High Energy Astrophysics and Cosmology in the era of all-sky surveys

A.Meshcheryakov & RU eRosita catalog group

Machine learning models for SRG/eRosita extragalactic sky: challenges, results and perspectives

SRG/eRosita status: 812 days all-sky X-ray survey

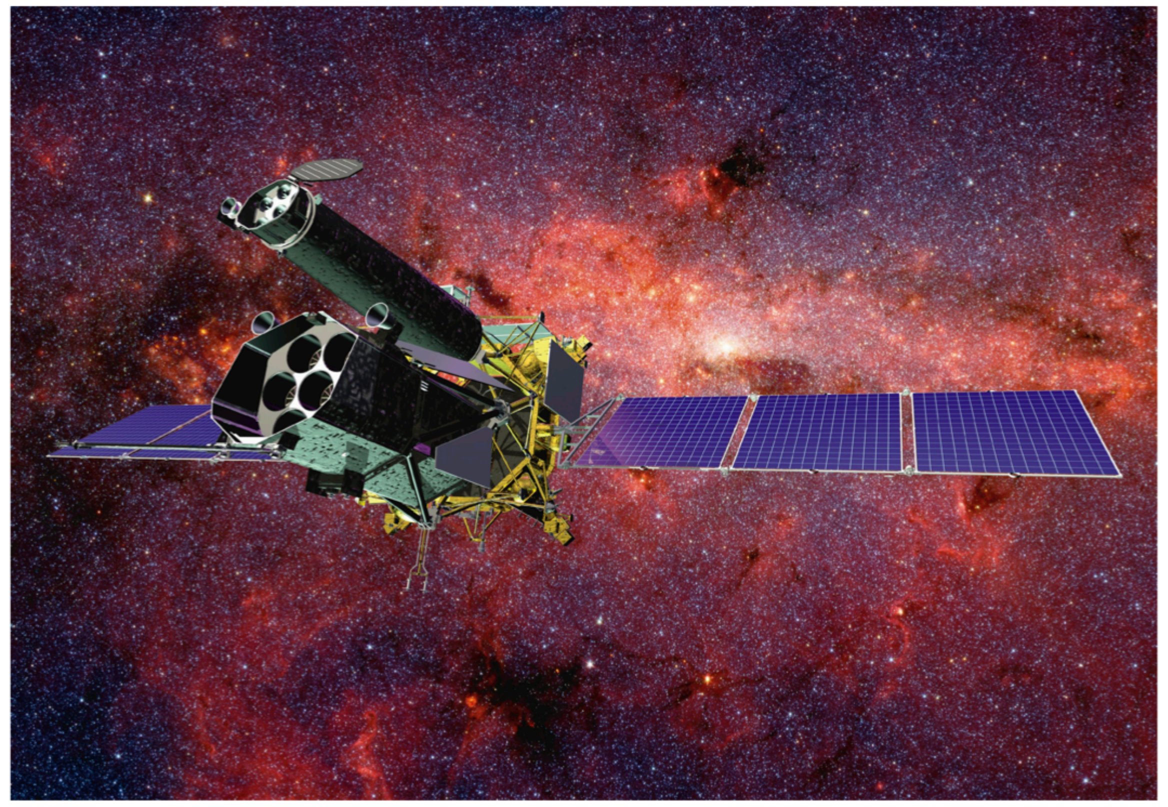
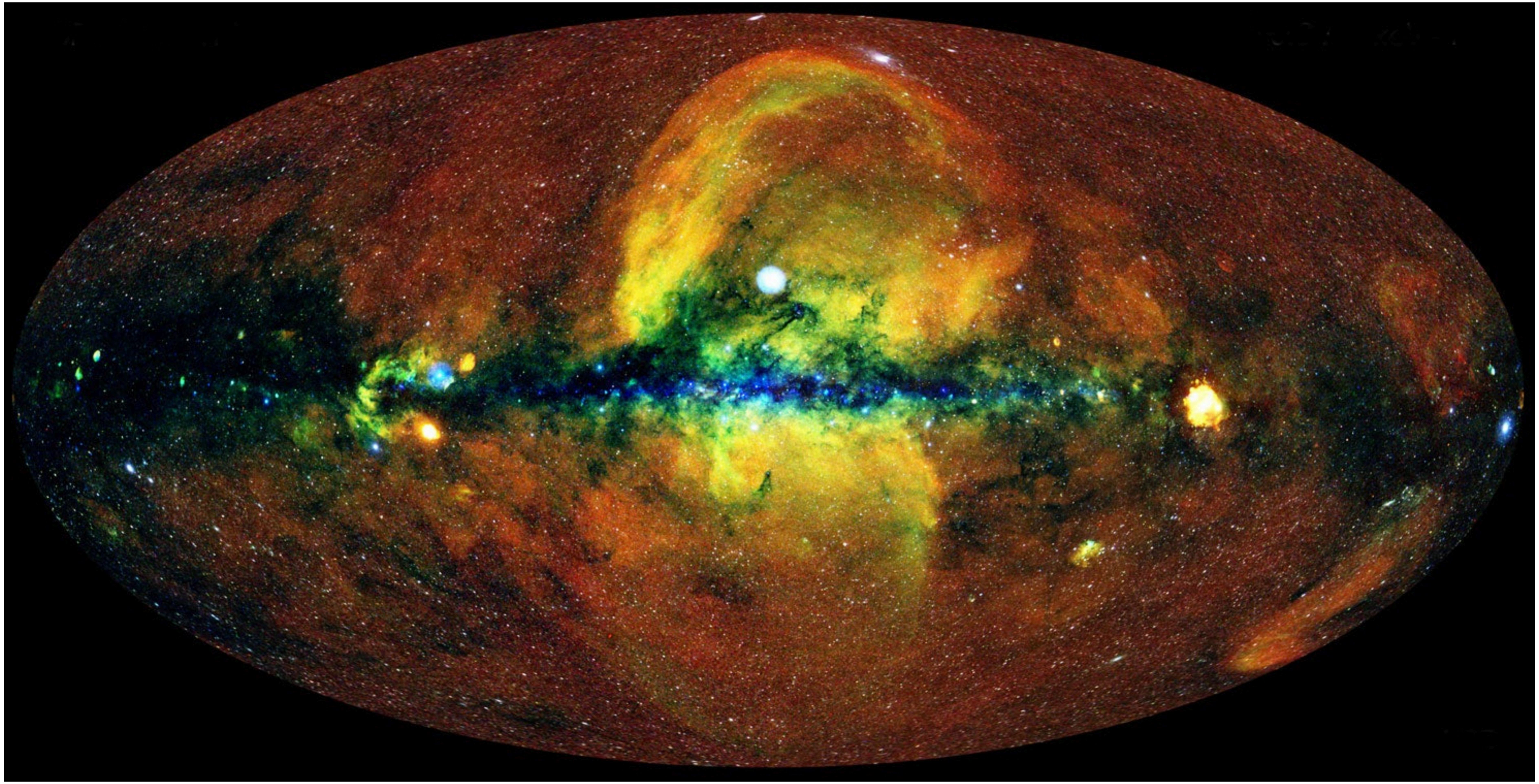


Fig. 3. SRG observatory in flight (artist's impression). Each X-ray telescope consists of seven independent mirror modules.

Fig. 1. Baikonur launch site (Kazakhstan): Proton rocket and the DM-03 upper stage with the SRG spacecraft.

Fig. 2. SRG orbital observatory with the folded solar panels in NPO Lavochkin's assembly hall before shipment to Baikonur.



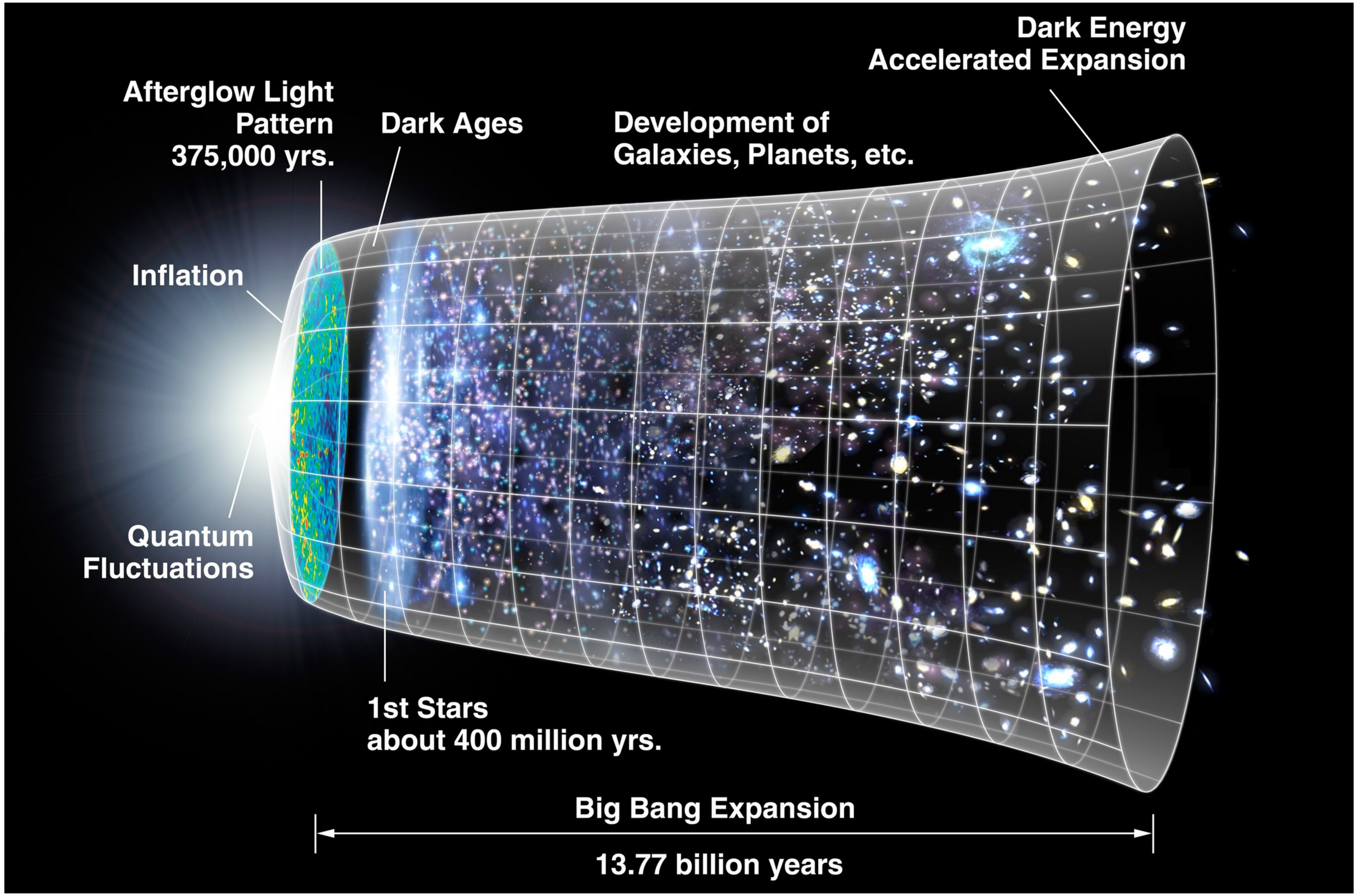
mid-Dec, 2019.
eROSITA completes the Calibration and Performance Verification (Cal-PV) program.

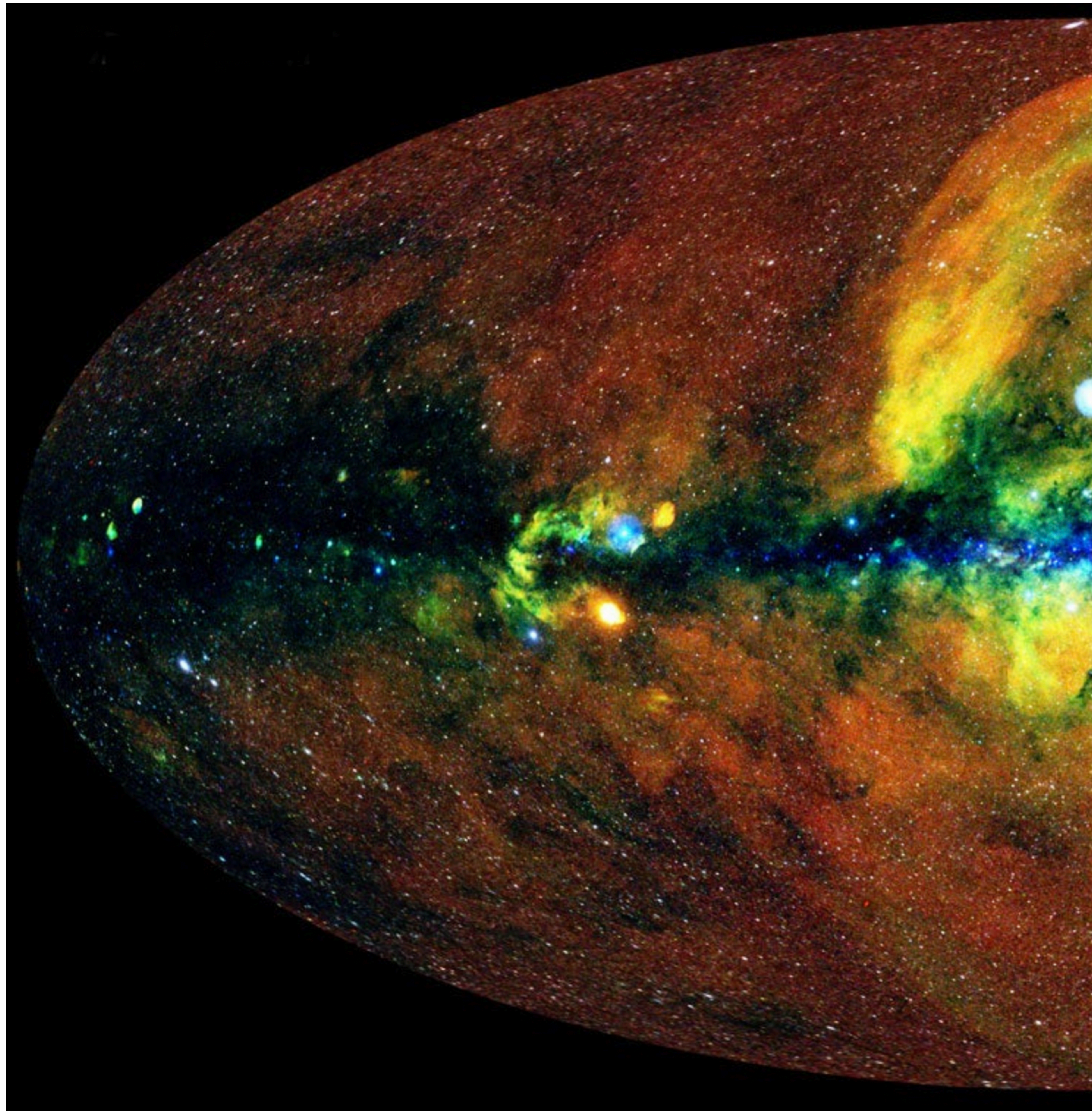
June 10, 2020
SRG completes first all-sky X-ray survey, which lasted from December 8, 2019 to June 10, 2020.

Feb 26, 2022
eROSITA suspended operations and placed in safe mode by German side.

..?

Why we need SRGz for eRosita ?





Data,
ML/DL

SRGz system (2018-2024):

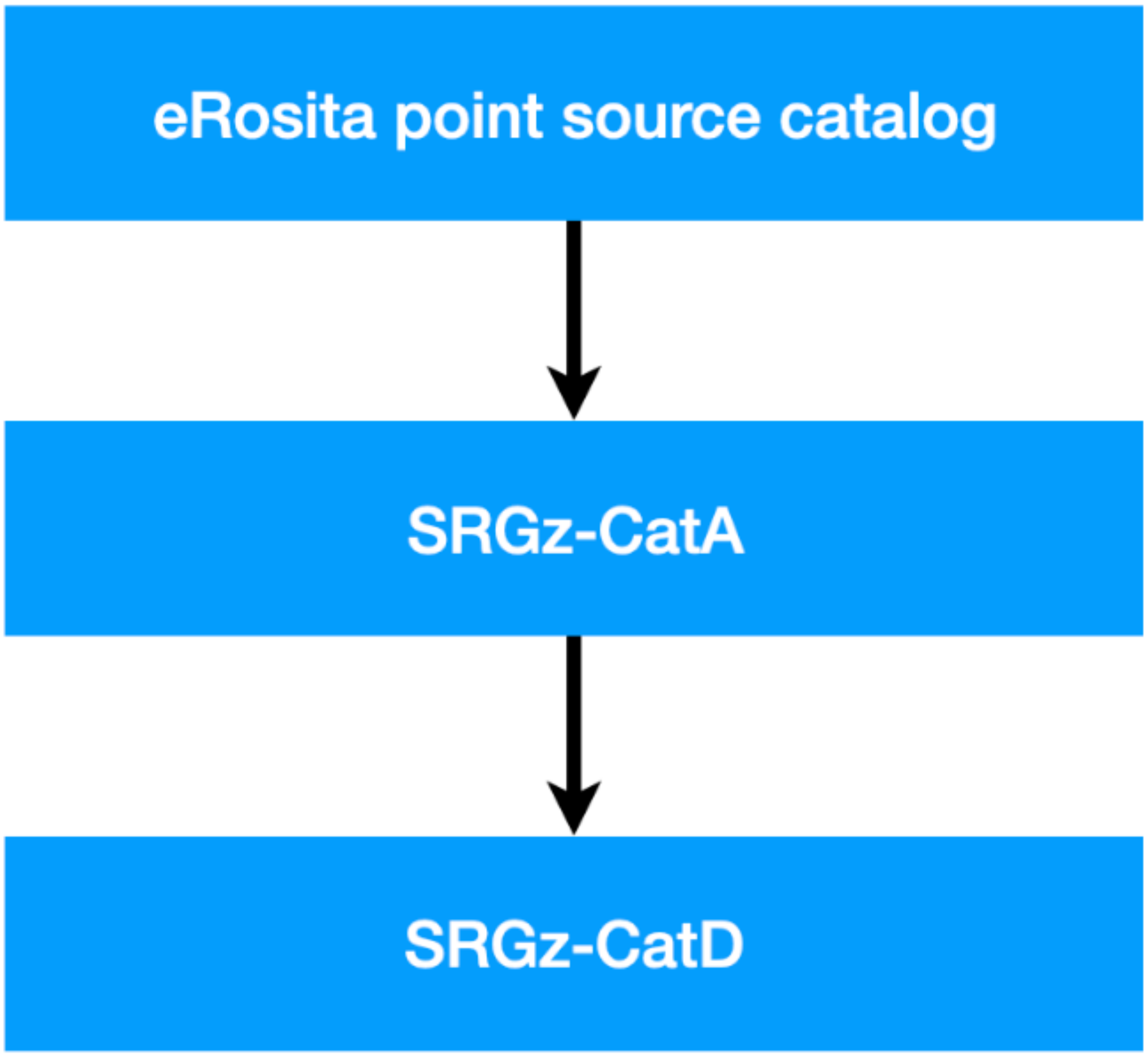
[1] Meshcheryakov, A. V., Khorunzhev, G. A., Voskresenskaya, S. A., Medvedev, P. S., Gilfanov, M. R., and Sunyaev, R. A., “**SRGz: Classification of eROSITA Point X-ray Sources in the 1%DESI Region and Calibration of Photometric Redshifts**”, Astronomy Letters, vol. 49, no. 11, Springer, pp. 646–661, 2023. doi:10.1134/S1063773723110129.

[2] Meshcheryakov, A. V., “**SRGz: Machine Learning Methods and Properties of the Catalog of SRG/eROSITA Point X-ray Source Optical Counterparts in the DESI Legacy Imaging Surveys Footprint**”, Astronomy Letters, vol. 49, no. 7, Springer, pp. 359–409, 2023. doi:10.1134/S1063773723070022.

[3] Borisov, V., Meshcheryakov, A., and Gerasimov, S., “**Probabilistic Photo-z Machine Learning Models for X-ray Sky Surveys**”, in Astronomical Data Analysis Software and Systems XXX, 2022, vol. 532, p. 231. doi:10.48550/arXiv.2107.01891.

[4] Khorunzhev, G. A., “**Discovery of the Most X-ray Luminous Quasar SRGE J170245.3+130104 at Redshift z≈5.5**”, Astronomy Letters, vol. 47, no. 3, Springer, pp. 123–140, 2021. doi:10.1134/S1063773721030026.

[5] Meshcheryakov, A. V., Glazkova, V. V., Gerasimov, S. V., and Mashechkin, I. V., “**Measuring the Probabilistic Photometric Redshifts of X-ray Quasars Based on the Quantile Regression of Ensembles of Decision Trees**”, Astronomy Letters, vol. 44, no. 12, Springer, pp. 735–753, 2018. doi:10.1134/S1063773718120058.



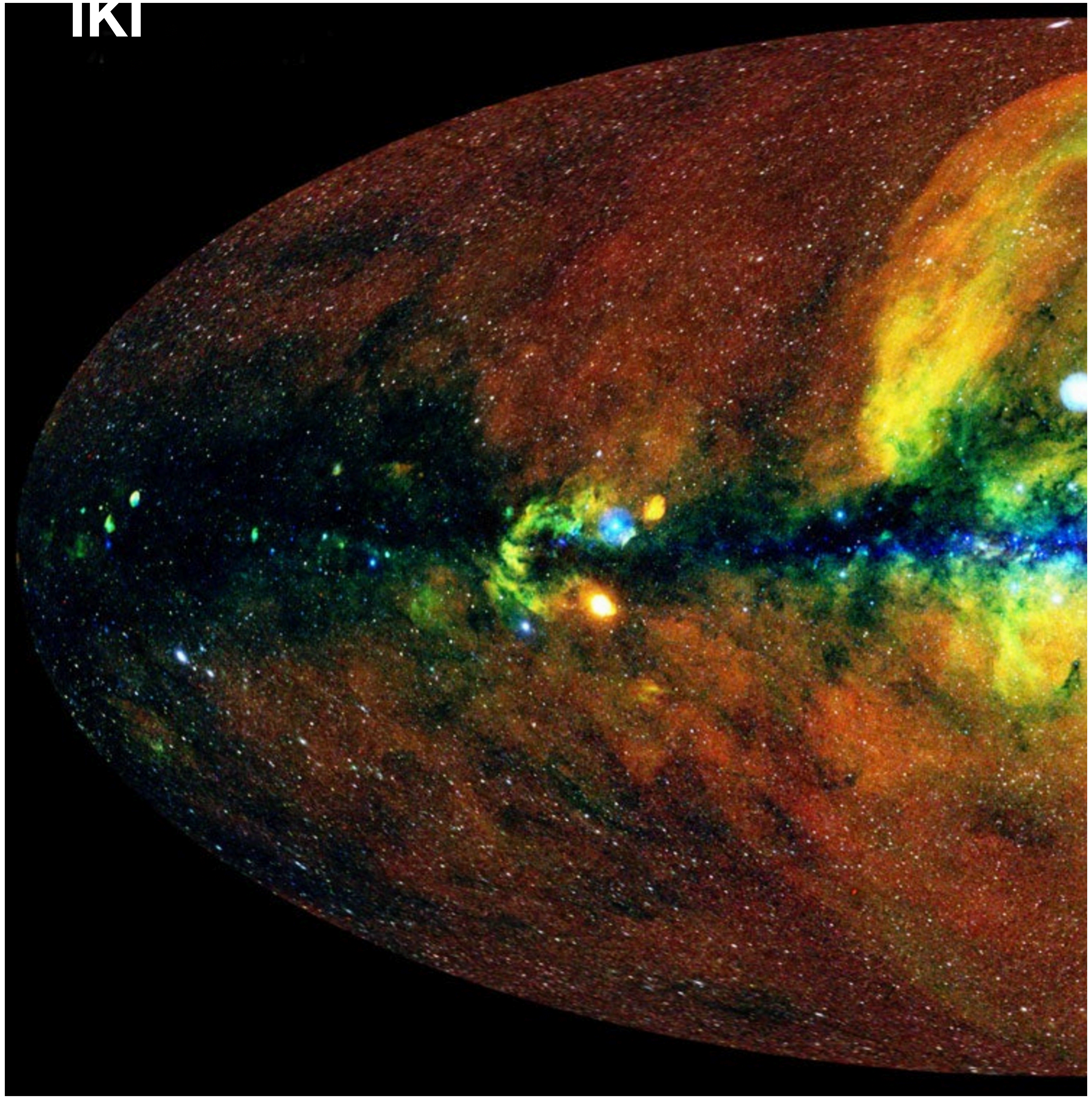
2-year all-sky survey RU eRosita catalog (eRASS4)

All optical objects from DESI Legacy Imaging Surveys in 30” radius from point X-ray sources (Eastern Galactic Hemisphere)

SRGz catalog of most probable optical counterparts of X-ray sources with photometric classification and photo-z

Fig. SRGz pipeline

SRGz-eRosita: 2-year all-sky survey, 0.5-2 keV, $0 < l < 180^{\circ}$, $|b| > 20^{\circ}$



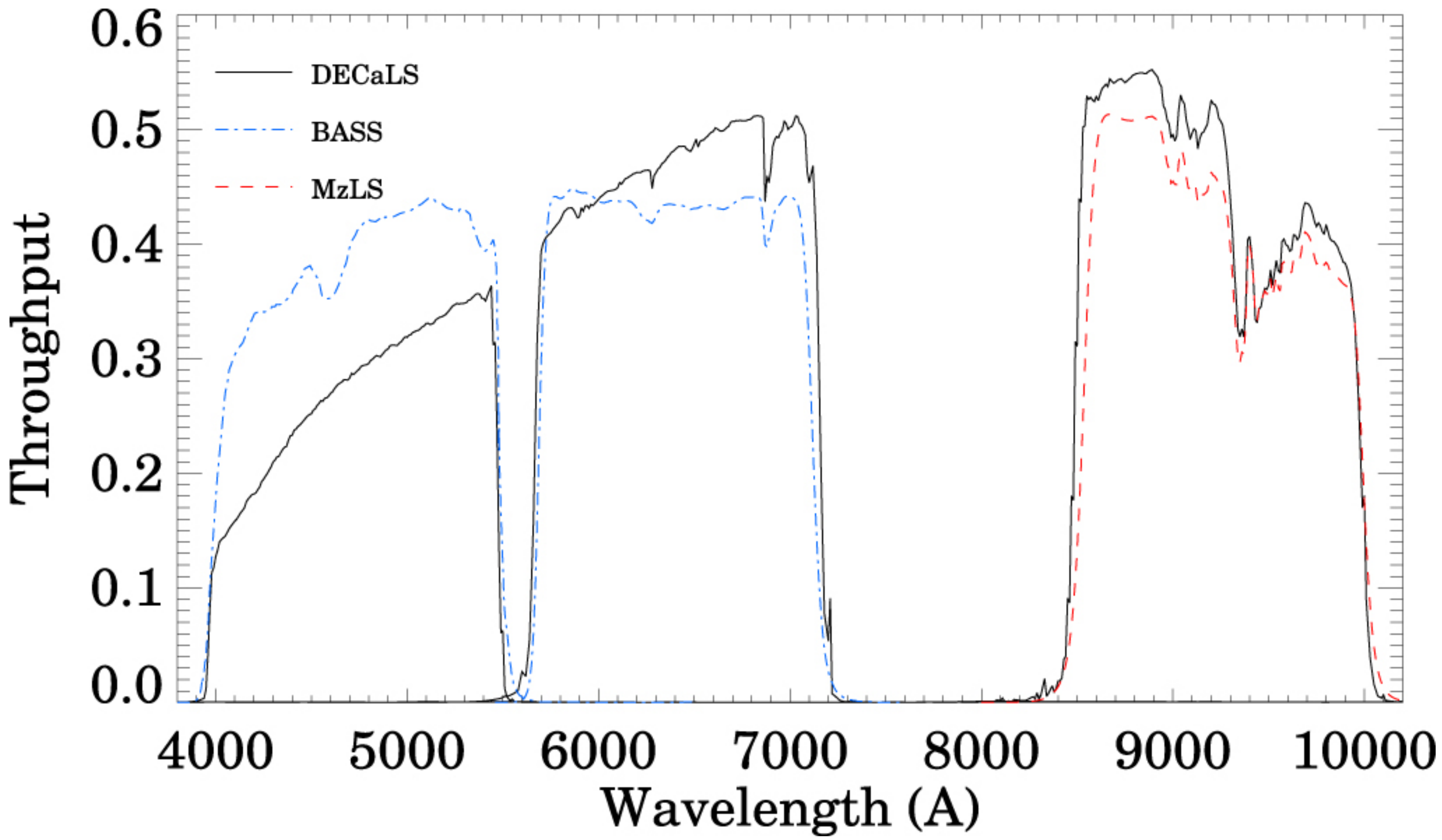
ero_srcname		X-ray object - RU eRosita
ero_RA		
ero_DEC		
ls_ra		optical object - DESI Legacy Imaging Surveys
ls_dec		
srg_match_flag		
srg_match_p	p_i	optical match probabilities
srg_match_pi	P_i	
srg_match_p0	$P_{\emptyset}$	
srg_match_pstar		photometric classification probabilities
srg_match_pqso		
srg_match_pgal		
srg_match_SQG.		
srg_match_warning		
srgz_z_max	z_{ph}	photo-z point estimate and its reliability
srgz_z_maxConf	$zConf$	
srgz_z_merr68		
srgz_z_perr68		
srgz_z_model		
srgz_z_warning		

SRGz measurements for 1.3×10^6 X-ray sources

Data & ML

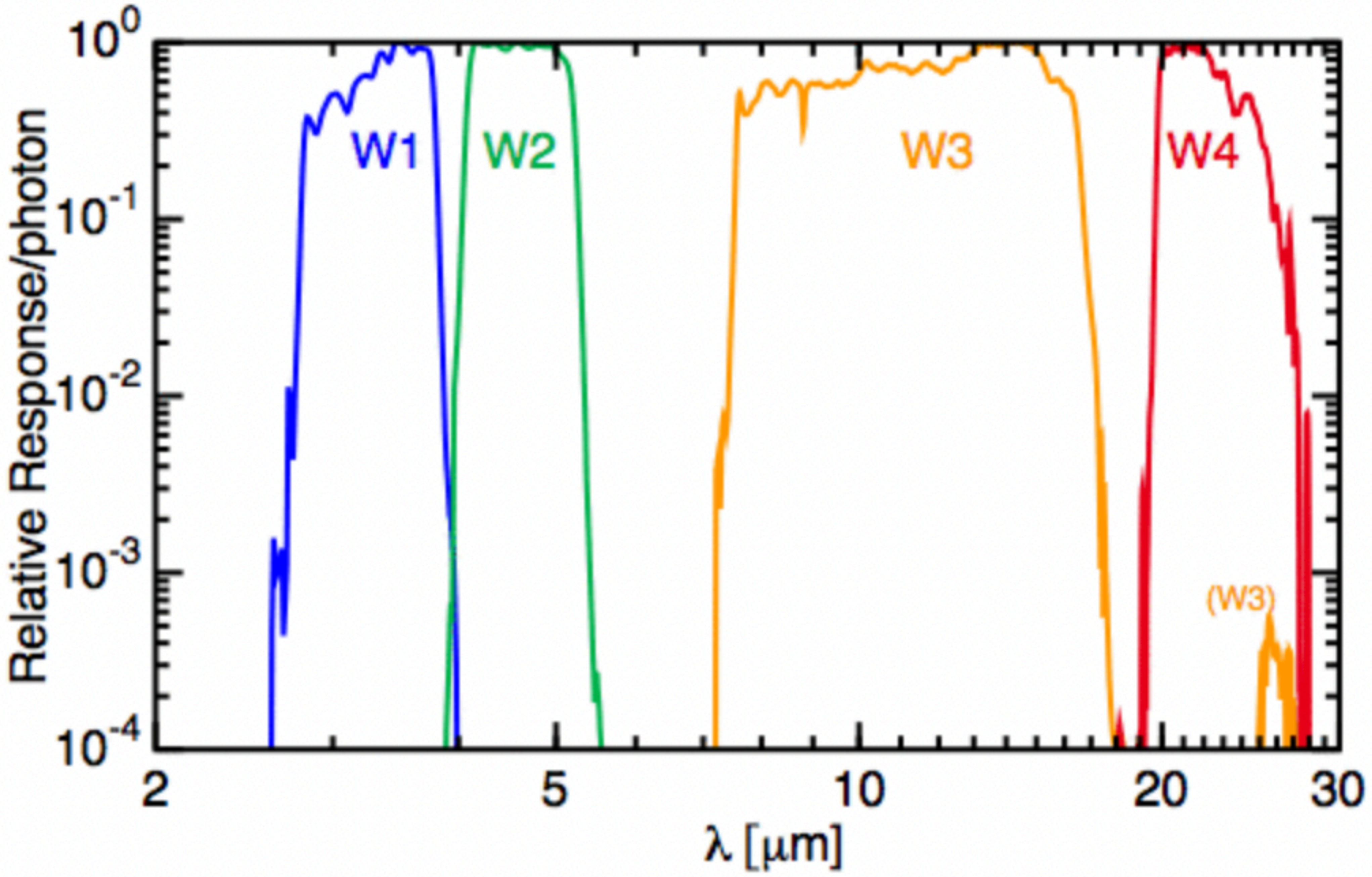
DESI Legacy Imaging Surveys

g r (i) z optical surveys



+

W1 W2 (W3 W4) WISE forced photometry



Optical/IR photometric data

Pan-STARRS

SDSS

DESI Legacy Imaging Surveys

NEOWISE

X-ray data

4XMM-DR13

CXC2

Neil Gehrels Swift Observatory

2SXPS

eROSITA

SRGz: 0.5 - 2 keV

Astrometric & spectroscopic data

esa gaia

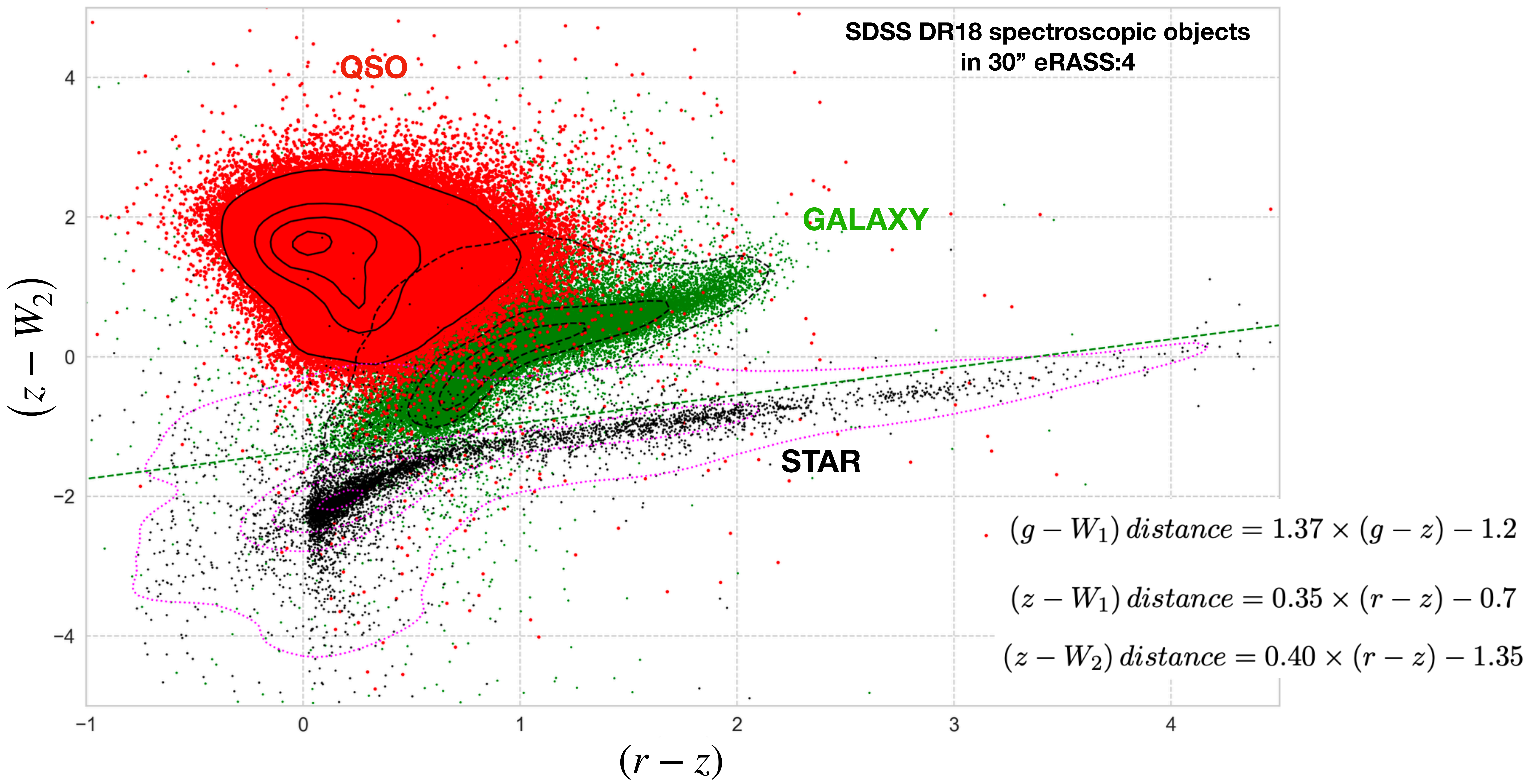
SDSS

DESI EDR

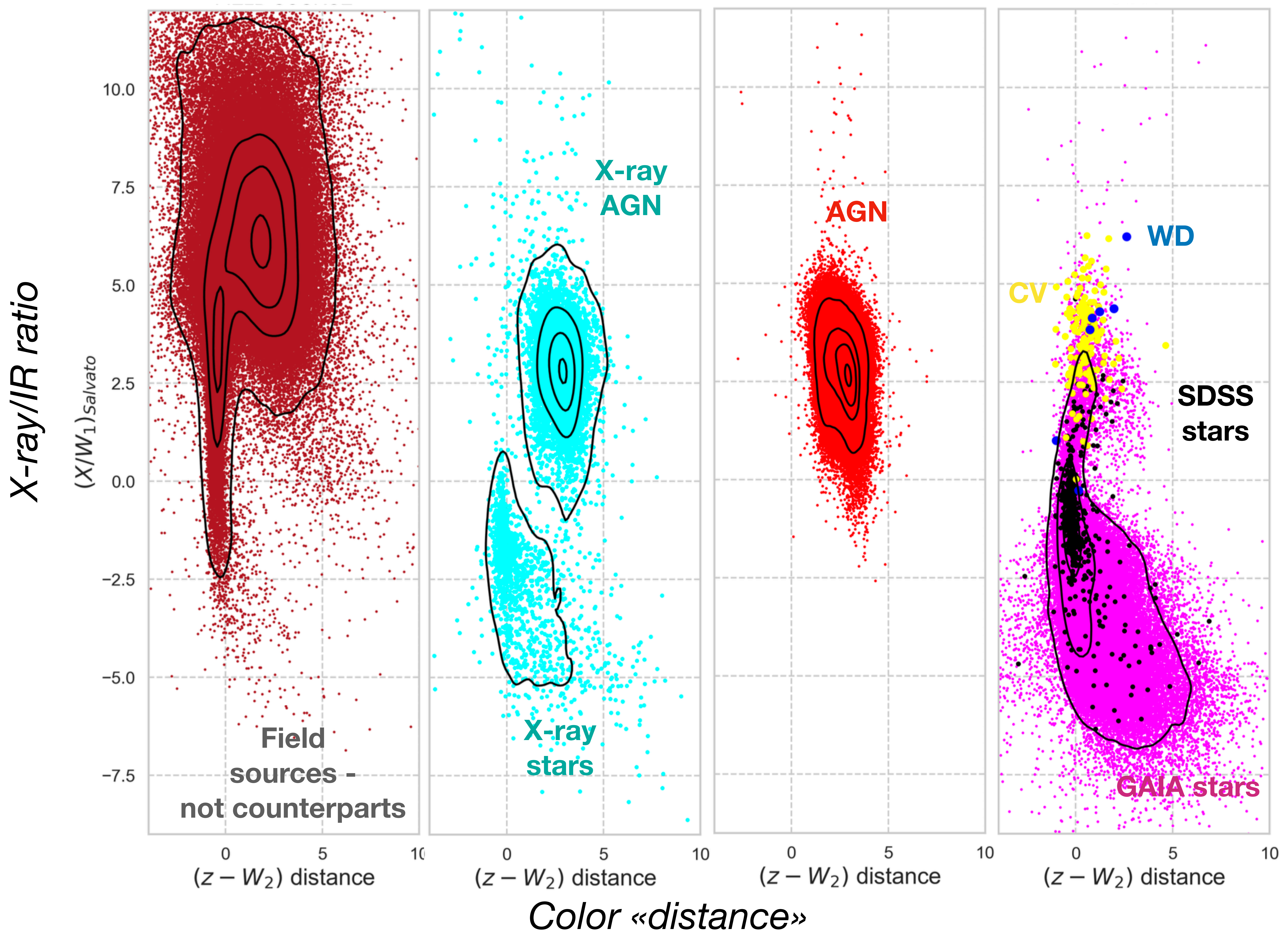
HELP

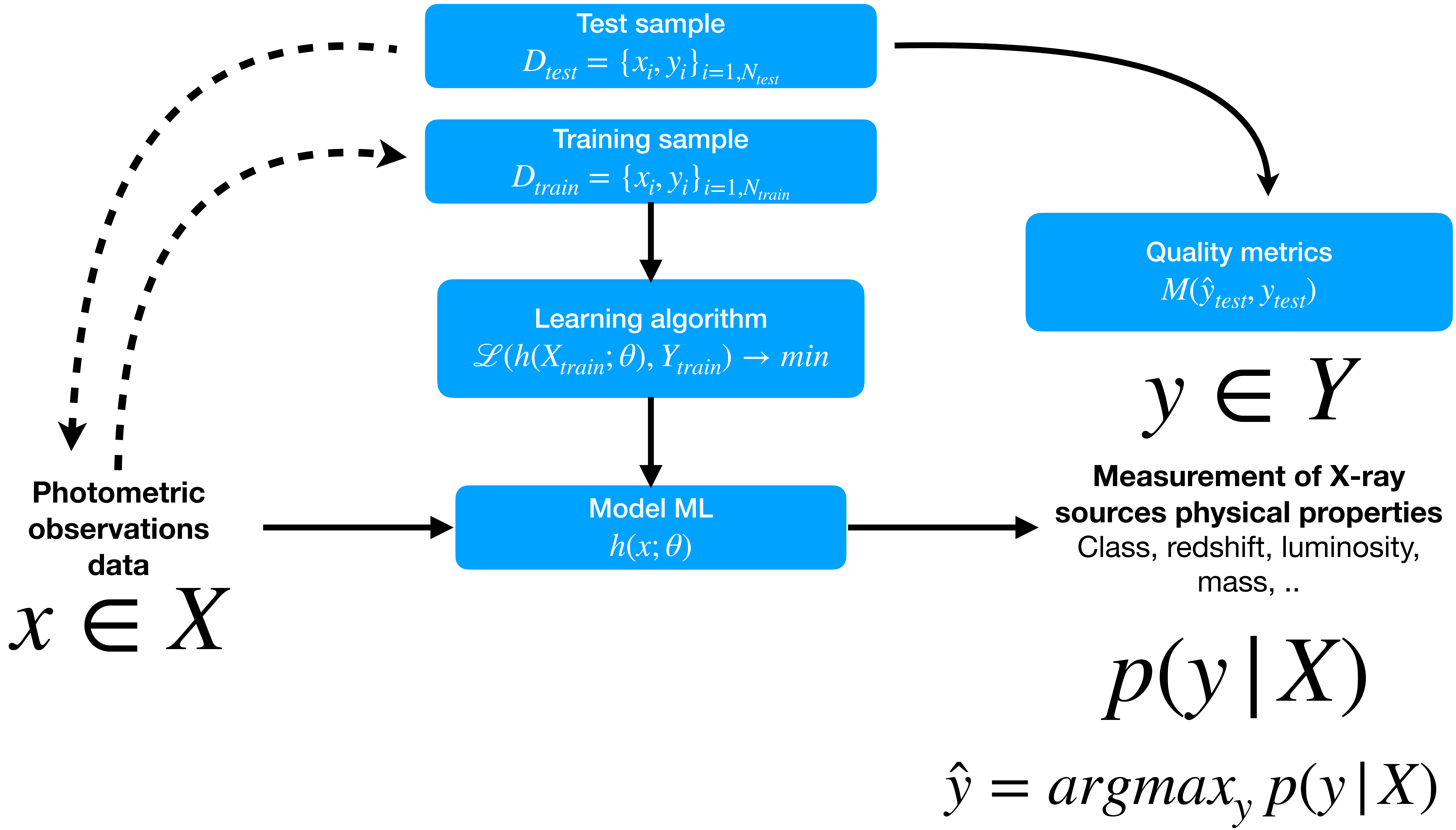
Herschel Extragalactic Legacy Project spectroscopic redshifts

Color «distances»



Point-like optical DESI LIS sources in the vicinity of eRosita X-ray objects




Test sample
$D_{test} = \{x_i, y_i\}_{i=1,N_{test}}$
Training sample
$D_{train} = \{x_i, y_i\}_{i=1,N_{train}}$
Learning algorithm
$\mathscr{L}(h(X_{train}; \theta), Y_{train}) \rightarrow min$
Model ML
$h(x; \theta)$
Quality metrics
$M(\hat{y}_{test}, y_{test})$
$y \in Y$
Photometric observations data
$x \in X$
Measurement of X-ray sources physical properties
Class, redshift, luminosity, mass, ..
$p(y|X)$
$\hat{y} = argmax_y\, p(y|X)$

Photometric features for optical match and classification models

$$P_{counterpart}, P_{fieldsource} = F^{match}(Xph)$$

	— $Model_{match}$
3_X	$offset$, n_{20}, n_{30}, g_{lim}, r_{lim}, z_{lim}, $W_{1,lim}$, $W_{2,lim}$, $type$, $sersic$, mag_X, X/g, X/r, X/z, X/W_1, X/W_2, $(X/W_1)_{salvato}$, g, r, z, W_1, W_2, $(g-r)$, $(g-z)$, $(g-W_1)$, $(g-W_2)$, $(r-z)$, $(r-W_1)$, $(r-W_2)$, $(z-W_1)$, $(z-W_2)$, (W_1-W_2), $(g-W_1)$ distance, $(z-W_1)$ distance, $(z-W_2)$ distance

$$P^{QvSG}_{QSO}, P^{QvSG}_{STAR+GAL} = F^{QvSG}(X_{ph})$$

$$P^{GvSQ}_{GAL}, P^{GvSQ}_{STAR+QSO} = F^{GvSQ}(X_{ph})$$

$$P^{SvQG}_{STAR}, P^{SvQG}_{QSO+GAL} = F^{SvQG}(X_{ph})$$

	— $Model1_{class}$, $Model2_{class}$
$3_{X,col}$	$type$, $sersic$, X/g, X/r, X/z, X/W_1, X/W_2, $(X/W_1)_{salvato}$, $(g-r)$, $(g-z)$, $(g-W_1)$, $(g-W_2)$, $(r-z)$, $(r-W_1)$, $(r-W_2)$, $(z-W_1)$, $(z-W_2)$, (W_1-W_2), $(g-W_1)$ distance, $(z-W_1)$ distance, $(z-W_2)$ distance

Photometric features for photo-z models

Признаки моделей измерения фотометрических красных смещений — $Model_{photoz}$	
2	g, r, z, W_1,W_2, $(g-r)$, $(g-z)$, $(g-W_1)$, $(g-W_2)$, $(r-z)$, $(r-W_1)$, $(r-W_2)$, $(z-W_1)$, $(z-W_2)$, (W_1-W_2)
3_{PS}	$features:2+$ $g_{PS,psf}$, $r_{PS,psf}$, $i_{PS,psf}$, $z_{PS,psf}$, $y_{PS,psf}$, $i_{PS,kron}$, $y_{PS,kron}$, $g_{PS,kron}$, $r_{PS,kron}$, $z_{PS,kron}$, $(g_{PS,psf}-i_{PS,psf})$, $(g_{PS,psf}-y_{PS,psf})$, $(r_{PS,psf}-i_{PS,psf})$, $(r_{PS,psf}-y_{PS,psf})$, $(i_{PS,psf}-z_{PS,psf})$, $(i_{PS,psf}-y_{PS,psf})$, $(z_{PS,psf}-y_{PS,psf})$, $(i_{PS,psf}-i_{PS,kron})$, $(y_{PS,psf}-y_{PS,kron})$, $(g_{PS,psf}-r_{PS,psf})$, $(g_{PS,psf}-z_{PS,psf})$, $(r_{PS,psf}-z_{PS,psf})$, $(g_{PS,psf}-g_{PS,kron})$, $(r_{PS,psf}-r_{PS,kron})$, $(z_{PS,psf}-z_{PS,kron})$, $(g_{PS,kron}-g)$, $(r_{PS,kron}-r)$, $(z_{PS,kron}-z)$, $(g_{PS,kron}-W_1)$, $(g_{PS,kron}-W_2)$, $(r_{PS,kron}-W_1)$, $(r_{PS,kron}-W_2)$, $(i_{PS,kron}-W_1)$, $(i_{PS,kron}-W_2)$, $(z_{PS,kron}-W_1)$, $(z_{PS,kron}-W_2)$, $(y_{PS,kron}-W1)$, $(y_{PS,kron}-W_2)$
3_{SDSS}	$features:2+$ u_{psf}, g_{psf}, r_{psf}, i_{psf}, z_{psf}, u_{cmodel}, i_{cmodel}, g_{cmodel}, r_{cmodel}, z_{cmodel}, $(u_{psf}-g_{psf})$, $(u_{psf}-r_{psf})$, $(u_{psf}-i_{psf})$, $(u_{psf}-z_{psf})$, $(u_{psf}-u_{cmodel})$, $(g_{psf}-i_{psf})$, $(g_{psf}-g_{cmodel})$, $(r_{psf}-i_{psf})$, $(i_{psf}-z_{psf})$, $(i_{psf}-i_{cmodel})$, $(g_{psf}-r_{psf})$, $(g_{psf}-z_{psf})$, $(r_{psf}-z_{psf})$, $(r_{psf}-r_{cmodel})$, $(z_{psf}-z_{cmodel})$, $(g_{cmodel}-g)$, $(r_{cmodel}-r)$, $(z_{cmodel}-z)$, $(u_{cmodel}-W_1)$, $(u_{cmodel}-W_2)$, $(g_{cmodel}-W_1)$, $(g_{cmodel}-W_2)$, $(r_{cmodel}-W_1)$, $(r_{cmodel}-W_2)$, $(i_{cmodel}-W_1)$, $(i_{cmodel}-W_2)$, $(z_{cmodel}-W_1)$, $(z_{cmodel}-W_2)$
4	SET$(features:2+features:3_{SDSS}+features:3_{PS})$

$\rightarrow P_2(z\,|\,X_{ph}), z_{2,ph}, zConf_2$

$\rightarrow P_{3PS}(z\,|\,X_{ph}), z_{3PS,ph}, zConf_{3PS}$

$\rightarrow P_{3SDSS}(z\,|\,X_{ph}), z_{3SDSS,ph}, zConf_{3SDSS}$

$\rightarrow P_4(z\,|\,X_{ph}), z_{4,ph}, zConf_4$

$$4 \rightarrow 3_{SDSS} \rightarrow 3_{PS} \rightarrow 2$$

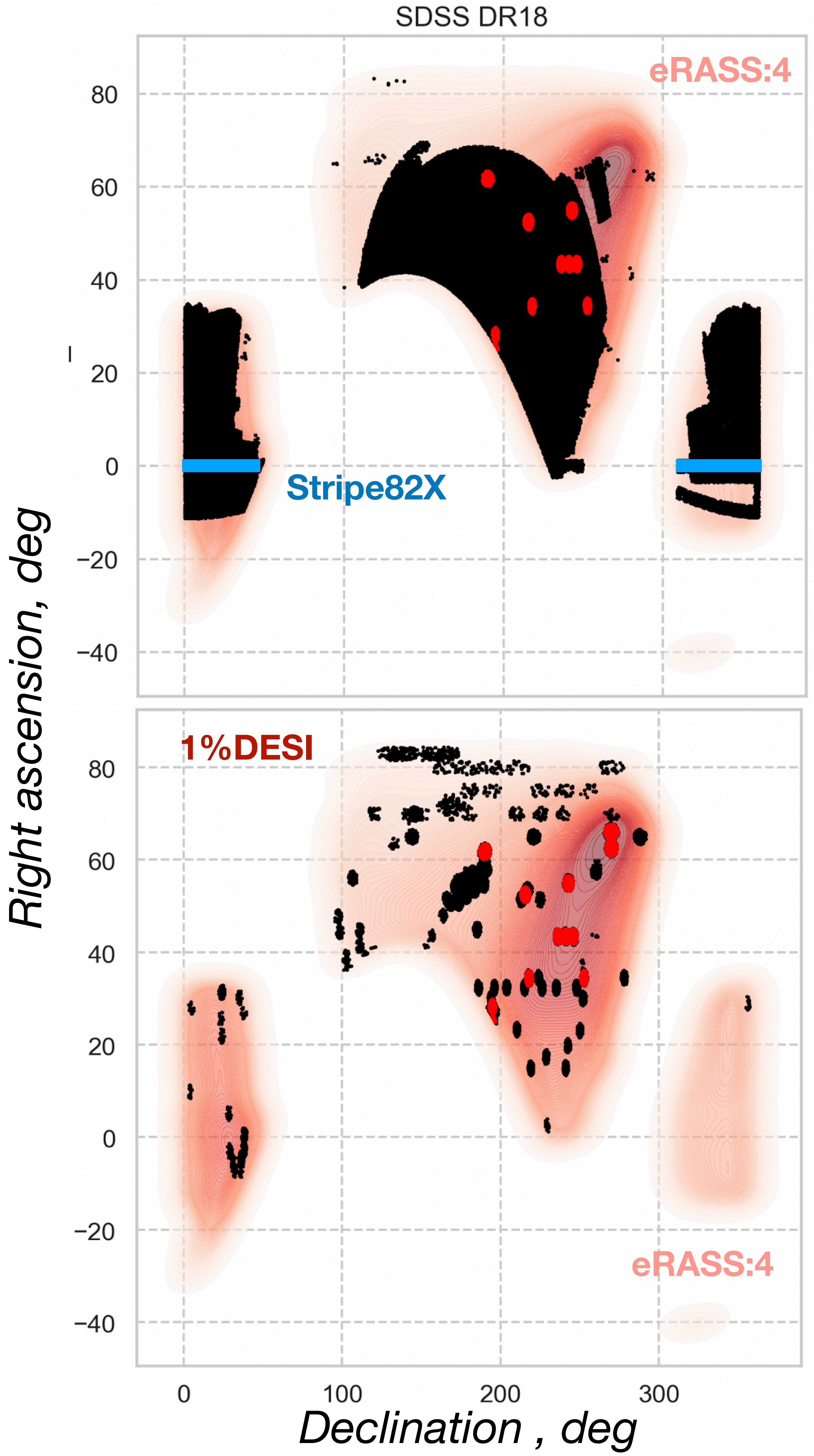
Training set:

586×10^3 **spectral objects -**

QSO: *SDSS DR14q, Ross&Cross (2020)*

GALAXY: SDSS DR16

Test samples



- SDSS field - 4179 sq.deg
- 1%DESI-East - 91.4 sq.deg
- Stripe82X - 31.3 sq.deg

Stripe 82X survey multiwavelength catalog (Ananna+, 2017)

Methods: ML

When Do Neural Nets Outperform Boosted Trees on Tabular Data?

Duncan McElfresh*[1,2], Sujay Khandagale[3], Jonathan Valverde[4], Vishak Prasad C[5], Ganesh Ramakrishnan[5], Micah Goldblum[6], Colin White[1,7]

[1] Abacus.AI, [2] Stanford, [3] Pinterest, [4] University of Maryland,
[5] IIT Bombay, [6] New York University, [7] Caltech

Abstract

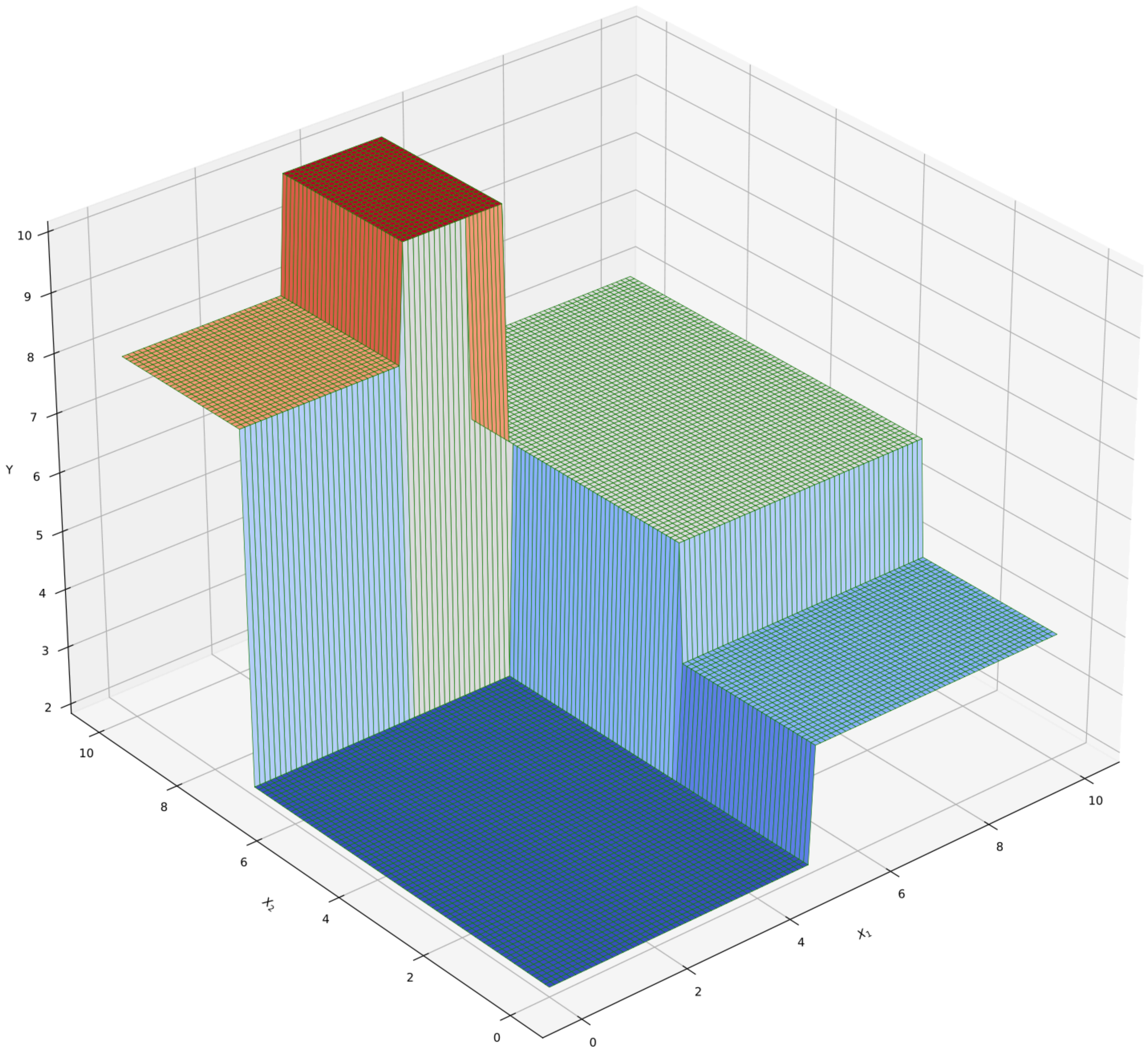
Tabular data is one of the most commonly used types of data in machine learning. Despite recent advances in neural nets (NNs) for tabular data, there is still an active discussion on whether or not NNs generally outperform gradient-boosted decision trees (GBDTs) on tabular data, with several recent works arguing either that GBDTs consistently outperform NNs on tabular data, or vice versa. In this work, we take a step back and question the importance of this debate. To this end, we conduct the largest tabular data analysis to date, comparing 19 algorithms across 176 datasets, and we find that the 'NN vs. GBDT' debate is overemphasized: for a surprisingly high number of datasets, either the performance difference between GBDTs and NNs is negligible, or light hyperparameter tuning on a GBDT is more important than choosing between NNs and GBDTs. Next, we analyze dozens of metafeatures to determine what *properties* of a dataset make NNs or GBDTs better-suited to perform well. For example, we find that GBDTs are much better than NNs at handling skewed or heavy-tailed feature distributions and other forms of dataset irregularities. Our insights act as a guide for practitioners to determine which techniques may work best on their dataset. Finally, with the goal of accelerating tabular data research, we release the TabZilla Benchmark Suite: a collection of the 36 'hardest' of the datasets we study. Our benchmark suite, codebase, and all raw results are available at https://github.com/naszilla/tabzilla.

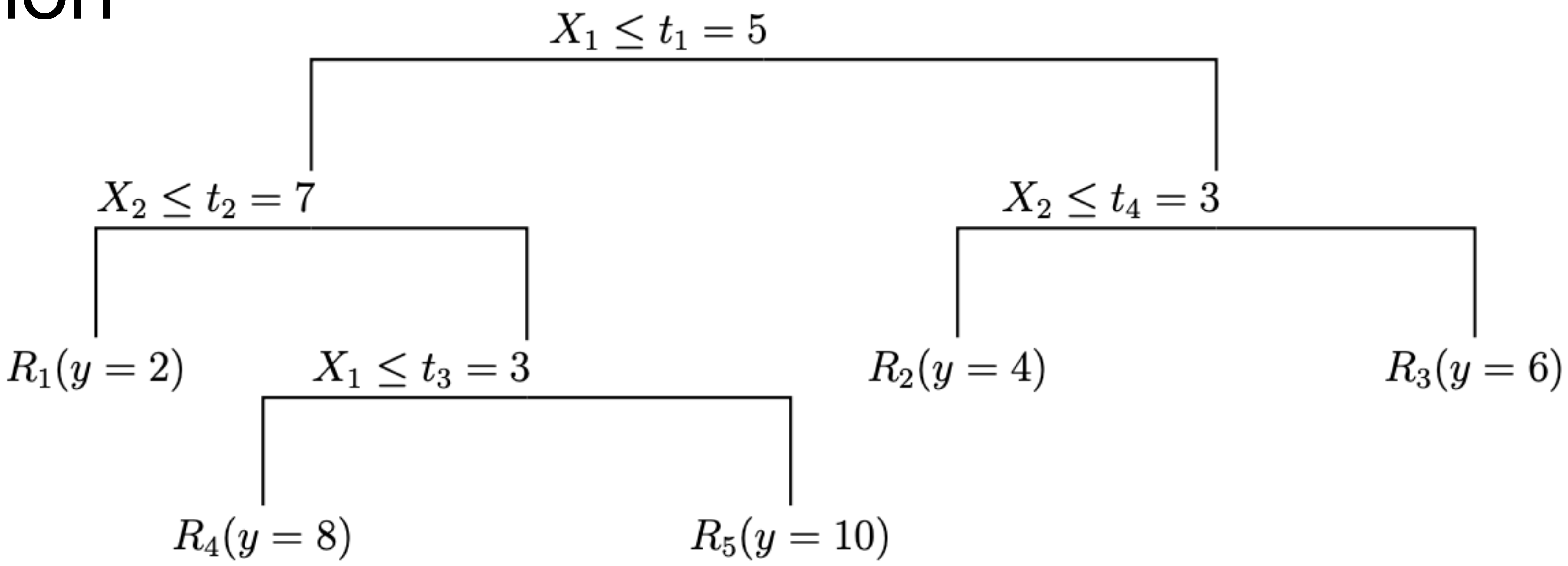
Advances in Neural Information Processing Systems 36 (2024), arXiv:2305.02997

	Hardness Metrics			Dataset Attributes			Top 3 Algs.		
Dataset	base	4th-best	GBDT	N	# feats.	Std. Kurtosis	1st	2nd	3rd
credit-g	**0.26**	**0.13**	**0.12**	1 000	21	1.92	ResNet	FTTransformer	CatBoost
jungle-chess	**0.30**	**0.18**	**0.17**	44 819	7	0.08	SAINT	TabNet	LightGBM
MiniBooNE	**0.20**	**0.09**	0.00	130 064	51	12162.65	LightGBM	XGBoost	CatBoost
albert	**0.42**	**0.28**	0.00	425 240	79	1686.90	CatBoost	XGBoost	ResNet
electricity	**0.46**	**0.38**	0.00	45 312	9	2693.51	LightGBM	XGBoost	FTTransformer
elevators	**0.36**	**0.08**	0.05	16 599	19	2986.50	TabNet	XGBoost	CatBoost
guillermo	**0.35**	**0.60**	0.00	20 000	4 297	NaN	XGBoost	RandomForest	TabNet
higgs	**0.41**	**0.10**	0.07	98 050	29	15.53	ResNet	XGBoost	LightGBM
nomao	**0.22**	**0.18**	0.00	34 465	119	1100.34	LightGBM	XGBoost	CatBoost
100-plants-texture	**0.20**	**0.11**	0.00	1 599	65	17.66	CatBoost	XGBoost	ResNet
poker-hand	**0.58**	**0.98**	0.00	1 025 009	11	0.08	XGBoost	CatBoost	KNN
profb	**0.39**	**0.38**	0.00	672	10	0.95	CatBoost	DeepFM	MLP-rtdl
socmob	**0.24**	**0.10**	0.00	1 156	6	NaN	XGBoost	CatBoost	ResNet
audiology	**0.43**	0.03	0.00	226	70	NaN	STG	XGBoost	ResNet
splice	**0.30**	0.03	0.00	3 190	61	NaN	LightGBM	XGBoost	CatBoost
vehicle	0.05	**0.10**	**0.10**	846	19	15.16	TabPFN	SVM	DANet
Australian	0.15	**0.08**	0.00	690	15	2.00	CatBoost	XGBoost	TabPFN
Bioresponse	0.07	**0.07**	0.00	3 751	1 777	328.77	LightGBM	XGBoost	CatBoost
GesturePhase	0.08	**0.08**	0.00	9 872	33	52.18	LightGBM	XGBoost	CatBoost
SpeedDating	0.18	**0.14**	0.00	8 378	121	36.43	XGBoost	CatBoost	LightGBM
ada-agnostic	0.12	**0.11**	0.00	4 562	49	NaN	XGBoost	CatBoost	LightGBM
airlines	**0.20**	**0.18**	0.00	539 382	8	2.01	LightGBM	XGBoost	CatBoost
artificial-characters	0.13	**0.11**	0.00	10 218	8	0.63	XGBoost	LightGBM	CatBoost
colic	0.13	**0.11**	0.00	368	27	4.00	CatBoost	XGBoost	FTTransformer
credit-approval	0.12	**0.08**	0.00	690	16	74.77	CatBoost	TabPFN	XGBoost
heart-h	0.10	**0.07**	0.08	294	14	NaN	DeepFM	TabTransformer	NAM
jasmine	0.13	**0.13**	0.00	2 984	145	47.60	CatBoost	XGBoost	LightGBM
kc1	0.14	**0.07**	0.00	2 109	22	28.34	CatBoost	XGBoost	FTTransformer
lymph	0.14	**0.08**	0.00	148	19	17.04	XGBoost	DANet	SAINT
mfeat-fourier	0.00	**0.07**	0.07	2 000	77	0.64	SVM	SAINT	STG
phoneme	0.10	**0.15**	0.00	5 404	6	1.23	XGBoost	LightGBM	RandomForest
qsar-biodeg	0.08	**0.08**	0.05	1 055	42	93.24	TabPFN	CatBoost	SAINT
balance-scale	0.07	0.05	**0.16**	625	5	0.02	TabPFN	SAINT	MLP
cnae-9	0.11	0.04	**0.10**	1 080	857	NaN	TabTransformer	STG	MLP-rtdl
mfeat-zernike	0.00	0.04	**0.10**	2 000	48	1.42	SVM	DANet	ResNet
monks-problems-2	0.04	0.00	**0.17**	601	7	NaN	SAINT	ResNet	MLP-rtdl

«on average, GBDTs do outperform NNs»

Decision trees for classification & regression



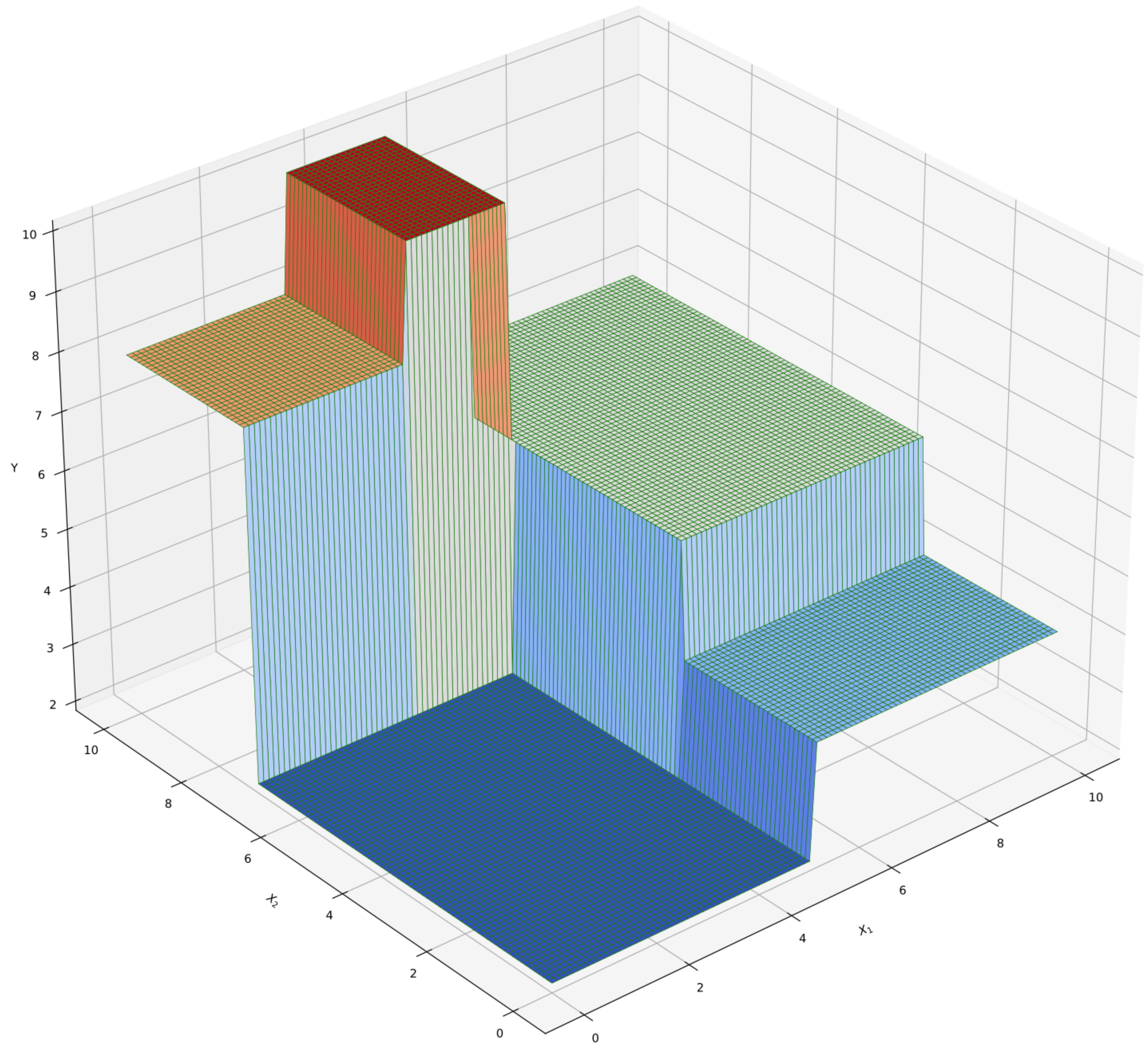


$$R_1 = \{\boldsymbol{x} : x_1 \leq t_1, x_2 \leq t_2\}$$

$$f(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{j=1}^{J} w_j \mathbb{I}\left(\boldsymbol{x} \in R_j\right)$$

$$w_j = \frac{\sum_{n=1}^{N} y_n \mathbb{I}\left(\boldsymbol{x}_n \in R_j\right)}{\sum_{n=1}^{N} \mathbb{I}\left(\boldsymbol{x}_n \in R_j\right)}$$

DT: model learning



Empirical Risc minimisation

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{n=1}^{N} \ell(y_n, f(\boldsymbol{x}_n; \boldsymbol{\theta})) = \sum_{j=1}^{J} \sum_{\boldsymbol{x}_n \in R_j} \ell(y_n, w_j) \longrightarrow \textbf{min}$$

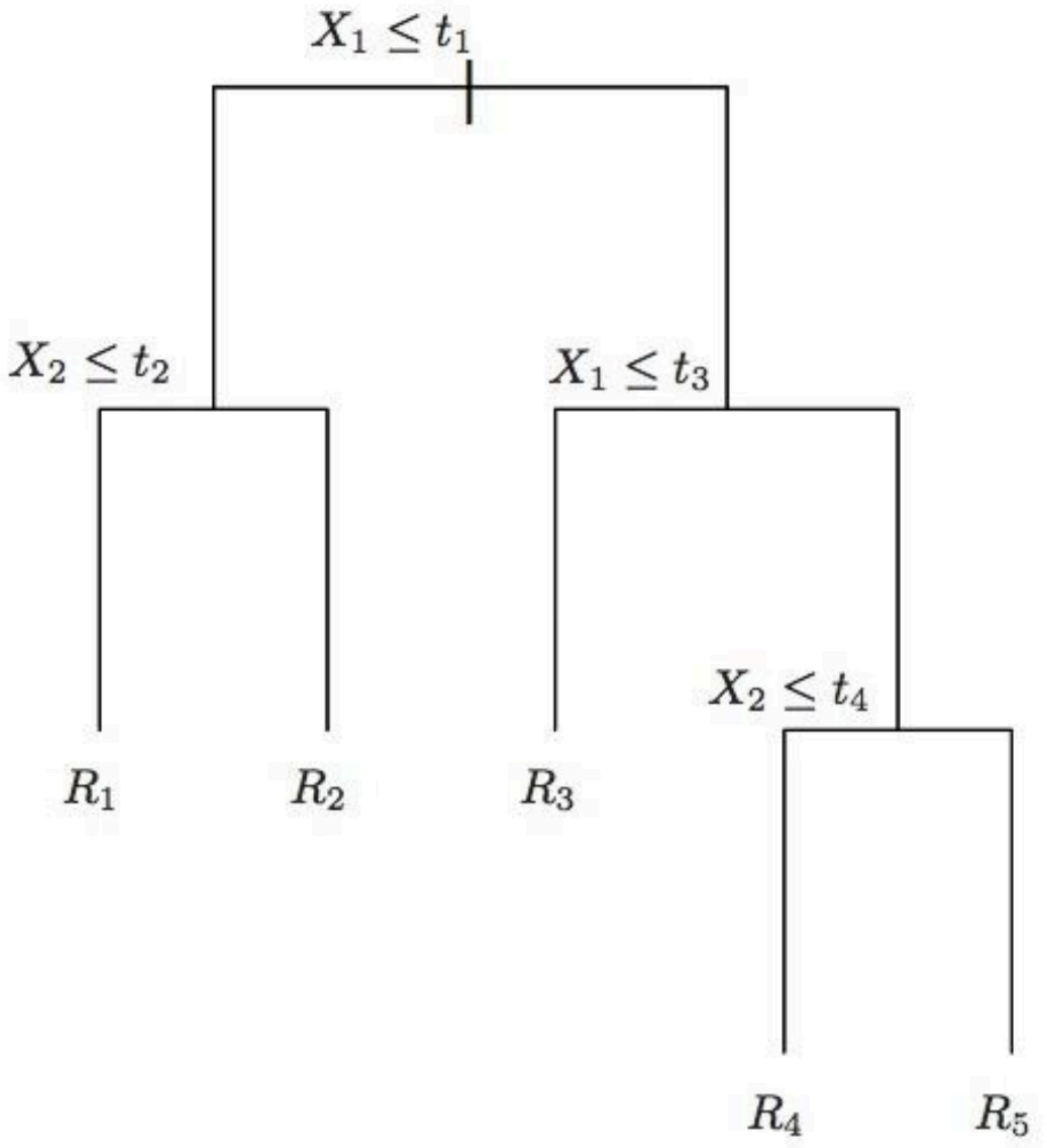
Regression:

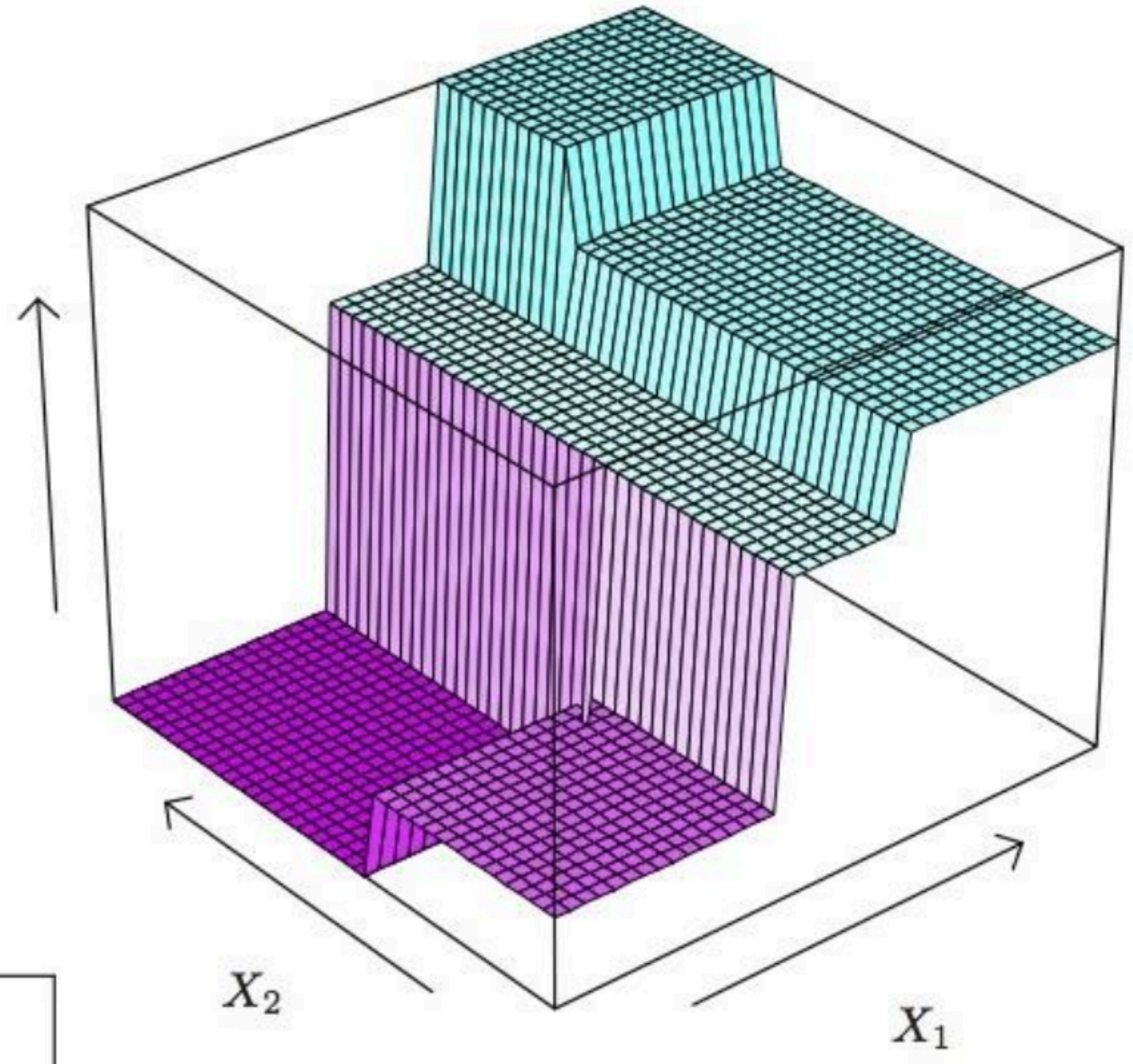
$$\mathrm{cost}(\mathcal{D}_i) = \frac{1}{|\mathcal{D}|} \sum_{n \in \mathcal{D}_i} (y_n - \overline{y})^2 \qquad \textbf{(MSE)}$$

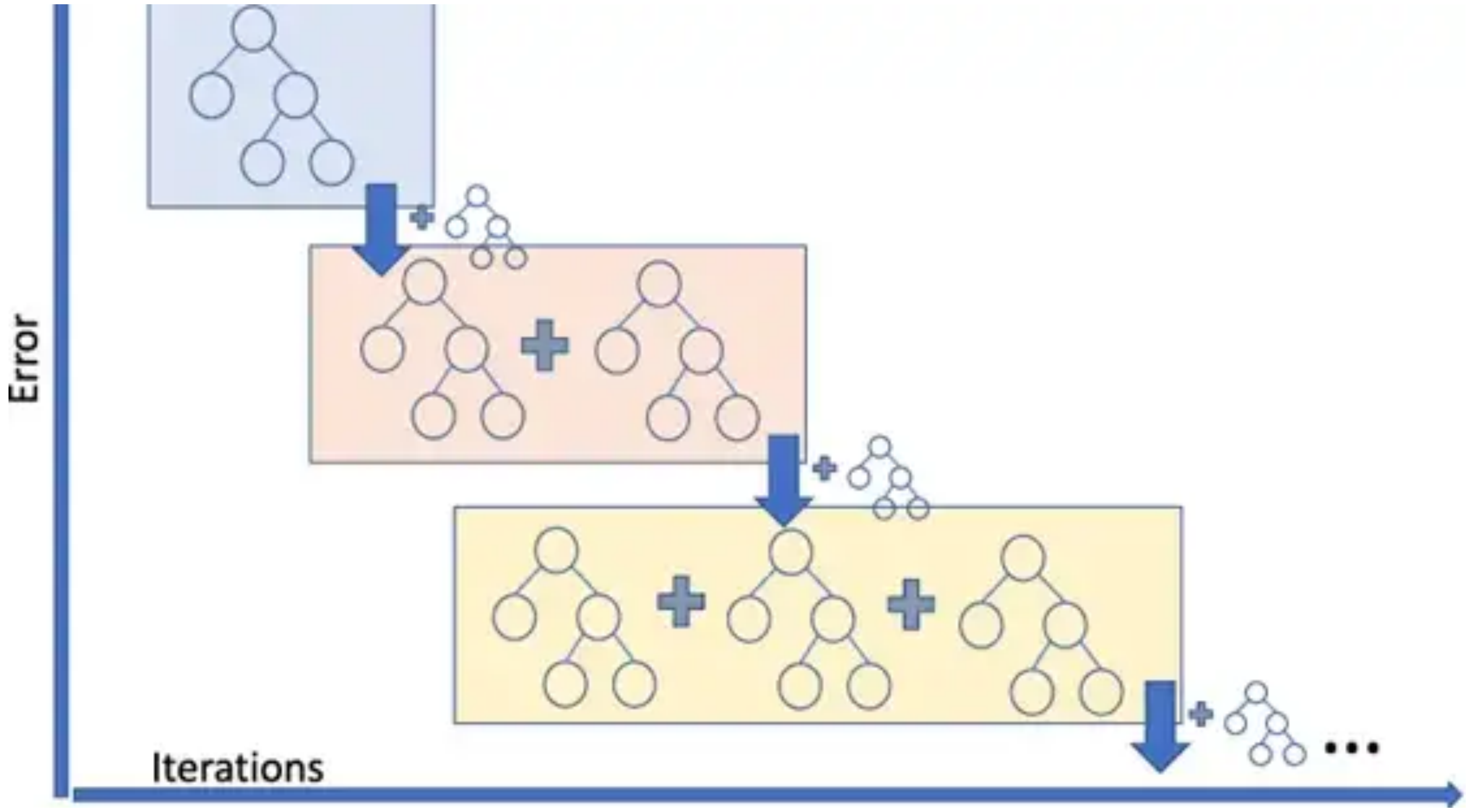
Classification:

$$H_i = \mathbb{H}(\hat{\boldsymbol{\pi}}_i) = -\sum_{c=1}^{C} \hat{\pi}_{ic} \log \hat{\pi}_{ic} \qquad \textbf{(entropy)}$$

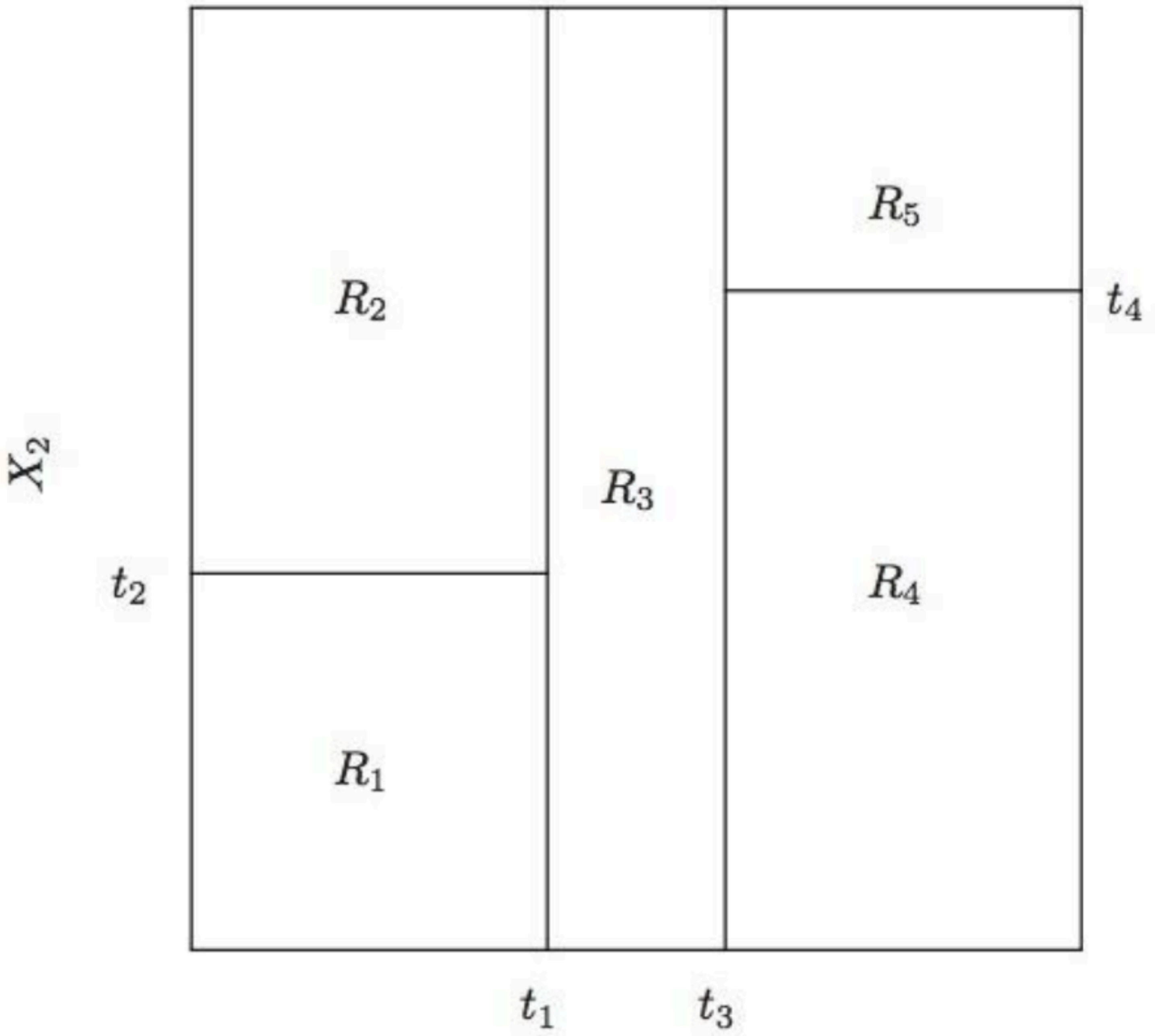
Gradient boosting







Gradient boosting machine

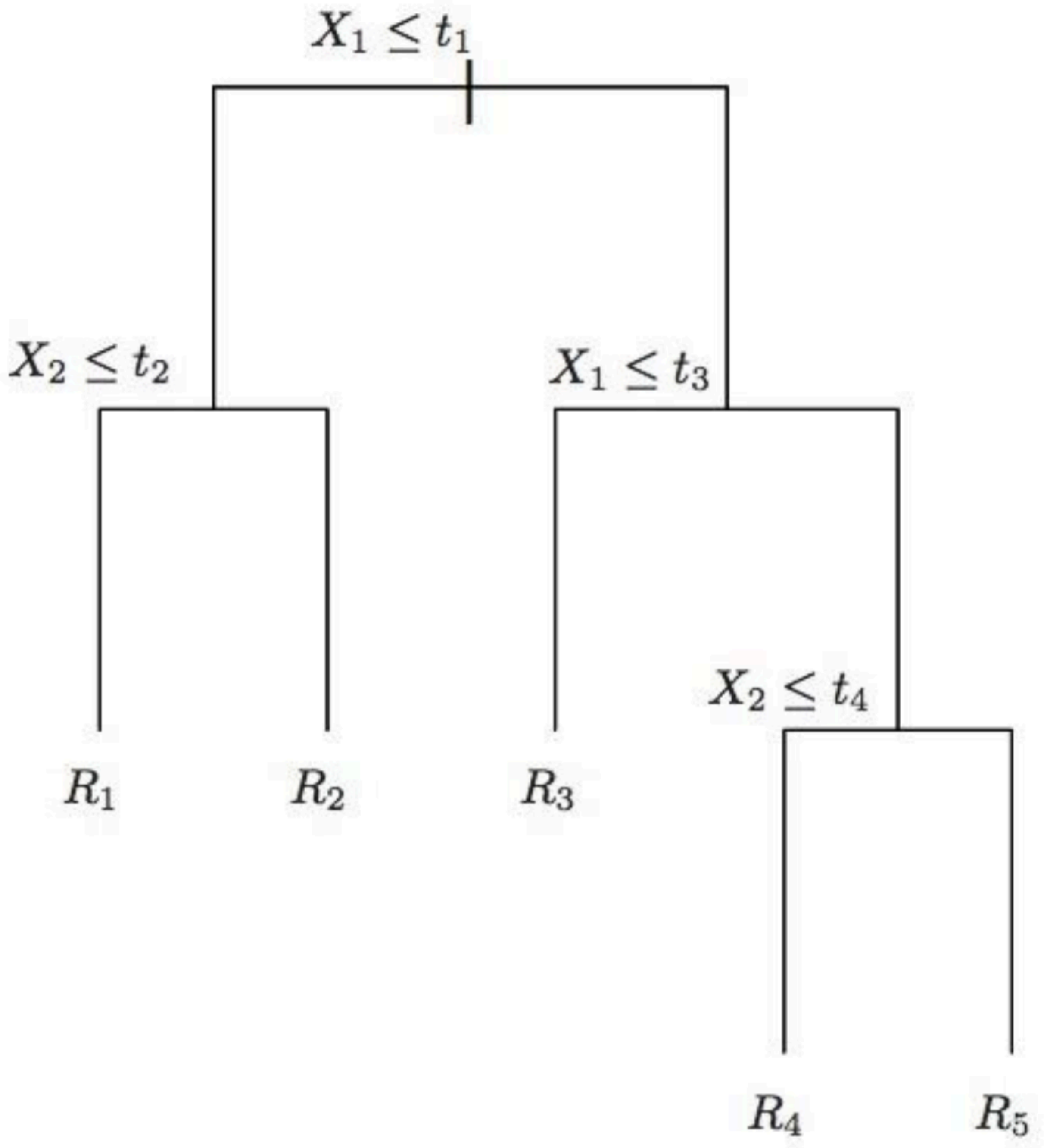


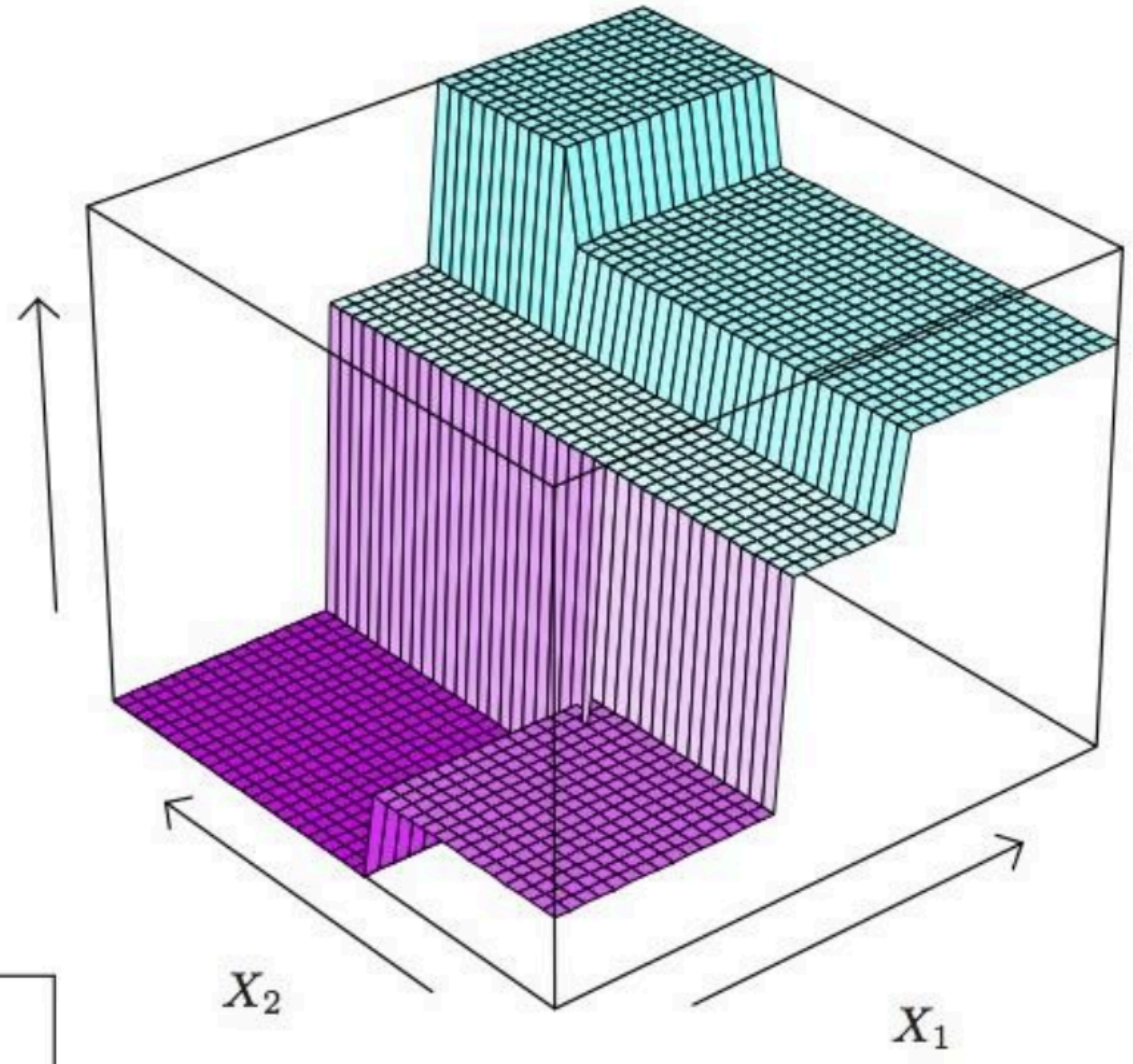
Industrial gradient boosting libraries:
XGBoost**, <u>LighGBM</u>,** CatBoost

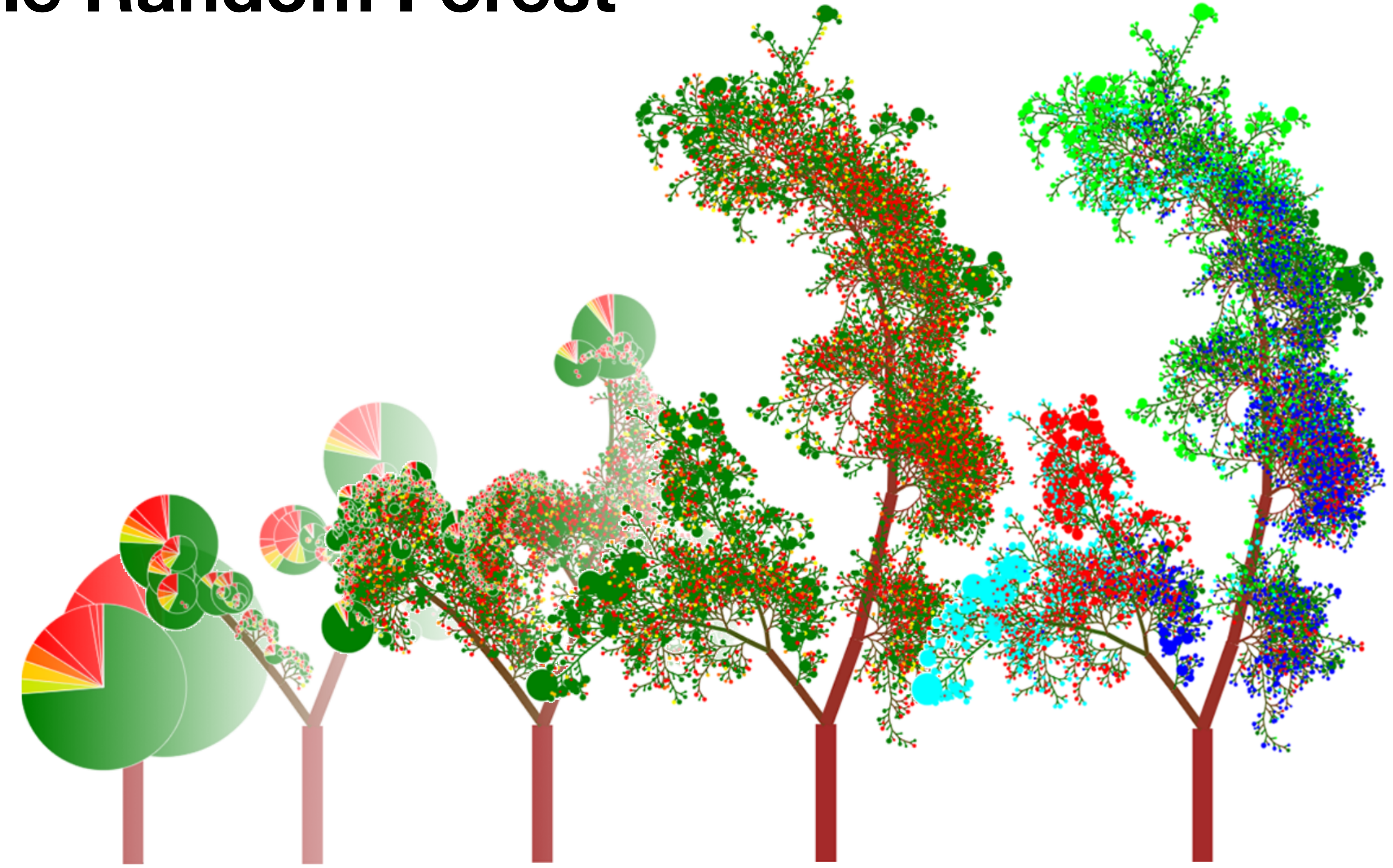
GB needs hyperparameters fine tuning

Randomized trees ensemble:
- «bagging» on spectral dataset objects
- «random subspaces» of photometric parameters

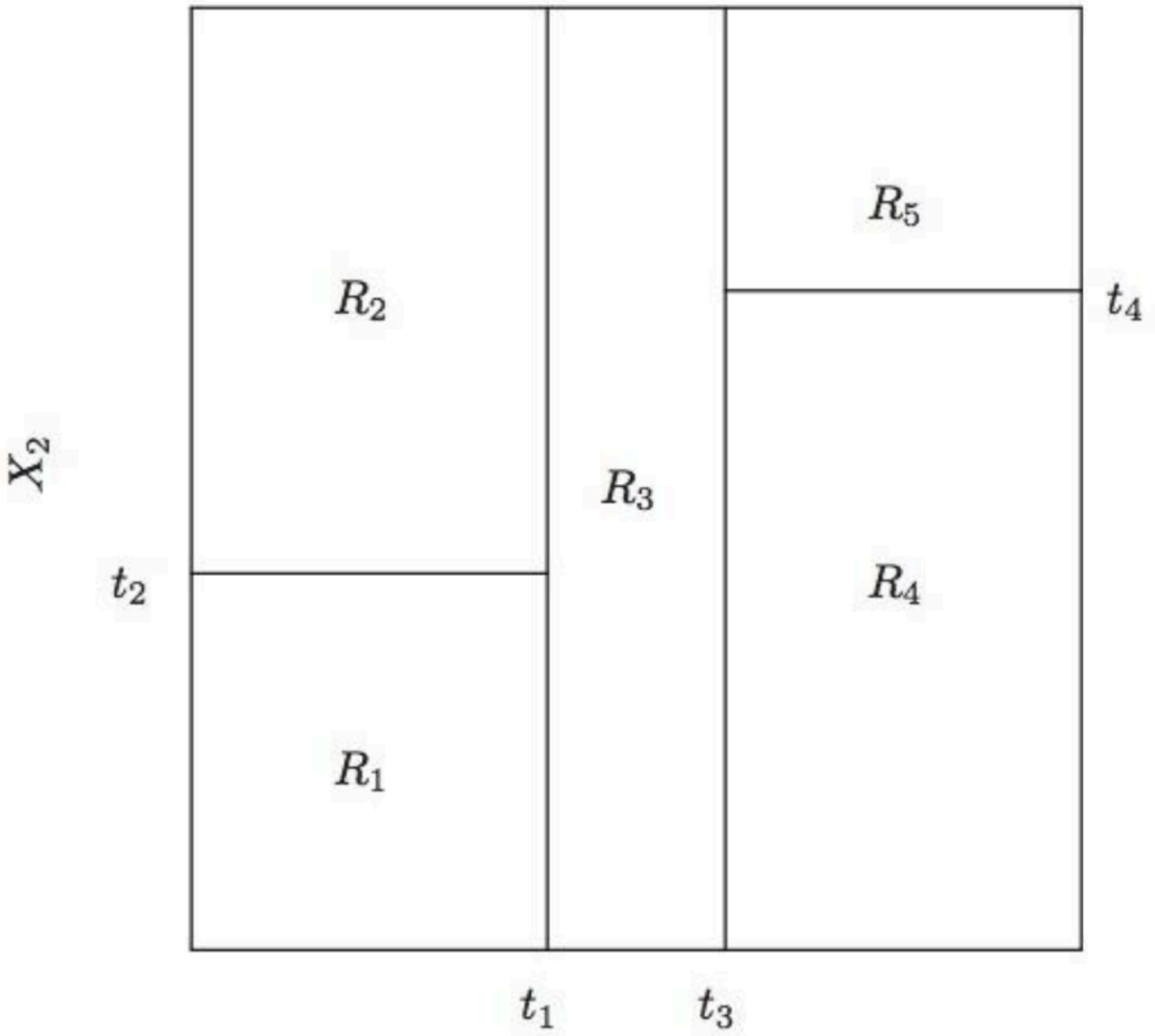
Quantile Random Forest







Credit: RFViz



Decision Tree with unlimited depth = <u>adopted nearest spectral neighbour of photometric object</u>

Random Forest = <u>P(z) empirical estimate</u>

RF (Breiman, 2001)

qRF (Meinshausen, 2003)

RF as adaptive kNN (Lin & Jeon, 2006)

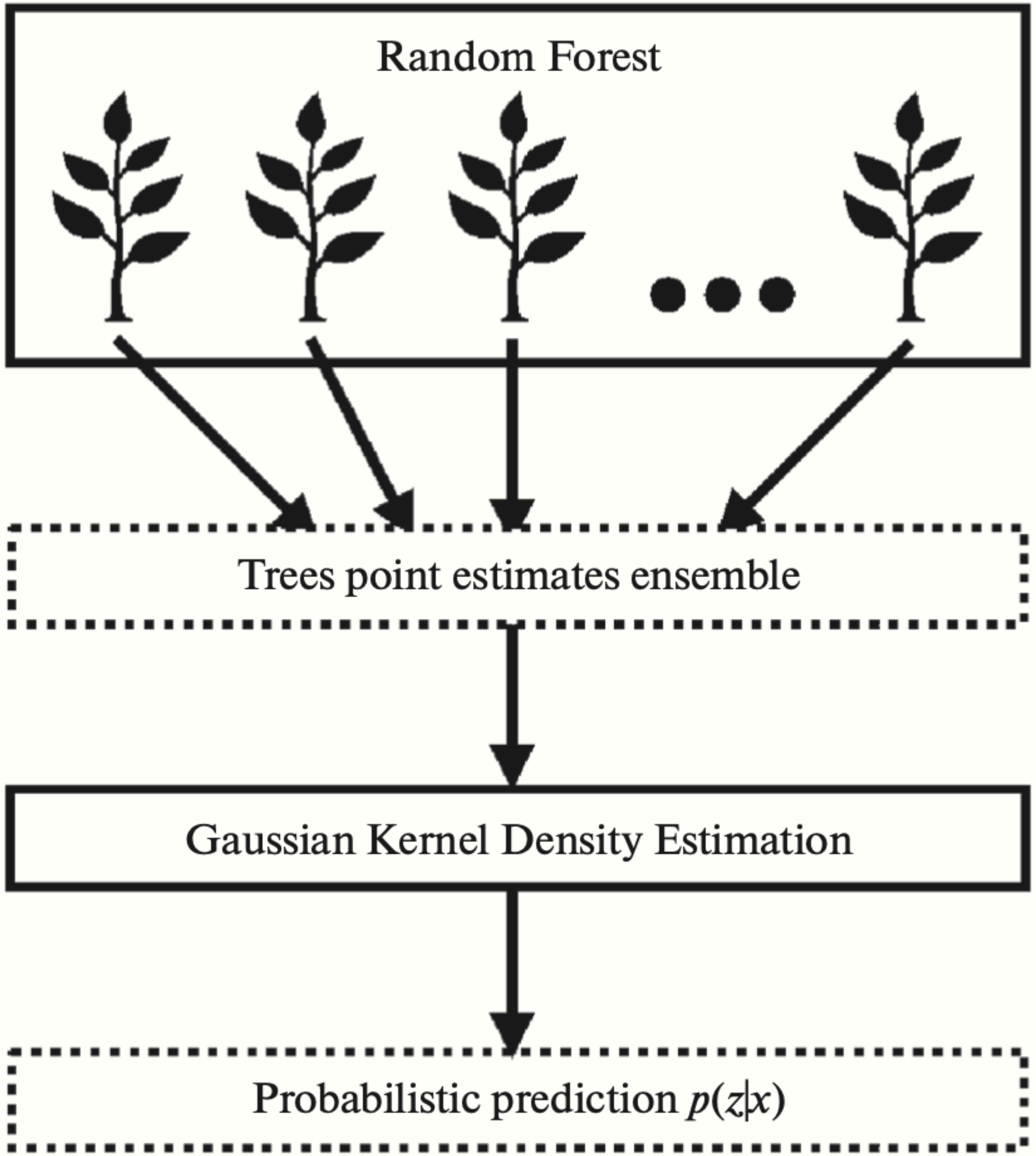
<u>*Astrophysics:*</u>

photo-z *(TPZ, Carrasco Kind, 2013)*

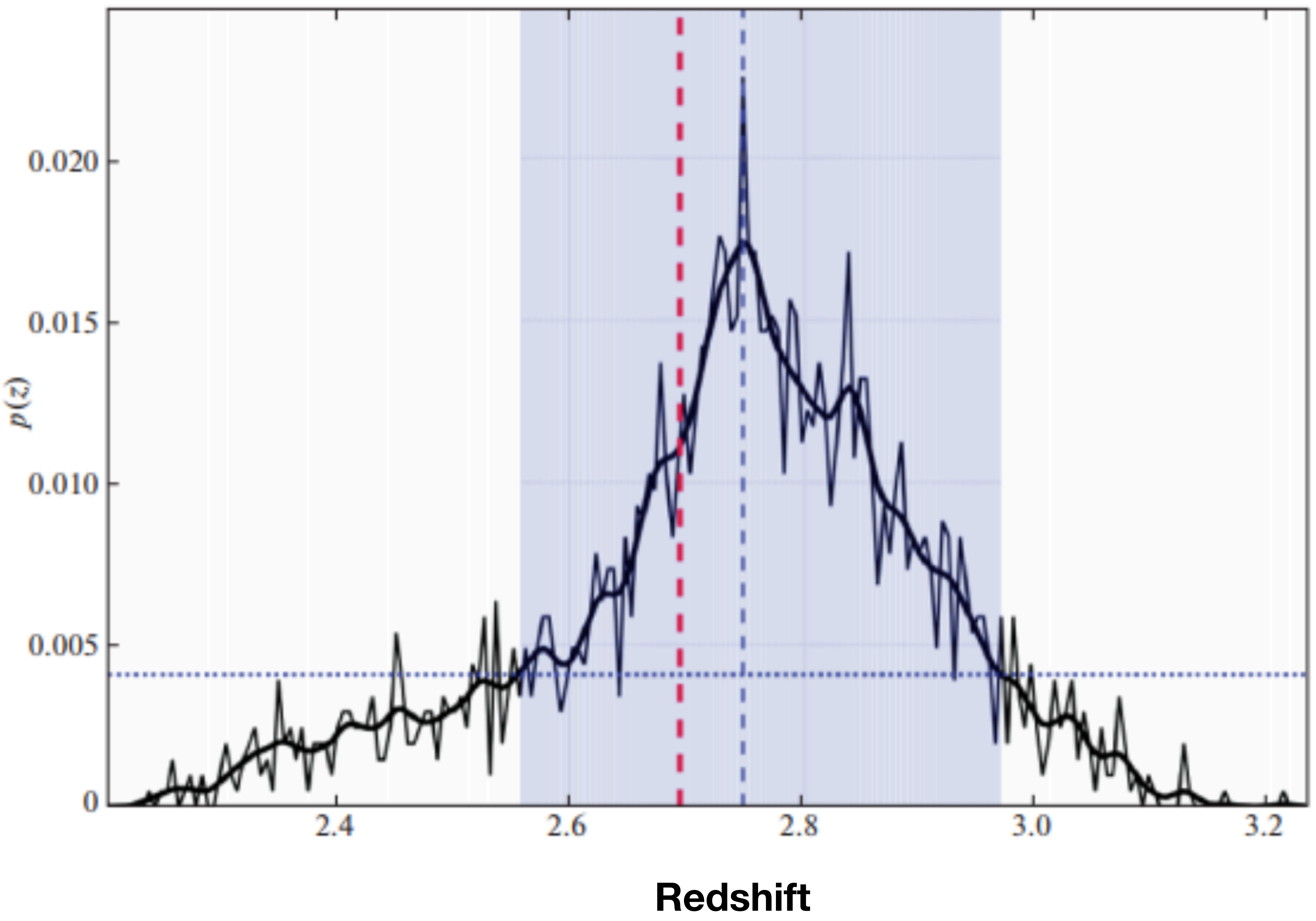
X-ray photo-z *(Meshcheryakov, 2018)*

SRGz *(Meshcheryakov, 2021-2023)*

Calibrated P(z) qRF *(Meshcheryakov, 2023)*

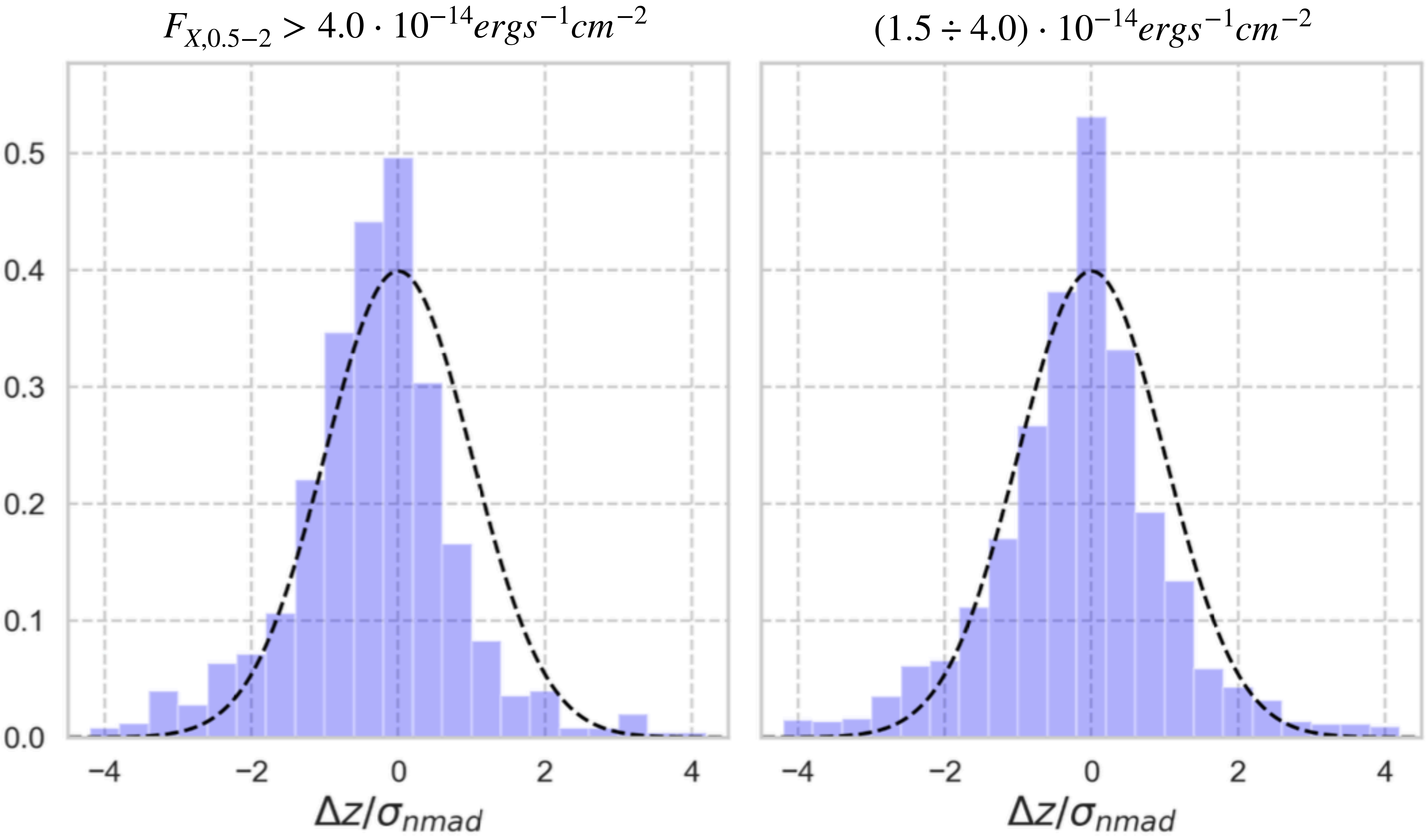


Random Forest
Trees point estimates ensemble
Gaussian Kernel Density Estimation
Probabilistic prediction $p(z|x)$

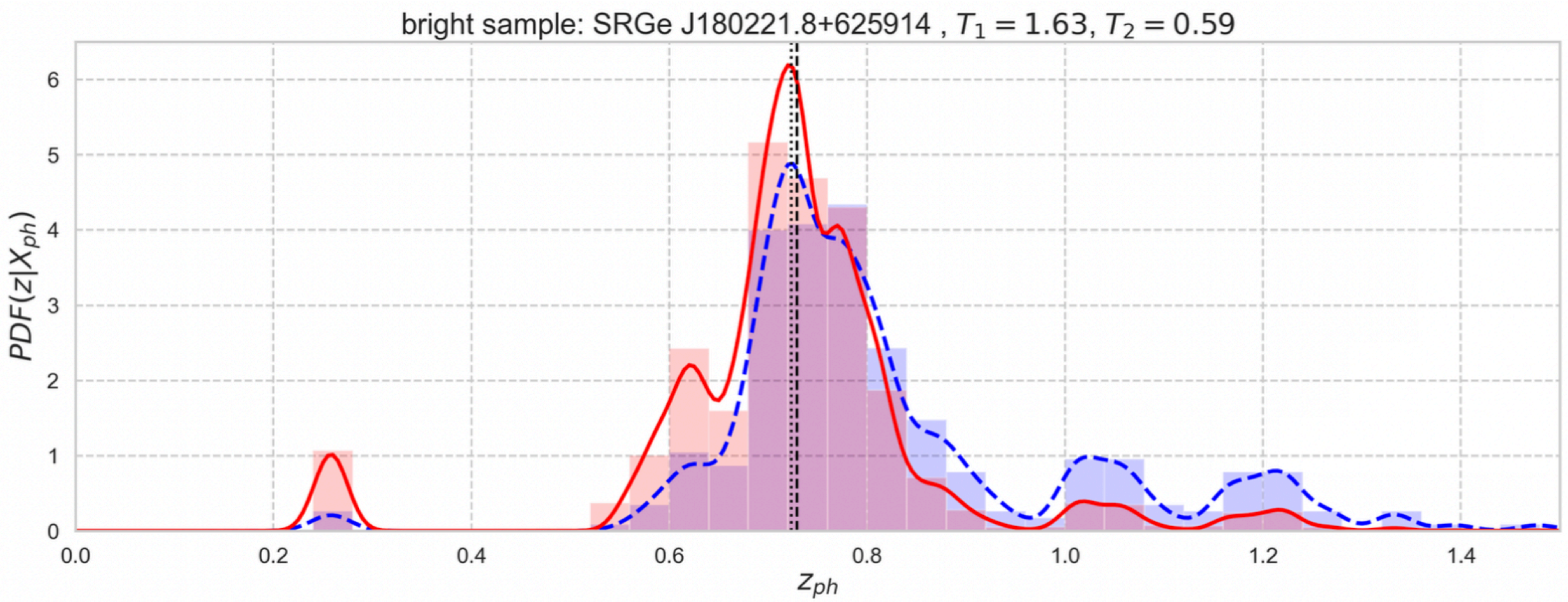


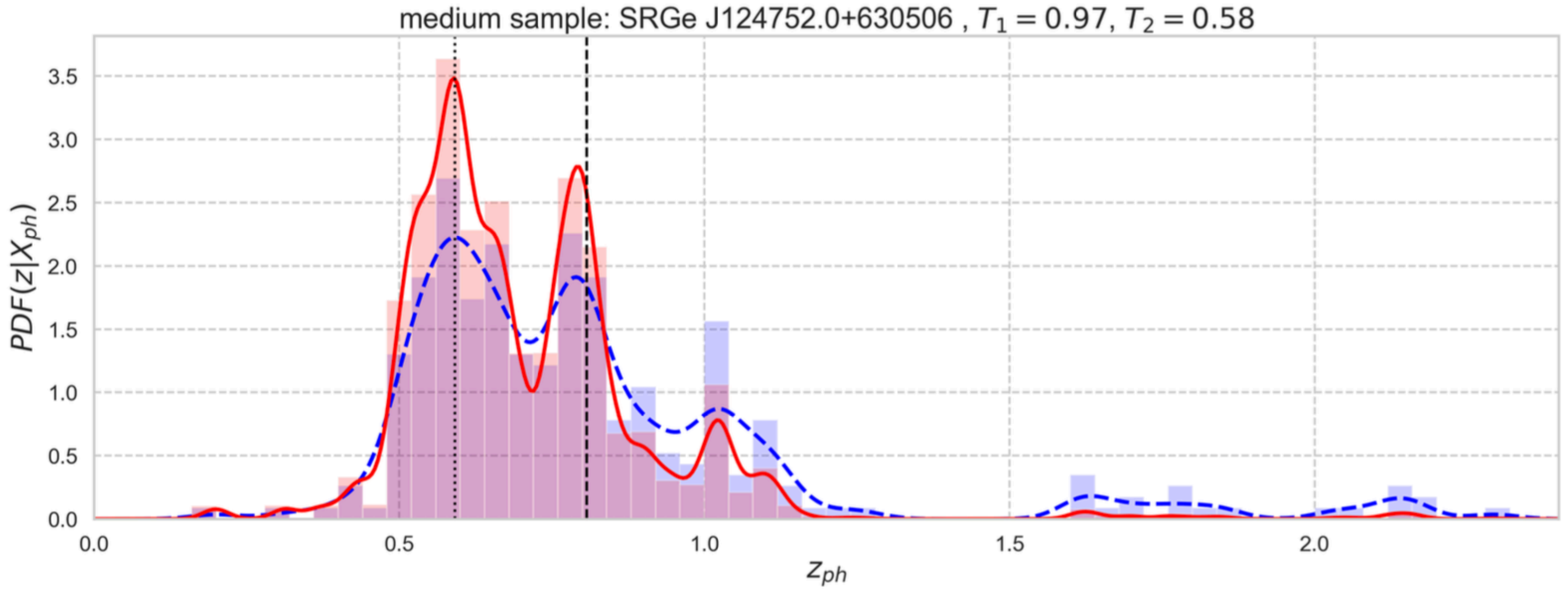
0.020
0.015
0.010
0.005
0
p(z)
2.4
2.6
2.8
3.0
3.2
Redshift

Non-gaussianity in P(z)



P(z) qRF calibration





Empirical & KDE-prediction of P(z) qRF:

$$\hat{P}_S(z) = \frac{\sum_{s=1}^{n} w_s \delta(z - z_s)}{\sum_{s=1}^{n} w_s}$$

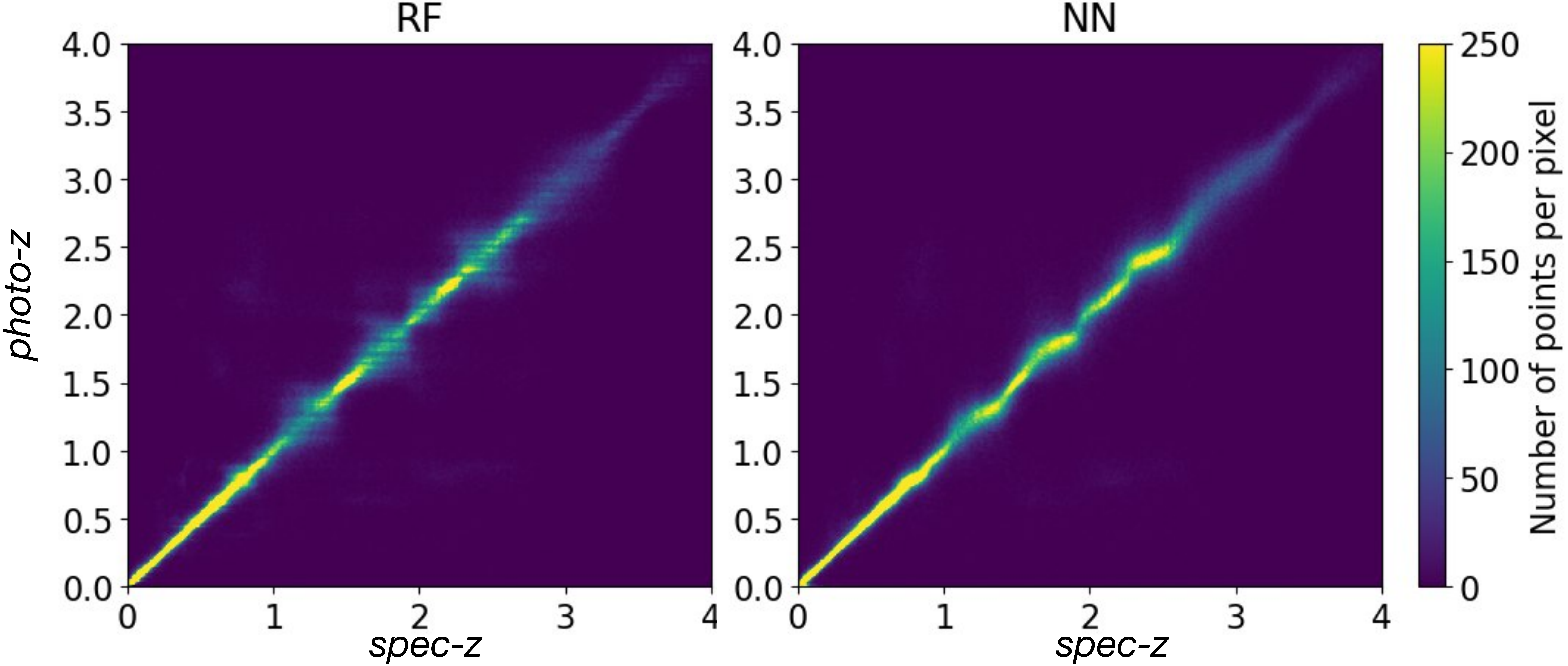
$$\hat{P}_{KDE}(z) = \frac{\sum_{s=1}^{n} w_s \mathcal{N}\left(\frac{z - z_s}{h}\right)}{h \sum_{s=1}^{n} w_s}$$

2-p temperature scaling method (Meshcheryakov et al. 2023b):

$$W(z, T) = \frac{\exp(\ln(\hat{P}_{KDE}(z; w_s = 1))/T)}{\hat{P}_{KDE}(z; w_s = 1)}$$

$$w_s = \begin{cases} \frac{W(z_s, T_1)}{W(z_{ph}, T_1)} & \text{if } z_s < z_{ph}; \\ \frac{W(z_s, T_2)}{W(z_{ph}, T_2)} & \text{if } z_s \geqslant z_{ph}. \end{cases}$$

X-ray AGN photo-z: **qRF** or NN?

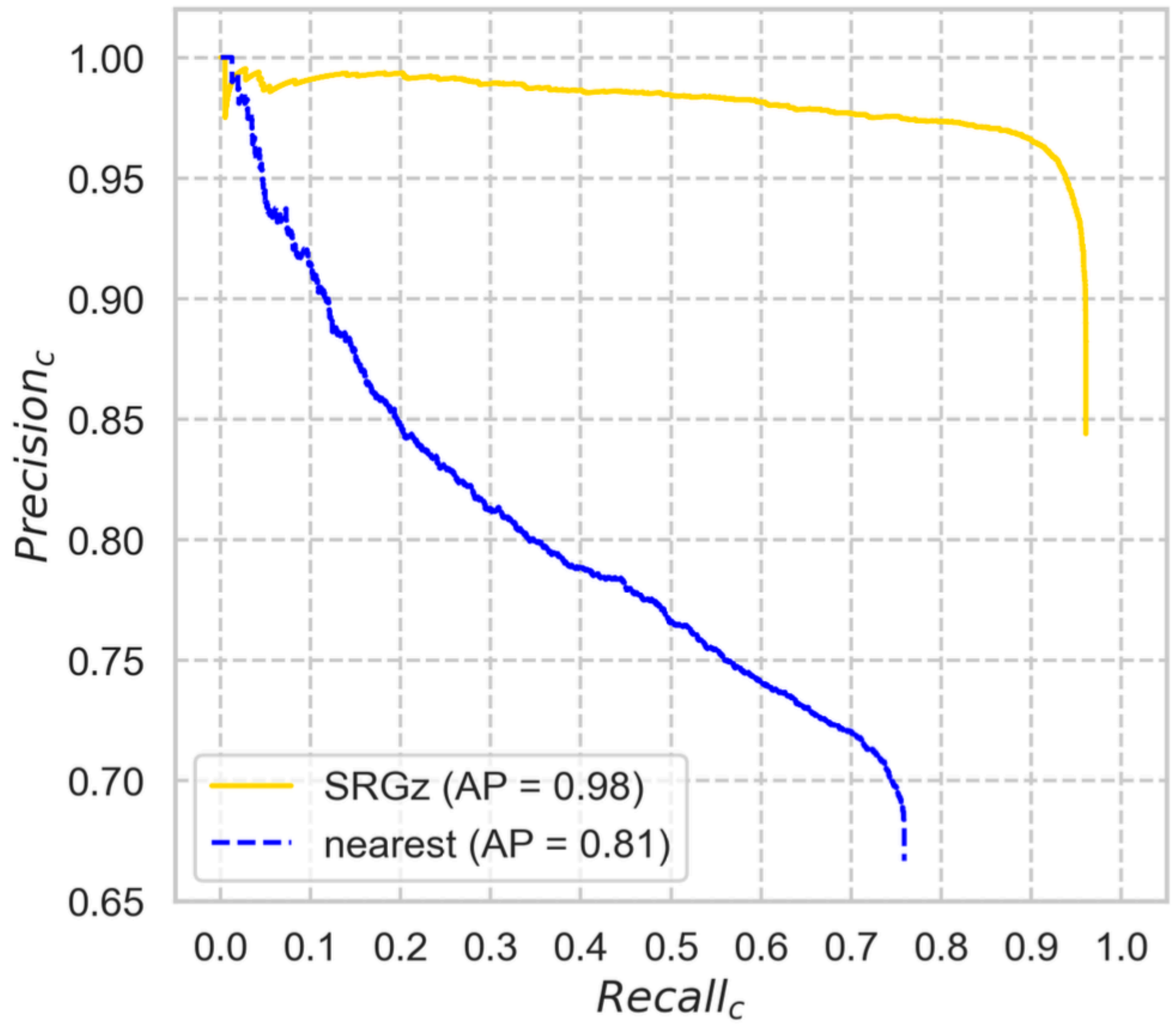


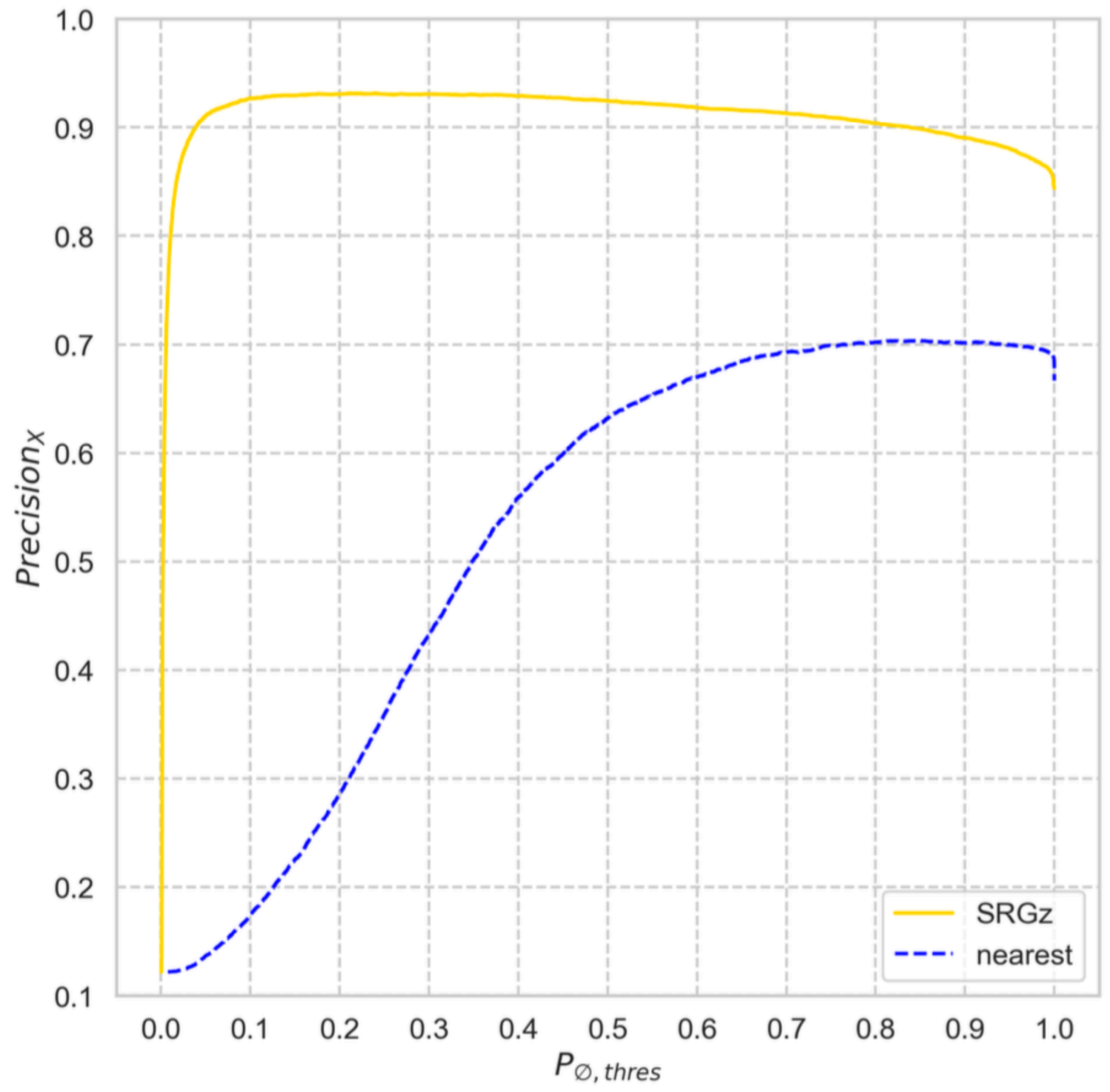
- **+ Precision on table data**
- **+ Work with correlated, categorical features and missed values**
- **+ Interpretability**

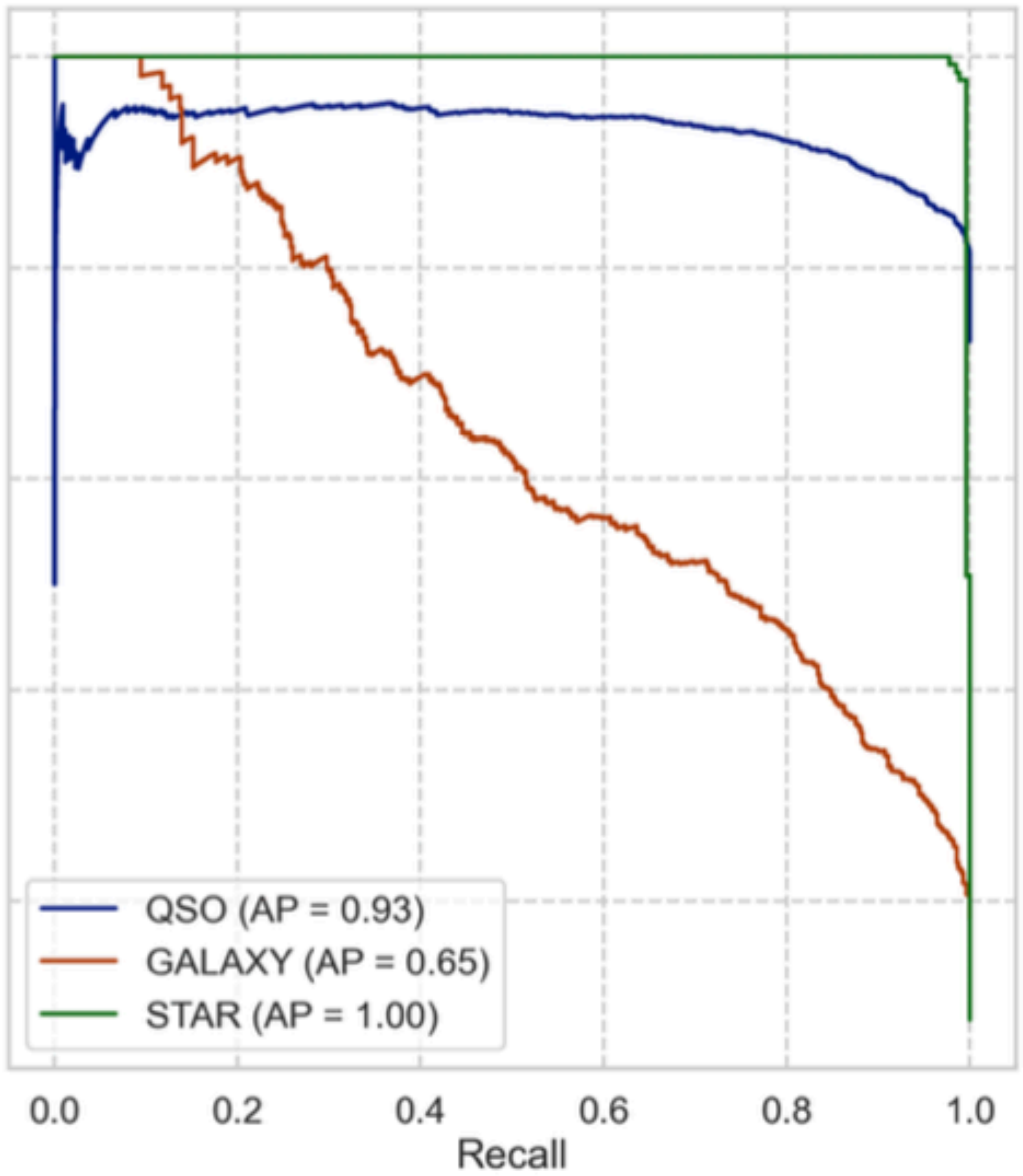
- **+ NN feature engineering**
- **+ Calibration of P(z)**
- **+ Transfer learning**
- **+ Multimodal encoders (table, image, time series)**

Results

Precision & completeness of X-ray sources optical match







Оптические компаньоны рентгеновских источников:

$$Recall_c^n = \frac{\hat{N}_c^{\star}(P_{\emptyset} \leqslant t_n)}{N_c},$$

$$Precision_c^n = \frac{\hat{N}_c^{\star}(P_{\emptyset} \leqslant t_n)}{\hat{N}_c(P_{\emptyset} \leqslant t_n)}$$

$$F_{X,0.5-2} > 1.5 \cdot 10^{-14} erg s^{-1} cm^{-2}$$

- SDSS field - 4179 sq.deg

Optical match precision of SRG/eROSITA X-ray sources

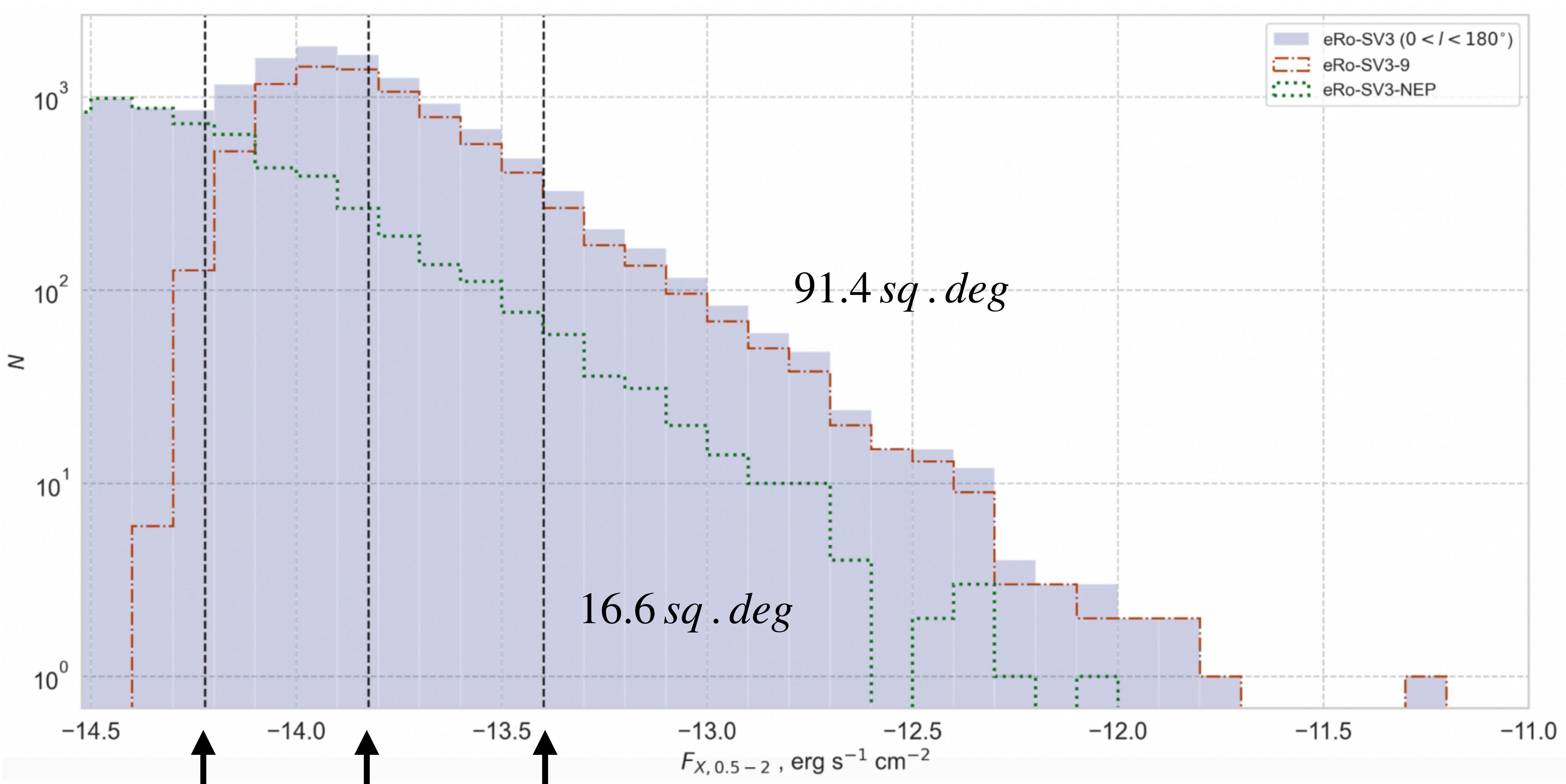
$$Precision_X^n(t) = \frac{\hat{N}_c^{\star}(P_{\emptyset} \leqslant t_n) + \hat{N}_h^{\star}(P_{\emptyset} > t_n)}{N_c + N_h}$$

$P_{\varnothing}$ - probability of hostless X-ray source

For standard SRGz - the completeness of **optical counterpart identification** in the DESI LIS survey area is **95%** (with an optical counterpart selection accuracy of 94%). SRGz achieves high quality of photometric classification of optical counterparts of X-ray sources: > **99% completeness of photometric classification of extragalactic objects** and stars on a test sample of sources with SDSS spectra and GAIA astrometric stars.

eRo-1%DESI-East

	N	N/N_X
bright: $F_{X,0.5-2} \geqslant 4 \cdot 10^{-14}$ эрг с$^{-1}$см$^{-2}$, $Area = 91.40$ кв.градусов		
eRosita point X-ray sources (N_X)	1080	100.00%
SRGz optical counterparts	1051	97.31%
astrometric stars GAIA DR2	129	11.94%
SDSS DR18	537	49.72%
SDSS DR16q	419	39.87%
SRGz photo-z training sample	207	19.70%
HELP	228	21.11%
DESI EDR	859	79.54%
DESI EDR/astrometric counterparts	984	91.11%
all spectroscopic/astrometric counterparts	1019	94.35%
medium: $1.5 \cdot 10^{-14} \leqslant F_{X,0.5-2} < 4 \cdot 10^{-14}$ эрг с$^{-1}$см$^{-2}$, $Area = 91.40$ кв.градусов		
eRosita point X-ray sources (N_X)	3701	100.00%
SRGz optical counterparts	3678	99.38%
astrometric stars GAIA DR2	261	7.05%
SDSS DR18	1431	38.67%
SDSS DR16q	1287	34.99%
SRGz photo-z training sample	529	14.38%
HELP	646	17.45%
DESI EDR	2832	76.52%
DESI EDR/astrometric counterparts	3067	82.87%
all spectroscopic/astrometric counterparts	3322	89.76%
faint: $6 \cdot 10^{-15} \leqslant F_{X,0.5-2} < 1.5 \cdot 10^{-14}$ эрг с$^{-1}$см$^{-2}$, $Area = 16.62$ кв.градусов		
eRosita point X-ray sources (N_X)	1824	100.00%
SRGz optical counterparts	1809	99.18%
astrometric stars GAIA DR2	185	10.14%
SDSS DR18	0	0.00%
SDSS DR16q	0	0.00%
SRGz photo-z training sample	0	0.00%
HELP	75	4.11%
DESI EDR	1169	64.09%
DESI EDR/astrometric counterparts	1316	72.15%
all spectroscopic/astrometric counterparts	1415	77.58%



bright : $F_{X,0.5-2} > 4 \cdot 10^{-14}\, erg\, cm^{-2}\, s^{-1}$

medium : $1.5 \cdot 10^{-14} < F_{X,0.5-2} \leq 4 \cdot 10^{-14}\, erg\, cm^{-2}\, s^{-1}$

faint : $6 \cdot 10^{-15} < F_{X,0.5-2} \leq 1.5 \cdot 10^{-14}\, erg\, cm^{-2}\, s^{-1}$

All spec/astrometric counterparts:

bright	94.4%
medium	89.8%
faint	77.6%

Confusion matrixes

Spec-class \ Photo-class	QSO	GALAXY	STAR
QSO	691	8	1
GALAXY	104	48	1
STAR	2	0	129

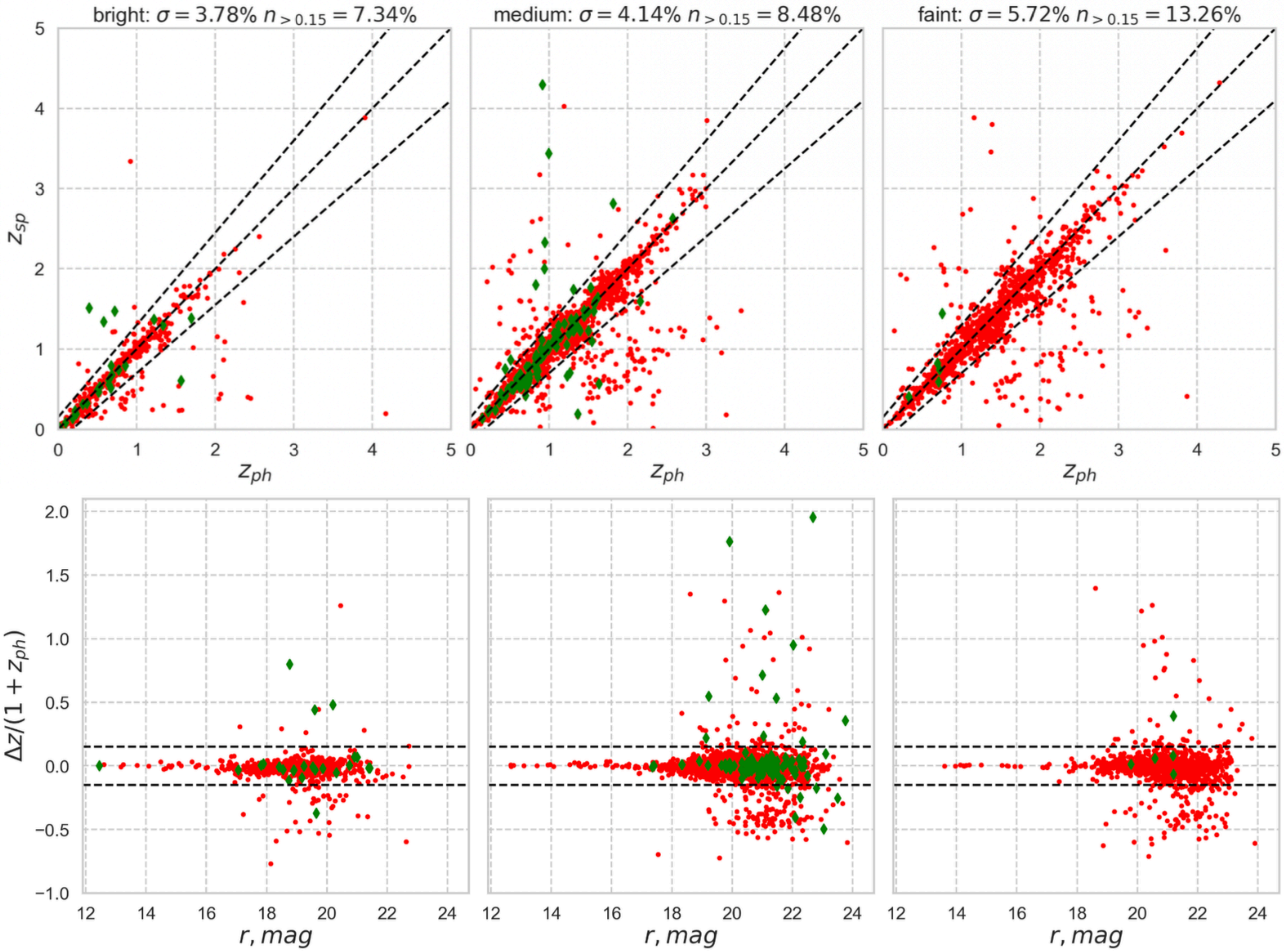
Spec-class \ Photo-class	QSO	GALAXY	STAR
QSO	2201	38	0
GALAXY	394	165	2
STAR	3	1	263

Spec-class \ Photo-class	QSO	GALAXY	STAR
QSO	854	19	2
GALAXY	148	103	1
STAR	1	0	188

bright : $F_{X,0.5-2} > 4 \cdot 10^{-14}\, erg\, cm^{-2}\, s^{-1}$

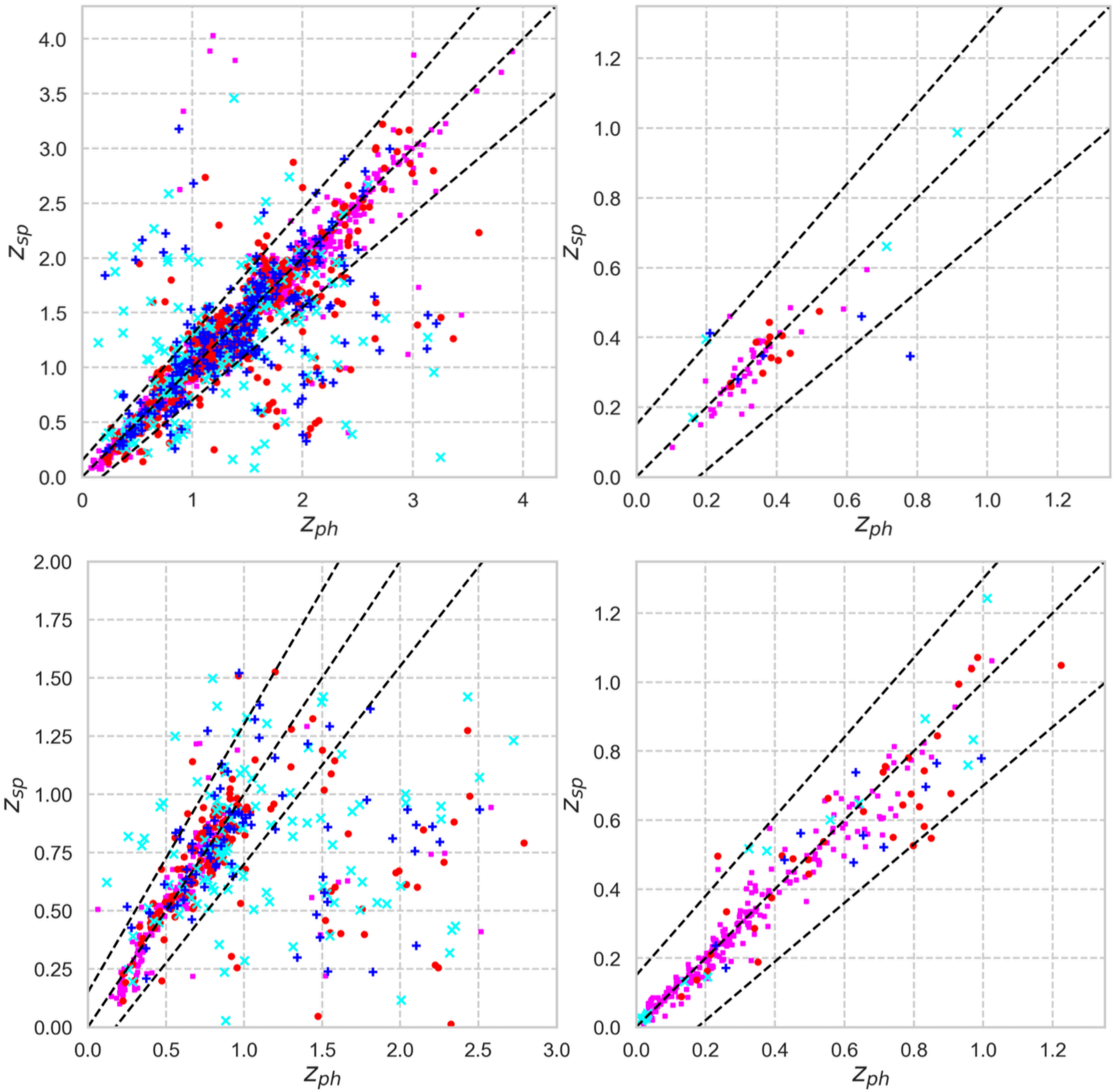
medium : $1.5 \cdot 10^{-14} < F_{X,0.5-2} \leq 4 \cdot 10^{-14}\, erg\, cm^{-2}\, s^{-1}$

faint : $6 \cdot 10^{-15} < F_{X,0.5-2} \leq 1.5 \cdot 10^{-14}\, erg\, cm^{-2}\, s^{-1}$


bright: $\sigma = 3.78\%$ $n_{>0.15} = 7.34\%$
medium: $\sigma = 4.14\%$ $n_{>0.15} = 8.48\%$
faint: $\sigma = 5.72\%$ $n_{>0.15} = 13.26\%$
z_{sp}
z_{ph}
$\Delta z/(1 + z_{ph})$
r, mag

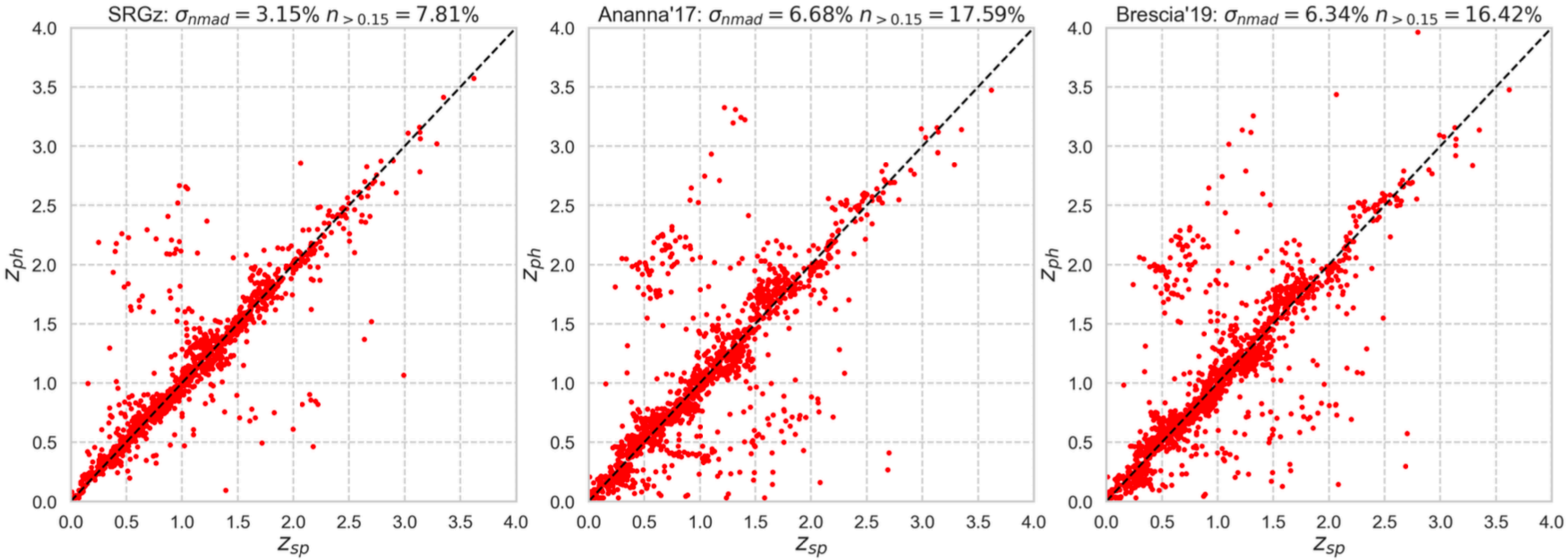
spec:QSO
phot: QSO

spec:QSO
phot: GALAXY



spec:GALAXY
phot: QSO

spec:GALAXY
phot: GALAXY



Metrics:

Normalized median absolute deviation

$$\sigma_{NMAD} = 1.48 \times median(|\delta z_{\text{norm,i}}|)$$

Catastrophic outliers fraction

$$n_{>0.15} = \frac{\#\{i = \overline{1, N} | \delta z_{\text{norm, i}} > 0.15\}}{N}$$

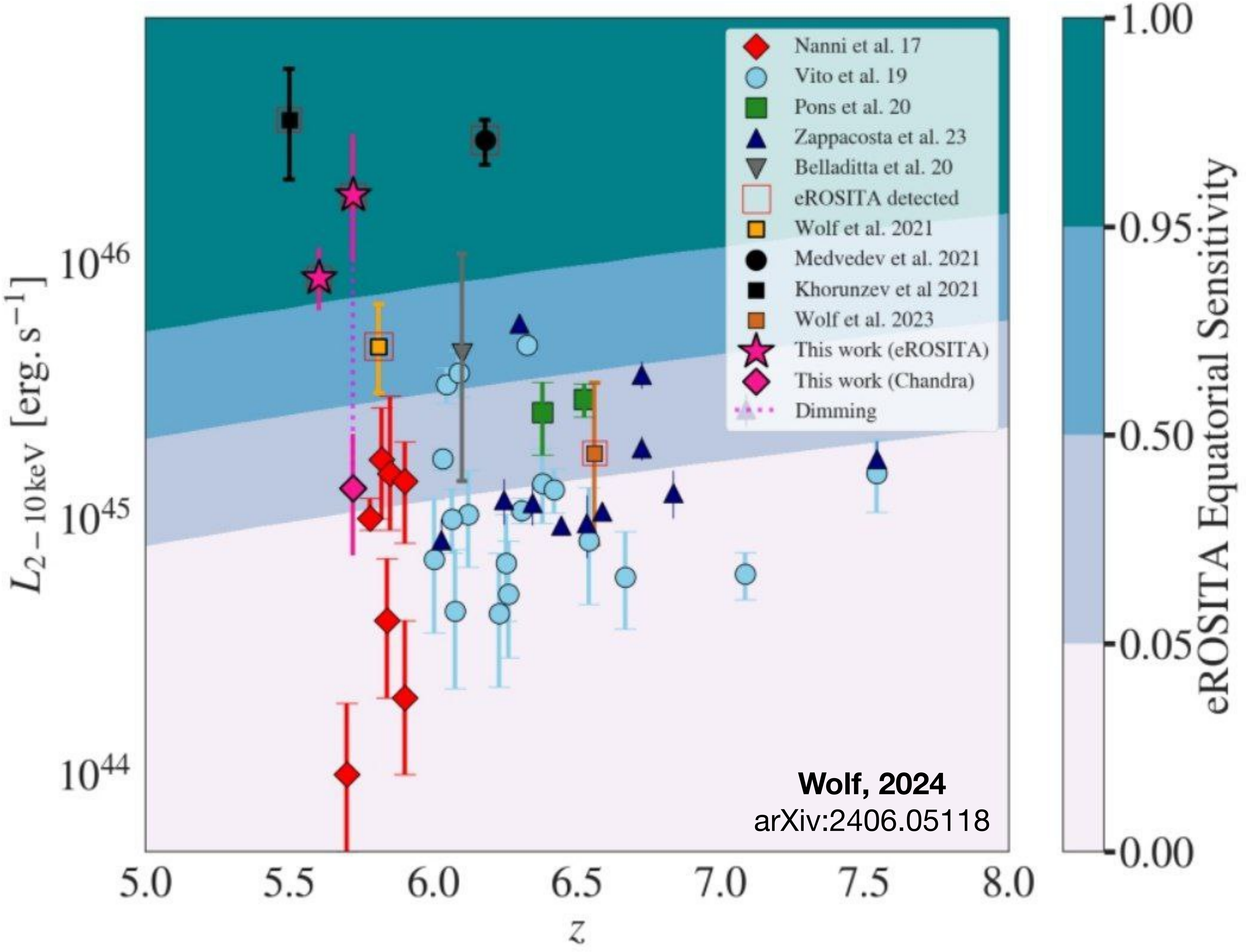
Ananna'17 - photo-z model based on SED templates

Brescia'19 - photo-z model based on MLP

The presented photo-z results for eROSITA X-ray sources in the Stripe82X field are more than a factor of 2 better in both metrics (σ_{NMAD} и $n_{>0.15}$) , compared to the photo-z results of other groups published in the Stripe82X field catalog.

■ Stripe82X - 31.3 кв.градусов

SRGz for distant X-ray quasars



Most luminous X-ray quasar at z=5.5 !!!

Khorunzhev, Meshcheryakov et al. **"Discovery of the Most X-ray Luminous Quasar SRGE J170245.3+130104 at Redshift z≈5.5"**, doi:10.1134/S1063773721030026

SRGz system (2018-2024):

[1] Meshcheryakov, A. V., Khorunzhev, G. A., Voskresenskaya, S. A., Medvedev, P. S., Gilfanov, M. R., and Sunyaev, R. A., “**SRGz: Classification of eROSITA Point X-ray Sources in the 1%DESI Region and Calibration of Photometric Redshifts**”, Astronomy Letters, vol. 49, no. 11, Springer, pp. 646–661, 2023. doi:10.1134/S1063773723110129.

[2] Meshcheryakov, A. V., “**SRGz: Machine Learning Methods and Properties of the Catalog of SRG/eROSITA Point X-ray Source Optical Counterparts in the DESI Legacy Imaging Surveys Footprint**”, Astronomy Letters, vol. 49, no. 7, Springer, pp. 359–409, 2023. doi:10.1134/S1063773723070022.

[3] Borisov, V., Meshcheryakov, A., and Gerasimov, S., “**Probabilistic Photo-z Machine Learning Models for X-ray Sky Surveys**”, in Astronomical Data Analysis Software and Systems XXX, 2022, vol. 532, p. 231. doi:10.48550/arXiv.2107.01891.

[4] Khorunzhev, G. A., “**Discovery of the Most X-ray Luminous Quasar SRGE J170245.3+130104 at Redshift z≈5.5**”, Astronomy Letters, vol. 47, no. 3, Springer, pp. 123–140, 2021. doi:10.1134/S1063773721030026.

[5] Meshcheryakov, A. V., Glazkova, V. V., Gerasimov, S. V., and Mashechkin, I. V., “**Measuring the Probabilistic Photometric Redshifts of X-ray Quasars Based on the Quantile Regression of Ensembles of Decision Trees**”, Astronomy Letters, vol. 44, no. 12, Springer, pp. 735–753, 2018. doi:10.1134/S1063773718120058.

Thank you!