

SRGz: ПОСТРОЕНИЕ МОДЕЛИ ОПТИЧЕСКОГО ОТОЖДЕСТВЛЕНИЯ РЕНТГЕНОВСКИХ ИСТОЧНИКОВ СРГ/ЕРОЗИТА НА ПРИМЕРЕ ДАННЫХ ОБЛАСТИ ДЫРЫ ЛОКМАНА

© 2022 г. М. И. Бельведерский^{1,2*}, А. В. Мещеряков^{1,3},
П. С. Медведев¹, М. Р. Гильфанов^{1,4}

¹Институт космических исследований РАН, Москва, Россия

²Национальный исследовательский университет “Высшая школа экономики”, Москва, Россия

³Московский государственный университет им. М.В. Ломоносова, Москва, Россия

⁴Институт астрофизики общества им. Макса Планка, Гархинг, Германия

Поступила в редакцию 03.12.2021 г.

После доработки 28.12.2021 г.; принята к публикации 28.12.2021 г.

Представлена вероятностная модель оптического отождествления рентгеновских источников СРГ/еРОЗИТА в области Дыры Локмана по данным фотометрического обзора DESI Legacy Imaging Surveys, которая учитывает как позиционную, так и фотометрическую информацию об оптических объектах в окрестности рентгеновских источников и позволяет производить отбор оптических компаньонов с точностью и полнотой на уровне $\approx 94\%$ (для $F_{X,0.5-2} > 10^{-14}$ эрг/с/см²). На основе данных модели отождествления мы откалибровали ошибку локализации при детектировании рентгеновских источников СРГ/еРОЗИТА в поле Дыры Локмана: $\sigma_{\text{corr}} = 0.87 \sqrt{\sigma_{\text{det}}^{2.53} + 1.12^2}$. Предложенная модель оптического отождествления в дальнейшем будет использована в составе системы SRGz анализа данных рентгеновского обзора всего неба СРГ/еРОЗИТА.

Ключевые слова: СРГ, еРОЗИТА, обзоры неба, оптическое отождествление, SDSS, DESI Legacy Surveys.

DOI: 10.31857/S0320010822020012

1. ВВЕДЕНИЕ

Рентгеновская обсерватория “Спектр-РГ” (СРГ) была запущена 13 июля 2019 г. и в настоящий момент находится на гало-орбите вокруг точки либрации L2 системы Земля — Солнце (Сюняев и др., 2021). Главная задача обсерватории — проведение обзора всего неба продолжительностью четыре года в диапазоне энергий 0.2–30 кэВ. На борту СРГ находятся два рентгеновских телескопа — еРОЗИТА (Предел и др., 2021) и ART-XC им. М.Н. Павлинского (Павлинский и др., 2021).

Для подавляющего большинства рентгеновских источников в обзоре СРГ/еРОЗИТА мы располагаем только весьма ограниченной информацией — это их положение на небе (в рамках ошибки локализации) и рентгеновский поток. Необходимая дополнительная информация об их физической природе, как правило, может быть получена из

анализа оптического спектра компаньона рентгеновского источника. Для проведения оптической спектроскопии рентгеновский источник необходимо соотнести с оптическим объектом в фотометрических обзорах неба, т.е. выполнить его оптическое отождествление. Отождествление рентгеновских источников является нетривиальной задачей, поскольку в области их локализации, как правило, оказывается более одного оптического объекта. Эта проблема становится особенно серьезной в плоскости Галактики, где наблюдается высокая плотность оптических объектов. В связи с этим необходима модель отождествления, которая позволит для каждого оптического объекта в области локализации рентгеновского источника оценить вероятность являться компаньоном (p_{match}) или позволит определить, что рентгеновский источник не имеет оптического компаньона в заданном фотометрическом обзоре (p_{\emptyset}).

К модели оптического отождествления могут быть выдвинуты следующие требования:

*Электронный адрес: mike.belveder@gmail.com

1. Прогноз компаньона моделью отождествления должен быть основан как на позиционной, так и на фотометрической информации об оптических источниках в поле рентгеновского объекта.
2. Оптический компаньон рентгеновского источника может отсутствовать в фотометрическом каталоге. В таких случаях модель должна сигнализировать об отсутствии компаньона высоким значением величины p_{\emptyset} .
3. При отождествлении рентгеновских объектов нас могут интересовать два принципиально разных режима: отбор объектов с высокой точностью (обычно выше 80%) и отбор объектов с высокой полнотой (выше 90%). Примером первой категории задач является поиск уникальных объектов в рентгеновском обзоре, второй — отождествление квазаров для построения функции светимости (где важно свести к минимуму влияние эффектов селекции). Модель отождествления, описываемая в данной работе, должна позволить производить отбор рентгеновских источников в обоих режимах при помощи изменения порога по величине p_{\emptyset} .

Как правило, для отождествления астрономических каталогов применяется подход, представленный в статье (Сазерленд, Сандерс, 1992) и основанный на методе отношения правдоподобий (Likelihood Ratio, LR). Этот метод применяется и развивается, например, в работах (Оябу и др., 2005; Нейлор, Брус, 2013). Метод отношения правдоподобий был успешно реализован для поиска компаньонов рентгеновских источников в таких обзорах, как XMM-COSMOS (Бруса и др., 2007), Chandra-COSMOS (Чивано и др., 2012; Марчеси и др., 2019), STRIPE-82X (Ламасса и др., 2016; Ананна и др., 2019). В качестве альтернативы для метода отношения правдоподобий (и в качестве его обобщения) можно привести алгоритмы отождествления, основанные на байесовской статистике (Будавари и Лоредо, 2015; Пино и др., 2017; Сальвато и др., 2018).

В данной работе мы представляем модель оптического отождествления точечных рентгеновских источников, открываемых телескопом SRG/eROZITA в области Дыры Локмана. Дыра Локмана (Lockman Hole, LH; Локман и др., 1986) представляет большой интерес для рентгеновских наблюдений внегалактических объектов, так как в направлении на эту область поглощение рентгеновского излучения межзвездным газом и пылью в нашей Галактике минимально (колонковая плотность нейтрального водорода на луче зрения

в Галактике составляет $N_H \approx 4.5 \times 10^{19} \text{ см}^{-2}$). Используя рентгеновские данные SRG/eROZITA и оптические из обзора DESI Legacy Imaging Surveys в этой уникальной области, мы протестировали подходы к оптическому отождествлению рентгеновских источников, которые затем будут использованы при обработке данных из обзора всего неба SRG/eROZITA.

Модели оптического отождествления точечных рентгеновских источников, которые описаны в этой статье, являются составной частью системы SRGz (Мещеряков и др., 2022), развиваемой в ИКИ РАН для интеллектуального анализа данных рентгеновского обзора SRG/eROZITA. Наша статья является первой в серии публикаций, представляющих результаты моделей и алгоритмов, предназначенных для отождествления рентгеновских источников SRG/eROZITA в многоволновых данных фотометрических обзоров неба и определения их физических характеристик методами машинного обучения.

Одной из основных характеристик модели оптического отождествления является зависимость ошибки локализации рентгеновского источника (σ) от правдоподобия его детектирования (\mathcal{L}); $\mathcal{L} = -\ln(p)$, где p — вероятность нулевой гипотезы о том, что наблюдаемый объект порожден фоновой компонентой на рентгеновском изображении. Мы будем измерять зависимость $\sigma(\mathcal{L})$ при построении модели отождествления на оптических данных в окрестности рентгеновских источников.

Стандартный алгоритм детектирования и характеристики точечных рентгеновских источников для каждого рентгеновского объекта, помимо определения его положения, вычисляет также ошибку локализации σ_{det} . Результаты построения модели оптического отождествления могут быть использованы для независимой калибровки ошибки локализации точечных рентгеновских источников.

Статья построена следующим образом. В разделе 2 описаны использованные рентгеновские и оптические данные, а также процедура получения тестовой выборки, на которой мы будем оценивать точность модели отождествления. В разделе 3 описана модель оптического отождествления рентгеновских источников SRG/eROZITA в области Дыры Локмана. В разделе 4 представлены основные результаты работы. В последнем разделе приведено заключение и сделаны выводы.

2. ДАННЫЕ

2.1. Рентгеновские данные

Наблюдения области Дыры Локмана являлись частью программы верификационных наблюдений телескопа SRG/eROZITA и были выполнены в

октябре 2019 г. Наблюдения производились в режиме растрового сканирования. Этот режим имеет заметные преимущества по сравнению со стандартным для большинства современных рентгеновских обсерваторий мозаичным режимом наблюдения, при котором поле обзора покрывается отдельными точечными наведениями. Режим растрового сканирования позволяет получать рентгеновские изображения участков неба большой площади с примерно постоянными по изображению чувствительностью и функцией отклика на точечный источник (Сюняев и др., 2021).

Схема растрового сканирования поля Дыры Локмана состояла из параллельных сканирований в чередующихся направлениях (с ориентацией, примерно соответствующей эклиптической системе координат) со смещением на 11.7 угл. мин. Скорость сканирования составила 9.1 угл. сек (что примерно на порядок меньше скорости сканирования в режиме обзора неба). Поле зрения телескопа SRG/ePOZITA составляет 1 градус, таким образом, каждый источник непрерывно наблюдался в течение ~ 6.6 мин. Область покрытия обзора LH составила ≈ 18.5 кв. град ($5^\circ \times 3.7^\circ$) с координатами центра $\alpha = 10^{\text{h}}35^{\text{m}}$ и $\delta = +57^\circ 38'$. Общая продолжительность обзора составила 180 кс, а среднее время экспозиции — около 8 кс на точку. Это позволило достигнуть глубины по потоку $\approx 3 \times 10^{-15}$ эрг/с/см² в диапазоне 0.5–2 кэВ.

Первичная обработка данных телескопа ePOZITA проводилась с помощью программного обеспечения, разработанного в ИКИ РАН с использованием компонент системы eSASS, созданной в Институте внеземной физики Общества им. Макса Планка (Германия). Регистрация источников проводилась с помощью аппроксимации распределения отсчетов с использованием функции отклика телескопа (программа `ermldet` системы eSASS). Всего было обнаружено 8309 рентгеновских источников (с правдоподобием рентгеновского детектирования $\mathcal{L} > 6$). Средняя поверхностная плотность рентгеновских источников в обзоре Дыры Локмана составила примерно ~ 370 ист./кв. град, что сравнимо с плотностями источников в известных рентгеновских обзорах обсерваторий Chandra и XMM-Newton с близкими к обзору Дыры Локмана характеристиками: XBootes (9.3 кв. град, Мюррей, 2015) и XMM-XXL-North (18.5 кв. град, Мензель и др., 2016).

Примерно 20% площади обзора Дыры Локмана было ранее исследовано обсерваториями ROSAT, Chandra и XMM-Newton. Здесь мы будем использовать данные наблюдений Chandra и XMM-Newton, представленные в каталогах CSC 2.0 (Эванс и др., 2010) и 4XMM DR10 (Вебб и др., 2020), для исследования точности отождествления

рентгеновских источников ePOZITA (как это описано ниже в подразделе 2.3).

2.2. Оптические данные

Для отождествления с рентгеновскими источниками ePOZITA будем использовать открытые данные двух фотометрических обзоров неба в оптическом диапазоне — DESI Legacy Imaging Surveys DR8 (сокр. DESI LIS, Дей и др., 2019) и SDSS DR14 (далее SDSS, Аболфати и др., 2018).

2.3. Тестовый каталог оптических отождествлений рентгеновских источников в области Дыры Локмана

Для исследования качества модели отождествления рентгеновских источников в области Дыры Локмана мы подготовили специальную тестовую выборку из источников, ранее наблюдавшихся в режиме точечных наведений обсерваториями Chandra и XMM-Newton. Мы получили несколько сотен источников ePOZITA, у которых оптический компаньон можно надежно определить по положению рентгеновских объектов из каталогов CSC 2.0 и 4XMM DR10 (далее — “тестовая выборка с компаньонами”). Для моделирования оптического окружения рентгеновских источников, у которых оптический компаньон отсутствует в фотометрическом обзоре DESI LIS, мы использовали оптические поля из тестовой выборки с компаньонами, предварительно удалив из них все оптические компаньоны.

Далее будет более подробно описана процедура получения тестовой выборки рентгеновских источников с оптическими компаньонами.

1. Всего в области Дыры Локмана обнаружено 8309 точечных источников SRG/ePOZITA. Сначала из них были отобраны такие, в $30''$ от которых присутствует либо единственный источник CSC (585 штук), либо единственный источник 4XMM (788). Из этих источников CSC и 4XMM были отобраны такие, на расстоянии $r < r_{\text{false}}$ от которых находится единственный объект DESI LIS. Радиус поиска r_{false} вычислялся как $r_{\text{false}} = (-\ln[1 - f_{\text{thresh}}]\pi^{-1}\rho_{\text{desi}}^{-1})^{1/2} = 1.43''$, где $\rho_{\text{desi}} \approx 17 \text{ arcmin}^{-2}$ — средняя плотность оптических источников в обзоре DESI LIS в области LH, $f_{\text{thresh}} = 0.03$ — пороговое значение вероятности случайно обнаружить один или более источник в кружке $r < r_{\text{false}}$, согласно пуассоновскому распределению со средним значением $\lambda = \rho_{\text{desi}}$. Таким образом, мы гарантируем чистоту создаваемой

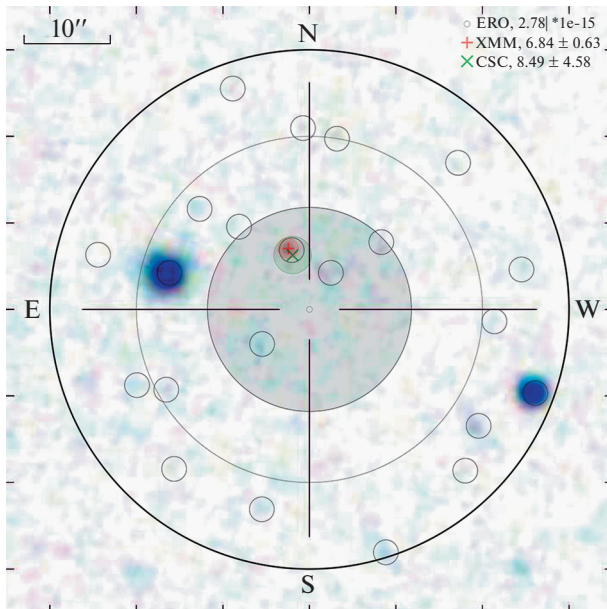


Рис. 1. Пример оптического поля рентгеновского источника СРГ/еРОЗИТА из тестовой выборки с компаньонами в области Дыры Локмана. Полупрозрачными кружками отмечены 95% области локализации для рентгеновских источников еРОЗИТА (большая серая область по центру рисунка), XMM-Newton (область вокруг знака “+”) и Chandra (область вокруг крестика). Данные XMM-Newton и Chandra однозначно указывают на оптический компаньон (подробнее см. текст).

тестовой выборки оптических компаньонов не хуже 97% (на практике чистота выборки компаньонов еще лучше, учитывая другие критерии отбора). Количество отобранных источников CSC на данном этапе составило 437, 4XMM — 473.

2. Затем в выборке были оставлены только такие источники CSC и 4XMM, рентгеновские потоки которых отличались не более чем в 5 раз от потоков источников еРОЗИТА (в поле которых они находятся). На этом этапе было отобрано в сумме 783 объекта (319 и 464 оптических компаньонов по данным CSC и 4XMM соответственно).
3. Далее мы отфильтровали поля, где, помимо единственного источника CSC, было обнаружено несколько источников 4XMM, и наоборот (сократив количество рентгеновских объектов до 577). На основе визуального просмотра оставшихся полей было принято решение убрать из выборки рентгеновские источники еРОЗИТА, в кружке с радиусом $40''$, от которых оказались объекты DESI LIS со звездными величинами в оптических

фильтрах r или z меньше 16. Это было сделано, чтобы в тестовую выборку не попали оптические поля DESI LIS, содержащие артефакты от лучей ярких звезд на оптическом изображении. Так была получена тестовая выборка, состоящая из 541 рентгеновского источника с надежным отождествлением оптического компаньона.

На рис. 1 показан пример оптического поля рентгеновского источника СРГ/еРОЗИТА из отобранной тестовой выборки компаньонов. Внутри окружности радиуса $R_{\text{match}} = 30''$ (показана сплошной жирной линией на рисунке) с центром на положении рентгеновского источника находится множество оптических объектов из каталога DESI LIS (показаны небольшими окружностями). Оптическое изображение на заднем плане рисунка было взято¹ из фотометрического обзора SDSS DR16. 95% область локализации рентгеновского источника СРГ/еРОЗИТА показана закрашенной серой областью в центре рисунка. На рисунке также показано положение рентгеновских источников из каталогов 4XMM DR10 и CSC 2.0 (знаком плюс и крестиком соответственно), и область их 95% локализации (небольшие закрашенные кружки). В легенде графика приведено значение рентгеновских потоков источников СРГ/еРОЗИТА, XMM-Newton и Chandra в эрг/с/см². Рентгеновские источники XMM-Newton и Chandra однозначно указывают на один оптический источник, который и является оптическим компаньоном рентгеновского объекта еРозита.

На рис. 2 показаны распределение по рентгеновскому потоку (0.5–2 кэВ) источников СРГ/еРОЗИТА в области Дыры Локмана и аналогичное распределение для тестовой выборки с компаньонами. Как можно видеть, форма распределения для тестовой выборки с компаньонами немного отличается от распределения по рентгеновскому потоку всех источников из области Дыры Локмана.

На основе выборки рентгеновских источников с надежно определенными оптическими компаньонами (описанной выше) был создан итоговый тестовый каталог Дыры Локмана. Тестовый каталог представляет собой набор всех точечных рентгеновских источников из области Дыры Локмана, которым при помощи процедуры бутстреппирования были поставлены в соответствие оптические поля из тестовой выборки компаньонов. Мы использовали зависимость вероятности наличия компаньона у рентгеновского источника от его рентгеновского потока (показанную на рис. 15, сплошная линия)

¹Мы использовали SciServer API для языка Python <https://github.com/sciserver/SciScript-Python>

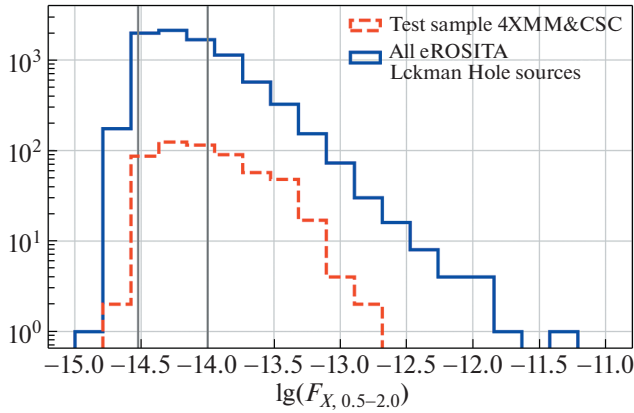


Рис. 2. Распределения по потоку всех рентгеновских источников SRG/eРОЗИТА в области Дыры Локмана (сплошная линия), распределение по потоку рентгеновских источников из тестовой выборки (штриховая линия). Вертикальными линиями отмечены пороговые значения по потоку 3×10^{-15} и 10^{-14} эрг/с/см² (соответствуют пороговой чувствительности области Дыры Локмана и четырехлетний обзор SRG/eРОЗИТА).

при случайном розыгрыше (“подбрасывании монетки” с вероятностью “орла”, равной $p_c(F_{X,0.5-2})$, см. подраздел 3.2) факта наличия оптического компаньона для каждого рентгеновского источника в выборке Дыры Локмана. Если розыгрыш показывал наличие у рентгеновского источника оптического компаньона, то оптическое поле с компаньоном случайно отбиралось из числа полей тестовой выборки с компаньонами с близким рентгеновским потоком (выборка с возвратом). Для каждого рентгеновского источника Дыры Локмана, у которого розыгрыш показывал отсутствие компаньона, оптическое поле случайно отбиралось из набора всех оптических полей тестовой выборки (независимо от рентгеновского потока), после удаления из них объектов, являющихся оптическими компаньонами.

3. МОДЕЛЬ

Модель оптического отождествления рентгеновского источника SRG/eРОЗИТА должна позволить определить, присутствует ли в окружении заданного рентгеновского источника оптический компаньон, а также выбрать для него наиболее вероятного оптического кандидата. Ложные оптические кандидаты и компаньон принадлежат различным порождающим распределениям как по угловому расстоянию от рентгеновского источника (см. подраздел 3.2), так и по звездным величинам (см. рис. 4, подраздел 3.4). Фотометрическая информация в предложенной нами модели отождествления содержится в эффективных величинах для оптических объектов (см. подраздел 3.4).

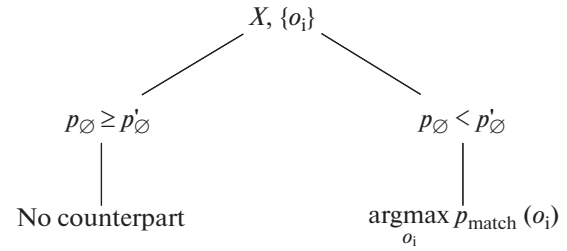


Рис. 3. Дерево поиска компаньона. На первом уровне в зависимости от порогового значения p'_\emptyset параметра p_\emptyset принимается решение о том, присутствует ли среди оптических кандидатов компаньон. В случае, если компаньон присутствует, на втором уровне для каждого кандидата вычисляется вероятность быть компаньоном p_{match} . В качестве компаньона выбирается кандидат с наибольшей вероятностью p_{match} .

Для каждого рентгеновского источника (X) будет рассчитана величина p_\emptyset — вероятность того, что заданное множество оптических кандидатов $\{o_i\}$ в его окружении *не* содержит компаньона источника X . Для каждого оптического объекта o_i будет найдена величина $p_{\text{match},i}$ — вероятность являться оптическим компаньоном выбранного рентгеновского источника (принимая во внимание расстояние до рентгеновского источника и фотометрическую информацию).

На рис. 3 показано дерево поиска оптического компаньона рентгеновского источника. Предлагаемый алгоритм поиска оптического компаньона рентгеновского источника SRG/eРОЗИТА состоит из двух уровней. На первом уровне на основе характеристик рентгеновского источника и его оптического окружения вычисляется величина p_\emptyset . Далее, в зависимости от выбранного порога p'_\emptyset , делается вывод, имеет ли заданный рентгеновский источник оптический компаньон. Если компаньон должен присутствовать (при $p_\emptyset < p'_\emptyset$), то на втором уровне модели для всех оптических объектов в поле вычисляются вероятности p_{match} и происходит выбор наиболее вероятного оптического кандидата.

В подразделе 3.1 будут введены эффективные звездные величины для оптических объектов и показано, как с их помощью может быть учтена фотометрическая информация об оптических кандидатах в поле рентгеновского источника. В подразделе 3.2 подробно описана модель позиционного отождествления. В подразделе 3.4 представлена модель отождествления объектов с учетом фотометрической информации. В подразделе 3.5 приведены выражения для вероятности оптического кандидата быть компаньоном рентгеновского источника (p_{match}) и вероятности отсутствия компаньона (p_\emptyset) у рентгеновского источника.

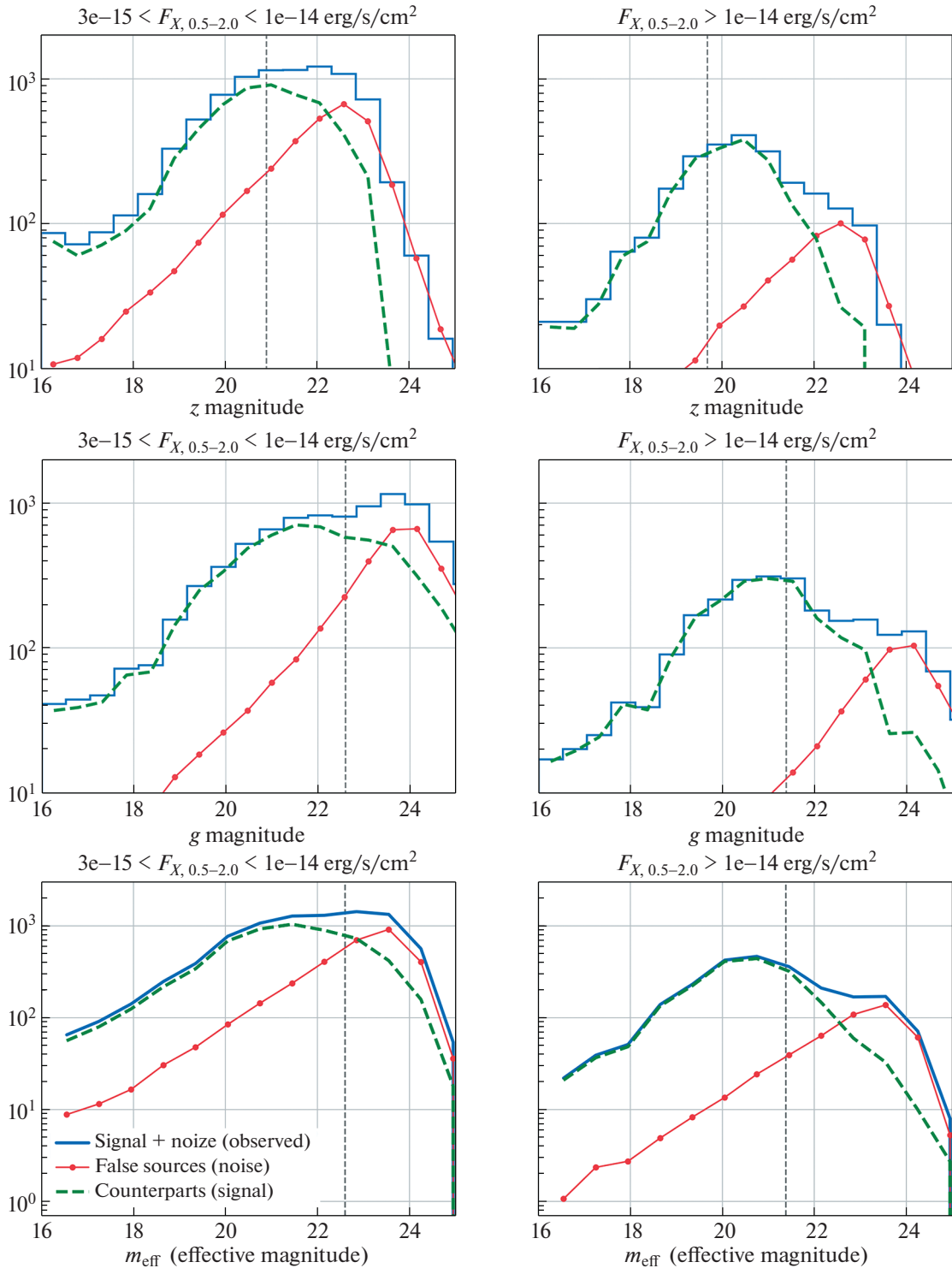


Рис. 4. Распределения всех оптических объектов, найденных внутри радиуса 2σ от рентгеновских источников в области Дыры Локмана с потоком $3 \times 10^{-15} < F_{X,0.5-2} < 10^{-14}$ (левый столбец графиков) и $F_{X,0.5-2} > 10^{-14}$ эрг/с/см² (правый столбец). Подробнее см. текст.

3.1. Эффективные звездные величины для оптических объектов

Выше было сказано, что оптические источники поля и оптические компаньоны точечных рентге-

новских объектов принадлежат различным порождающим распределениям не только по угловому расстоянию от рентгеновского источника, но и по своим фотометрическим признакам (в силу извест-

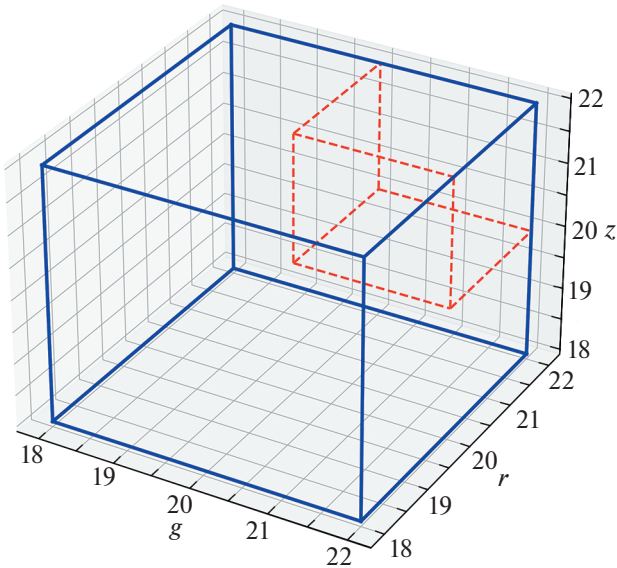


Рис. 5. Графическое представление эффективной звездной величины m_{eff} . На ребрах куба показаны звездные величины в фильтрах g , r и z . Объекты с заданной m_{eff} лежат на трех гранях куба, ограниченных штриховыми линиями на рисунке, и их продолжениях в сторону больших звездных величин.

ной корреляции между оптическим и рентгеновским потоком у квазаров).

Для иллюстрации на рис. 4 мы построили распределения всех оптических объектов, найденных внутри радиуса 2σ от источников в области Дыры Локмана, с различным рентгеновским потоком: $3 \times 10^{-15} < F_{X,0.5-2} < 10^{-14}$ (левый столбец графиков) и $F_{X,0.5-2} > 10^{-14}$ эрг/с/см² (правый столбец). Сплошная линия — наблюдаемое распределение всех оптических источников, вклад оптических объектов поля и компаньонов показан сплошной линией с точками и штриховой линией соответственно. Число объектов поля измерялось в кольце $20''-30''$ центра рентгеновских источников; кривая для компаньонов на рисунках получена путем вычитания из общего числа источников количества источников поля (с учетом площадей кружка радиуса 2σ и кольца $20''-30''$). На рисунках хорошо видно, что с увеличением рентгеновского потока распределения оптических компаньонов и оптических объектов, не имеющих отношения к рентгеновским источникам, начинают все более заметно отличаться. На верхней панели рис. 4 показаны распределения по звездной величине в красном фильтре z , на средней панели — в голубом фильтре g . На нижней панели рис. 4 показаны распределения оптических объектов по эффективной звездной величине (см. далее). Вертикальной штриховой линией на нижней панели показан 85%

квантиль распределения компаньонов рентгеновских источников по эффективной звездной величине ($m_{\text{eff}} = 22.6$), на панелях для фильтров z и g вертикальными штриховыми линиями отмечены значения, полученные через преобразование (1) для $m_{\text{eff}} = 22.6$.

Для учета фотометрической информации об оптических объектах мы введем эффективные звездные величины, которые будут выражаться относительно трех оптических фильтров, использующихся в фотометрическом обзоре DESI LIS:

$$m_{\text{eff}} = (m_g + c_g)|(m_r + c_r)|(m_z + c_z), \quad (1)$$

где c_g , c_r , c_z — выбранные константы, величины в фильтрах g , r , z заданы в АВ-системе. Графическое представление эффективной звездной величины m_{eff} приведено на рис. 5. Если оптический объект имеет эффективную звездную величину m_{eff} , то его АВ-величина в фильтре g равна $(m_g + c_g)$, либо его величина в фильтре r равна $(m_r + c_r)$, либо его величина в фильтре z равна $(m_z + c_z)$. Идея эффективной звездной величины (для $c_g = 0$, $c_r = 0$, $c_z = 0$) состоит в том, чтобы задать уровень спектральной плотности потока, который может достигаться объектом в любом из рассматриваемых оптических фильтров. В фотометрических обзорах (таких как SDSS, DESI LIS, Pan-STARRS) предельная величина в красных оптических фильтрах (z), как правило, заметно выше чем в голубых (g). Поэтому коэффициенты c_g , c_r , c_z можно определить таким образом, чтобы они соответствовали близкому отношению сигнал/шум в оптических фильтрах. Мы выбрали $c_g = 0$, $c_r = 0$, $c_z = -1.7$ (что соответствует разнице в предельных величинах для фотометрических данных SDSS, Аболфати и др. 2018). В общем случае величины c_g , c_r , c_z можно рассматривать как гиперпараметры (аналогично моделям машинного обучения) и обучать их значения, оптимальные с точки зрения точности оптического отождествления рентгеновских источников.

Необходимо отметить, что эффективная звездная величина вводится для того, чтобы свести всю фотометрическую информацию об оптическом объекте к одному параметру — m_{eff} . Таким образом, мы “загрубляем” фотометрическую информацию об оптическом объекте, но одновременно делаем модели отождествления более простыми, интерпретируемыми и устойчивыми к шуму и систематическим ошибкам в оптических данных. В будущем мы планируем также рассмотреть более сложные подходы на основе алгоритмов машинного обучения для более точного учета фотометрической информации об оптических объектах.

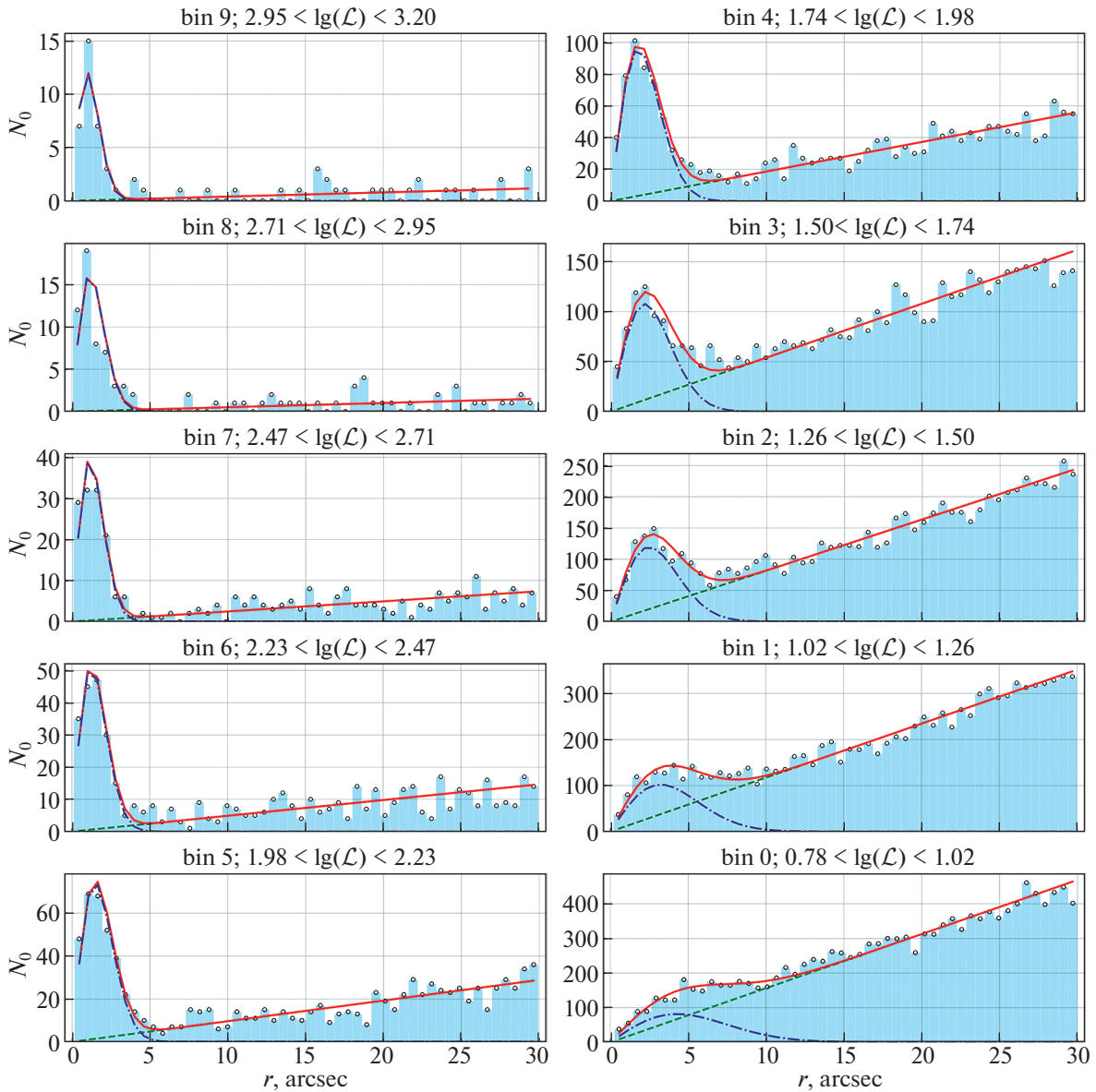


Рис. 6. Гистограммы распределения оптических кандидатов ($p_{\text{lim}} = 0.85$, см. подраздел 3.4) по угловому расстоянию от рентгеновского источника. Штрихпунктирной линией отмечено ожидаемое распределение компаньонов, пунктирной — ожидаемое распределение ложных источников. Сплошная линия обозначает ожидаемое совокупное распределение.

3.2. Позиционная модель

Оптический кандидат может принадлежать одному из двух распределений: распределению компаньонов или распределению ложных (фоновых) оптических источников.

Введем математическую модель оптического отождествления рентгеновских объектов, ограниченную следующими (стандартными для подобных моделей) предположениями (см., например, Буда-вари, Лоредо, 2015):

- Функция ошибки на положения рентгеновских источников описывается двумерным

нормальным распределением с ковариационной матрицей $\begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$. Тогда величина

на угловое расстояние (r) между направлением на истинное положение небесного объекта и направлением на рентгеновский источник при $r \ll \pi$ будет описываться распределением Рэлея:

$$\phi(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right). \quad (2)$$

- Ошибка локализации оптических источни-

ков DESI LIS пренебрежимо мала по сравнению с ошибками локализации рентгеновских источников SRG/eROZITA.

- Распределение оптических источников поля (не связанных с рентгеновскими объектами) для небольшой площадки в области Дыры Локмана имеет вид пуассоновского распределения с определенной поверхностной плотностью объектов (ρ).

Пусть в рентгеновском каталоге содержится N_X источников. Тогда среднее количество $N_o(r)$ всех оптических кандидатов в кружке радиуса r определяется следующим выражением:

$$N_o(r) = N_X \left(p_c \int_0^r \phi(r) 2\pi r dr + \rho \pi r^2 \right) = \quad (3)$$

$$= N_X \left(p_c \left[1 - \exp\left(\frac{-r^2}{2\sigma^2}\right) \right] + \rho \pi r^2 \right),$$

где $p_c(F_{X,0.5-2})$ — вероятность наличия компаньона у рентгеновского источника с заданным рентгеновским потоком (и в заданном фотометрическом обзоре). Важно отметить различие между величинами p_c и p_\varnothing . Величина p_\varnothing представляет собой вероятность отсутствия компаньона в заданном оптическом каталоге с учетом его оптического окружения (т.е. конкретного набора оптических объектов в окрестности рентгеновского источника, подробнее см. подраздел 3.5).

Запишем теперь выражение для среднего количества кандидатов в кольце i , ограниченном окружностями с радиусами r_i, r_{i+1} :

$$\lambda_i(r_i, r_{i+1}; \sigma, p_c, \rho) = N_o(r_{i+1}) - N_o(r_i). \quad (4)$$

Функция вероятности для распределения Пуассона:

$$\Pr(\lambda_i, k_i) = \frac{e^{-\lambda_i} \lambda_i^{k_i}}{k_i!} \quad (5)$$

и соответствующая ей функция правдоподобия:

$$\mathcal{Z} = \sum_i k_i \ln \lambda_i - \lambda_i, \quad (6)$$

где k_i — наблюдаемое количество оптических объектов в кольце i . Параметры σ, p_c, ρ модели (3) находятся путем оптимизации функции максимального правдоподобия (6) для данных по рентгеновским источникам в выбранном интервале правдоподобия рентгеновского детектирования \mathcal{L} и эффективных величин оптических объектов m_{eff} . Чтобы учесть фотометрическую информацию, мы построили семейство моделей отождествления для оптических объектов, отфильтрованных с разной предельной

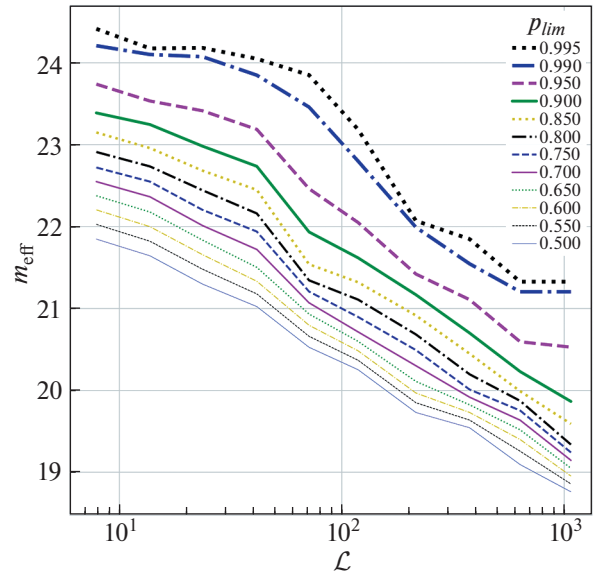


Рис. 7. Зависимость эффективной звездной величины от правдоподобия рентгеновского детектирования $m_{\text{eff}}(\mathcal{L})$. Показаны кривые для различных значений выбора максимального квантиля распределения оптических компаньонов (p_{lim}).

эффективной величиной (см. подраздел 3.4). При этом сами значения m_{eff} для указанных моделей были выбраны по измеренным квантилям распределения оптических компаньонов для каждого интервала по \mathcal{L} .

3.3. Зависимость параметров модели отождествления от правдоподобия детектирования и потока рентгеновских источников

Здесь в целях иллюстрации метода мы покажем результаты аппроксимации данных моделью оптического отождествления при выборе предельной эффективной звездной величины m_{eff} на уровне $p_{\text{lim}} = 0.85$ квантиля распределения оптических компаньонов. В каждом интервале по правдоподобию рентгеновского детектирования мы определим свое значение m_{eff} , согласно выбранному p_{lim} . Также мы обсудим, как меняются параметры моделей для разных значений p_{lim} .

На рис. 6 для $p_{\text{lim}} = 0.85$ показаны распределения оптических объектов по угловому расстоянию ($R < 30''$) до рентгеновского источника в разных интервалах по правдоподобию детектирования \mathcal{L} . Также на графиках приведены наилучшие модели, которые, как можно видеть, довольно хорошо аппроксимируют данные. Дополнительная информация приведена в табл. 1.

На рис. 7 показаны зависимости эффективной звездной величины от правдоподобия рентгеновского детектирования для различных значений квантиля распределения оптических компаньонов.

Таблица 1. Границы по \mathcal{L} для каждого бина, а также медианные значения \mathcal{L} в бинах указаны в столбцах $\mathcal{L}_{\text{left}}$, $\mathcal{L}_{\text{right}}$ и $\tilde{\mathcal{L}}$. В столбце $\tilde{\sigma}_{\text{det}}$ приведены медианные значения ошибок σ_{det} . В столбцах σ , ρ , p_c представлены значения параметров модели с ошибками; \mathcal{Z} — значения функции правдоподобия, использованной в подразделе 3.2; m_{eff} — пороги по эффективной звездной величине; n_{xray} — количество рентгеновских источников в бинах; n_{optic} — суммарное количество оптических кандидатов из окружения этих источников ($30''$). Все значения приведены для $p_{\text{lim}} = 0.85$ (см. подраздел 3.4)

№	$\mathcal{L}_{\text{left}}$	$\mathcal{L}_{\text{right}}$	$\tilde{\mathcal{L}}$	$\tilde{\sigma}_{\text{det}}$	σ	ρ	p_c	m_{eff}	\mathcal{Z}	n_{optic}	n_{xray}
0	6.00	10.44	7.88	3.32	$4.28^{+0.22}_{-0.22}$	$9.11^{+0.06}_{-0.16}$	$0.59^{+0.03}_{-0.05}$	23.15	-59362.77	12796	1653
1	10.44	18.17	13.70	2.70	$3.22^{+0.11}_{-0.14}$	$7.84^{+0.05}_{-0.09}$	$0.63^{+0.03}_{-0.04}$	22.96	-42495.91	9762	1438
2	18.17	31.62	23.63	2.20	$2.43^{+0.06}_{-0.08}$	$6.42^{+0.05}_{-0.10}$	$0.66^{+0.02}_{-0.03}$	22.68	-27814.70	6950	1219
3	31.62	55.24	41.48	1.79	$2.18^{+0.07}_{-0.09}$	$5.55^{+0.07}_{-0.14}$	$0.70^{+0.02}_{-0.04}$	22.45	-16972.17	4716	933
4	55.24	96.50	70.98	1.45	$1.76^{+0.05}_{-0.07}$	$2.74^{+0.04}_{-0.11}$	$0.72^{+0.03}_{-0.04}$	21.54	-5088.33	1868	651
5	96.50	168.58	118.30	1.17	$1.45^{+0.05}_{-0.07}$	$2.35^{+0.07}_{-0.14}$	$0.77^{+0.04}_{-0.05}$	21.32	-2245.76	1025	394
6	168.58	294.49	216.14	0.95	$1.23^{+0.07}_{-0.07}$	$2.03^{+0.11}_{-0.18}$	$0.77^{+0.06}_{-0.06}$	20.92	-864.96	544	231
7	294.49	514.45	374.95	0.78	$1.10^{+0.06}_{-0.06}$	$1.62^{+0.09}_{-0.13}$	$0.84^{+0.08}_{-0.08}$	20.45	-360.85	307	146
8	514.45	898.70	626.75	0.62	$1.14^{+0.09}_{-0.10}$	$0.81^{+0.12}_{-0.16}$	$0.88^{+0.08}_{-0.12}$	19.99	-23.92	90	60
9	898.70	1569.94	1075.39	0.50	$0.96^{+0.12}_{-0.11}$	$0.81^{+0.17}_{-0.19}$	$0.86^{+0.08}_{-0.14}$	19.59	4.97	62	39

Для фиксированного p_{lim} кривая $m_{\text{eff}}(\mathcal{L})$ представляет собой монотонно убывающую зависимость. Кривые для $p_{\text{lim}} = 50$ и 99.5% квантилей в $\sim 4-5$ раз отличаются по плотности оптических объектов поля.

Поскольку время экспозиции СРГ/еРОЗИТА слабо менялось внутри поля Дыры Локмана, мы можем использовать следующую формулу для пересчета правдоподобия рентгеновского детектирования \mathcal{L} в рентгеновский поток (хорошо аппроксимирующую медианные значения потока в интервалах по \mathcal{L}): $\lg(F_{X,0.5-2}) = 0.66 \times \lg(\mathcal{L}) - 15.09$.

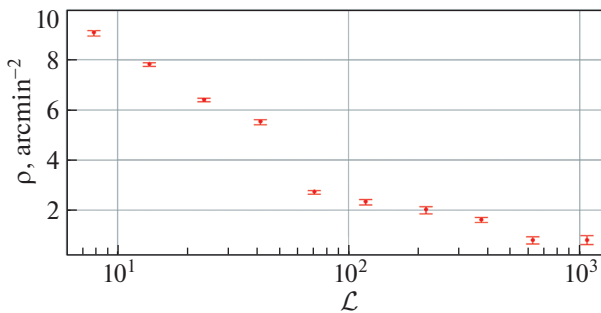


Рис. 8. Значения плотности ложных источников ρ (в угл. мин $^{-2}$) в разных бинах по \mathcal{L} для $p_{\text{lim}} = 0.85$ (см. подраздел 3.4).

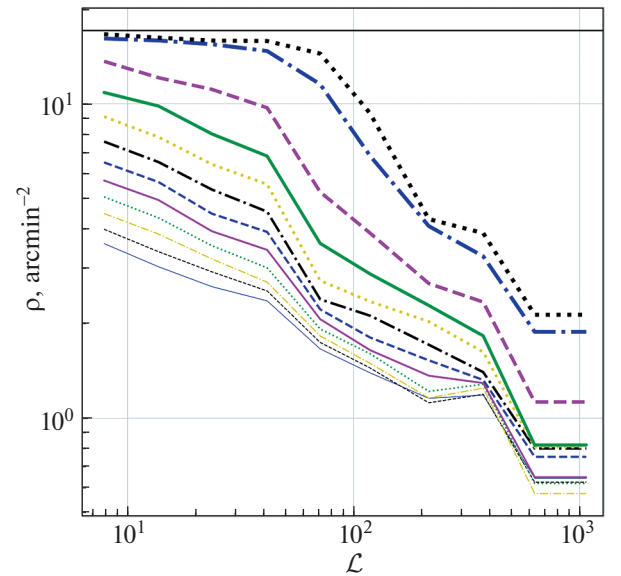


Рис. 9. Зависимость поверхностной плотности оптических объектов поля от правдоподобия рентгеновского детектирования — $\rho(\mathcal{L})$. Показаны кривые для различных значений выбора максимального квантиля распределения оптических компаньонов (p_{lim}). Горизонтальной сплошной линией отмечены значения плотности по всем источникам DESI LIS (легенду для кривых см. рис. 7).

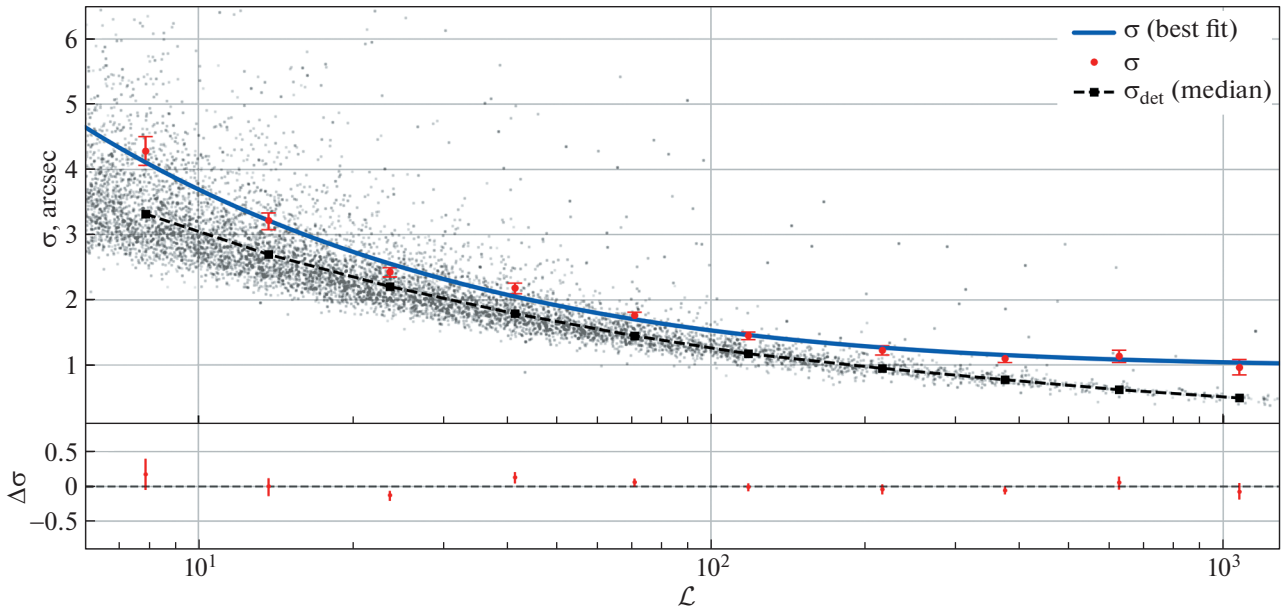


Рис. 10. Зависимость ошибки локализации рентгеновских источников Дыры Локмана от правдоподобия их детектирования. Сплошной линией (σ best fit) показана наилучшая параметрическая аппроксимация зависимости $\sigma(\mathcal{L})$, на нижней панели приведены ее невязки. Облаком точек на графике показаны ошибки локализации рентгеновских источников Дыры Локмана σ_{det} , измеренные алгоритмом детектирования источников. Квадратами, соединенными пунктирной линией (σ_{det} median), показаны их медианные значения в интервалах по \mathcal{L} .

На рис. 8, 10 и 12 показаны наилучшие значения параметров моделей отождествления в зависимости от правдоподобия рентгеновского детектирования (для ρ , σ) и от рентгеновского потока (для p_c). Мы также измерили ошибки на параметры ρ , σ , p_c в каждом интервале, воспользовавшись стандартным статистическим тестом отношения правдоподобия (англ. likelihood ratio test, см., например, Джеймс 2006). Мы варьировали значение выбранного параметра в окрестности оптимального значения функции правдоподобия ($\mathcal{Z}^{\text{best}}$) и вычисляли максимальное правдоподобие по остальным параметрам. Значения 68% доверительных интервалов по выбранному параметру соответствовали правдоподобию $\mathcal{Z}^{\text{best}} - \chi_{0.68}^2$.

3.3.1. Зависимость $\rho(\mathcal{L})$. Зависимость плотности оптических объектов поля (отобранных до пороговой эффективной звездной величины) от правдоподобия детектирования рентгеновских источников $\rho(\mathcal{L})$ показана рис. 8 для $p_{\text{lim}} = 0.85$. Она имеет характерную особенность: с увеличением правдоподобия детектирования рентгеновских источников от $\mathcal{L} = 6$ до $\mathcal{L} \approx 10^3$ плотность оптических объектов поля падает на порядок. Это происходит вследствие того, что более яркие рентгеновские источники, в среднем, имеют более яркие оптические компаньоны, а для каждого интервала по \mathcal{L} мы выбирали свою предельную эффективную звездную величину m_{eff} (соответствующую в данном случае

$p_{\text{lim}} = 85\%$ квантилю распределения оптических компаньонов).

На рис. 9 показаны зависимости плотности оптических источников поля для разных уровней квантиля распределения оптических компаньонов p_{lim} . Без фильтрации по эффективной звездной величине плотность оптических источников поля не зависит от правдоподобия детектирования рентгеновских объектов (горизонтальная линия на рисунке). Кривые $\rho(\mathcal{L})$ для $p_{\text{lim}} = 50$ и 99.5% квантилей отличаются в $\sim 4-5$ раз по плотности оптических объектов поля, что также отражает различие в пороговых значениях m_{eff} при отборе оптических объектов.

3.3.2. Зависимость $\sigma(\mathcal{L})$. На рис. 10 показана зависимость ошибки локализации рентгеновских источников от правдоподобия их детектирования $\sigma(\mathcal{L})$ (для $p_{\text{lim}} = 0.85$). Оптимальные значения параметра σ , найденные в интервалах по \mathcal{L} , показаны закрашенными кружками с ошибками. Точками на графике показаны значения ошибки локализации σ_{det} (измеряемой алгоритмом детектирования) для всех точечных рентгеновских источников в области Дыры Локмана. Как можно видеть, алгоритм детектирования точечных рентгеновских источников заметно занижает реальную ошибку локализации объекта и требует калибровки по оптическим данным в окрестности рентгеновских источников (см. подраздел 4.1).

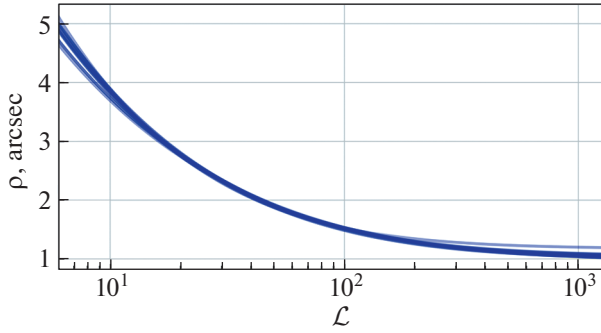


Рис. 11. Зависимость ошибки локализации объекта от правдоподобия рентгеновского детектирования — $\sigma(\mathcal{L})$. Показаны кривые, измеренные при различных значениях максимального квантиля распределения оптических компаньонов (p_{lim}). Из графика следует, что выбор p_{lim} не влияет на измерение зависимости $\sigma(\mathcal{L})$.

Измеренные нами ошибки локализации рентгеновских источников в интервалах по правдоподобию детектирования хорошо описываются следующей трехпараметрической зависимостью (для $p_{\text{lim}} = 0.85$, см. рис. 10):

$$\sigma(\mathcal{L}; a, b, \sigma_0) = \sqrt{a\mathcal{L}^b + \sigma_0^2}, \quad (7)$$

с оптимальными значениями параметров $a = 114.12 \pm 17.16$, $b = -0.96 \pm 0.04$, $\sigma_0 = 0.96 \pm 0.12$.

На рис. 11 показаны наилучшие аппроксимации зависимости ошибки локализации рентгеновских источников для разного выбора уровня квантиля распределения оптических компаньонов p_{lim} . Из графика видно, что выбор p_{lim} практически не влияет на измерение зависимости $\sigma(\mathcal{L})$.

3.3.3. Зависимость $p_c(F_{0.5-2})$. На рис. 12 (нижний график) показана зависимость вероятности наличия у рентгеновского источника оптического компаньона от рентгеновского потока $p_c(F_{X,0.5-2})$ для $p_{\text{lim}} = 0.85$. Точками на графике показаны измеренные значения параметра p_c для рентгеновских объектов в интервалах по \mathcal{L} .

Полученные нами ошибки на параметр p_c не являются гауссовыми, так как сами вероятности p_c лежат в диапазоне $[0, 1]$. Для аппроксимации данных $p_c(\mathcal{L})$ параметрической зависимостью, мы предварительно сделали преобразование значений p_c из интервала $[0, 1]$ на всю числовую ось $(-\infty, +\infty)$ с помощью простого логистического преобразования

$$\text{logit}(p_c) = \ln\left(\frac{p_c}{1-p_c}\right). \quad (8)$$

На рис. 12 (верхний график) показана зависимость $\text{logit}(p_c)$ от рентгеновского потока (для $p_{\text{lim}} = 0.85$). Ошибки на $\text{logit}(p_c)$ близки к гауссовым.

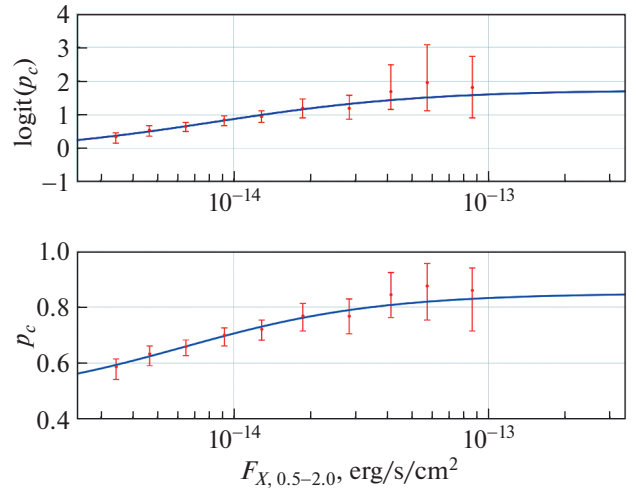


Рис. 12. Зависимость вероятности наличия оптических компаньонов в каталоге DESI LIS p_c от потока рентгеновских источников $F_{X,0.5-2}$ для $p_{\text{lim}} = 0.85$. На верхней панели изображены значения $\text{logit}(p_c)$, на нижней — p_c .

Мы аппроксимировали зависимость $\text{logit}(p_c)$ — $\log F_{X,0.5-2}$ следующей моделью:

$$\text{logit}(p_c) = p_1 + \frac{\text{logit}(p_{\text{lim}}) - p_1}{1 + \exp(-p_3[\lg(F_{X,0.5-2}) - p_2])}, \quad (9)$$

где p_1, p_2, p_3 — параметры логистической кривой, которые были найдены с помощью оптимизации по критерию χ^2 .

На рис. 12 сплошной линией (на верхней и нижней панели) показана наилучшая аппроксимация данных моделью (9). Отметим, что для перехода от значений $\text{logit}(p_c)$ в логистическом пространстве к вероятностям p_c мы использовали следующую формулу обратного логистического преобразования

$$p_c = \frac{\exp(\text{logit})}{1 + \exp(\text{logit})}. \quad (10)$$

На рис. 13 показаны наилучшие аппроксимации вероятности наличия оптического компаньона от рентгеновского потока источника $F_{X,0.5-2}$ для разного выбора уровня квантиля распределения оптических компаньонов p_{lim} . Видно, что с уменьшением p_{lim} в выборке оптических кандидатов остаются только все более яркие объекты, при этом уменьшается значение p_c .

3.4. Позиционно-фотометрическая модель

Выше было показано, как строятся позиционные модели оптического отождествления рентгеновских источников СРГ/еРОЗИТА с фильтрацией оптических объектов по эффективной звездной

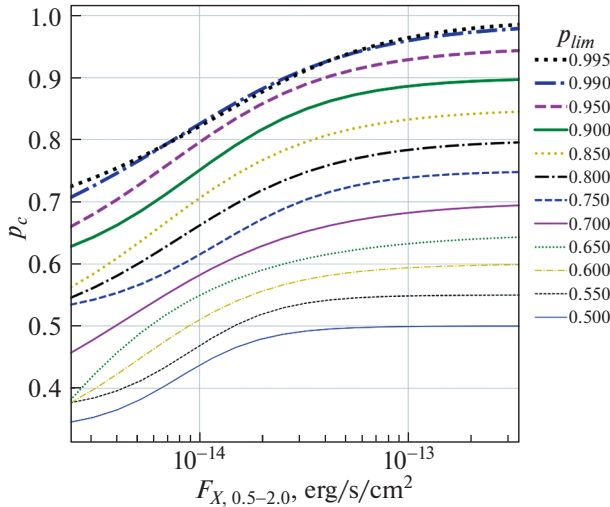


Рис. 13. Вероятность наличия у источника SRG/eРОЗИТА оптического компаньона как функция от рентгеновского потока — $p_c(F_{X,0.5-2})$. Показаны кривые для различных значений выбора максимального квантиля распределения оптических компаньонов (p_{lim}) (описание кривых см. рис. 7).

величине (соответствующей заданному квантилю распределения оптических компаньонов p_{lim}).

Для того чтобы учитывать при отождествлении объектов фотометрическую информацию, мы создали сетку из 12 позиционных моделей оптического отождествления для значений p_{lim} , 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.99, 0.995 и соответствующих им значений m_{eff} . Каждое значение m_{eff} для сетки моделей может быть представлено графически в пространстве звездных величин как объемная фигура с заданным положением граней (для иллюстрации см. рис. 5, область, ограниченная штриховыми линиями). Зависимости величины m_{eff} и параметров модели отождествления от вероятности рентгеновского детектирования (или рентгеновского потока) для 12 рассматриваемых значений p_{lim} показаны на рис. 7, 9, 11, 13.

Семейство моделей оптического отождествления для разных p_{lim} вместе с моделью, полученной на всех оптических данных (без фильтрации по m_{eff}), мы будем называть позиционно-фотометрической моделью отождествления.

3.5. Зависимости p_{match} и p_{\emptyset}

Здесь мы рассмотрим, как будут вычисляться вероятность отсутствия оптического компаньона (p_{\emptyset}) для рентгеновского объекта с заданным оптическим окружением и вероятность быть компаньоном (p_{match}) для каждого из оптических объектов.

Вероятность того, что выбранный оптический кандидат i , находящийся на расстоянии r_i от рентгеновского источника (с параметрами σ и $p_{c,i}^{eff}$) в

оптическом поле с плотностью источников ρ_i^{eff} , является компаньоном, определяется следующим выражением:

$$p_{match} = \frac{p_{c,i}^{eff} \exp\left(\frac{-r_i^2}{2\sigma^2}\right)}{p_{c,i}^{eff} \exp\left(\frac{-r_i^2}{2\sigma^2}\right) + 2\pi\rho_i^{eff}\sigma^2}. \quad (11)$$

Значения $p_{c,i}^{eff}$, ρ_i^{eff} вычислены здесь при помощи позиционно-фотометрической модели (см. подраздел 3.4). Для заданного кандидата из сетки позиционных моделей с разным значением p_{lim} подходящая модель выбирается с наиболее близким значением m_{eff} к эффективной звездной величине кандидата (при этом значение m_{eff} модели должно быть больше m_{eff} кандидата). Для кандидатов, эффективная звездная величина которых превышает значение m_{eff} (соответствующее максимальному выбранному нами значению $p_{lim} = 0.995$), были использованы параметры модели, полученной на всех оптических данных (без фильтрации по m_{eff}): p_c^{pos} , ρ_i^{pos} .

Вероятность отсутствия оптического компаньона для заданного рентгеновского источника и его оптического поля определяется как

$$p_{\emptyset} = \frac{1}{1 + \frac{1}{2\pi\sigma} \sum_{i=1}^{n_o} \exp\left(\frac{-r_i^2}{2\sigma^2}\right) \frac{p'_{c,i}}{\rho'_i}}, \quad (12)$$

где n_o — количество оптических кандидатов DESI LIS в окрестности источника SRG/eРОЗИТА (в нашем случае в $30''$), $p'_{c,i} = p_c^{pos} - \overline{p_{c,i}^{eff}}$, $\rho'_i = \rho_i^{pos} - \overline{\rho_i^{eff}}$. Параметры $\overline{p_{c,i}^{eff}}$, $\overline{\rho_i^{eff}}$ также как и $p_{c,i}^{eff}$, ρ_i^{eff} получены с помощью модели, m_{eff} которой имеет наименьшее отличие от m_{eff} кандидата, но при этом m_{eff} выбранной модели должен быть меньше m_{eff} кандидата.

3.6. Метрики оценки качества отождествления и отбора источников

Для оценки качества отождествления рентгеновских источников мы будем использовать метрику точности

$$\text{Precision}_X = \frac{\hat{N}_{c+h}^*}{N_X}, \quad (13)$$

где N_X — общее количество всех рентгеновских источников в выборке, $\hat{N}_{c+h}^* = \hat{N}_c^* + \hat{N}_h^*$ — количество рентгеновских источников, у которых оптический компаньон найден правильно (\hat{N}_c^*) или правильно указан факт отсутствия компаньона в выбранном фотометрическом обзоре (\hat{N}_h^*).

Для оценки качества отбора оптических компаньонов мы будем использовать следующие метрики:

- Полнота отбора компаньонов

$$\text{Recall}_c = \frac{\hat{N}_c^*}{N_c}, \quad (14)$$

где N_c — полное число рентгеновских источников с оптическими компаньонами в выборке.

- Точность отбора компаньонов

$$\text{Precision}_c = \frac{\hat{N}_c^*}{\hat{N}_c}, \quad (15)$$

где \hat{N}_c — число рентгеновских источников, отобранных алгоритмом отождествления как источники с компаньонами.

Для оценки качества отбора рентгеновских источников без оптических компаньонов (в обзоре DESI LIS) мы используем следующие метрики:

- Полнота отбора рентгеновских объектов без компаньонов

$$\text{Recall}_h = \frac{\hat{N}_h^*}{N_h}, \quad (16)$$

где N_h — полное число рентгеновских источников без оптических компаньонов в выборке.

- Точность отбора рентгеновских источников без оптических компаньонов

$$\text{Precision}_h = \frac{\hat{N}_h^*}{\hat{N}_h}, \quad (17)$$

где \hat{N}_h — число рентгеновских источников, отобранных алгоритмом отождествления как источники без компаньонов.

4. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В разделе представлены основные результаты работы.

В подразделе 4.1 приведена зависимость ошибки локализации рентгеновских источников от правдоподобия рентгеновского детектирования $\sigma(\mathcal{L})$. Зависимость была измерена при помощи информации об оптических объектах в окрестности рентгеновских источников СРГ/еРОЗИТА в поле LH. Анализируя найденную зависимость $\sigma(\mathcal{L})$ и усредненные данные об ошибках детектирования рентгеновских источников $\sigma_{\text{det}}(\mathcal{L})$, мы найдем калибровочную зависимость $\sigma(\sigma_{\text{det}})$ и сравним ее с результатами в литературе.

В подразделе 4.2 описана зависимость вероятности наличия у рентгеновского источника оптического компаньона (в фотометрических обзорах DESI LIS и SDSS в поле LH) от его рентгеновского потока — $p_c(F_{X,0.5-2})$.

В подразделе 4.3 показано, насколько позволяет улучшить процедуру поиска компаньонов привлечение фотометрической информации о кандидатах.

В подразделе 4.4 приведена информация о качестве оптического отождествления рентгеновских источников на всей выборке LH.

В подразделе 4.5 описаны результаты, которые позволяет достичь наша модель отождествления при поиске оптических компаньонов и отборе рентгеновских источников без компаньона.

4.1. Калибровка ошибки локализации точечных рентгеновских источников СРГ/еРОЗИТА

Одной из основных характеристик модели оптического отождествления является зависимость ошибки локализации рентгеновского источника от его правдоподобия детектирования $\sigma(\mathcal{L})$. Эта зависимость была измерена нами при поиске наилучших параметров модели отождествления в интервалах по \mathcal{L} на основе оптических данных DESI LIS в окрестности рентгеновских источников из области Дыры Локмана.

Нами была найдена калибровочная зависимость между ошибкой локализации точечных рентгеновских источников на основе данных об их оптических компаньонах (σ , в угловых секундах) и полученной алгоритмом детектирования σ_{det} (в рамках стандартного пайплайна eSASS). Мы определили медианные средние значения величины σ_{det} для рентгеновских источников в интервалах по \mathcal{L} (см. квадратные значки, соединенные штриховой линией на рис. 10). Полученная калибровочная зависимость $\sigma_{\text{corr}}(\sigma_{\text{det}})$ показана на рис. 14. Сплошной и штриховой линиями на графике показаны наилучшие (оптимизация χ^2) двух- и трехпараметрические калибровочные модели соответственно:

$$\sigma_{\text{corr},1} = a\sqrt{\sigma_{\text{det}}^2 + c^2}, \quad (18)$$

где $a = 1.11 \pm 0.03$, $c = 0.68 \pm 0.07$;

$$\sigma_{\text{corr},2} = a\sqrt{\sigma_{\text{det}}^b + c^2}, \quad (19)$$

где $a = 0.87 \pm 0.08$, $b = 2.53 \pm 0.20$, $c = 1.12 \pm 0.19$.

Обе модели одинаково хорошо описывают данные с учетом ошибок измерений. Пунктирной линией показана линия $\sigma_{\text{corr}} = \sigma_{\text{det}}$. Отметим, что ошибки локализации рентгеновских источников по

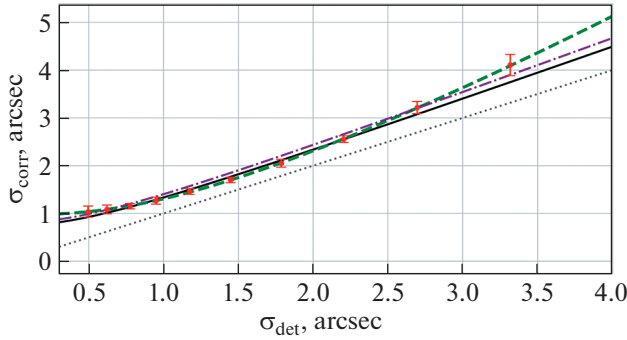


Рис. 14. Калибровочная зависимость между средней ошибкой локализации точечных рентгеновских источников, полученной на основе данных об их оптических компаньонах (σ_{corr}), и на этапе рентгеновского детектирования (σ_{det}). Сплошной и штриховой линиями на рисунке показаны наилучшие двух- и трехпараметрические модели соответственно. Штрихпунктирной линией показана зависимость из работы (Брюннер и др., 2021), измеренная по данным в поле eFEDS. Пунктирной линией показана линия $\sigma_{\text{corr}} = \sigma_{\text{det}}$.

данным об оптических компаньонах лежат в среднем выше ошибок, измеренных алгоритмом детектирования.

Для сравнения, штрихпунктирной линией на рис. 14 показана калибровочная зависимость из работы Брюннер и др. 2021 (измеренная по данным в поле eFEDS):

$$\sigma_{\text{corr,eFEDS}} = 1.15 \sqrt{\sigma_{\text{det}}^2 + 0.7^2}. \quad (20)$$

Можно сделать вывод, что полученные нами калибровочные зависимости $\sigma_{\text{corr}}(\sigma_{\text{det}})$ для области Дыры Локмана хорошо согласуются с калибровкой в поле eFEDS (Брюннер и др., 2021).

4.2. Зависимость $p_c(F_{X,0.5-2})$

При оптическом отождествлении рентгеновского источника важно учитывать, что оптический компаньон рентгеновского объекта может отсутствовать в выбранном фотометрическом каталоге. В этой работе мы измерили для области Дыры Локмана зависимость между вероятностью присутствия компаньона в обзорах DESI LIS и SDSS от рентгеновского потока объекта SRG/eROSITA (рис. 15, верхняя и нижняя панели соответственно). Наилучшие аппроксимации (см. формулы (9), (10)) измеренных зависимостей $p_c(F_{X,0.5-2})$ для оптических данных DESI LIS и SDSS показаны на рис. 15 сплошной и штриховой линией соответственно.

Как следует из рисунка, на рентгеновском потоке $F_{X,0.5-2} \approx 3 \times 10^{15}$ эрг/с/см² (что соответствует пороговой чувствительности в области Дыры Локмана) вероятность иметь компаньон p_c составляет

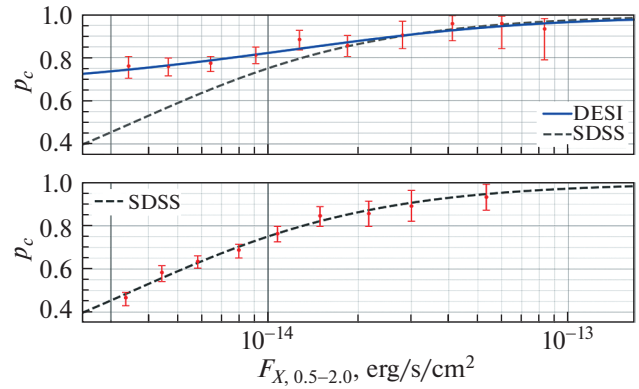


Рис. 15. Зависимость $p_c(F_{X,0.5-2})$. Сплошной линией показана зависимость для обзора DESI LIS, штриховой — для SDSS. Вертикальными линиями отмечены пороговые значения по потоку 3×10^{-15} и 10^{-14} эрг/с/см².

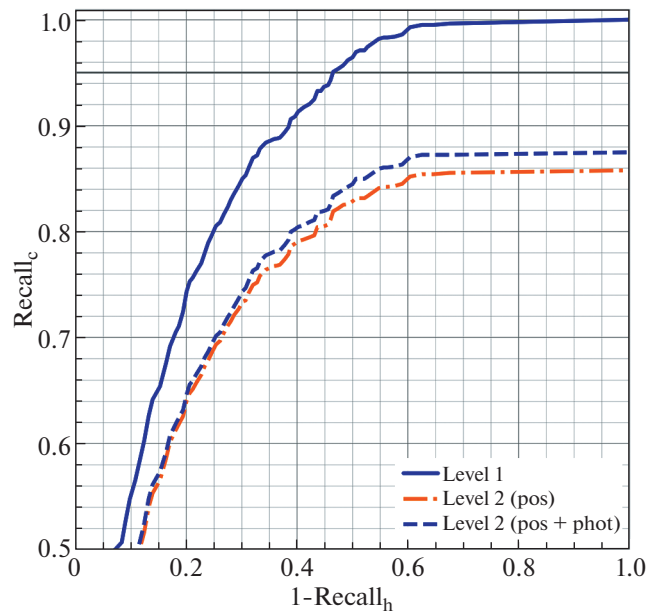


Рис. 16. Зависимость полноты отбора оптических компаньонов рентгеновских источников (Recall_c) от доли неотобранных полей без рентгеновских компаньонов ($1 - \text{Recall}_h$) для рентгеновских источников с потоком выше 3×10^{-15} эрг/с/см². Сплошной линией показаны результаты для уровня 1 модели отождествления (рис. 3). Штриховая линия показывает результат отбора с учетом фотометрической и позиционной информации (уровень 2 pos+phot). Штрихпунктирная линия показывает результат отбора с учетом только позиционной информации при выборе оптического компаньона, т.е. при выборе ближайшего оптического объекта (уровень 2 pos).

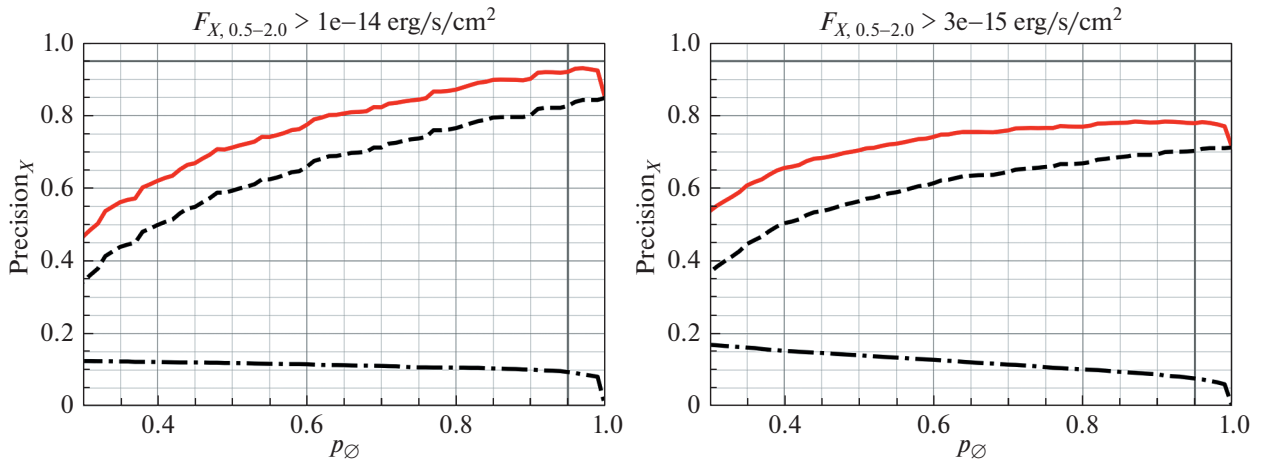


Рис. 17. Точность отождествления (сплошная линия). При определении точности корректным отождествлением был назван либо правильно найденный компаньон (\hat{N}_c^*/N_X , штриховая линия), либо правильно выявленный рентгеновский источник без компаньона (\hat{N}_b^*/N_X , штрихпунктирная линия). Метрики были вычислены относительно обзора DESI LIS для двух различных порогов по рентгеновому потоку. Вертикальной линией отмечено значение $p_0 = 0.95$.

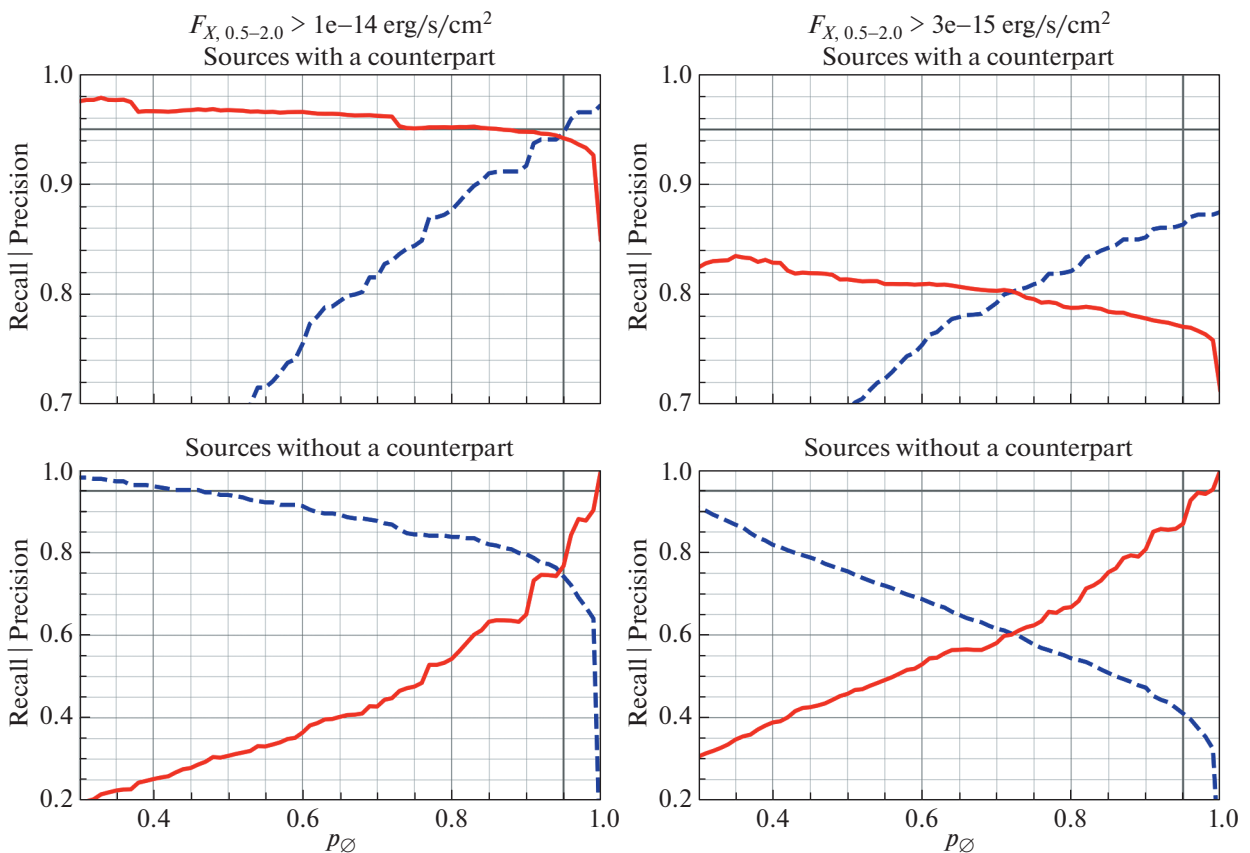


Рис. 18. Метрики качества отбора для рентгеновских источников с найденным компаньоном и рентгеновских источников без компаньона. Сплошная линия соответствует точности для источников с компаньоном ($Precision_c$, верхние панели) и для источников без компаньоном ($Precision_b$, нижние панели). Штриховая линия соответствует полноте отбора для источников с компаньоном ($Recall_c$, верхние панели) и для источников без компаньона ($Recall_b$, нижние панели). Метрики были вычислены относительно обзора DESI LIS для двух различных порогов по рентгеновому потоку.

$\approx 74\%$ и $\approx 45\%$ для обзора DESI LIS и SDSS соответственно. Для $F_{X,0.5-2} \approx 10^{-14}$ эрг/с/см² (что соответствует ожидаемой пороговой чувствительности четырехлетнего обзора SRG/ePOZITA в экваториальной плоскости), значение p_c для DESI LIS и SDSS практически сравниваются и составляют $\approx 82\%$ и $\approx 75\%$ соответственно. На больших рентгеновских потоках $F_{X,0.5-2} \gtrsim 10^{-14}$ эрг/с/см² доля объектов с оптическими компаньонами в обоих обзорах сравнивается (в пределах ошибок измерений на параметр p_c).

4.3. Оптическое отождествление с использованием эффективной звездной величины источников

Ложные источники и компаньоны принадлежат различным порождающим распределениям как по угловому расстоянию от рентгеновского источника, так и по звездным величинам. Фотометрическая информация в модели содержится в эффективных величинах для оптических объектов. Введение фотометрической информации позволяет сократить количество ненайденных компаньонов. На рис. 16 приведены кривые для полноты при отборе источников, имеющих компаньон (сплошная линия); при отборе источников, компаньон которых идентифицирован на основе позиционной информации (штрихпунктирная линия); при отборе источников, компаньон которых идентифицирован на основе позиционной и фотометрической информации (штриховая линия). Для построения сплошной линии в качестве объектов положительного класса были выбраны поля, содержащие компаньон. Для остальных линий объектами положительного класса являются поля, в которых компаньон содержится и найден на основе позиционной информации или на основе позиционной информации с привлечением фотометрической.

Видно, что для каталога DESI LIS с привлечением фотометрической информации количество ненайденных компаньонов сокращается на $\approx 13\%$.

4.4. Качество оптического отождествления рентгеновских источников на всей выборке Дыры Локмана

Точность отождествления всей выборки рентгеновских объектов из области Дыры Локмана с потоком выше 3×10^{-15} эрг/с/см² составляет 78% (рис. 17, правая панель, сплошная линия). С потоком выше 10^{-14} — 93% (рис. 17, левая панель, сплошная линия). При определении точности корректным отождествлением был назван либо правильно найденный компаньон, либо правильно выявленный рентгеновский источник без компаньона. Точность и полнота отождествления варьируются в

зависимости от выбора параметра p_\emptyset (вероятность отсутствия компаньона в каталоге DESI LIS для заданного рентгеновского источника и его оптического окружения).

4.5. Поиск оптических компаньонов и отбор рентгеновских источников без компаньона

При отождествлении рентгеновских объектов нас могут интересовать два принципиально разных режима: отбор объектов с высокой точностью (поиск уникальных объектов в обзоре) и отбор объектов с высокой полнотой (отождествление квазаров для построение функции светимости, где важно свести к минимуму влияние эффектов селекции). Модель отождествления, описанная в этой работе, позволяет производить отбор рентгеновских источников в обоих режимах при помощи изменения порога по параметру p_\emptyset .

- Представленная модель отождествления позволяет производить поиск компаньонов для рентгеновских источников SRG/ePOZITA в области Дыры Локмана с потоком выше 10^{-14} эрг/с/см² с точностью 94% и полнотой 94% (сплошная и штриховая линии соответственно на рис. 18, левый верхний график). Для источников с потоком выше 3×10^{-15} эрг/с/см² при точности 77% достигается полнота 86% (сплошная и штриховая линии соответственно на рис. 18, правый верхний график). Значения приведены для $p_\emptyset = 0.95$ (соответствует вертикальной линии на графике) и относительно оптического обзора DESI LIS.
- Модель отождествления позволяет производить отбор рентгеновских источников без оптического компаньона в обзоре DESI LIS. Для рентгеновских источников с потоком выше 10^{-14} эрг/с/см² при точности 77% достигается полнота 74% (сплошная и штриховая линии соответственно на рис. 18, левый нижний график). Для источников с потоком выше 3×10^{-15} эрг/с/см² при точности 87% достигается полнота 41% (сплошная и штриховая линии соответственно на рис. 18, правый нижний график). Значения приведены для $p_\emptyset = 0.95$.

5. ВЫВОДЫ

В работе представлена вероятностная модель оптического отождествления рентгеновских источников SRG/ePOZITA в области Дыры Локмана по данным фотометрического обзора DESI Legacy

Imaging Surveys (DESI LIS). Для измерения эффективности модели были использованы данные каталогов CSC 2.0 и 4XMM DR10 в поле Дыры Локмана.

Получены следующие результаты:

1. По данным об оптических источниках (DESI LIS) в полях рентгеновских объектов была измерена зависимость ошибки локализации рентгеновских источников СРГ/еРОЗИТА в области Дыры Локмана от правдоподобия их детектирования $\sigma(\mathcal{L})$. Мы нашли, что средняя ошибка локализации рентгеновского источника в угловых секундах σ_{det} (полученная при помощи стандартного пайплайна детектирования точечных рентгеновских источников eSASS) может быть уточнена следующим образом: $\sigma_{\text{corr}} = a\sqrt{\sigma_{\text{det}}^b + c^2}$, где $a = 0.87 \pm 0.08$, $b = 2.53 \pm 0.20$, $c = 1.12 \pm 0.19$.
2. Измерена зависимость вероятности наличия оптического компаньона от рентгеновского потока источника $p_c(F_{X,0.5-2})$ для фотометрических обзоров различной глубины (DESI LIS и SDSS). На рентгеновском потоке 3×10^{-15} эрг/с/см² (чувствительность обзора еРОЗИТА в области Дыры Локмана) значение p_c составляет $\approx 74\%$ для DESI LIS и $\approx 45\%$ для SDSS. На $F_{X,0.5-2} = 10^{-14}$ эрг/с/см² (соответствует чувствительности четырехлетнего обзора СРГ/еРОЗИТА в экваториальной плоскости) — $\approx 82\%$ и $\approx 75\%$ для DESI LIS и SDSS соответственно.
3. Построена модель оптического отождествления рентгеновских источников СРГ/еРОЗИТА в поле Дыры Локмана, которая учитывает как позиционную, так и фотометрическую информацию об отождествляемых оптических объектах. Фотометрическая информация в модели содержится в эффективных величинах для оптических объектов. Показано, что учет фотометрической информации таким способом позволяет сократить количество ошибок на 13% при поиске компаньонов рентгеновских источников с рентгеновским потоком $F_{X,0.5-2} > 3 \times 10^{-15}$ эрг/с/см² (относительно метода, в котором используется только позиционная информация). Точность отождествления всей выборки рентгеновских объектов из области Дыры Локмана с потоком выше

3×10^{-15} эрг/с/см² составляет 78%. С потоком выше 10^{-14} — 93%. При определении точности корректным отождествлением был назван либо правильно найденный компаньон, либо правильно выявленный рентгеновский источник без компаньона. Точность и полнота отождествления варьируются в зависимости от выбора параметра p_{\emptyset} (вероятность отсутствия компаньона в каталоге DESI LIS для заданного рентгеновского источника и его оптического окружения).

4. Представленная модель отождествления позволяет производить поиск компаньонов для рентгеновских источников СРГ/еРОЗИТА в области Дыры Локмана с потоком выше 10^{-14} эрг/с/см² с точностью 94% и полнотой 94% в данных обзора DESI LIS. Для источников с потоком выше 3×10^{-15} эрг/с/см² при точности 77% достигается полнота 86%. Значения приведены для $p_{\emptyset} = 0.95$.
5. Модель отождествления позволяет производить отбор рентгеновских источников без оптического компаньона в обзоре DESI LIS. Для рентгеновских источников с потоком выше 10^{-14} эрг/с/см² при точности 77% достигается полнота 74%. Для источников с потоком выше 3×10^{-15} эрг/с/см² при точности 87% достигается полнота 41%. Значения приведены для $p_{\emptyset} = 0.95$.

Представленная модель будет использована в обзоре СРГ/еРОЗИТА всего неба для оптического отождествления рентгеновских источников.

Работа основана 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/еРОЗИТА изготовлен консорциумом германских институтов во главе с Институтом внеземной астрофизики Общества им. Макса Планка (МРЕ) 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, Усурийске и Байконуре и финансируется Роскосмосом. Используемые в настоящей работе данные телескопа еРОЗИТА обработаны с помощью

программного обеспечения eSASS, разработанного германским консорциумом eРОЗИТА, и программного обеспечения, разработанного российским консорциумом телескопа SRG/eРОЗИТА. Система создана в научной рабочей группе по поиску рентгеновских источников, их отождествлению и составлению каталога по данным телескопа SRG/eРОЗИТА в отделе астрофизики высоких энергий ИКИ РАН.

В работе было использовано следующее программное обеспечение. Научные библиотеки Python: Numpy (Харрис и др., 2020), Scipy (Виртанен и др., 2020), Astropy (Робиталли и др., 2013; Прайс-Вилан и др., 2018), Pandas (МакКинни, 2010), Matplotlib (Хантер, 2007). Также использовались TOPCAT (Тейлор, 2005), CSCview² и научная платформа SciServer³.

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

СПИСОК ЛИТЕРАТУРЫ

- Аболфати и др. (B. Abolfathi, D.S. Aguado, G. Aguilar, C. Allende Prieto, A. Almeida, T.T. Ananna, et al.), *Astrophys. J. Suppl. Ser.* **235**, 42 (2018).
- Ананна и др. (T.T. Ananna, M. Salvato, S. Lamassa, C.M. Urry, N. Cappelluti, C. Cardamone, et al.), *VizieR Online Data Catalog*, p. J/ApJ/850/66 (2019).
- Бруса и др. (M. Brusa, G. Zamorani, A. Comastri, G. Hasinger, N. Cappelluti, F. Civano, et al.), *Astrophys. J. Suppl. Ser.* **172**, 353 (2007).
- Брюннер и др. (H. Brunner, T. Liu, G. Lamer, A. Georgakakis, A. Merloni, M. Brusa, et al.), arXiv e-prints, p. arXiv:2106.14517 (2021).
- Будавари и Лоредо (T. Budav'ari and T.J. Loredo), *Annual Review of Statistics and Its Application* **2**, 113 (2015).
- Вебб и др. (N.A. Webb, M. Coriat, I. Traulsen, J. Ballet, C. Motch, F.J. Carrera, et al.), *Astron. Astrophys.* **641**, A136 (2020).
- Виртанен и др. (P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, et al.), *Nature Methods* **17**, 261 (2020).
- Джеймс (F. James), *Statistical Methods in Experimental Physics: 2nd Edition* (2006).
- Ламасса и др. (S.M. Lamassa, C.M. Urry, N. Cappelluti, H. Bohringer, A. Comastri, E. Glikman, et al.), *VizieR Online Data Catalog*, p. J/ApJ/817/172 (2016).
- МакКинни (Wes McKinney), in St'efan van der Walt and Jarrod Millman (eds.), *Proceedings of the 9th Python in Science Conference*, pp. 56–61 (2010).
- Марчеси и др. (S. Marchesi, F. Civano, M. Elvis, M. Salvato, M. Brusa, A. Comastri, et al.), *VizieR Online Data Catalog*, p. J/ApJ/817/34 (2019).
- Мещеряков и др. (in preparation), (2022).
- Нейлор, Брус (T. Naylor, P.S. Broos, and E.D. Feigelson), *Astrophys. J. Suppl. Ser.* **209**, 30 (2013).
- Оябу и др. (S. Oyabu, M.S. Yun, T. Murayama, D.B. Sanders, K. Kawara, Y. Taniguchi, et al.), *Astron. J.* **130**, 2019 (2005).
- Павлинский и др. (M. Pavlinsky, A. Tkachenko, V. Levin, N. Alexandrovich, V. Arefiev, V. Babyshkin, et al.), *Astron. Astrophys.* **650**, A42 (2021).
- Пино и др. (F.X. Pineau, S. Derriere, C. Motch, F.J. Carrera, F. Genova, L. Michel, et al.), *Astron. Astrophys.* **597**, A89 (2017).
- Прайс-Вилан и др. (Astropy Collaboration, A.M. Price-Whelan, B.M. Sipőcz, H.M. Günther, P.L. Lim, S.M. Crawford, et al.), *Astron. J.* **156**, 123 (2018).
- Предель и др. (P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, et al.), *Astron. Astrophys.* **647**, A1 (2021).
- Робиталли и др. (Astropy Collaboration, T.P. Robitaille, E.J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, et al.), *Astron. Astrophys.* **558**, A33 (2013).
- Сазерленд, Сандерс (W. Sutherland and W. Saunders), *MNRAS* **259**, 413 (1992).
- Сальвато и др. (M. Salvato, J. Buchner, T. Budav'ari, T. Dwelly, A. Merloni, M. Brusa, et al.), *MNRAS* **473**, 4937 (2018).
- Сюняев и др. (R. Sunyaev, V. Arefiev, V. Babyshkin, A. Bogomolov, K. Borisov, M. Buntov, et al.), arXiv e-prints, p. arXiv:2104.13267 (2021).
- Тейлор (M.B. Taylor), in P. Shopbell, M. Britton, and R. Ebert (eds.), *Astronomical Data Analysis Software and Systems XIV*, Vol. 347 of *Astronomical Society of the Pacific Conference Series*, p. 29 (2005).
- Хантер (J.D. Hunter), *Computing in Science & Engineering* **9**, 90 (2007).
- Харрис и др. (C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, et al.), *Nature* **585**, 357 (2020).
- Чивано и др. (F. Civano, M. Elvis, M. Brusa, A. Comastri, M. Salvato, G. Zamorani, et al.), *VizieR Online Data Catalog*, p. J/ApJS/201/30 (2012).
- Эванс и др. (I.N. Evans, F.A. Primini, K.J. Glotfelty, C.S. Anderson, N.R. Bonaventura, J.C. Chen, et al.), *Astrophys. J. Suppl. Ser.* **189**, 37 (2010).

²<http://cda.cfa.harvard.edu/cscview/>

³www.sciserver.org