

SRGZ: КЛАССИФИКАЦИЯ ТОЧЕЧНЫХ РЕНТГЕНОВСКИХ ИСТОЧНИКОВ ЕРОЗИТА В ОБЛАСТИ 1%DESI И КАЛИБРОВКА ФОТОМЕТРИЧЕСКИХ КРАСНЫХ СМЕЩЕНИЙ

© 2024 г. А. В. Мещеряков^{1, 2 *}, Г. А. Хорунжев¹, С. А. Воскресенская^{1, 3}, П. С. Медведев¹,
М. Р. Гильфанов^{1, 4}, Р. А. Сюняев^{1, 4}

¹Институт космических исследований РАН, Москва, Россия

²Московский государственный университет им. М. В. Ломоносова, Москва, Россия

³НИУ Высшая школа экономики, Москва, Россия

⁴Институт им. Макса Планка, Гаршинг, Германи

Поступила в редакцию 06.12.2023 г.

После доработки 26.12.2023 г.; принята к публикации 26.12.2023 г.

Рассматривается популяция точечных рентгеновских источников двухлетнего обзора СРГ/еРОЗИТА в области 1%-го спектроскопического обзора DESI в восточной галактической полусфере (eROSITA-1%DESI-East). Рассматриваемые данные сочетают в себе большую площадь обзора (91.4 кв. градусов) и рекордно высокую полноту (90–95%) спектроскопии оптических компаньонов рентгеновских источников с потоком $F_{X,0.5-2} \geq 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$. Мы сравниваем результаты фотометрических (SRGz) и спектральных/астрометрических измерений (DESI EDR, SDSS, HELP, GAIA) классов и красных смещений объектов в зависимости от их рентгеновского потока. Нами отмечается высокая точность фотометрических красных смещений, полученных моделями SRGz для рентгеновских источников в двухлетнем обзоре неба еРОЗИТА (в области покрытия фотометрических обзоров DESI Legacy Imaging Surveys/Pan-STARRS1/SDSS): стандартное отклонение $\sigma_{NMAD} \approx 4\%$ и доля выбросов на уровне $n_{>0.15} = 7-8.5\%$ (для оптических компаньонов рентгеновских источников с $F_{X,0.5-2} \geq 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$). Прогнозы photo-z рентгеновских источников в SRGz имеют негауссовый характер; качество калибровки $PDF(z)$ важно для точной оценки доверительных интервалов прогноза красных смещений. Предложен новый метод постобработки вероятностных прогнозов photo-z, основанный на двухтемпературной коррекции распределения $PDF(z)$. Подход позволяет значительно улучшить калибровку вероятностных прогнозов и доверительных интервалов фотометрических красных смещений рентгеновских источников еРОЗИТА.

Ключевые слова: машинное обучение, СРГ/еРОЗИТА, обзоры неба, рентгеновская астрономия, фотометрические красные смещения, галактики, квазары, звезды.

DOI: 10.31857/S0320010824010031, **EDN:** OQYMBI

1. ВВЕДЕНИЕ

12 декабря 2019 г. рентгеновские телескопы еРОЗИТА (Предель и др., 2021) и АРТ-ХС им. Павлинского (Павлинский и др., 2021) на борту космической обсерватории СРГ (Сюняев и др., 2021) начали рентгеновский обзор всего неба. К концу 2021 г. еРОЗИТА завершила двухлетние обзорные наблюдения, в которых зарегистрировала более 2 млн точечных рентгеновских источников (с лог-правдоподобием детектирования $L > 6$) в восточной галактической полусфере $0 < l < 180^\circ$ ¹ (Медведев и др., 2022).

* Электронный адрес: mesch@cosmos.ru

¹ За анализ данных еРОЗИТА в восточной половине неба (в галактических координатах) отвечает российский консорциум еРОЗИТА.

Анализ данных о рентгеновских источниках в еРОЗИТА в восточной галактической полусфере в значительной степени опирается на исследование всей доступной информации из крупнейших публичных фотометрических обзоров неба большой площади — DESI Legacy Imaging Surveys (Дей и др., 2019), Pan-STARRS1 (Чамберс и др., 2016), а также связанных с ними данных принудительной ИК-фотометрии WISE (Мейнзер и др., 2011; Райт и др., 2010) оптических источников DESI LIS и Pan-STARRS1 (Дей и др., 2019; Буренин, 2022). Для наиболее эффективного использования доступной фотометрической информации в окрестности рентгеновских источников еРОЗИТА, мы создали систему SRGz (Мещеряков и др., 2023), в которой используются методы машинного обучения для решения задач оптического отождествления, фотометрической классификации и прогнозов

photo-z рентгеновских источников еРОЗИТА. Модели SRGz основываются на древовидных ансамблевых алгоритмах случайного леса и градиентного бустинга и показали высокую точность измерений физических характеристик рентгеновских источников.

Ранний выпуск данных спектроскопического обзора DESI (DESI коллаборация и др., 2023), вышедший летом 2023 г., представляет собой уникальный наблюдательный материал для исследования популяций рентгеновских источников и их оптических компаньонов. В составе DESI EDR наибольший интерес представляют данные однопроцентного обзора DESI (далее 1%DESI, DESI EDR SV3), где достигается рекордно высокая полнота однородной спектроскопии слабых оптических объектов (в несколько раз больше плотности спектроскопии, достигнутой в SDSS) в сочетании с площадью 200 кв. градусов (при этом около половины площадок 1%DESI располагаются на восточной галактической полусфере). Комбинирование обзоров еРОЗИТА и 1%DESI позволяет получить рентгеновскую выборку с рекордно высокой полнотой однородной спектроскопии оптических компаньонов рентгеновских источников (по сравнению, с современными среднеформатными рентгеновскими обзорами, например, XMM–XXL и Stripe 82X; рентгеновский обзор eFEDS превосходит по площади, но значительно уступает по полноте оптической спектроскопии). В настоящей работе мы будем использовать объединенные данные двухлетнего обзора еРОЗИТА в восточной галактической полусфере и спектроскопического обзора 1%DESI для оценки точности моделей системы SRGz: фотометрической классификации и измерения красных смещений рентгеновских источников в обзоре неба еРОЗИТА.

Измерения photo-z рентгеновских источников в SRGz выполняются алгоритмом квантильного случайного леса без ограничения глубины, обученном на большой тренировочной выборке оптических квазаров и галактик SDSS (подробнее см. Мещеряков и др., 2023). Качество точечных прогнозов photo-z в работе (Мещеряков и др., 2023) оценивалось по стандартным метрикам σ_{NMAD} и $n_{>0.15}$ (нормализованное медианное абсолютное отклонение и доля катастрофических выбросов с $|z_{ph} - z_{sp}| / (1 + z_{sp}) > 0.15$) на рентгеновских источниках еРОЗИТА в области Stripe82X и в ~2 раза превосходит результаты других групп (Брешиа и др., 2019; Ананна и др., 2017), опубликованных в каталоге Stripe82X.

Несмотря на высокое качество точечных прогнозов photo-z рентгеновских источников еРОЗИТА в системе SRGz, используемый алгоритм случайного леса не гарантирует хорошую калибровку индивидуальных вероятностных прогнозов фотометрических красных смещений $PDF(z)$ (см.,

например, обсуждение различных методов в Ньюман и Груен, 2022; на рис. 5–6 в работе Мещерякова и др., 2018, можно видеть недостатки калибровки прогнозов photo-z рентгеновских источников). Необходимо также отметить, что сложности калибровки индивидуальных вероятностных прогнозов $PDF(z)$ возникают как в моделях photo-z на основе спектральных шаблонов, так и в моделях photo-z на основе машинного обучения (см., например, сравнение качества калибровки прогнозов $PDF(z)$ галактик, полученных разными моделями photo-z, рис. 7 в работе Коллаборация Euclid и др., 2020). Для того чтобы исправить ситуацию, ряд авторов предлагают различные методы постобработки вероятностных прогнозов (см., например, Гомес и др., 2018), однако данные подходы вместе с улучшением калибровки или отдельных показателей вероятностных прогнозов, в то же время, ухудшают другие показатели точности photo-z. В данной работе мы предложим новый метод постобработки прогнозов $PDF(z)$, который позволяет существенно улучшить калибровку доверительных интервалов photo-z рентгеновских источников, при этом не оказывает влияния на точность измерений photo-z (например, по метрикам σ_{NMAD} и $n_{>0.15}$).

Статья организована следующим образом. В разделе 2 мы опишем используемые нами данные обзоров неба. В разделе 3 представлены выборки ярких, средних и слабых рентгеновских источников еРОЗИТА в области 1% обзора DESI в восточной галактической полусфере. В разделе 4 мы рассмотрим результаты SRGz фотометрической классификации и измерения фотометрических красных смещений рентгеновских источников в сравнении с данными спектроскопических/астрометрических каталогов в этой области неба. В разделе 5 мы предложим новый метод калибровки вероятностных прогнозов фотометрических красных смещений рентгеновских объектов. Наконец, в разделе 6 представлены наши выводы.

2. ДАННЫЕ

2.1. Рентгеновский обзор СРГ/еРОЗИТА

В настоящей работе будут рассматриваться точечные рентгеновские источники СРГ/еРОЗИТА, обнаруженные со значимостью детектирования $L \geq 10$ в двухлетнем обзоре неба в восточной галактической полусфере, в области покрытия DESI EDR SV3.

2.2. Фотометрические и астрометрические каталоги

Мы будем использовать данные фотометрического каталога DESI Legacy Imaging Surveys DR9 (DESI LIS). DESI LIS представляет собой фотометрический обзор неба в оптическом диапазоне

в трех фильтрах (g , r , z), полученный по данным трех обзоров неба, сделанных различными телескопами (Beijing-Arizona Sky Survey (BASS), Dark Energy Camera Legacy Survey (DECaLS), Mayall z -band Legacy Survey (MzLS)) и приведенных к единой фотометрической системе (Дей и др., 2019).

DESI LIS DR9 дополнен фотометрическими измерениями в ИК-диапазоне по данным спутника WISE (принудительная фотометрия в четырех фильтрах). Мы будем использовать ИК-данные в фильтрах W_1 и W_2 (после шести лет обзора NEOWISE-Reactivation). Также для значительного числа источников в каталоге DESI LIS DR9 содержатся астрометрические данные GAIA DR2, которые мы будем использовать для классификации источников Галактики (астрометрические звезды отбираются нами по $S/N > 5$ по параллаксу и/или собственному движению в GAIA DR2).

Все оптические звездные величины в данной работе приводятся в AB-системе.

2.3. SRGz

В статье Мещерякова и др. (2023) описаны методы системы SRGz (версия 2.1) для физического отождествления точечных рентгеновских источников обзора ePOZITA в восточной галактической полусфере по фотометрическим данным в области покрытия обзора DESI Legacy Imaging Surveys. В работе также рассмотрены модели машинного обучения, позволившие получить измерения космологического красного смещения и класса рентгеновского объекта (квазар/галактика/звезда) по данным многоволновых фотометрических обзоров неба.

Для всех точечных рентгеновских источников SRGz отбирает наиболее вероятный оптический компаньон (отмечен как $srg_match_flag = 1$ в каталоге SRGz), для которого далее делаются прогнозы фотометрического класса и красного смещения (Мещеряков и др., 2023).

Дополнительно, для разбиения выборки точечных рентгеновских источников ePOZITA на объекты с оптическим компаньоном в данных DESI LIS и “бездомные” источники (рентгеновские объекты, не имеющие оптического компаньона в DESI LIS), в работе Мещерякова и др., 2023) предложено использовать следующий критерий:

$$srg_match_p0 \leq 0.24, \quad (1)$$

где srg_match_p0 — вероятность “бездомного” рентгеновского объекта, вычисляемая моделью отождествления SRGz (подробнее см. § 5.1 в работе Мещерякова и др., 2023). Для данного критерия отбора оптических компаньонов и “бездомных” рентгеновских объектов в работе (Мещеряков и др., 2023)

оценивались показатели точности и полноты. На выборке рентгеновских источников с рентгеновским потоком $F_{X,0.5-2} > 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$ оценка полноты и точности отождествления оптических компаньонов составила $\approx 95\%$ и $\approx 94\%$ соответственно; а оценка полноты и точности отождествления рентгеновских источников, не имеющих оптических партнеров (“бездомных” в обзоре DESI LIS), составила 82% (точность — 85%).

В данной работе мы будем использовать наиболее полную выборку наиболее вероятных оптических компаньонов рентгеновских источников (т.е. без использования дополнительного критерия (1)).

2.4. Спектроскопические каталоги

В области 1% DESI на восточной галактической полусфере, мы объединили данные нескольких спектроскопических каталогов:

- Спектроскопический каталог DESI EDR (DESI коллаборация и др., 2023) содержит спектроскопические классы оптических объектов (SPECTYPE — QSO, GALAXY, STAR) и красные смещения для внегалактических источников. В настоящей работе будут рассматриваться спектроскопические объекты DESI EDR с ZWARN = 0. Кроме того, для объектов SPECTYPE = GALAXY мы дополнительно требуем выполнения условий DELTACHI2 > 15 и $Z < 1.6$, что позволяет получить более надежную спектроскопическую выборку (см. DESI коллаборация и др., 2023).

- Спектроскопический каталог HELP (Ширли и др., 2021).

- Спектроскопический каталог квазаров SDSS DR16Q (Люк и др., 2020).

- Спектроскопический каталог SDSS DR18 (Алмейда и др., 2023).

В настоящей работе будут рассматриваться спектроскопические объекты SDSS с ZWARNING = 0.

Если у выбранного оптического объекта имеются спектроскопические измерения красного смещения в нескольких каталогах, то мы выбираем красные смещения согласно приоритету: DESI EDR → HELP → SDSS DR16Q → SDSS DR18. Данный выбор приоритетов обусловлен более надежными измерениями z_{sp} в конвейере обработки спектров DESI EDR по отношению к измерениям в каталоге SDSS: см. сравнение красных смещений в каталогах DESI EDR и SDSS на выборке спектров оптических квазаров подвергнутых визуальной проверке (§ 7.3 и рис. 17–18 в работе Шоссидон и др., 2023). Каталог квазаров SDSS DR16Q превосходит по качеству определения z_{sp} выборку квазаров из основного каталога SDSS, но также содержит заметную долю ошибочных спектральных измерений для далеких объектов на $z \gtrsim 4$ (Бу, Шен, 2022).

Таблица 1. Общие характеристики обзора eRO-1DESI-East для трех ограничений по рентгеновскому потоку (соответствующих яркой, средней и слабой рентгеновским выборкам)

	N	N/N_X
Bright: $F_{X,0.5-2} \geq 4 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$, $Area = 91.40$ 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: $1.5 \times 10^{-14} \leq F_{X,0.5-2} < 4 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$, $Area = 91.40$ 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: $6 \times 10^{-15} \leq F_{X,0.5-2} < 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$, $Area = 16.62$ 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%

Примечание. Приведено общее число точечных рентгеновских источников в выборках с ограничением по $F_{X,0.5-2}$; число вероятных оптических компаньонов, найденных SRGz, включая число астрометрических звезд GAIA и спектроскопических объектов. Выборка слабых рентгеновских объектов приведена на меньшей площади обзора в двух полях вблизи северного полюса эклиптики.

3. ВЫБОРКИ ТОЧЕЧНЫХ РЕНТГЕНОВСКИХ ИСТОЧНИКОВ EROSITA-1%DESI-EAST

Как было описано выше в п. 2.4, ранний выпуск спектроскопического обзора DESI (EDR)

содержит в себе результаты наблюдений по трем программам SV1,2,3. Для нас наибольший интерес представляют данные однопроцентного обзора DESI EDR SV3. Это результаты много-объектной спектроскопии в 20 специально отобранных

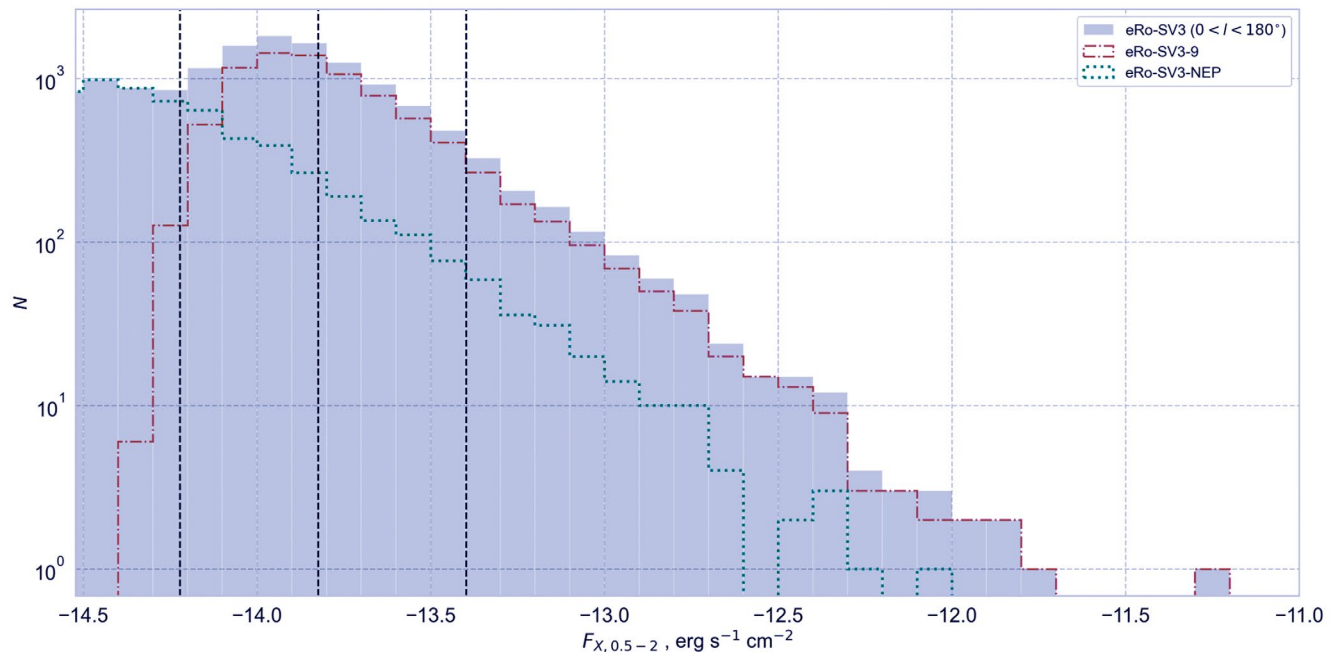


Рис. 1. Распределение точечных рентгеновских объектов ePOZITA в области обзора eRo-1DESI-East по рентгеновскому потоку в диапазоне 0.5–2 кэВ. Отдельно показаны распределения для двух площадок вблизи северного полюса эклиптики (eRo-SV3-NEP) и для всех остальных площадок обзора (eRo-SV3-9). Вертикальными штриховыми линиями отмечены пороговые рентгеновские потоки для яркой, средней и слабой выборки обзора: (“bright”) $F_{X,0.5-2} \geq 4 \times 10^{-14} \text{ эрг с}^{-1} \text{ см}^{-2}$, (“medium”) $1.5 \times 10^{-14} \leq F_{X,0.5-2} < 4 \times 10^{-14} \text{ эрг с}^{-1} \text{ см}^{-2}$, (“faint”) $6 \times 10^{-15} \leq F_{X,0.5-2} < 1.5 \times 10^{-14} \text{ эрг с}^{-1} \text{ см}^{-2}$.

площадках на небе (R0-R19, см. табл. 4 в DESI коллаборация и др., 2023). Расположение многих площадок пересекается с известными глубокими полями, по которым уже имеются высококачественные данные фотометрических и/или спектроскопических обзоров в публичном доступе. 11 полей DESI EDR SV3 полностью попадают в область покрытия рентгеновского обзора ePOZITA в восточной галактической полусфере. Площадки DESI EDR SV3 имеют круглую форму и площадь 8.48 кв. градусов каждая (в рамках данной площади наблюдается высокая плотность спектроскопических наблюдений DESI). Максимальная полнота спектроскопии для каждой площадки SV3 достигается в ее центральной области с площадью 6.48 кв. градусов. Для каждого поля мы вычислили максимальный радиус спектроскопического покрытия $R_{\text{sp,max}} \approx \sqrt{8.48 / \pi}$. Радиус поля, в котором мы исследовали рентгеновские источники ePOZITA для каждой области DESI EDR SV3, мы выбрали равным $R_X = R_{\text{sp,max}} - 1 = 1.626^\circ$ (данный выбор радиуса площадок обусловлен нашим желанием иметь все возможные оптические компаньоны рентгеновских источников внутри областей с высокой плотностью спектроскопических наблюдений). Таким образом, площадь каждой площадки DESI EDR SV3, в которой мы исследовали точечные рентгеновские источники, составила 8.31 кв. градусов.

Далее мы будем обозначать eRosita-1%DESI-East полученную выборку точечных рентгеновских источников (с лог-вероятностью детектирования $L > 10$) в 11 полях DESI EDR SV3 в восточной галактической полусфере. На рис. 3 показано распределение источников eRosita-1%DESI-East по рентгеновскому потоку в диапазоне 0.5–2 кэВ.

Данные eRosita-1%DESI-East можно разбить на две группы полей. Девять полей R3, 4, 6, 7, 11, 12, 13, 14, 19 (обозначенные eRo-SV3-9 на рис. 3) общей площадью 74.78 кв. градусов наблюдались в обзоре ePOZITA с умеренной экспозицией и имеют спектроскопическое и фотометрическое покрытие обзора SDSS. Поля R15, 18 (обозначенные eRo-SV3-NEP на рис. 3) общей площадью 16.62 кв. градусов находятся вблизи северного полюса эклиптики и наблюдались в обзоре неба ePOZITA с большой экспозицией. Спектральные данные SDSS не покрывают данные поля, а фотометрический обзор SDSS покрывает их лишь частично. 11 полей вместе имеют общую площадь 91.40 кв. градусов.

В рассматриваемых данных eRosita-1%DESI-East мы выделим три выборки источников с ограничением по рентгеновскому потоку в диапазоне 0.5–2 кэВ:

$F_{X,0.5-2} \geq 4 \times 10^{-14} \text{ эрг с}^{-1} \text{ см}^{-2}$, площадь покрытия – 91.40 кв. градусов.

$1.5 \leq F_{X,0.5-2} < 4 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$, площадь покрытия — 91.40 кв. градусов.

$0.6 \leq F_{X,0.5-2} < 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$, площадь покрытия — 16.62 кв. градусов.

Выбранные пороги по рентгеновскому потоку примерно соответствуют чувствительности полугодового (1), двухлетнего обзора еРОЗИТА в экваториальной области (2) и двухлетнего обзора еРОЗИТА в области северного полюса эклиптики (3). На рис. 3 вертикальными штриховыми линиями отмечены значения рентгеновских потоков, соответствующие выбранным границам выборок “bright”, “medium” и “faint”.

4. ФОТОМЕТРИЧЕСКИЕ ИЗМЕРЕНИЯ КЛАССОВ И КРАСНЫХ СМЕЩЕНИЙ НА ВЫБОРКАХ EROSITA-1%DESI-EAST

Общие характеристики полученных рентгеновских выборок и информация о полноте спектроскопической и астрометрической классификации оптических компаньонов в них приведены в табл. 1.

Необходимо отметить, что (см. табл. 1) для небольшой части (0.6–2.7%) рентгеновских источников SRGz не обнаружил оптические компаньоны в обзоре DESI LIS. В выборках “bright”, “medium” и “faint” имеется 29, 23 и 15 таких рентгеновских объектов, соответственно. Мы провели ручную проверку каждого из них и делаем вывод, что 64 из 67 объектов данного списка ассоциируется с яркой ($V = 6\text{--}13$ зв. вел.) оптической звездой в области локализации рентгеновского источника.

4.1. Фотометрическая классификация SRGz

В недавней работе (Мещеряков и др., 2023) точность моделей фотометрической классификации SRGz оценивалась на тестовой области SDSS площадью несколько тысяч кв. градусов с наибольшей плотностью спектроскопических квазаров SDSS DR16Q. Тестовая выборка в поле SDSS включала в себя ≈ 163 тыс. рентгеновских квазаров, ≈ 17 тыс. галактик и 34 тыс. рентгеновских звезд GAIA DR2. Модель классификации SRGz правильно классифицировала (см. табл. 7 в Мещеряков и др., 2023) $>99\%$ всех объектов (при классификации на классы рентгеновских звезд (STARS) и рентгеновских внегалактических источников (объединенный класс QSO+GALAXY)). Необходимо отметить, что данная выборка (в которой использовались данные об оптических объектах, имеющих спектроскопическую классификацию SDSS, и звездах с паралаксами / собственными движениями GAIA DR2) составила только $\approx 49\%$ от полной выборки оптических компаньонов рентгеновских источников еРОЗИТА в выбранной авторами тестовой области SDSS.

Данные eRosita-1%DESI-East, рассматриваемые здесь, обладают существенно большей (по сравнению с обзорами Stripe82X, XMM-XXL-N, eFEDS) долей спектроскопических измерений для оптических компаньонов рентгеновских источников еРОЗИТА. Для результатов, представленных ниже, мы будем использовать только DESI EDR и GAIA DR2. При этом доля оптических компаньонов объектов РОЗИТА со спектроскопической/астрометрической классификацией составляет (см. табл. 1): 91.1% (“bright”), 82.9% (“medium”), 72.2% (“faint”). Таким образом, данные 1%-го обзора DESI и астрометрия GAIA дают нам возможность исследовать точность и полноту фотометрической классификации SRGz на существенно более репрезентативной выборке рентгеновских источников, чем это было сделано ранее в работе (Мещеряков и др., 2023).

На верхней панели рис. 2 приведены матрицы путаницы фотометрической классификации для трех выборок eRosita-1%DESI-East: “bright” (слева), “medium” (в центре) и “faint” (справа). Для каждого объекта здесь выбран наиболее вероятный фотометрический класс (содержится в столбце *srg_match_SQG* в каталоге SRGz, см. табл. 2 в статье Мещеряков и др., 2023). Строчки матриц путаницы соответствуют разным классам объектов (согласно спектральной классификации DESI EDR и данным GAIA DR2), в столбцах таблицы — фотометрические классы SRGz.

Объекты выбранного класса (рентгеновские звезды (STAR), рентгеновские объекты, имеющие оптические спектры квазаров (QSO) или галактик (GALAXY) также можно отбирать на основе отдельных моделей классификации по каждому классу. Соответствующие вероятности *srg_match_pstar*, *srg_match_pqso*, *srg_match_pstar* приведены в каталоге SRGz (см. формулы (32)–(34) в статье Мещеряков и др., 2023). На рис. 2 (нижняя панель) показаны кривые точность–полнота (Precision–Recall) для моделей фотометрической классификации разных классов рентгеновских объектов. Каждая точка на этих кривых соответствует выборке объектов с определенным порогом по вероятности соответствующей модели. В подписях к кривым приведены значения средней точности, достигаемой при таком отборе объектов каждого класса.

Мы делаем следующие выводы в отношении фотометрической классификации рентгеновских источников на выборках eRosita-1%DESI:

1. SRGz показывает высокую точность фотометрической классификации рентгеновских звезд на всех рассматриваемых рентгеновских потоках. В общей сложности по всем трем рентгеновским выборкам мы видим малое число ошибок I и II рода при фотометрической классификации звезд. SRGz ошибочно записывает в фотометрические

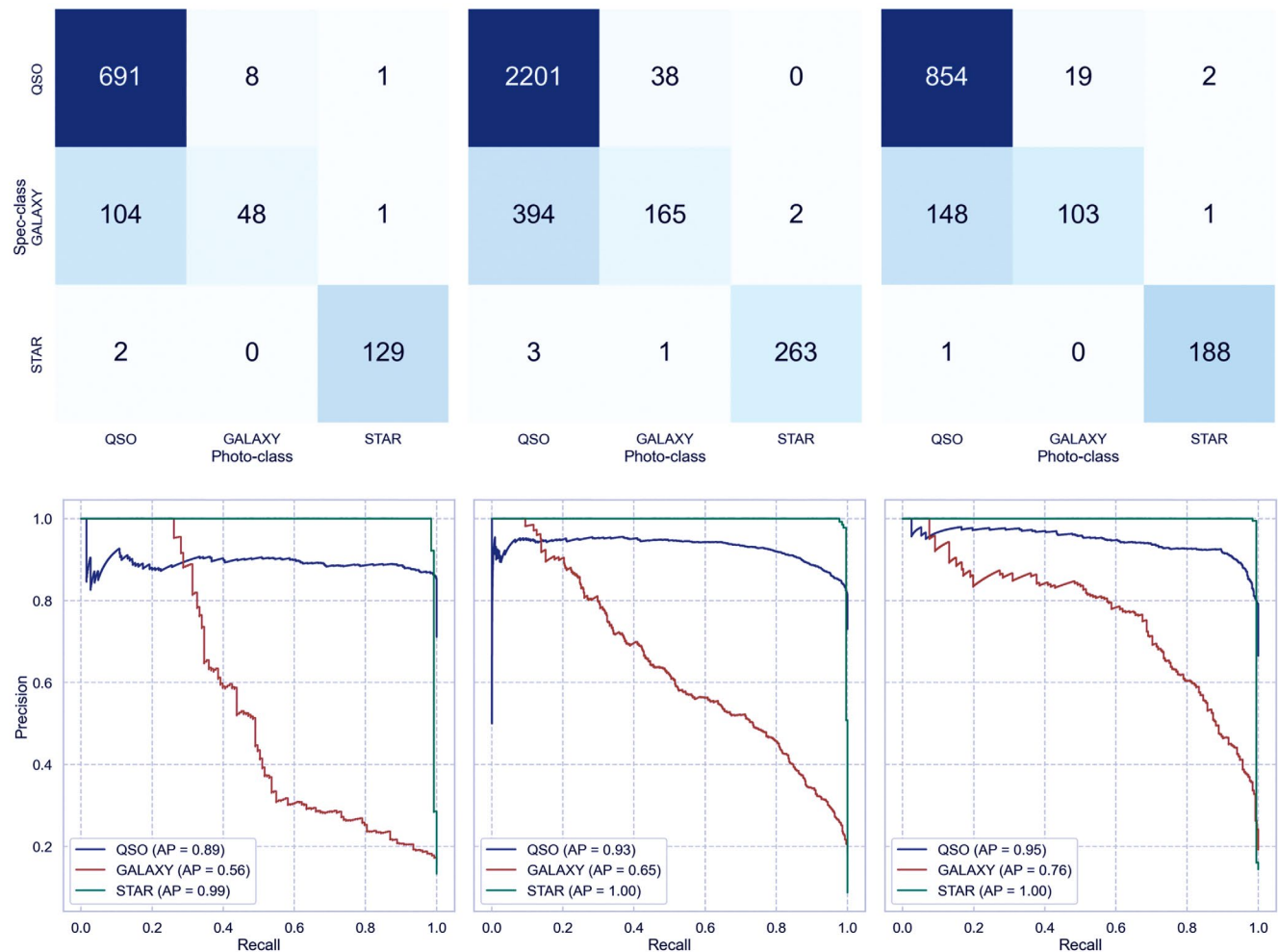


Рис. 2. Матрицы путаницы фотометрической классификации (верхняя панель) и графики точность—полнота (Precision—Recall, нижняя панель) для трех выборок eRosita-1%DESI-East: “bright” (слева), “medium” (в центре) и “faint” (справа). На графиках Precision—Recall показаны кривые классификации рентгеновских звезд (STAR), рентгеновских объектов, имеющих оптический спектр квазара (QSO), и спектр галактики (GALAXY). В подписи к кривым приведены значения средней точности, достигаемой при отборе объектов каждого класса.

звезды только 7 внегалактических источников (из 4780) и неправильно классифицирует только 7 рентгеновских звезд (из 587).

2. Большая часть внегалактических рентгеновских источников, не показывающих квазарных особенностей в своем оптическом спектре, отождествляется фотометрическим классификатором SRGz как квазары: 2/3 объектов (из 966 рентгеновских источников eRosita-1%DESI-East), имеющих оптический спектр GALAXY по классификации DESI EDR, показывают фотометрический класс QSO. С другой стороны, среди 382 фотометрических галактик только $\approx 17\%$ имеют спектр оптического квазара, согласно DESI EDR.

В следующих разделах статьи мы подробно рассмотрим результаты измерения фотометрических красных смещений оптических

компаньонов рентгеновских источников на выборке eRosita-1%DESI-East.

4.2. Фотометрические красные смещения

В работе (Мещеряков и др., 2023) качество точных прогнозов фотометрических красных смещений SRGz оценивалось для выборки рентгеновских источников ePOZITA в области Stripe 82X на площади 31.3 кв. градусов и сравнивалось с результатами других групп в этом поле (см., рис. 24 в указанной работе). Тестовая выборка включала в себя 1379 внегалактических рентгеновских источников с измерениями спектрального красного смещения их оптических компаньонов, что составило $\approx 71\%$ всей выборки точечных источников ePOZITA в этом поле. В настоящей статье

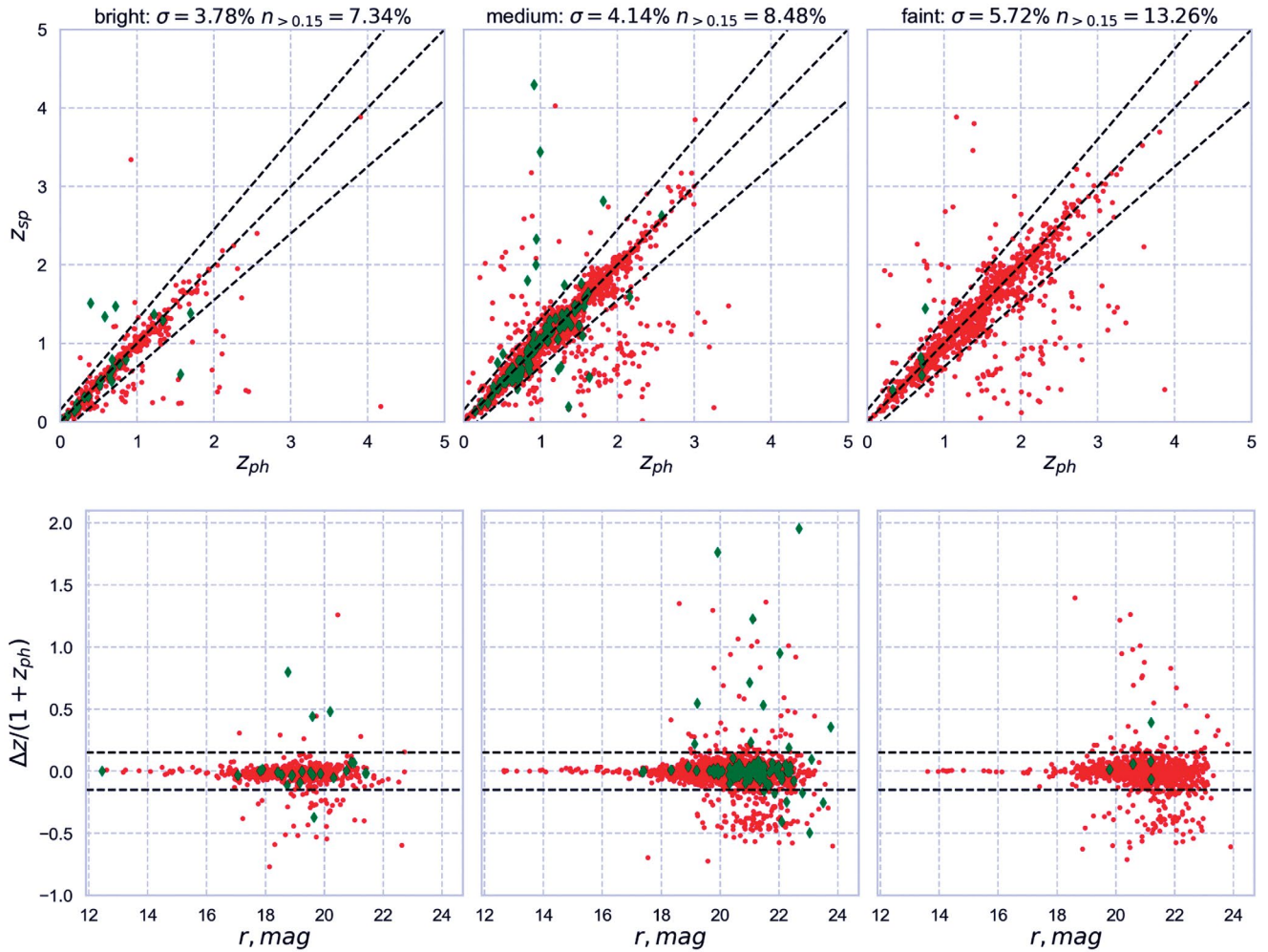


Рис. 3. Верхняя панель: диаграмма рассеяния фотометрических красных смещений для трех выборок оптических компаньонов рентгеновских источников. Нижняя панель: зависимость нормированного отклонения фотометрического красного смещения (Δz_{norm}) от звездной величины в фильтре r . Кружки и ромбы на графиках — объекты со спектроскопическими измерениями DESI EDR и других обзоров соответственно. За пределами черных штриховых линий находится область катастрофических выбросов в измерениях photo- z . На графиках можно видеть, как меняется точность photo- z (по величине разброса Δz_{norm} и появлению выбросов) в зависимости от рентгеновского и оптического потоков объекта.

мы имеем возможность исследовать точность моделей photo- z на разных рентгеновских потоках и на более репрезентативной спектроскопической выборке оптических компаньонов рентгеновских источников eRosita-1%DESI-East.

Для исследования точности измерения photo- z на выборках eRosita-1%DESI-East мы будем использовать данные всех рассматриваемых спектроскопических каталогов в соответствии с их приоритетом (подробнее см. п. 2.4). Часть рентгеновских объектов eRosita-1%DESI-East участвовали в обучении моделей photo- z SRGz (14–20% выборок “bright” и “medium”, см. табл. 1), и эти объекты были исключены из соответствующих тестовых выборок.

На верхней панели рис. 3 представлены диаграммы рассеяния фотометрических красных смещений для трех тестовых выборок рентгеновских источников: “bright”, “medium” и “faint”. Кружками и ромбами (разным цветом) на графиках показаны объекты со спектроскопическими измерениями DESI EDR и из всех остальных спектроскопических каталогов, соответственно. За пределами штриховых линий на графиках находится область катастрофических выбросов в измерениях фотометрических красных смещений, которые будут определяться следующим (стандартным для photo- z квазаров в литературе) образом:

$$|\Delta z_{\text{norm}}| > 0.15, \quad (2)$$

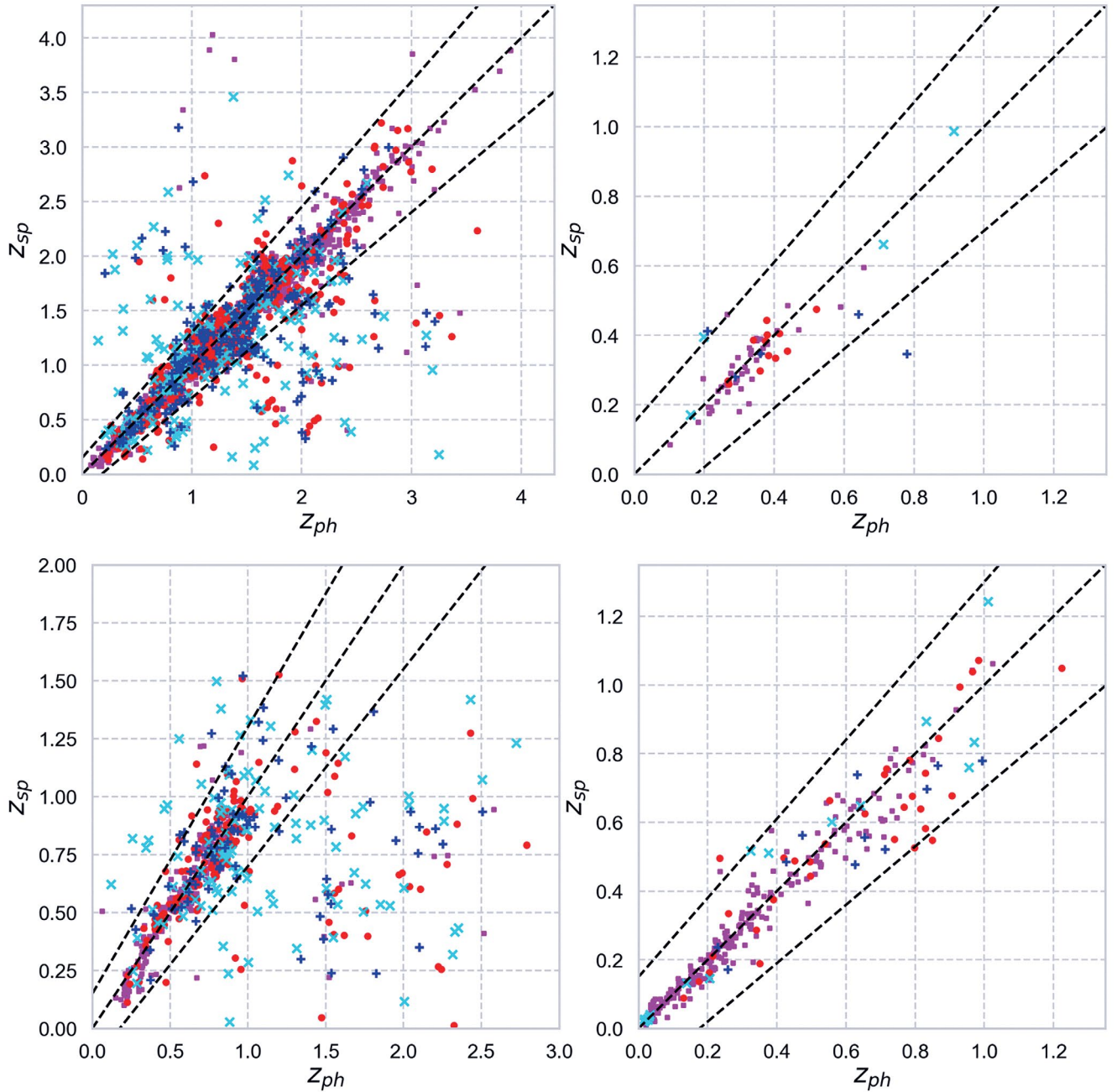


Рис. 4. Диаграммы рассеяния фотометрических красных смещений для рентгеновских источников eRosita-1%DESI-East для разных фотометрических и спектроскопических классов внегалактических рентгеновских источников (которые рассматривались в п. 4.1). На графиках показаны только объекты со спектральной классификацией DESI LIS (QSO — верхняя панель графиков, GALAXY — нижняя панель графиков). Фотометрические квазары показаны на графиках слева, фотометрические галактики — справа. На всех графиках разными символами и цветом показаны объекты в разных диапазонах по параметру надежности оценок photo-z (z_{Conf}): $z_{Conf} > 0.6$ (малиновые треугольники), $0.4 \leq z_{Conf} \leq 0.6$ (красные кружки), $0.3 < z_{Conf} \leq 0.4$ (синие плюсы), $z_{Conf} \leq 0.3$ (голубые кресты). За пределами черных штриховых линий находится область катастрофических выбросов ($\Delta z_{norm} > 0.15$) в измерениях photo-z.

где $\Delta z_{norm} = (\Delta z) / (1 + z_{ph})$, $\Delta z = z_{sp} - z_{ph}$, z_{sp} — спектральное красное смещение, z_{ph} — измерение SRGz (точный прогноз фотометрического красного смещения объекта по наиболее вероятному

значению распределения $PDF(z)$). Поверх графиков верхней панели рис. 3 подписаны значения метрик (нормализованное медианное отклонение σ_{NMAD} и доля катастрофических выбросов $n_{>0.15}$, см.

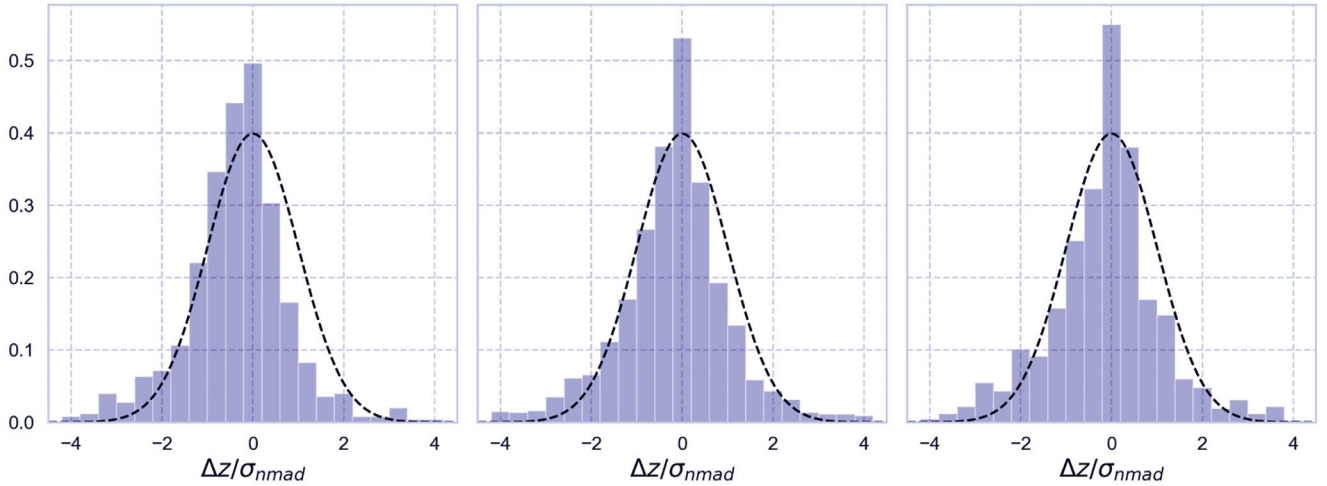


Рис 5. Распределения отклонений $\Delta z = z_{sp} - z_{ph}$ точечных прогнозов SRGz фотометрических красных смещений рентгеновских источников, нормированных на оценку стандартного отклонения σ_{NMAD} , для выборок источников eRosita-1%DESI-East в разных интервалах по рентгеновскому потоку: “bright” (слева), “medium” (в центре) и “faint” (справа). Штриховыми линиями показаны нормальные распределения с нулевым средним и единичной дисперсией. Из графиков следует, что измерения фотометрических красных смещений рентгеновских источников в SRGz имеют существенно негауссовый характер. Обращает на себя внимание как избыток объектов в крыльях распределения, так и их существенно несимметричная форма.

также Мещеряков и др., 2023) для каждой рентгеновской подвыборки.

На нижней панели рис. 3 представлена зависимость нормированного отклонения фотометрического красного смещения (Δz_{norm}) от звездной величины (в фильтре r DESI LIS) оптического компаньона рентгеновского источника. На этих графиках можно видеть, как меняется точность photo-z (по величине разброса Δz_{norm} и появлению выбросов) в зависимости от рентгеновского и оптического потока объекта.

4.2.1 Отличия в прогнозах photo-z для разных фотометрических / спектроскопических классов объектов. На рис. 4 показаны диаграммы рассеяния фотометрических красных смещений для рентгеновских источников eRosita-1%DESI-East для разных фотометрических и спектроскопических классов внегалактических рентгеновских источников (которые рассматривались в п. 4.1). На графиках показаны только объекты со спектральной классификацией DESI LIS (QSO – верхняя панель графиков, GALAXY – нижняя панель графиков). Фотометрические квазары показаны на графиках слева, фотометрические галактики – справа. На всех графиках разными символами и цветом показаны объекты в разных диапазонах по параметру надежности оценок photo-z ($zConf$ – см. формулу (38) в статье Мещеряков и др., 2023). Штриховыми и пунктирными прямыми на графиках показаны границы областей катастрофических выбросов.

Мы хотим обратить внимание читателя на несколько деталей:

1. Рентгеновские источники с фотометрическим классом GALAXY демонстрируют заметно более высокую точность измерения photo-z и практически не имеют катастрофических выбросов, в отличие от рентгеновских объектов с фотометрическим классом QSO.

2. Более половины катастрофических выбросов среди измерений photo-z попадают в объекты с низким значением $zConf$ и могут быть убраны путем соответствующей фильтрации.

3. Мы видим, что часть объектов фотометрического класса QSO, являющихся катастрофическими выбросами с точки зрения измерений красных смещений с $z_{sp} < z_{ph}$, обладают высоким значением $zConf$ и не могут быть отфильтрованы по этому параметру.

4.2.2. Негауссовость прогнозов SRGz. Далее мы коснемся вопроса, насколько прогнозы photo-z, полученные для рентгеновских источников моделями SRGz, можно считать гауссовыми. Для каждой выборки eRosita-1%DESI-East мы рассчитали величину нормированного медианного абсолютного отклонения σ_{NMAD} , являющуюся робастным способом оценки стандартного отклонения для колокола распределения ошибок измерений photo-z. На рис. 5 приведены распределения величины $(z_{sp} - z_{ph})/\sigma_{NMAD}$ для трех рассматриваемых рентгеновских выборок. Для сравнения, на графиках приведены функции нормального распределения с единичной дисперсией.

Из графиков рис. 5 следует, что измерения фотометрических красных смещений в SRGz имеют существенно негауссовый характер. Обращает на себя внимание как избыток объектов в крыльях распределения, так и их существенно несимметричная форма. Таким образом, полученные вероятностные прогнозы photo-z рентгеновских источников не могут характеризоваться только двумя параметрами (среднее и дисперсия прогноза), имеют сложную форму (иногда с несколькими пиками) и требуют рассмотрения полных распределений $PDF(z)$, которые приводятся системой SRGz для каждого объекта.

В следующем разделе мы подробно остановимся на особенностях полных вероятностных прогнозов SRGz для фотометрических красных смещений рентгеновских объектов и методе их калибровки.

4.3. Точечные прогнозы SRGz: выводы

Мы обращаем внимание на следующие особенности точечных измерений SRGz фотометрических красных смещений рентгеновских источников ePOZITA:

1. Нами отмечается высокая точность фотометрических красных смещений, полученных моделями SRGz, для рентгеновских источников в двухлетнем обзоре неба ePOZITA: стандартное отклонение ($\sigma_{NMAD} \approx 4\%$) для оптических компаньонов рентгеновских источников с $F_{X,0.5-2} \geq 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$.

2. В измерениях photo-z рентгеновских квазаров присутствуют, так называемые, катастрофические выбросы — значительные отклонения измерения photo-z от красных смещений линий в спектрах объектов. Присутствие выбросов в прогнозах вызвано неоднозначностью самого преобразования фотометрических признаков рентгеновских объектов в красные смещения ($X_{\text{ph}} \rightarrow z$). При этом, доля выбросов зависит от различных факторов: от набора использованных фотометрических фильтров (Мещеряков и др., 2018), от выбранного алгоритма photo-z, репрезентативности и объема тренировочных выборок (или набора шаблонов), на которых модель обучается. Модели SRGz в области покрытия обзоров DESI Legacy Imaging Surveys/Pan-STARRS1/SDSS, в настоящее время, позволяют получить прогнозы photo-z с долей выбросов на уровне $n_{>0.15} = 7-8.5\%$ для рентгеновских источников с $F_{X,0.5-2} \geq 1.5 \times 10^{-14}$ эрг $\text{с}^{-1} \text{см}^{-2}$.

3. Еще одной важной особенностью прогнозов фотометрических красных смещений рентгеновских квазаров (связанной напрямую с наличием выбросов) является их негауссовость, что не позволяет при описании ошибки прогноза photo-z ограничиться только значением его стандартного отклонения. Функция распределения вероятностного

прогноза $PDF(z)$ может иметь мультимодальный характер. Восстановление этой функции для каждого рентгеновского объекта является значительно более сложной задачей, чем получение точечного прогноза z_{ph} .

5. ВЕРОЯТНОСТНЫЕ ПРОГНОЗЫ ФОТОМЕТРИЧЕСКИХ КРАСНЫХ СМЕЩЕНИЙ И ИХ КАЛИБРОВКА

В этом разделе статьи будет рассмотрен метод восстановления полного вероятностного прогноза $PDF(z)$ фотометрического красного смещения рентгеновского объекта (который используется в системе SRGz версии 2.1, см. Мещеряков и др., 2023) и предложена модель его калибровки, а также выполнен анализ качества калибровки прогнозов SRGz на рентгеновских выборках eRosita-1%DESI-East. Отметим, что описанная ниже модель калибровки вероятностных прогнозов может также применяться для других приложений квантильной регрессии за пределами астрофизической предметной области.

5.1. Измерения фотометрических красных смещений рентгеновских источников в SRGz

Для вероятностных прогнозов фотометрических красных смещений в системе SRGz версии 2.1 (рассматриваемой в данной работе) применяется метод квантильного случайного леса (см. подробное описание модели применительно к астрономии в Мещеряков и др., 2018, 2023). Алгоритм случайного леса был предложен Лео Брейманом (Брейман, 2001) и впервые применен для задач квантильной регрессии в работе (Майнсхаузен, 2006). В астрономии, случайный лес является, в настоящее время, одним из самых широко используемых алгоритмов машинного обучения. В частности, для измерения фотометрических красных смещений галактик он используется в модели TPZ (Карраско Кинд, Бруннер, 2013). В системе SRGz мы адаптировали случайный лес для задачи измерения photo-z рентгеновских источников. В SRGz алгоритм случайного леса используется в варианте деревьев решений построенных без ограничения глубины. Свойства подобных моделей подробно изучались в работе (Танг и др., 2018). Каждое дерево решений случайного леса на выходе указывает на объект спектральной тренировочной выборки, который оно определило как «ближайший» в процессе своего разбиения (представление о методе случайного леса как об адаптивном алгоритме ближайших соседей рассматривалось в статье Лин, Джеон, 2006). Дерево случайного леса выбирает спектральный объект из случайной подвыборки (полученной путем бутстреп агрегации, так называемый бэггинг) тренировочных объектов;

построение дерева ведется на случайной подвыборке фотометрических признаков (так называемый подход случайных подпространств). Такой процесс выбора случайных подвыборок объектов и признаков для построения деревьев решений называется их рандомизацией. Таким образом, алгоритм случайного леса SRGz позволяет получить ансамбль прогнозов фотометрических красных смещений сделанными рандомизированными деревьями решений для каждого рентгеновского объекта.

Полученный ансамбль прогнозов деревьев решений случайного леса $\{z_1, \dots, z_n\}$ (где $n = 288$ для моделей SRGz, описанных в работе Мешеряков и др., 2023), является эмпирическим представлением вероятностного прогноза красного смещения $PDF(z)$ для заданного рентгеновского объекта с набором фотометрических признаков X_{ph} :

$$P_S(z) = \frac{1}{\sum_{s=1}^n w_s} \sum_{s=1}^n w_s \delta(z - z_s), \quad (3)$$

где $\sigma(x)$ — дельта-функция Дирака, а w_s — вес прогноза (z_s) дерева решений s . Для модели photo-z без использования дополнительной калибровки (см. п. 5.2) веса всех деревьев равны $w_s = 1$.

Для измерения фотометрического красного смещения объекта мы вычисляем взвешенную (см., например, Динардо и др., 1996) функцию ядерной оценки плотности распределения (3) с подобранной эмпирически величиной параметра сглаживания $h = 0.1$:

$$P_{KDE}(z) = \frac{1}{h \sum_{s=1}^n w_s} \sum_{s=1}^n w_s \mathcal{N}\left(\frac{z - z_s}{h}\right), \quad (4)$$

где $\mathcal{N}(x)$ — функция нормального распределения с нулевым средним $\mu = 1$ и единичной дисперсией $\sigma^2 = 1$. Фотометрическое красное смещение объекта соответствует максимуму функции ядерной оценки плотности на сетке прогнозов деревьев $\{z_1, \dots, z_n\}$:

$$z_{ph} = \underset{z_s}{\operatorname{argmax}} P_{KDE}(z_s). \quad (5)$$

Модель калибровки вероятностных прогнозов фотометрических красных смещений рентгеновских источников описана ниже.

5.2. Калибровка вероятностных прогнозов фотометрических красных смещений

Прогнозы $PDF(z)$ рентгеновских источников, полученные SRGz методом квантильного случайного леса (см. п. 5.1), могут нуждаться в дополнительной калибровке по нескольким причинам.

Во-первых, прогноз z_{ph} получается из эмпирического распределения прогнозов независимых деревьев решений $P_S(z)$ путем некоторой процедуры усреднения и распределение $PDF(z_{ph})$, таким образом, должно иметь меньший разброс, чем $PDF(z_s)$ для независимых деревьев решений. Во-вторых, деревья решений в составе случайного леса не являются полностью независимыми (так как строятся на пересекающихся выборках спектральных объектов). Мы можем сделать вывод, что текущие модели измерения фотометрических красных смещений SRGz (как и подавляющее большинство предложенных в литературе подходов photo-z на основе машинного обучения и с использованием шаблонов, см. обсуждение в § 4.1.1 обзора Ньюман, Груен, 2022) не гарантируют калибровку полного вероятностного прогноза $PDF(z)$ рентгеновских источников.

Сильной стороной прогнозов SRGz на основе случайного леса является их высокая точность по метрикам σ_{NMAD} и $n_{>0.15}$ по отношению к другим предложенным моделям photo-z рентгеновских источников (Мешеряков и др., 2023). Необходимо также отметить, что качество калибровки вероятностных прогнозов красных смещений никак не гарантирует точности photo-z. Так, в работе Шмидт и др. (2020) продемонстрирована тривиальная модель *trainZ* (представляющая собой распределение спектральных красных смещений тренировочной выборки), которая имеет идеальную калибровку и при этом посредственную точность прогнозов. Ниже мы предложим метод калибровки вероятностных прогнозов красного смещения, который не меняет положение максимума распределения вероятностного прогноза (и, таким образом, сохраняет высокую точность измерений z_{ph}) и, при этом, существенно улучшает калибровку $PDF(z)$.

Современный обзор лучших подходов к проблеме калибровки моделей машинного (глубокого) обучения приведен в статье (Ванг, 2023). Одним из алгоритмов, сочетающим в себе точность калибровки и низкую вычислительную сложность, является метод температурного масштабирования вероятностей (Temperature Scaling, см. Мозафари и др., 2018; Платт, 2000). Основываясь на данном подходе, мы предлагаем следующую модель калибровки вероятностного прогноза красных смещений.

Введем вспомогательную функцию $W(z, T)$:

$$W(z, T) = \frac{\exp(\ln(P_{KDE}(z; w_s = 1)) / T)}{P_{KDE}(z; w_s = 1)}, \quad (6)$$

где $P_{KDE}(z; w_s = 1)$ — функция ядерной оценки плотности распределения вероятностного прогноза красного смещения по формуле (3) с весами равными единице; T — параметр модели температурного масштабирования.

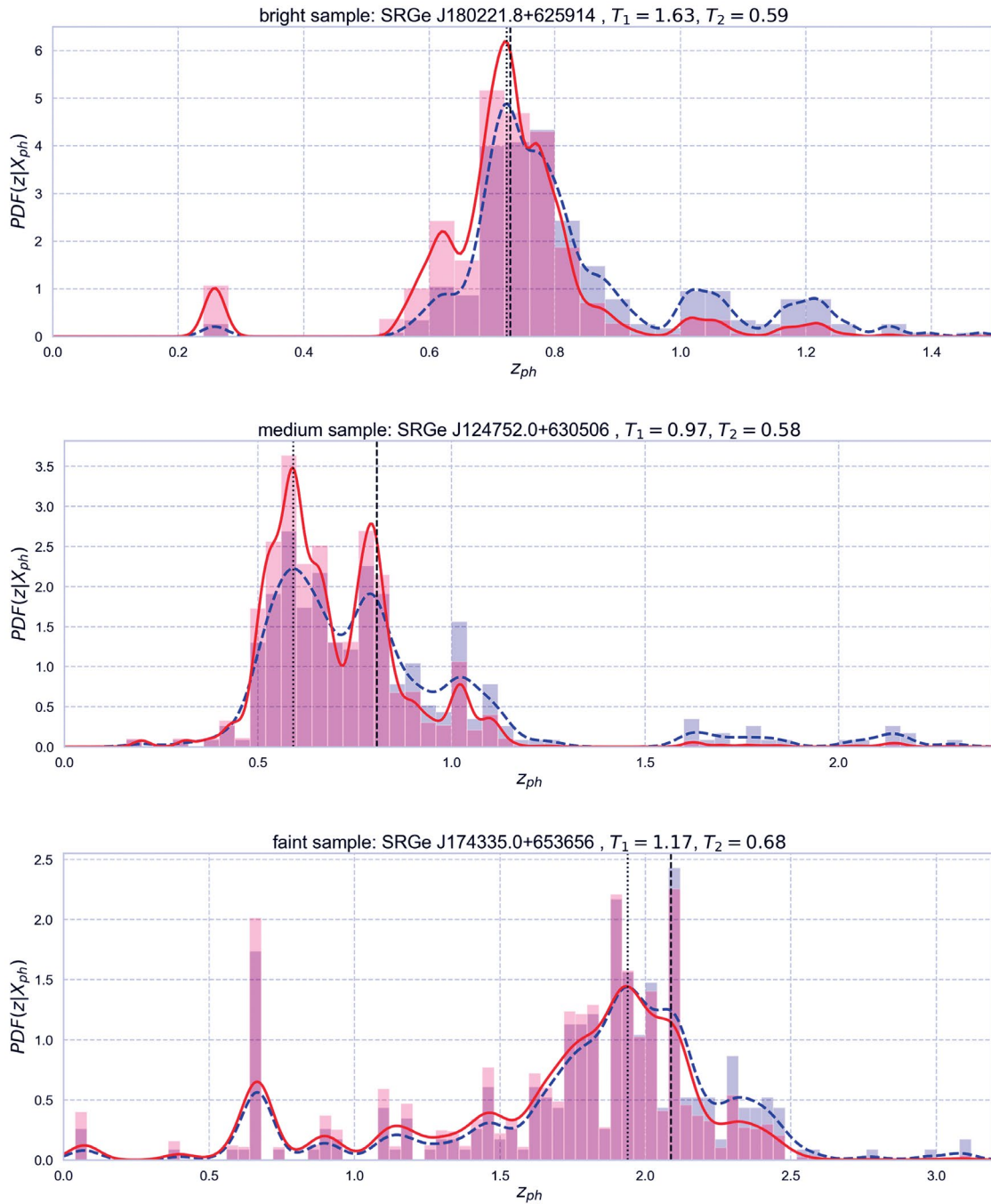


Рис. 6. Примеры вероятностных прогнозов $PDF(z)$, сделанных моделями SRGz для объектов выборок “bright” (верхняя панель), “medium” (средняя панель) и “faint” (нижняя панель). Эмпирическое распределение $P_S(z)$ показано в виде гистограммы голубого цвета, $P_{KDE}(z)$ – штриховая синяя линия на графиках. Калиброванные эмпирические распределения $P_S(z)$ показаны в виде гистограмм красного цвета, калиброванные функции ядерной оценки плотности распределения $P_{KDE}(z)$ показаны сплошной красной линией на графиках. Параметры модели двухтемпературной калибровки подписаны в заголовке графиков. Вертикальные пунктирная и штриховая линии обозначают точечный прогноз фотометрического красного смещения (z_{ph}) и спектральное красное смещение (z_{sp}) объекта соответственно. На верхнем графике фотометрическое красное смещение объекта очень хорошо согласуется с результатом спектроскопических измерений. Два других источника демонстрируют расхождения между измерениями z_{ph} и z_{sp} . Наибольшее расхождение между фотометрическим и спектральным измерениями демонстрирует объект на средней панели. Также здесь хорошо виден мультимодальный характер прогнозов photo-z рентгеновских квазаров по выбранной системе широкополосных фильтров: $PDF(z)$ имеет в этом случае несколько пиков, одному (наиболее вероятному) из которых следует прогноз photo-z, в другом лежит спектральное измерение красного смещения.

Для каждого прогноза z_s дерева решений (входящего в состав ансамбля прогнозов случайного леса $s = 1, \dots, n$ деревьев, см. (3)) мы будем вычислять вес w_s по следующим формулам (отдельно для левого и правого крыла функции распределения по отношению к z_{ph}):

$$w_s = \begin{cases} \frac{W(z_s, T_1)}{W(z_{ph}, T_1)}, & \text{если } z_s < z_{ph}, \\ \frac{W(z_s, T_2)}{W(z_{ph}, T_2)}, & \text{если } z_s \geq z_{ph}. \end{cases} \quad (7)$$

Далее, калиброванный вероятностный прогноз красного смещения может быть рассчитан по приведенным ранее формулам (3)–(4), подставляя в них веса (7).

Параметры модели калибровки T_1 и T_2 будут определяться нами на рассматриваемых в данной работе рентгеновских выборках с известными спектроскопическими красными смещениями для оптических компаньонов рентгеновских объектов. На заданной выборке для поиска параметров модели калибровки мы будем минимизировать значение статистики (KS) в тесте Колмогорова–Смирнова:

$$KS = \max_{|\Delta z|} |CDF(|\Delta z|) - CDF(|\Delta z|)|, \quad (8)$$

где $CDF(|\Delta z|)$ и $CDF(|\Delta z|)$ – теоретические и наблюдаемые значения функции распределения для величины $|\Delta z| = |z_{sp} - z_{ph}|$. Для оптимизации параметров T_1 и T_2 мы будем отдельно рассматривать подвыборки объектов со спектральными красными смещениями по левую ($z_{sp} < z_{ph}$) и правую ($z_{sp} \geq z_{ph}$) сторону от точечных прогнозов z_{ph} соответственно. Отметим, что предлагаемый здесь метод позволяет откалибровать прогнозы доверительных интервалов для величины z_{ph} .

На рис. 6 приведены примеры вероятностных прогнозов $PDF(z)$, сделанных моделями SRGz для трех рентгеновских источников eRosita-1%DESI-East (случайно взятых из выборок “bright”, “medium” и “faint”). Эмпирическое распределение $P_S(z)$ показано в виде гистограммы голубого цвета (с шагом $\Delta z_{bin} = 0.04$), $P_{KDE}(z)$ – штриховая синяя линия на графиках. Калиброванные эмпирические распределения $P_S(z)$ показаны в виде гистограмм красного цвета, калиброванные функции ядерной оценки плотности распределения $P_{KDE}(z)$ показаны сплошной красной линией на графиках. Параметры модели двух-температурной калибровки подписаны в заголовке графиков. Вертикальные пунктирная и штриховая линии обозначают точечный прогноз фотометрического красного смещения (z_{ph}) и спектральное красное смещение (z_{sp}) объекта соответственно. На верхнем графике фотометрическое

красное смещение объекта очень хорошо согласуется с результатом спектроскопических измерений. Два других источника демонстрируют расхождения между измерениями z_{ph} и z_{sp} . Наибольшее расхождение между фотометрическим и спектральным измерением демонстрирует объект на средней панели рис. 5.1. Также здесь хорошо виден мультимодальный характер прогнозов photo-z рентгеновских квазаров по выбранной системе широкополосных фильтров: $PDF(z)$ имеет в этом случае несколько пиков, одному (наиболее вероятному) из которых следует прогноз photo-z, в другом лежит спектральное измерение красного смещения.

Необходимо отметить ряд особенностей калибровки вероятностных прогнозов фотометрических красных смещений, сделанных методом двух-температурного масштабирования (которые также хорошо видны на рис. 6):

- В результате калибровки не изменяется положение пиков на графике $PDF(z)$, меняется только их относительная высота и ширина. Предлагаемый метод постобработки вероятностных прогнозов не изменяет порядок пиков по высоте по одну сторону от главного пика z_{ph} – если один пик был выше другого, то он так и останется выше после калибровки.

- Значение параметра “температуры” $T = 1$ в предлагаемом методе соответствует случаю вероятностных прогнозов без дополнительной калибровки. При $T > 1$ вероятность в соответствующем крыле $PDF(z)$ увеличивается, при $T < 1$ – масса вероятности уменьшается.

- Левое и правое крылья (относительно z_{ph}) распределения $PDF(z)$ калибруются независимо в соответствии со своими параметрами (“температурами”) T_1 и T_2 . Это позволяет обладать большей гибкостью при калибровке несимметричных профилей $PDF(z)$, характерных для квазаров.

В следующей части мы рассмотрим результаты калибровки вероятностных смещений рентгеновских квазаров eRosita-1%DESI-East предложенным здесь методом.

5.3. Результаты калибровки вероятностных прогнозов SRGz

Можно задаться вопросом, в какой степени измерения $P_S(z)$ (см. формулу (3)), полученные SRGz для объектов из рассматриваемых рентгеновских выборок, в действительности, можно считать прогнозами вероятности обнаружить объект на том или ином красном смещении? Степень калибровки вероятностного прогноза красного смещения можно наглядно представить с помощью так называемых PIT-диаграмм (PIT, англ. Probability Integral Transform, см., например, Д’Инсанто, Полстерер, 2018; Мещеряков и др., 2018; Дэвид, 1984) и квантильных графиков (QQ-диаграмм) на

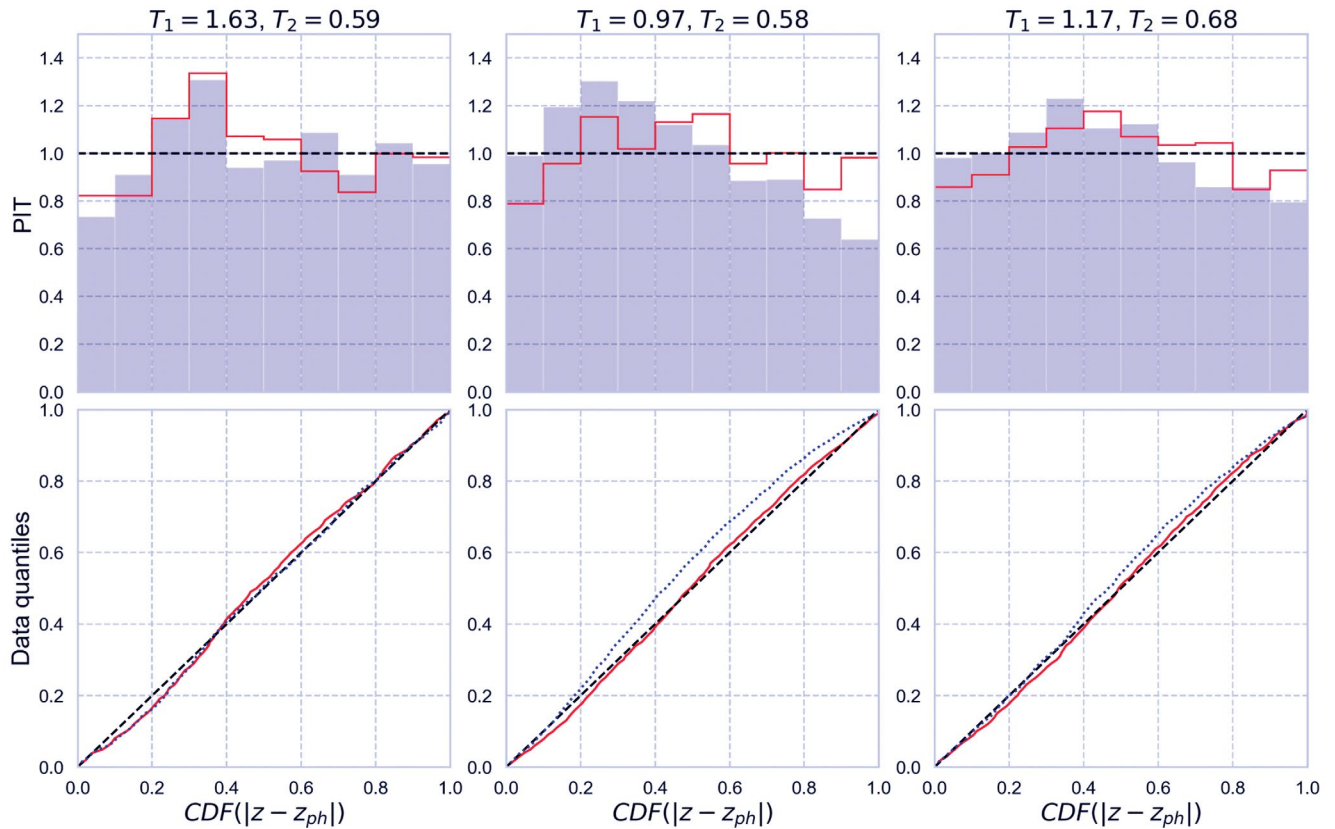


Рис. 7. PIT-диаграммы (вверху) и QQ-диаграммы (внизу) иллюстрируют калибровку вероятностных прогнозов фотометрических красных смещений SRGz для трех выборок рентгеновских источников eRosita-1%DESI-East: “bright” (слева), “medium” (в центре) и “faint” (справа). Вероятностные прогнозы с идеальной калибровкой соответствуют горизонтальной прямой линии на PIT-диаграмме и диагонали графика квантилей (показаны черными штриховыми линиями на графиках). PIT-диаграмма для вероятностных прогнозов SRGz до и после калибровки показана соответственно в виде закрашенной синей гистограммы и сплошной красной линией на верхней панели графиков. Соответствующие QQ-диаграммы прогнозов SRGz показаны синими пунктирными линиями (до калибровки) и сплошными красными линиями (после калибровки) на нижней панели рисунка. Значения параметров метода калибровки $PDF(z)$ приведены над верхней панелью графиков.

их основе. Отметим, что мы рассматриваем здесь PIT и QQ-диаграммы, построенные для величины $|\Delta z| = |z_{sp} - z_{ph}|$, что будет отражать качество доверительных интервалов на измерение z_{ph} .

На рис. 7 показаны PIT (верхняя панель) и QQ-диаграммы для трех тестовых выборок источников eRosita-1%DESI-East в разных диапазонах по рентгеновскому потоку (см. п. 4.2): “bright” (слева), “medium” (в центре) и “faint” (справа). Вероятностные прогнозы с идеальной калибровкой соответствуют горизонтальной прямой линии на PIT-диаграмме и диагонали графика квантилей (показаны черными штриховыми линиями на графиках). PIT-диаграмма для вероятностных прогнозов SRGz до и после калибровки показана соответственно в виде закрашенной синей гистограммы и сплошной красной линией на верхней панели графиков. Соответствующие QQ-диаграммы прогнозов SRGz показаны синими пунктирными

линиями (до калибровки) и сплошными красными линиями (после калибровки) на нижней панели рисунка. Выбранные значения параметров калибровки для разных рентгеновских подвыборок приведены на рис. 6.

Как видно на рис. 7, вероятностные прогнозы SRGz без калибровки демонстрируют расхождения в значениях теоретического и наблюдаемого значений $CDF(|\Delta z|)$ (особенно заметные для выборки “medium”). Постобработка вероятностных прогнозов, предложенная в п. 5.2, позволяет значительно улучшить калибровку доверительных интервалов прогнозов фотометрических красных смещений рентгеновских источников.

6. ВЫВОДЫ

В настоящей работе рассматривается популяция точечных рентгеновских источников двухлетнего

обзора СРГ/еРОЗИТА в области 1%-го спектроскопического обзора DESI в восточной галактической полусфере. Мы исследуем фотометрические и спектральные измерения классов и красных смещений рентгеновских объектов в трех выборках с ограничением по рентгеновскому потоку: $F_{X,0.5-2} \geq 4 \times 10^{-14}$ эрг с⁻¹ см⁻² (яркая), $F_{X,0.5-2} = (1.5 - 4) \times 10^{-14}$ эрг с⁻¹ см⁻² (средняя), $F_{X,0.5-2} = (0.6 - 1.5) \times 10^{-14}$ эрг с⁻¹ см⁻² (слабая) с общей площадью 91.4, 91.4 и 16.62 кв. градусов соответственно. Представленные рентгеновские выборки сочетают в себе большую площадь (значительно превышающую площадь среднеформатных рентгеновских обзоров неба XMM–XXL, Stripe82X на потоках $F_{X,0.5-2} \geq 1.5 \times 10^{-14}$ эрг с⁻¹ см⁻² и сравнимую с рентгеновским обзором eFEDS) и рекордно высокую полноту спектроскопических измерений оптических компаньонов. Мы приводим общую информацию о популяциях рентгеновских источников, их фотометрической классификации и измерении красных смещений системой SRGz в сравнении со спектральными/астрометрическими измерениями DESI EDR, SDSS, HELP, GAIA.

Нами отмечается высокая точность фотометрических красных смещений полученных моделями SRGz для рентгеновских источников в двухлетнем обзоре неба еРОЗИТА (в области покрытия фотометрических обзоров DESI Legacy Imaging Surveys/Pan-STARRS1/SDSS): стандартное отклонение $\sigma_{NMAD} \approx 4\%$ и доля выбросов на уровне $n_{>0.15} = 7-8.5\%$ (для оптических компаньонов рентгеновских источников с $F_{X,0.5-2} \geq 1.5 \times 10^{-14}$ эрг с⁻¹ см⁻²). Прогнозы photo-z рентгеновских источников в SRGz имеют негауссовый характер; качество калибровки $PDF(z)$ важно для точной оценки доверительных интервалов прогноза красных смещений.

Нами предложен метод постобработки вероятностных прогнозов фотометрических красных смещений внегалактических объектов, основанный на алгоритме двух-температурного масштабирования вероятностей. Применение метода позволяет уточнить калибровку доверительных интервалов красных смещений источников, не оказывая влияния на полученные наиболее вероятные значения их photo-z. Мы используем предложенный подход для калибровки вероятностных прогнозов фотометрических красных смещений точечных рентгеновских источников еРОЗИТА, полученных моделью SRGz. Следует отметить, что данный метод также может быть использован для калибровки $PDF(z)$ астрономических объектов полученных различными моделями photo-z (как на основе шаблонов, так и на основе машинного обучения).

Настоящее исследование основано на наблюдениях телескопа еРОЗИТА на борту обсерватории СРГ. Обсерватория СРГ изготовлена Роскосмосом в интересах Российской академии наук в лице

Института космических исследований (ИКИ) в рамках Российской федеральной научной программы с участием Германского центра авиации и космонавтики (DLR). Рентгеновский телескоп СРГ изготовлен консорциумом германских институтов во главе с Институтом внеземной астрофизики Общества им. Макса Планка (MPE) при поддержке DLR. Космический аппарат СРГ спроектирован, изготовлен, запущен и управляется НПО им. Лавочкина и его субподрядчиками. Прием научных данных осуществляется комплексом антенн дальней космической связи в Медвежьих озерах, Уссурийске и Байконуре и финансируется Роскосмосом. Использованные в настоящей работе данные телескопа еРозита обработаны с помощью программного обеспечения eSASS, разработанного германским консорциумом еРОЗИТА, и программного обеспечения, разработанного российским консорциумом телескопа СРГ/еРОЗИТА. Система SRGz создана в научной рабочей группе по поиску рентгеновских источников, их отождествлению и составлению каталога по данным телескопа СРГ/еРОЗИТА в отделе астрофизики высоких энергий ИКИ РАН.

ФИНАНСИРОВАНИЕ

Работа выполнена при поддержке гранта РНФ 21-12-00343.

СПИСОК ЛИТЕРАТУРЫ

1. DESI коллаборация и др. (DESI Collaboration, G. Adame, J. Aguilar, S. Ahlen, S. Alam, G. Aldering, et al.), arXiv e-prints, p. arXiv:2306.06308 (2023).
2. Алмейда и др. (A. Almeida, S.F. Anderson, M. Argudo-Fernández, C. Badenes, K. Barger, J.K. Barrera-Ballesteros, et al.), *Astrophys. J. Suppl. Ser.* **267**, 44 (2023).
3. Ананна и др. (T.T. Ananna, M. Salvato, S. LaMassa, C.M. Urry, N. Cappelluti, C. Cardamone, et al.), *Astrophys. J.* **850**, 66 (2017).
4. Брейман (L. Breiman), *Machine Learn.* **45**, 5 (2001).
5. Брешиа и др. (M. Brescia, M. Salvato, S. Cavuoti, T.T. Ananna, G. Riccio, S.M. LaMassa, et al.), *MNRAS* **489**, 663 (2019).
6. Буренин Р.А., Письма в Астрон. журн. **48**, 167 (2022) [R.A. Burenin, *Astron. Lett.* **48**, 153 (2022)].
7. Ванг (C. Wang), arXiv e-prints, p. arXiv:2308.01222 (2023).
8. Ву, Шен (Q. Wu and Y. Shen), *Astrophys. J. Suppl. Ser.* **263**, 42 (2022).
9. Гомес и др. (Z. Gomes, M.J. Jarvis, I.A. Almosallam, and S.J. Roberts), *MNRAS* **475**, 331 (2018).
10. Д'Инсанта, Полстерер (A. D'Isanto and K.L. Polsterer), *Astron. Astrophys.* **609**, A111 (2018).

11. Дей и др. (A. Dey, D.J. Schlegel, D. Lang, R. Blum, K. Burleigh, X. Fan, et al.), *Astron. J.* **157**, 168 (2019).
12. Динардо и др. (J. Dinardo, N. Fortin, and T. Lémieux), *Econometrica* **64**, 1001 (1996).
13. Дэвид (A.P. Dawid), *J. Royal Statistic. Soc.: Ser. A (General)* **147**, 278 (1984).
14. Карраско Кинд, Бруннер (M. Carrasco Kind and R.J. Brunner), *MNRAS* **432**, 1483 (2013).
15. Коллаборация Euclid и др. (Euclid Collaboration, G. Desprez, S. Paltani, J. Coupon, I. Almosallam, A. Alvarez-Ayllon, et al.), *Astron. Astrophys.* **644**, A31 (2020).
16. Лин и Джеон (Y. Lin and Y. Jeon), *J. Am. Statistic. Associat.* **101**, 578 (2006).
17. Люк и др. (B.W. Lyke, A.N. Higley, J.N. McLane, D.P. Schurhammer, A.D. Myers, A.J. Ross, et al.), *Astrophys. J. Suppl. Ser.* **250**, 8 (2020).
18. Майнсхаузен (N. Meinshausen), *J. Machine Learn. Res.* **7**, 983 (2006).
19. Медведев П.С., Гильфанов М.Р., Сазонов С.Ю. и др., *Письма в Астрон. журн.* **49**, 35 (2023) [P.S. Medvedev, M.R. Gilfanov, S.Y. Sazonov, R.A. Sunyaev, and G.A. Khorunzhev, *Astron. Lett.* **48**, 735 (2022)].
20. Мейнзер и др. (A. Mainzer, J. Bauer, T. Grav, J. Masiero, R.M. Cutri, J. Dailey, et al.), *Astrophys. J.* **731**, 53 (2011).
21. Мещеряков А.В., Глазкова В.В., Герасимов С.В., Машечкин И.В., *Письма в Астрон. журн.* **44**, 801 (2018) [A.V. Meshcheryakov, V.V. Glazkova, S.V. Gerasimov, and I.V. Mashechkin, *Astron. Lett.* **44**, 735 (2018)].
22. Мещеряков А.В., Борисов В.Д., Хорунжев Г.А. и др., *Письма в Астрон. журн.* **49**, 441 (2023) [A.V. Meshcheryakov, V.D. Borisov, G.A. Khorunzhev, P.A. Medvedev, M.R. Gilfanov, M.I. Belvedersky, et al., *Astron. Lett.* **49**, 359 (2023)].
23. Мозафари и др. (A. Sadat Mozafari, H. Siqueira Gomes, W. Leao, S. Janny, and C. Gagn'e), *arXiv e-prints*, p. arXiv:1810.11586 (2018).
24. Ньюман, Груен (J.A. Newman and D. Gruen), *Ann. Rev. Astron. Astrophys.* **60**, 363 (2022).
25. Павлинский и др. (M. Pavlinsky, A. Tkachenko, V. Levin, N. Alexandrovich, V. Arefiev, V. Babyshkin, et al.), *Astron. Astrophys.* **650**, A42 (2021).
26. Платт (J. Platt), *Adv. Large Margin Classif.* **10** (2000).
27. Предель и др. (P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, et al.), *Astron. Astrophys.* **647**, A1 (2021).
28. Райт и др. (E.L. Wright, P.R.M. Eisenhardt, K. Mainzer, M.E. Ressler, R.M. Cutri, T. Jarrett, et al.), *Astron. J.* **140**, 1868 (2010).
29. Сюняев и др. (R. Sunyaev, V. Arefiev, V. Babyshkin, A. Bogomolov, K. Borisov, M. Buntov, et al.), *Astron. Astrophys.* **656**, A132 (2021).
30. Танг и др. (C. Tang, D. Garreau, and U. von Luxburg), in S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Es.), *Adv. in Neural Information Processing Systems*, Vol. 31, Curran Associates, Inc. (2018).
31. Чамберс и др. (K.C. Chambers, E.A. Magnier, N. Metcalfe, H.A. Flewelling, M.E. Huber, C.Z. Waters, et al.), *arXiv e-prints*, p. arXiv:1612.05560 (2016).
32. Ширли и др. (R. Shirley, K. Duncan, M.C. Campos Varillas, P.D. Hurley, K. Malek, Y. Roehlly, et al.), *MNRAS* **507**, 129 (2021).
33. Шмидт и др. (S.J. Schmidt, A.I. Malz, J.Y.H. Soo, I.A. Almosallam, M. Brescia, S. Cavaoti, et al.), *MNRAS* **499**, 1587 (2020).
34. Шоссидон и др. (E. Chaussidon, C. Yèche, N. Palanque-Delabrouille, D.M. Alexander, J. Yang, S. Ahlen, et al.), *Astrophys. J.* **944**, 107 (2023).