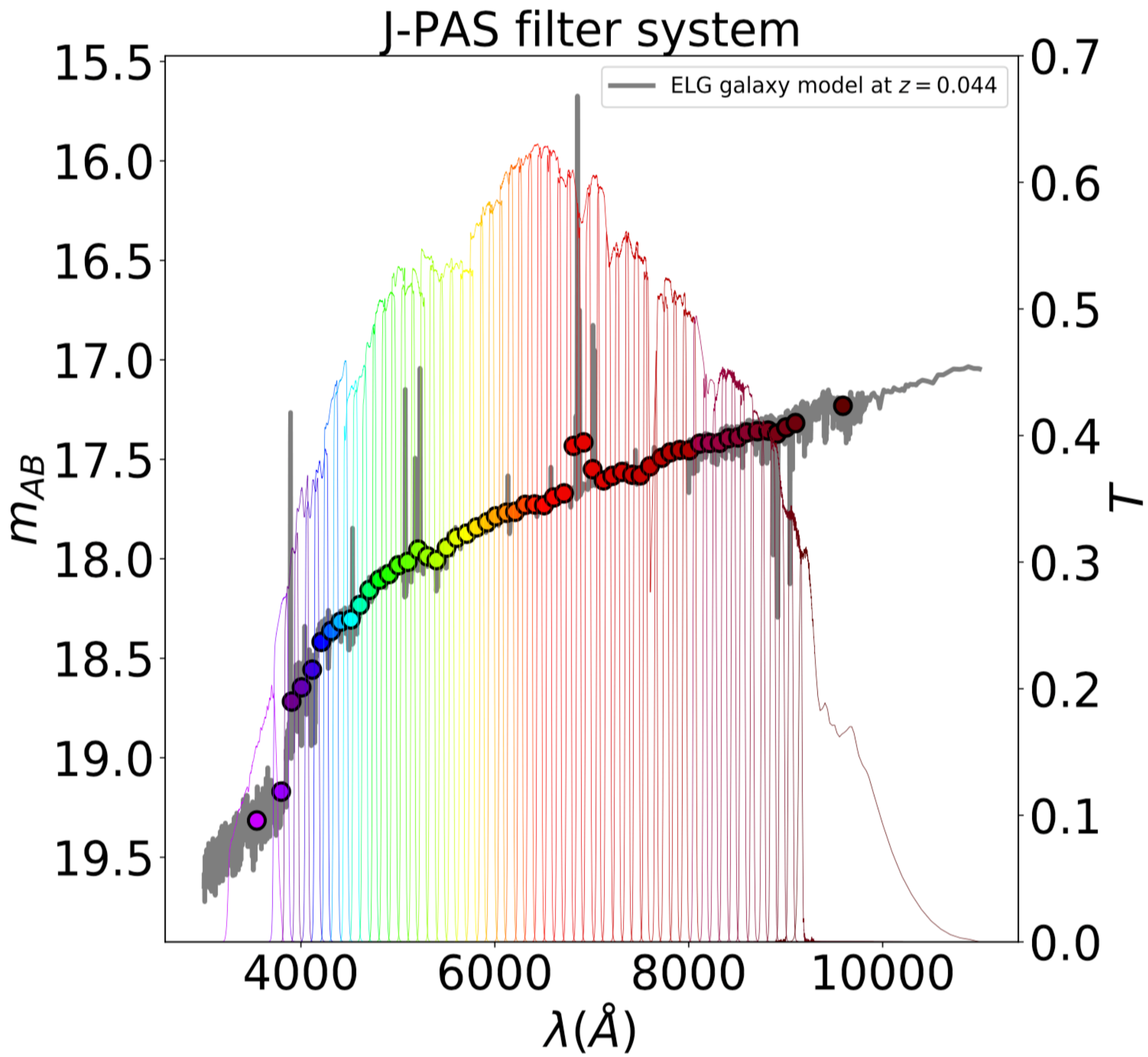
Age
Emission lines
SFR
...

Javalambre Physics of the Accelerating Universe (J-PAS)

OAJ - Observatorio Astrofísico de Javalambre, Teruel (España)

Javalambre Physics of the Accelerating Universe (J-PAS)

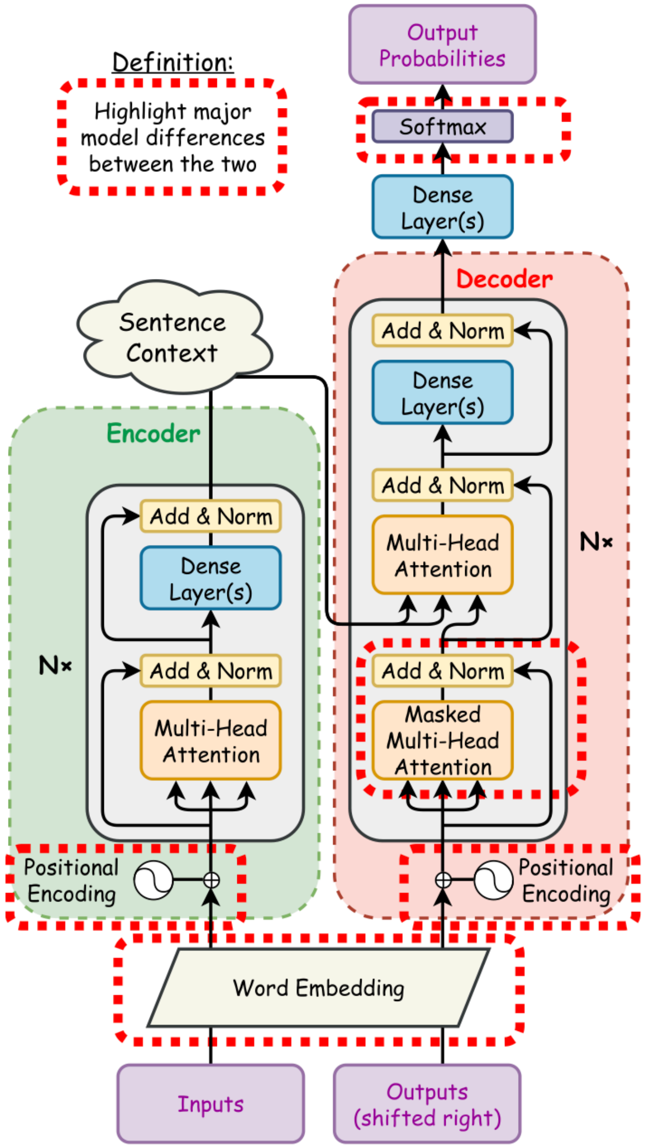




Andromeda galaxy
fits in the FoV of the JPCAM

56 narrow band filters capture
the SED of galaxies up to z ~ 1

Transformer-based models



"God does not play dice with the universe", Albert Einstein

"I am no longer accepting the things I cannot change, I am changing the things I cannot accept", Angela Davis

"Be the change that you wish to see in the world.", Mahatma Gandhi

Pre-trained to predicted missing words and sentences!

ChatGPT Claude Gemini

Training goal, the loss function and the context

uSDSS, rSDSS, Stellar mass, Hα

uSDSS, rSDSS, Stellar mass, Hα

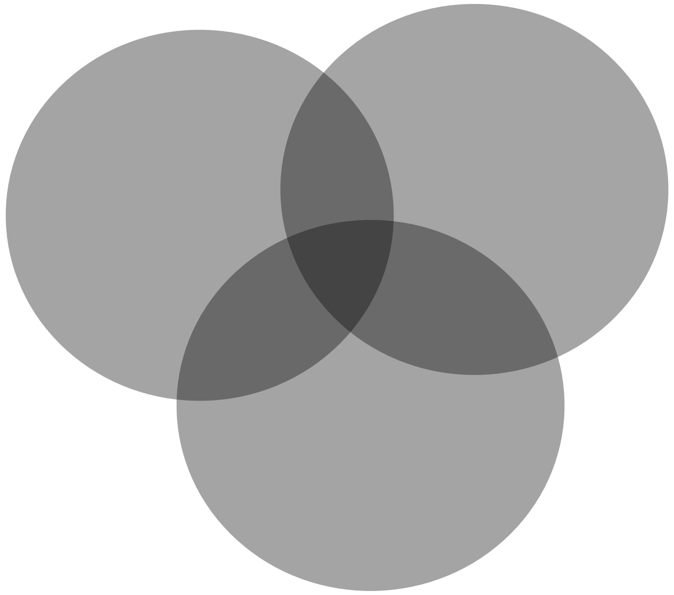
uSDSS, rSDSS, Stellar mass, Hα

uSDSS, rSDSS, Stellar mass, Hα

uSDSS, ~~rSDSS~~, Stellar mass, Hα

$$\mathcal{L} = \sum_i \left(\frac{(y^i_{\text{true}} - y^i_{\text{pred}})^2}{2\sigma_i^2} + \frac{1}{2}\log(\sigma_i^2) \right)$$

$$\sigma_i^2 = (\sigma^i_{pred})^2 + (\sigma^i_{true})^2$$

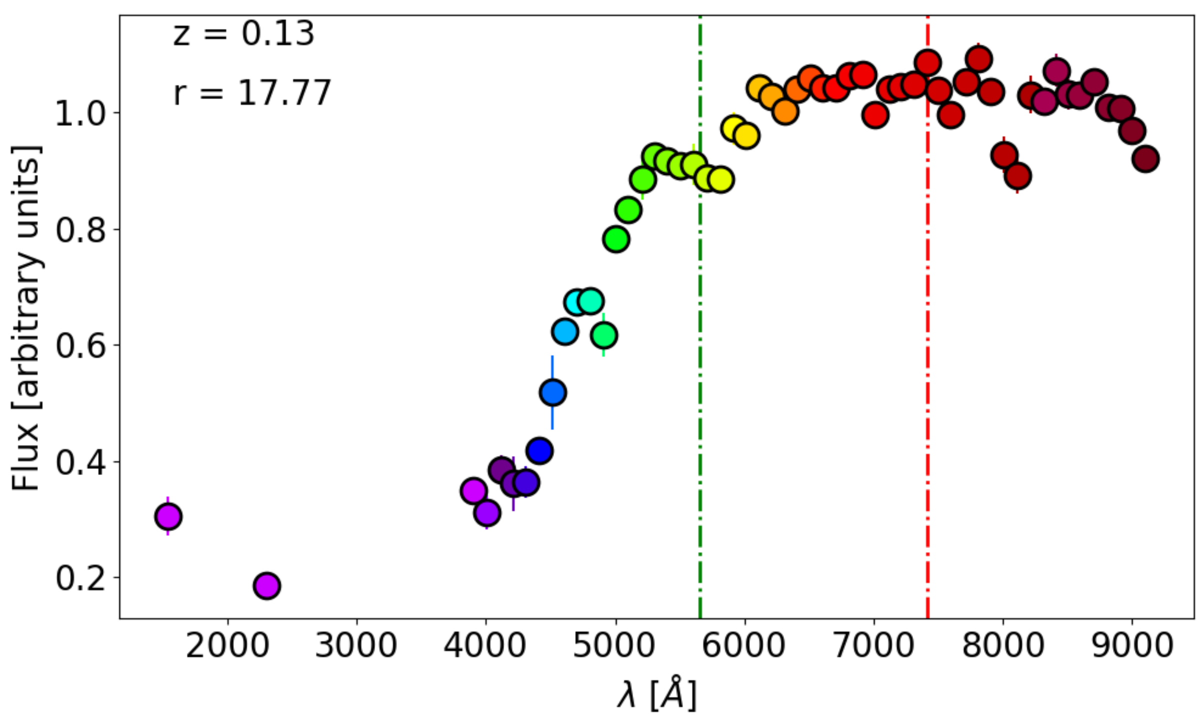


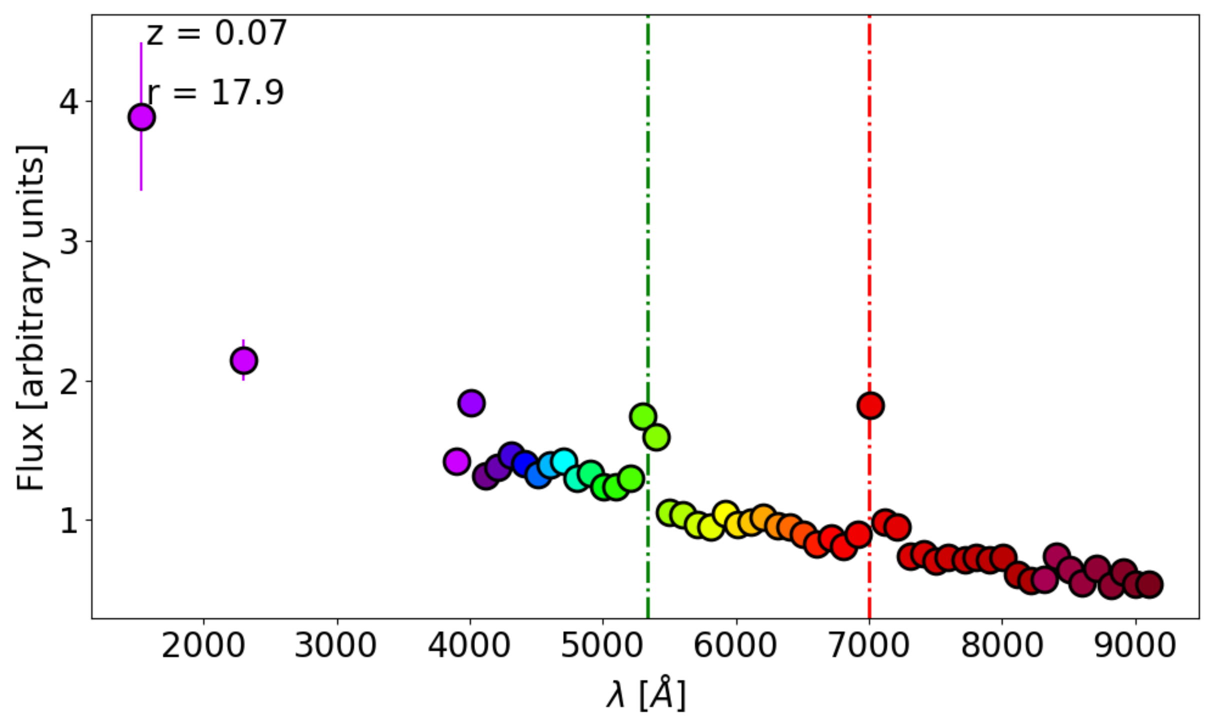
Towards an astronomical foundation model for stars with a transformer-based model

Henry W. Leung[1★] and Jo Bovy[1,2]
[1]David A. Dunlap Department of Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario, M5S 3H4, Canada
[2]Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario, M5S 3H4, Canada

The model takes into account the uncertainty of target variables, but also learn to predict uncertainties given the context

Building the training set, retrieving the SED





1.- We generate synthetic j-spectra from DESI and SDSS spectra and crossmatch with GALEX photometry

2.- We use miniJPAS observations to model the error so the training set emulate the depth of miniJPAS

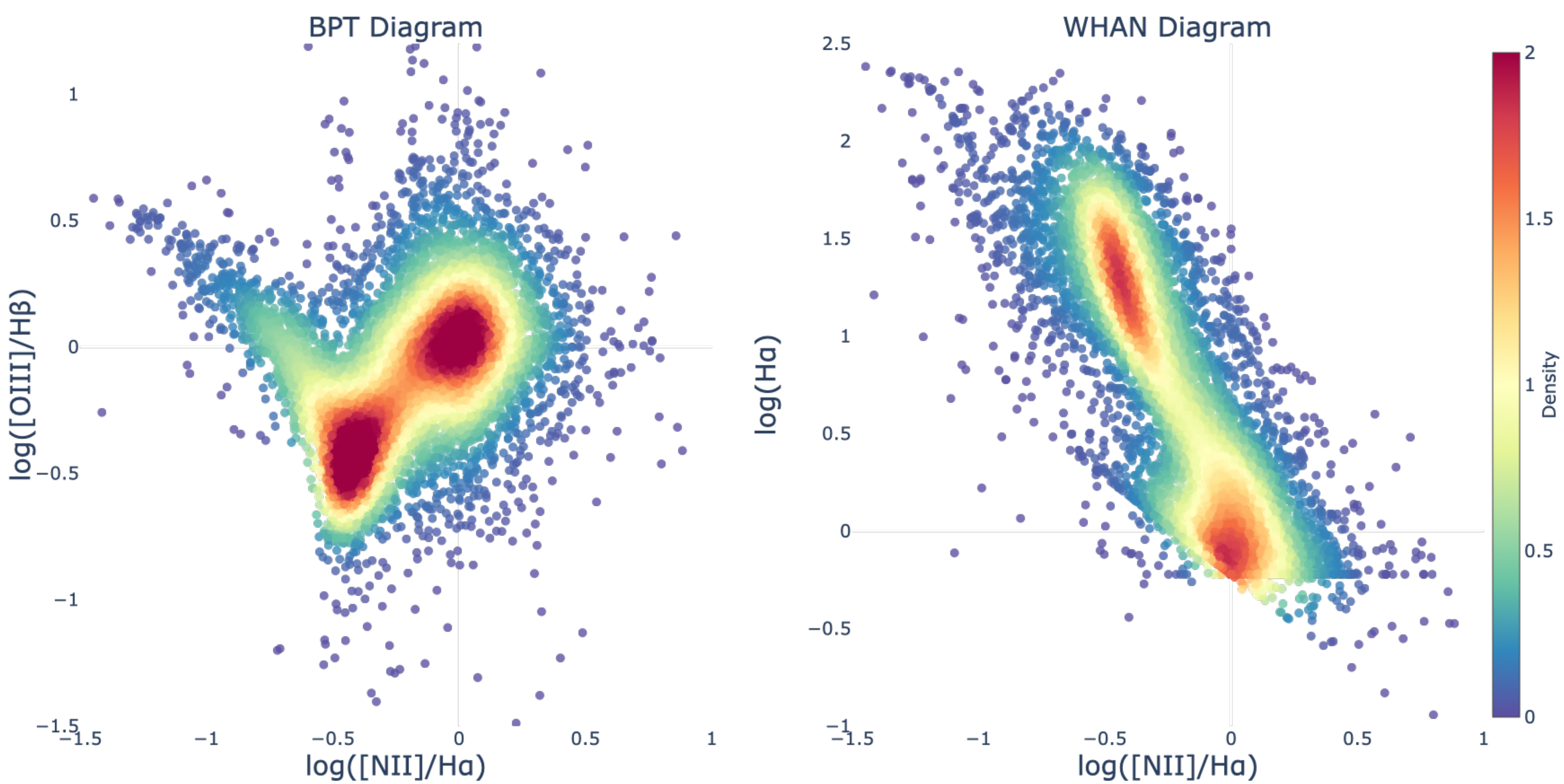
Building the training set, emission lines properties

SDSS

DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

[OII], Hβ, [OIII], [NII], Hα, [SII] doublet

$z < 0.4$



1.- We estimate a probability as function of the density, so galaxies that are underrepresent are more likely to appear during training

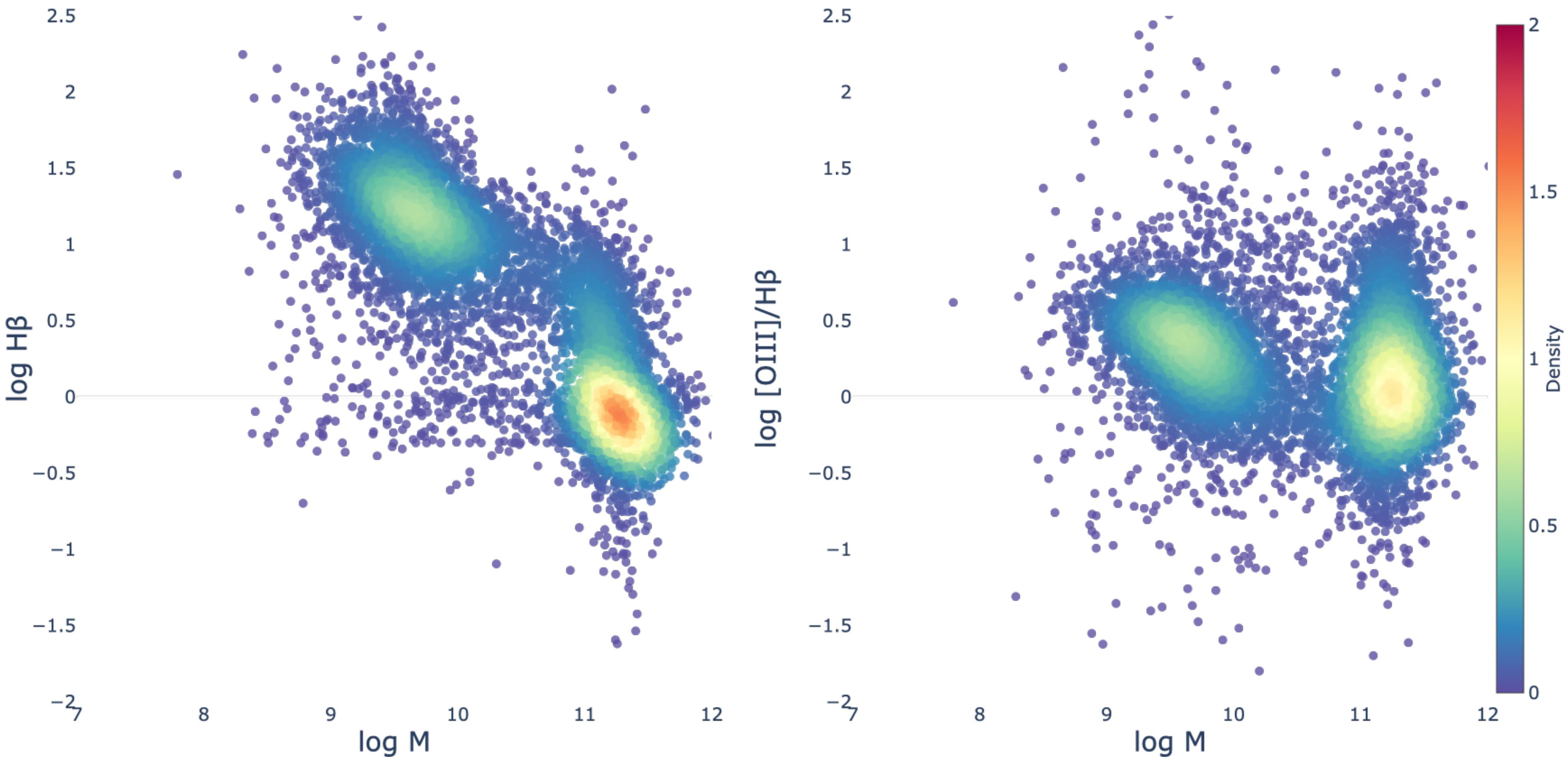
Building the training set, emission lines properties

SDSS

DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

[OII], Hβ, [OIII]

z > 0.4



1.- We estimate a probability as function of the density, so galaxies that are underrepresent are more likely to appear during training

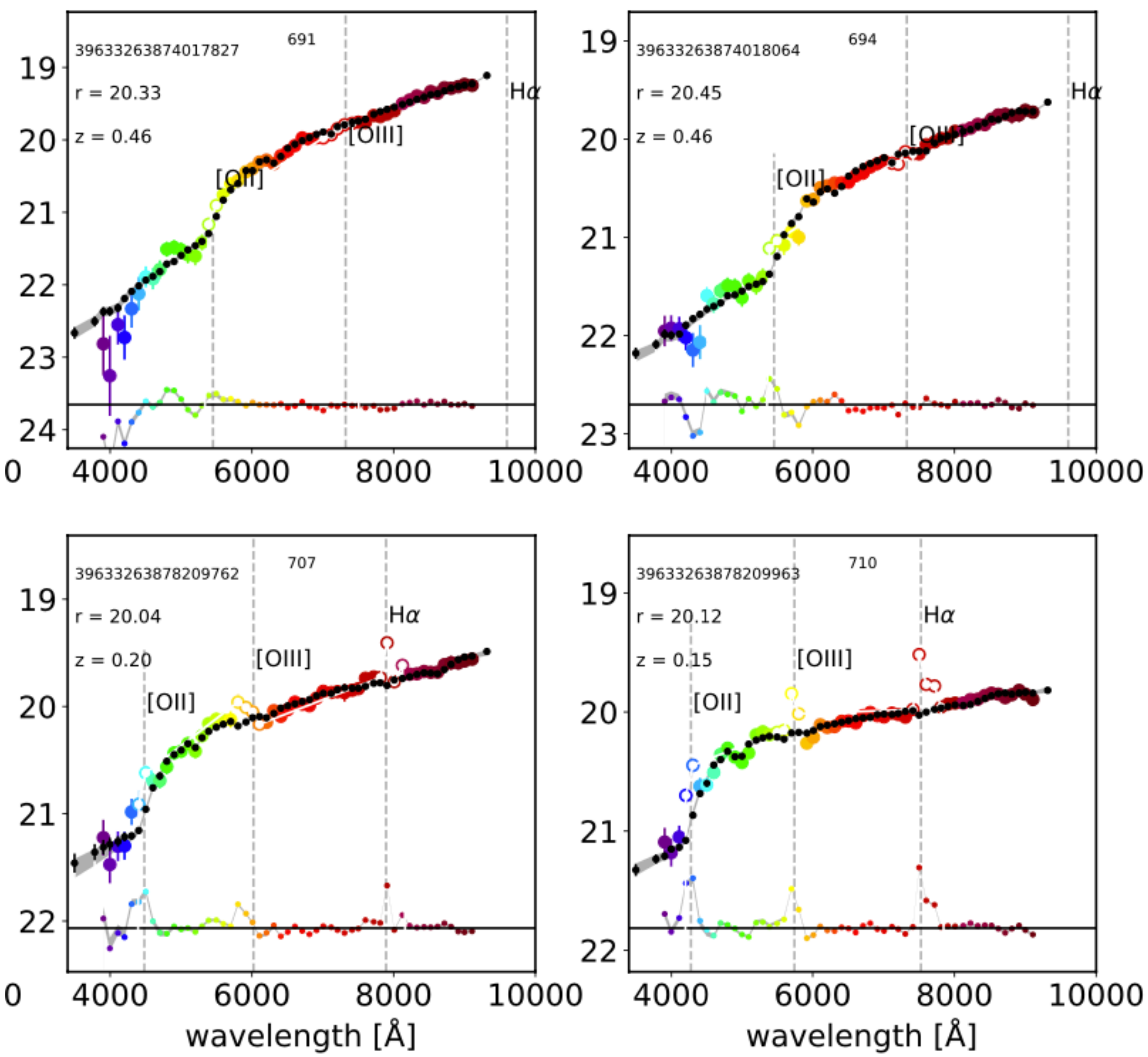
Building the training set, stellar population properties

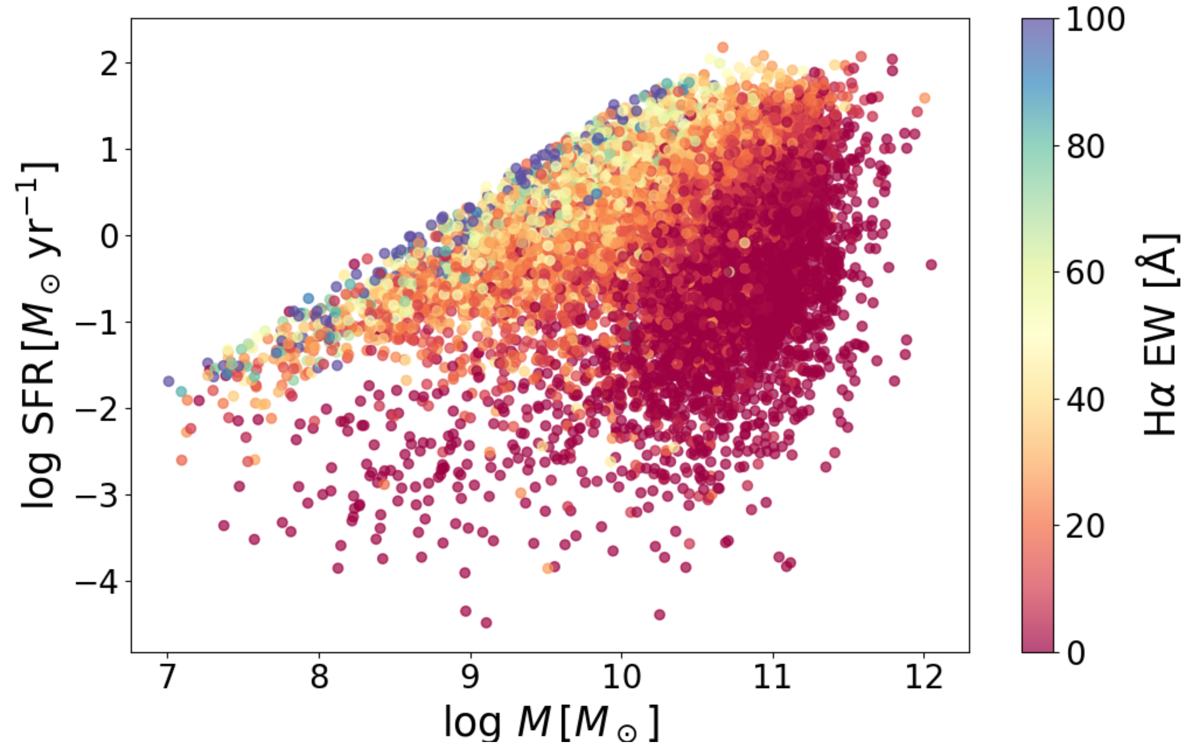
SDSS

DARK ENERGY SPECTROSCOPIC INSTRUMENT
U.S. Department of Energy Office of Science

$$Z_\star \quad t_0 \quad \tau \quad A_V$$

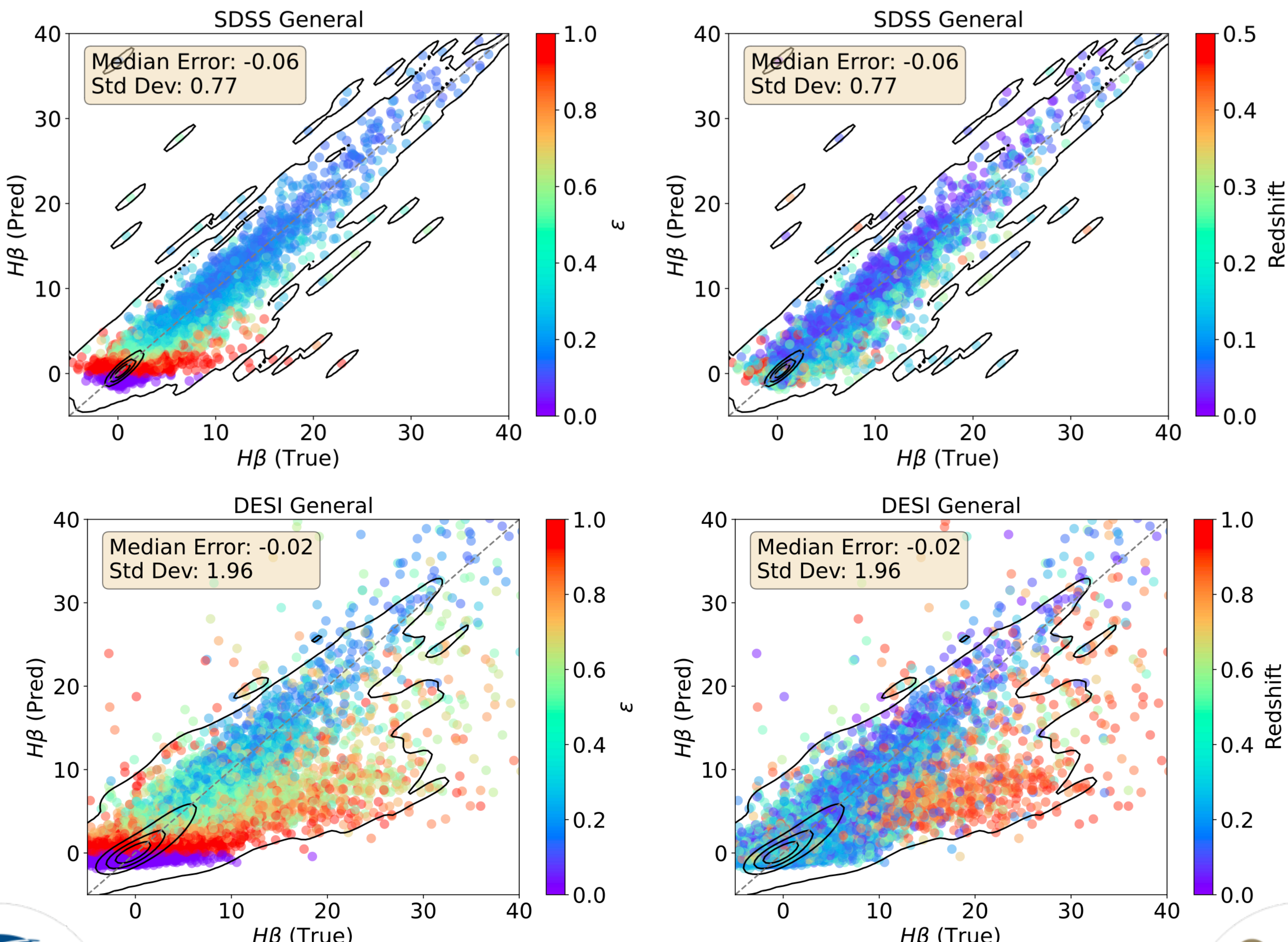
$$\Psi(t) = \phi \frac{t_0 - t}{\tau} \exp\left[-(t_0 - t)/\tau\right]$$



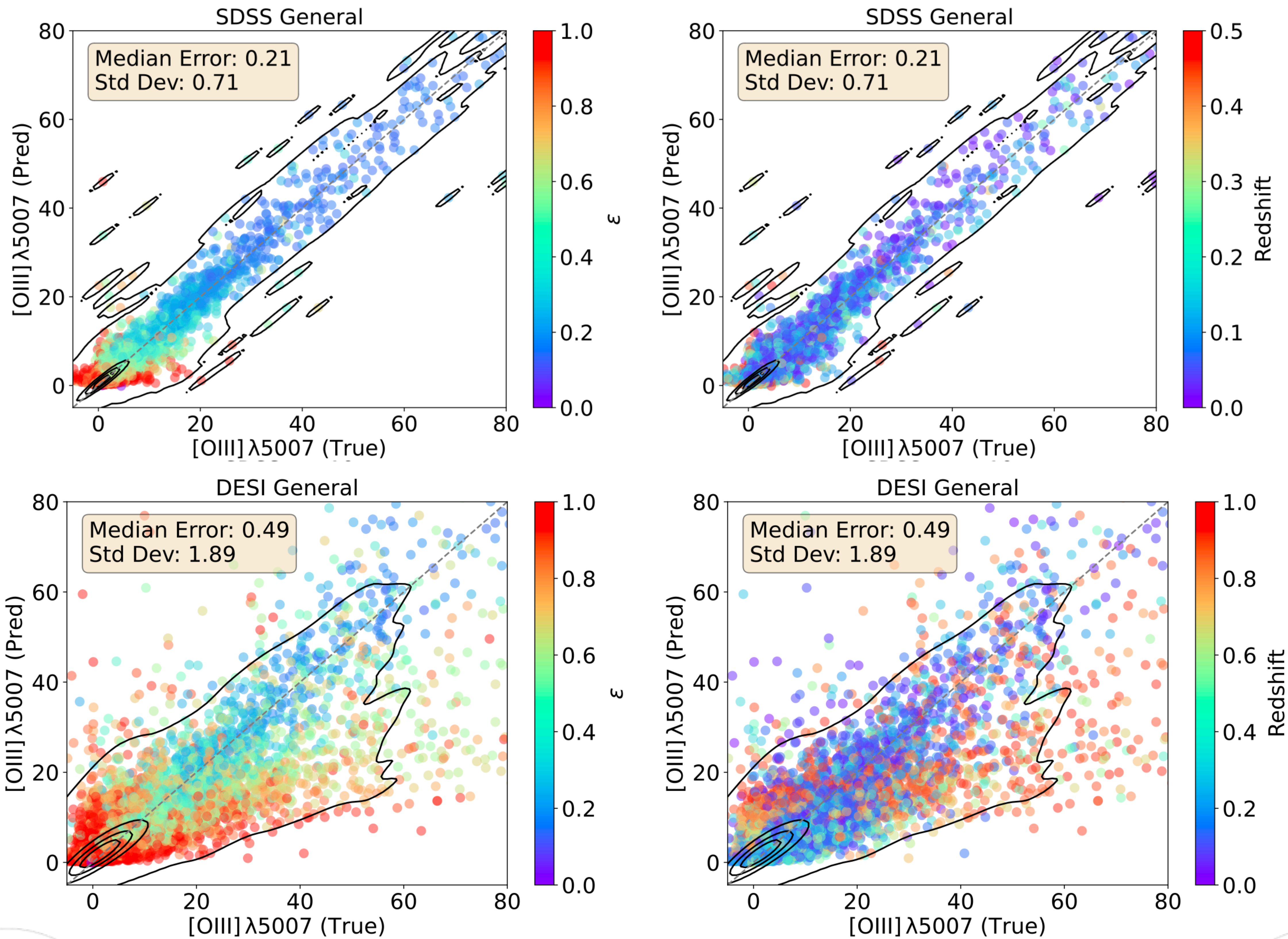


We fit the SSP properties with BaySeAGal, this is bayesian parametric code, that model the SFH with a tau model

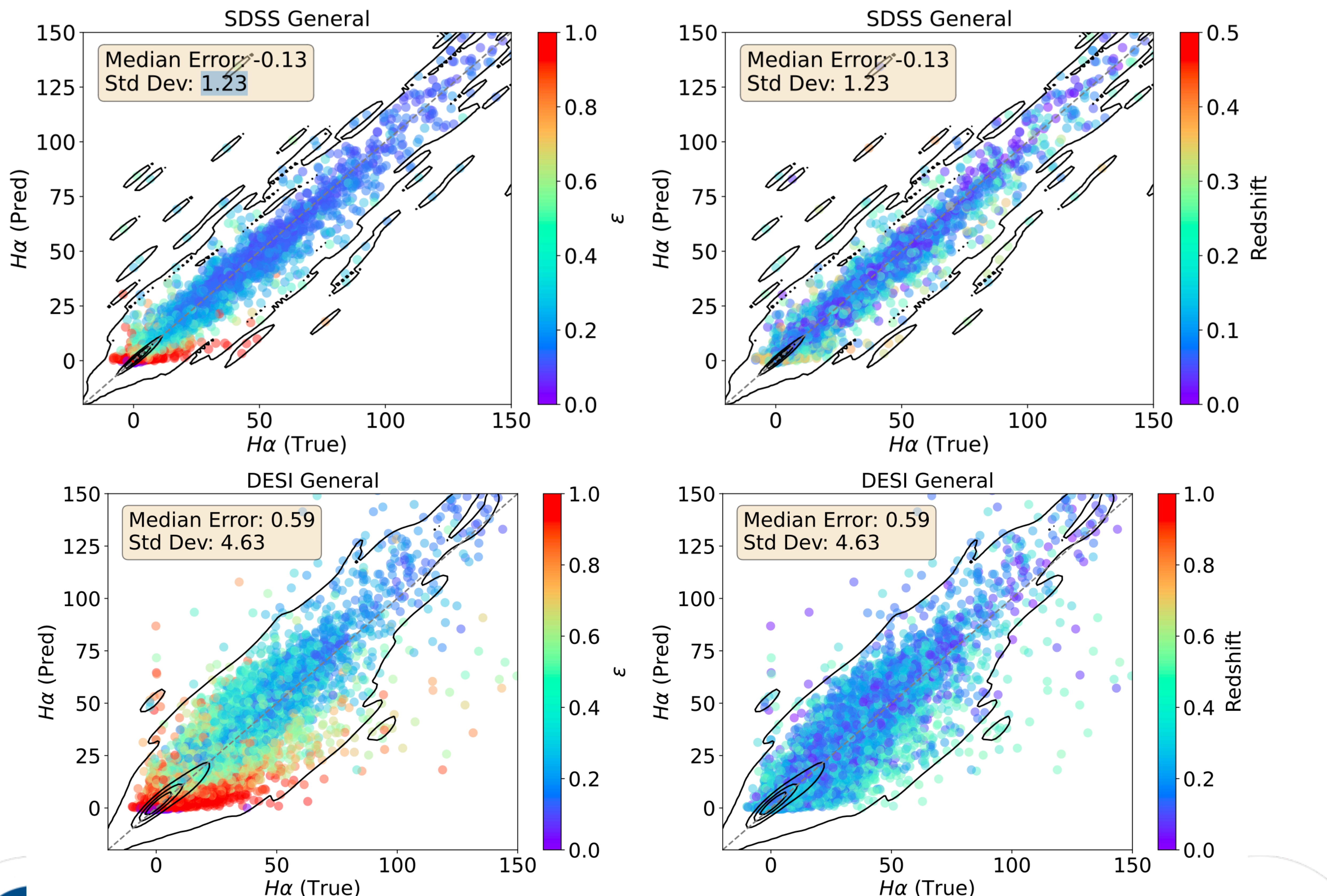
Results on SDSS and DESI test samples, emission lines



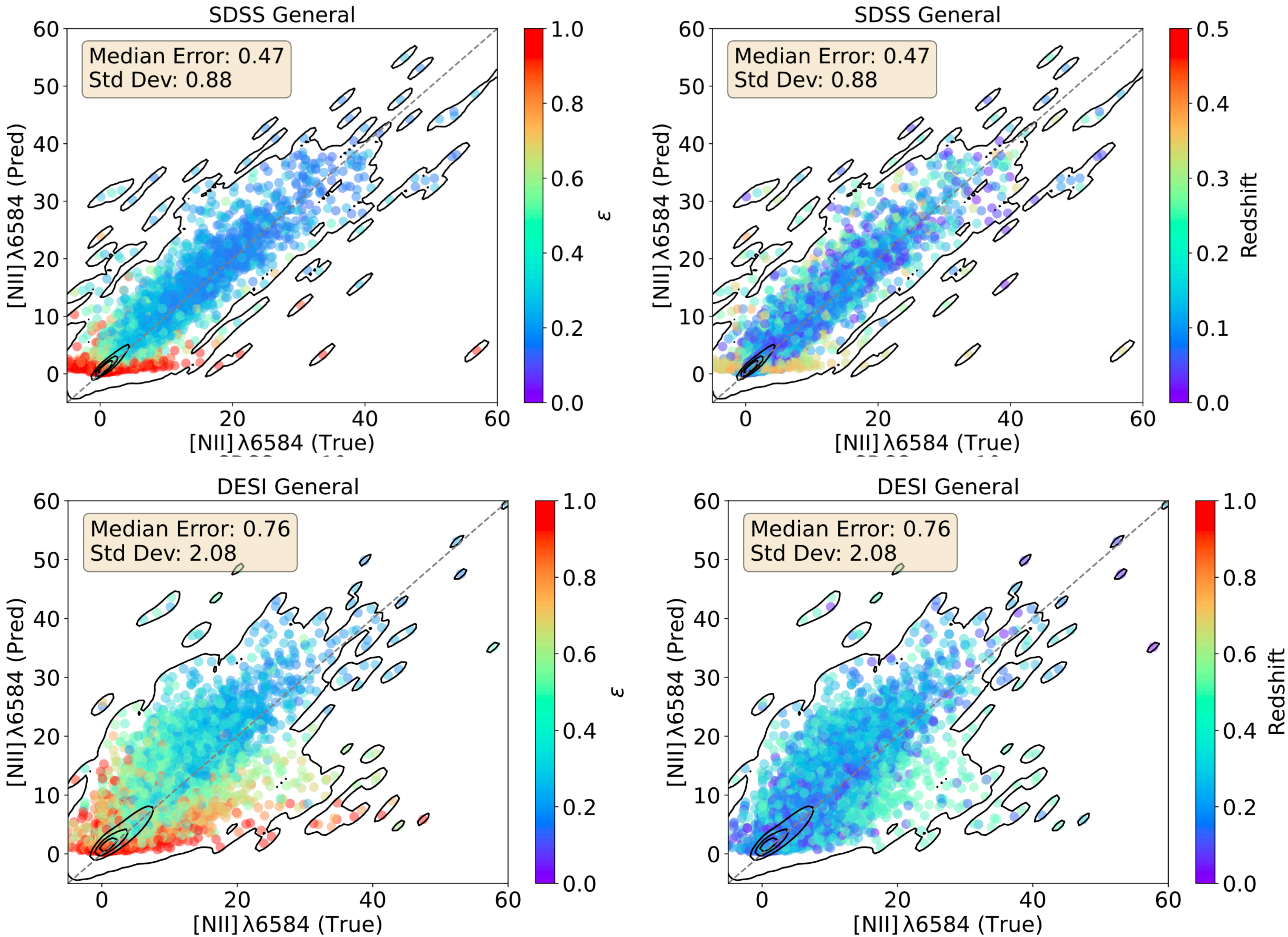
Results on SDSS and DESI test samples, emission lines



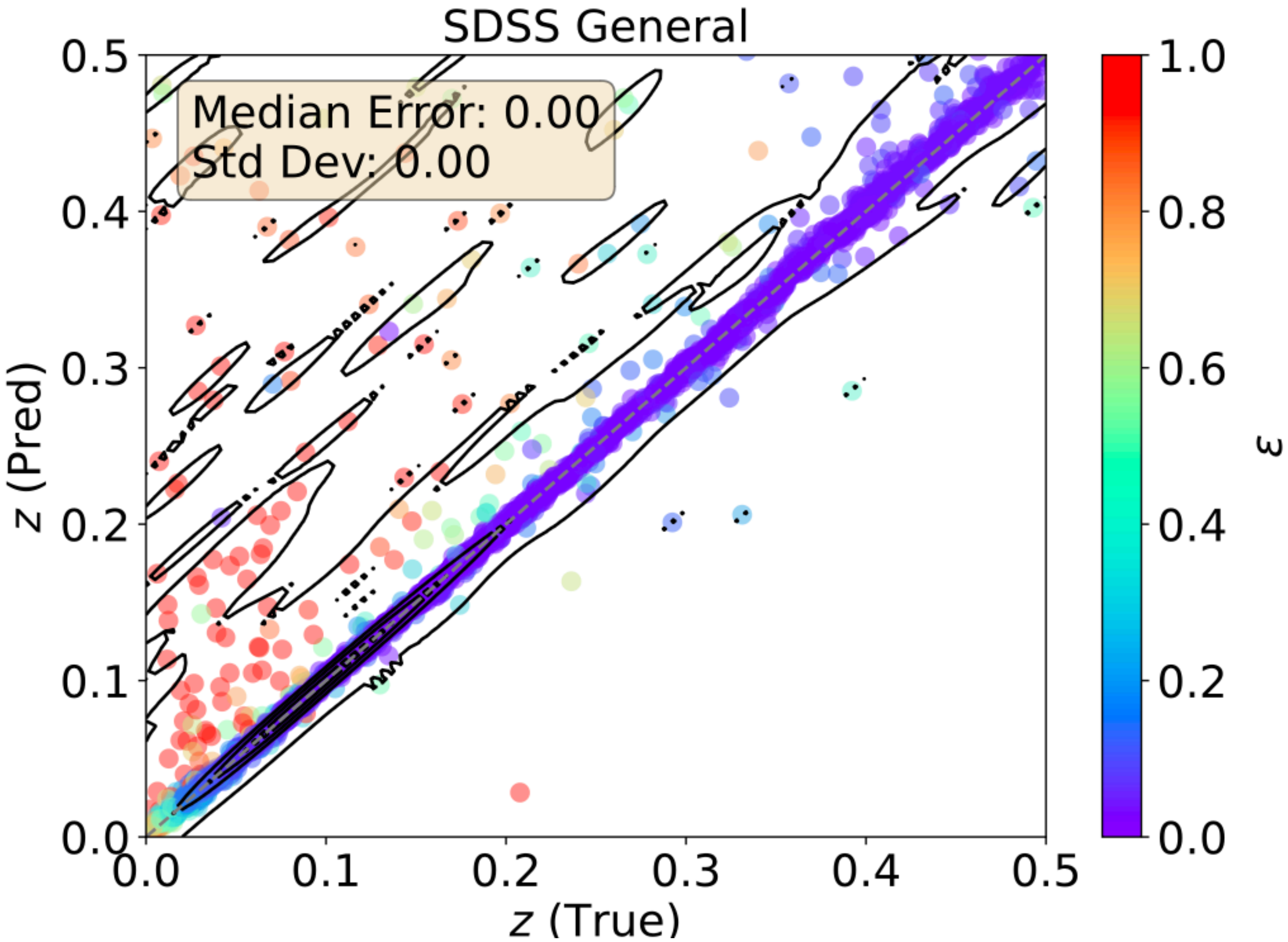
Results on SDSS and DESI test samples, emission lines

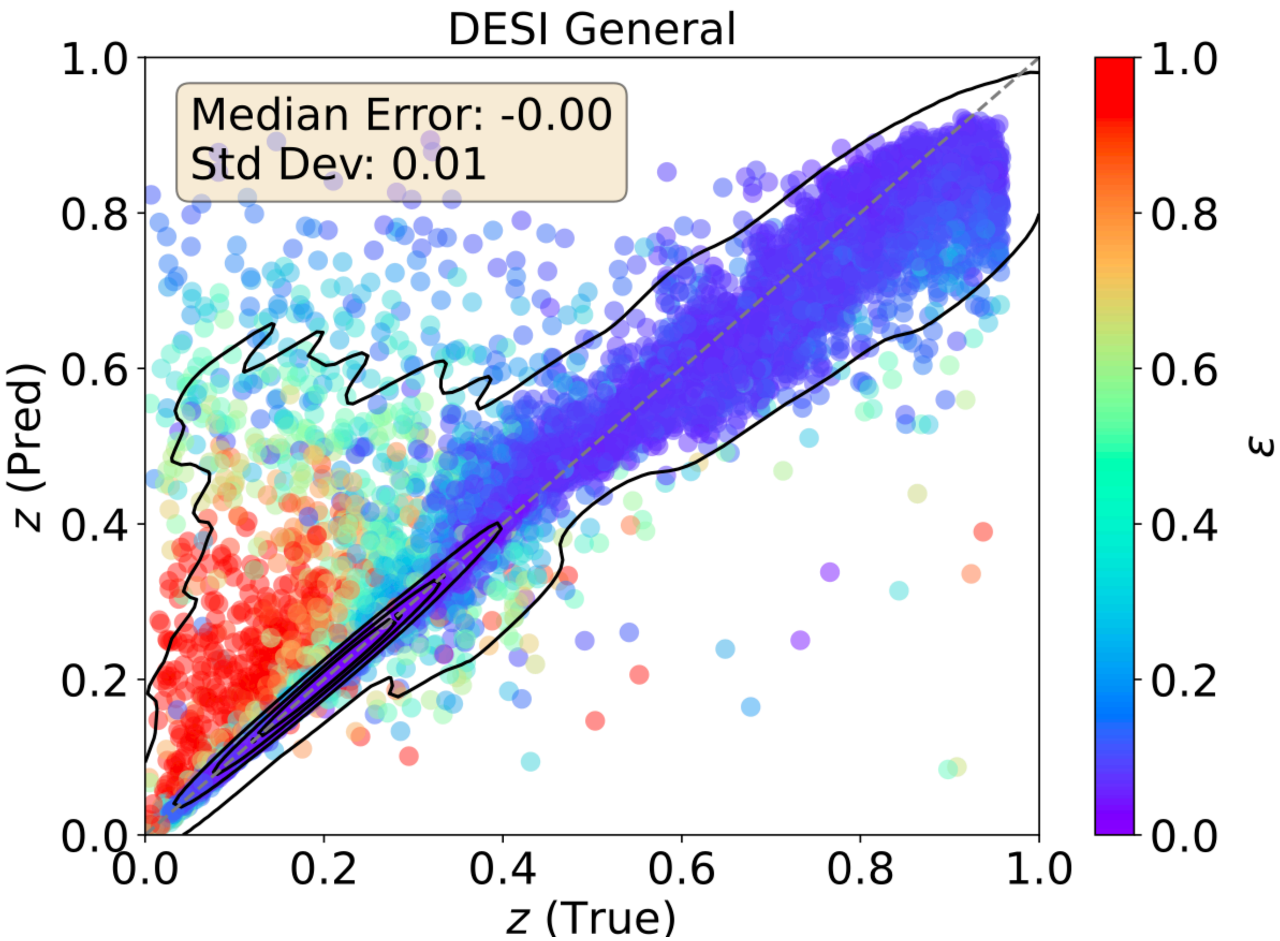


Results on SDSS and DESI test samples, emission lines

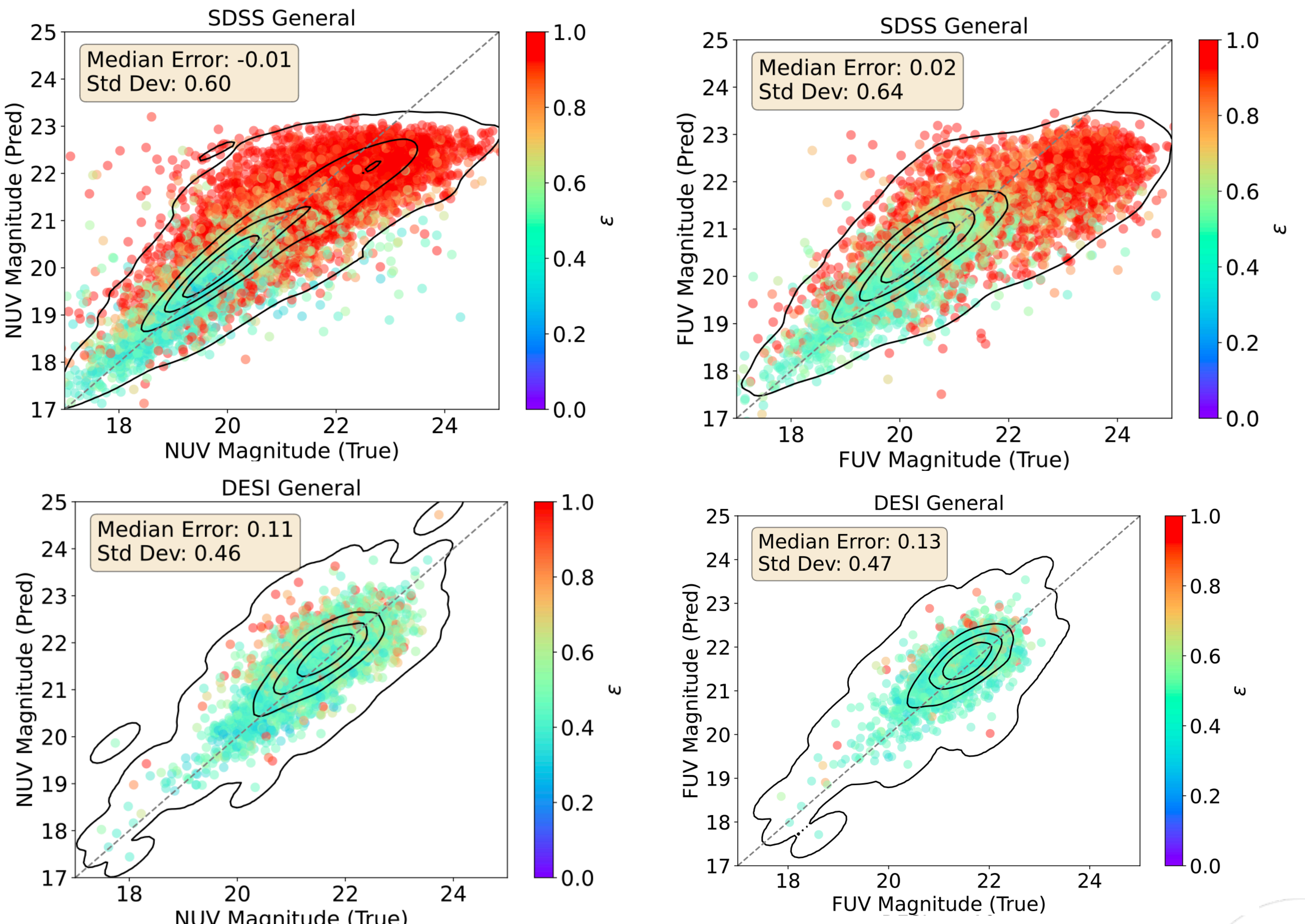


Results on SDSS and DESI test samples, photo-z

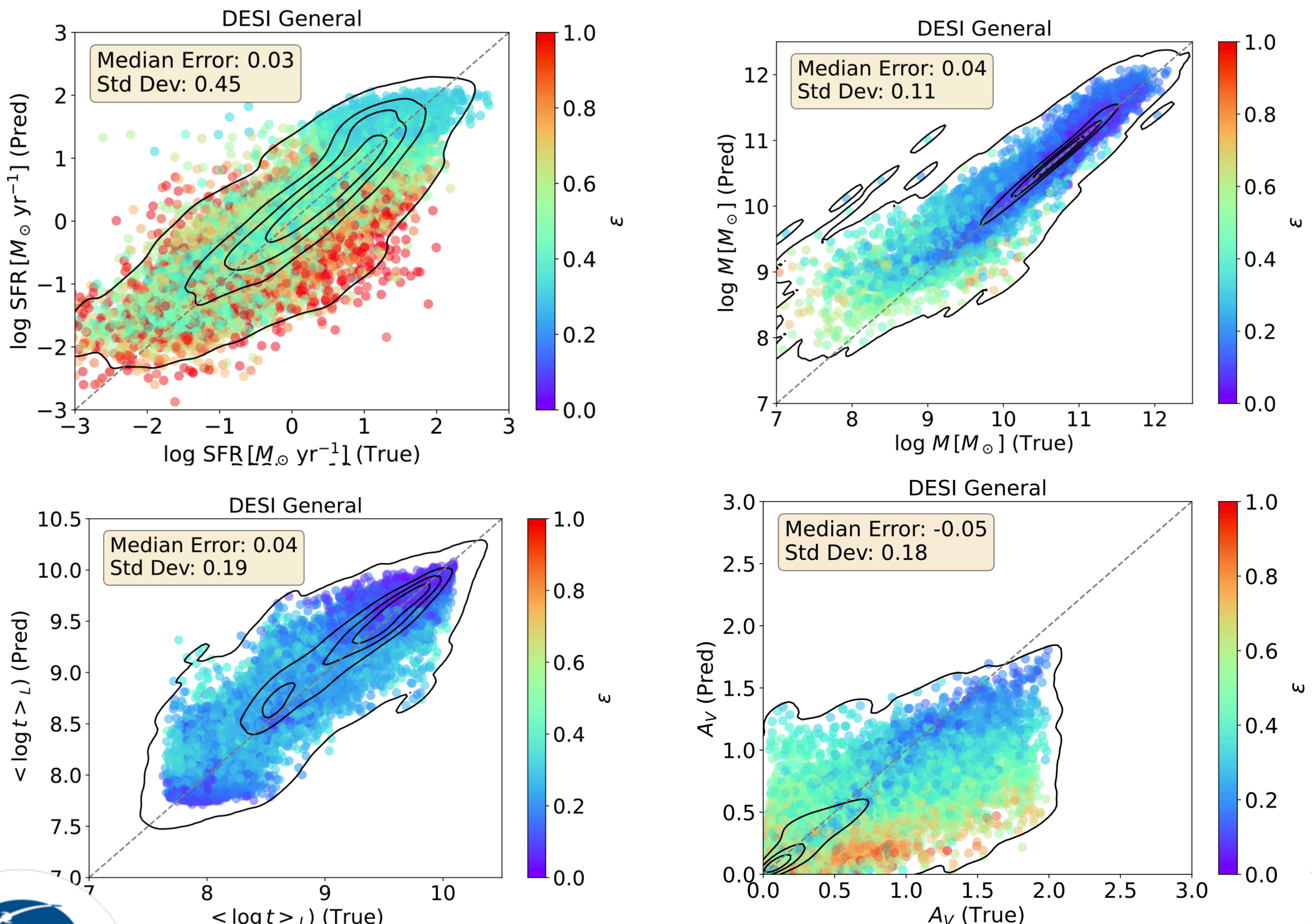




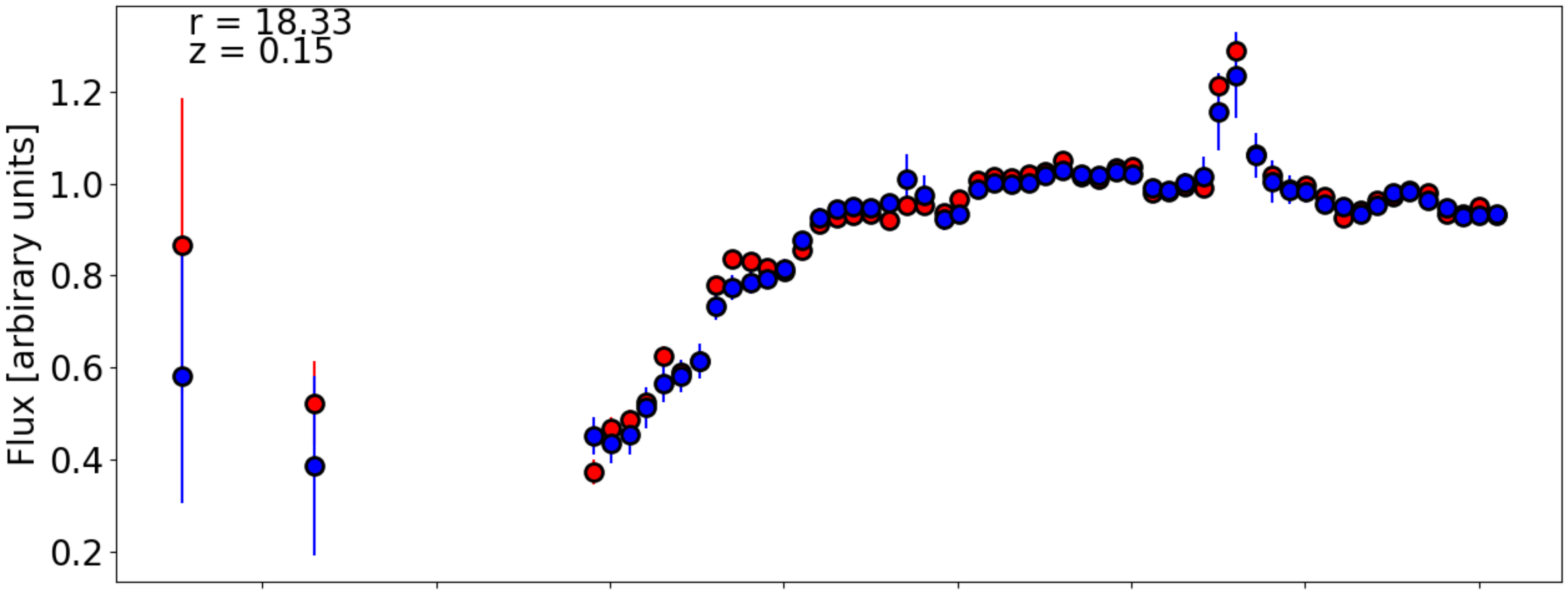
Results on SDSS and DESI test samples, FUV and NUV



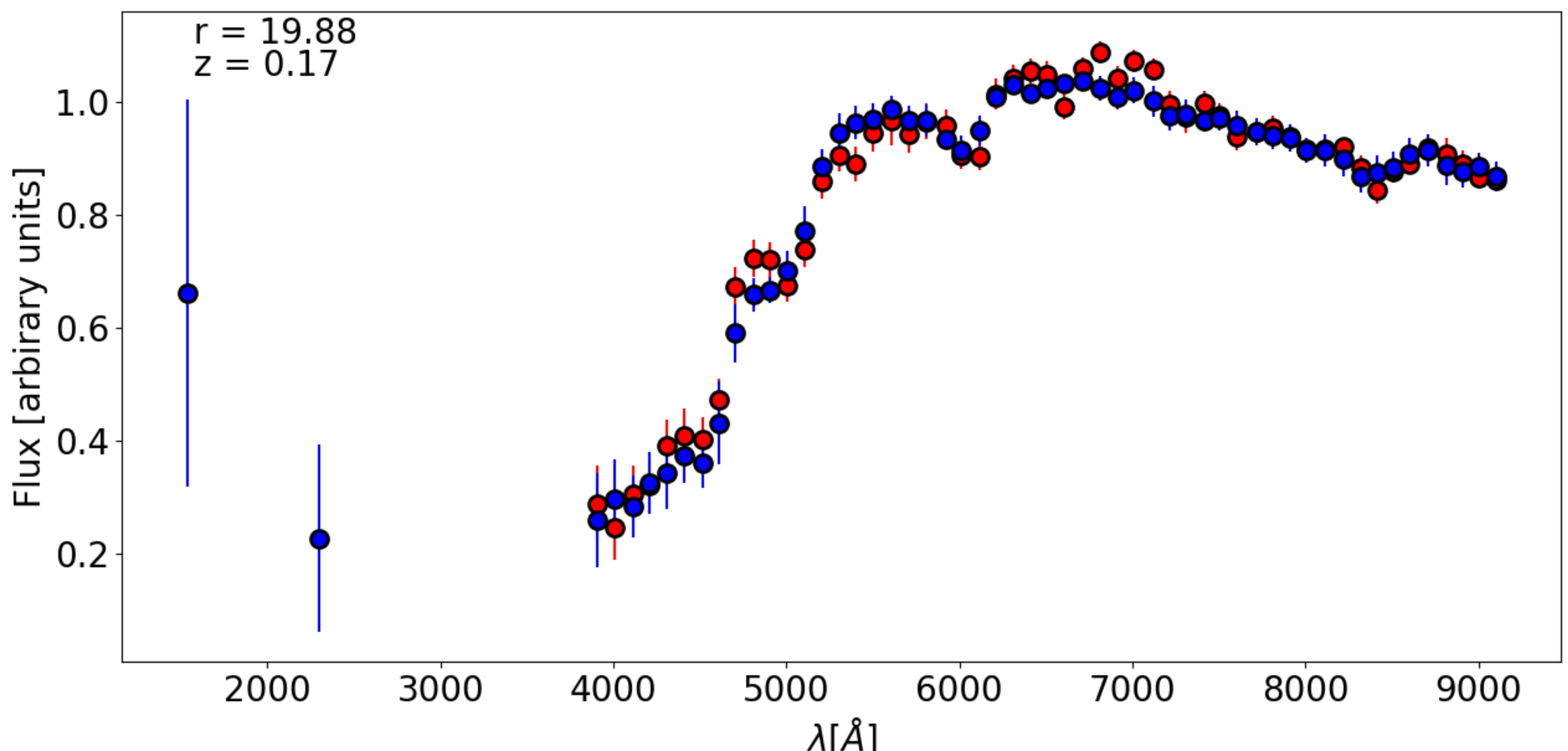
Results on DESI test samples, SP properties



Results on DESI test samples, inverse process



$(u-r)_{int}$	1.72
$(u-r)_{rest}$	2.12
$\log tL$	9.45
$\log tM$	9.75
$\log Z$	-2.57
$\log M$	10.48
$\log SFR$	-0.04
t_0	9.42
τ	1.66
A_v	1.62
Hα	17.0
Hβ	2.76
[OIII]	1.44
[OII]	4.28
[NII]	11.4



$(u-r)_{int}$	2.26
$(u-r)_{rest}$	2.32
$\log tL$	9.23
$\log tM$	9.27
$\log Z$	-1.28
$\log M$	9.75
$\log SFR$	-1.58
t_0	2.35
τ	0.22
A_v	1.08
Hα	-0.71
Hβ	-1.14
[OIII]	-0.75
[OII]	2.4
[NII]	-0.58

Conclusion, limitations, and future improvements

1. We are building a foundational model for galaxy SEDs with the potential to obtain SP properties, emission lines, and redshift of galaxies within a single code.

2. Furthermore, the model can incorporate observations from multiple surveys.

3. Currently, the model is restricted to galaxies; stars and QSOs are not included in the training.

4. The predicted errors assume that the solution can be described as Gaussian.

5. At present, we only use data from miniJPAS to emulate the S/N and depth of our training set. We have not yet trained with real data. However, we will soon incorporate data from the Science Verification Point of J-PAS, enabling the model to be trained with real data, which is crucial for domain adaptation.

6. The model currently incorporates only one modality, integrated photometry. It is possible to develop a model that also trains with galaxy images, helping to capture other parameters such as galaxy morphology.