Построение модели оптического отождествления рентгеновских источников СРГ/еРОЗИТА на примере данных области Дыры Локмана

Бельведерский М., Мещеряков А., Гильфанов М., Медведев П.

ИКИ РАН, Москва 2021



Пример оптического поля рентгеновского источника СРГ/еРОЗИТА

- Для подавляющего большинства рентгеновских источников в обзоре СРГ/еРОЗИТА мы располагаем ограниченной информацией
- В области локализации рентгеновского источника находится множество оптических объектов из каталога DESI LIS (показаны небольшими окружностями)
- Для проведения спектроскопии требуется отождествить рентгеновский источник с его компаньоном

*p*_{match,i} — вероятность того, что оптический кандидат является компаньоном заданного рентгеновского источника

*p*_Ø — вероятность того, что рентгеновский источник не имеет оптического компаньона в заданном фотометрическом обзоре с учетом его оптического окружения





Задача: построить модель оптического отождествления рентгеновских источников СРГ/еРОЗИТА в области LH

Требования к модели:

- фотометрической информации
- Модель должна сообщать об отсутствии компаньона высоким значением параметра p_{o}

• Модель отождествления должна быть основана на позиционной и

• Должны быть доступны два разных режима отбора объектов: с высокой точностью (как правило выше 80%) и с высокой полнотой (выше 90%)



Задача: построить модель оптического отождествления рентгеновских источников СРГ/еРОЗИТА в области LH



- $\sigma = \sigma(\mathscr{L})$ ошибка локализации рентгеновского источника
- фоновой компонентой на рентгеновском изображении

• $\mathscr{L} = -\ln(p)$, где p —вероятность нулевой гипотезы о том, что наблюдаемый объект порожден

• Стандартный алгоритм eSASS вычисляет также ошибку локализации $\sigma_{
m det}$ (изображены на графике)

Результаты Выводы

- использование позиционной и фотометрической информации

Модель

- рентгеновские и оптические, тестовая выборка

Данные

Содержание



Данные Модель Результаты Выводы

6

Рентгеновские данные

Дыра Локмана

Lockman Hole, (LH) (Локман и др., 1986)

- Область покрытия обзора LH составила
 ≈ 18.5 кв. град (5° × 3.7°)
- Всего в области LH обнаружено 8309 точечных рентгеновских источников с правдоподобием детектирования выше 6
- Глубина по потоку $\approx 3 \times 10^{-15}$ эрг/с/см² в диапазоне 0.5 2 кэВ

Оптические данные

- DESI Legacy Imaging Surveys DR8 (DESI)
- SDSS DR14 (SDSS)

ESA/XMM-Newton/G. Hasinger (MPE Garching, Germany)

This colour view combines X-ray data collected at energies of 0.5–2 keV (shown in red), 2–4.5 keV (green) and 4.5–10 keV (blue). The image spans half a degree on the short side; north is up and east to the left.



Тестовая выборка

На основе данных XMM & CSC

Всего в области LH обнаружено 8309 точечных рентгеновских источников

	4XMM-DR10	CSC 2.0
Источник XMM / CSC единственный в поле ERO (30")	788	585
Источник DESI единственный в окрестности XMM / CSC	473	437
Поток XMM/CSC отличается не более чем в 5 раз от потоков источников ERO	464	319

- Исключены поля, где помимо единственного источника CSC было обнаружено несколько источников 4XMM и наоборот (сократив количество рентгеновских объектов до 577)
- Исключены 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 звезд на оптическом изображении. Так была получена тестовая выборка, состоящая из 541 рентгеновского источника с надежным отождествлением оптического компаньона



Тестовые данные

Тестовый каталог

- Распределение по рентгеновскому потоку (0.5 2 кэВ) источников СРГ/еРОЗИТА отличается от распределения для тестовой выборки
- На основе тестовой выборки с компаньонами был создан тестовый каталог
- При розыгрыше учитывалось значение p_c(F_{X,0.5-2}) вероятности наличия в оптическом каталоге компаньона рентгеновского источника



четырехлетнего обзора СРГ / еРОЗИТА)

Тестовые данные

Тестовый каталог

- Распределение по рентгеновскому потоку (0.5 2 кэВ) источников СРГ/еРОЗИТА отличается от распределения для тестовой выборки
- На основе тестовой выборки с компаньонами был создан тестовый каталог
- При розыгрыше учитывалось значение p_c(F_{X,0.5-2}) вероятности наличия в оптическом каталоге компаньона рентгеновского источника

Метрики качества отождествления и отбора источников:

- Точность отождествления: **Precision** $_X$
- Полнота Recall_c и точность Precision_c отбора источников с компаньоном
- Полнота Recall_h и точность Precision_h отбора рентгеновских источников без компаньонов



четырехлетнего обзора СРГ / еРОЗИТА)

Данные Модель Результаты Выводы

11

Данные Иодель Результаты Выводы



Математическая модель отождествления

Пусть в некотором диапазоне по правдоподобию детектирования $\mathscr L$ содержится $N_{\mathbf{X}}$ источников. Тогда среднее количество $N_{0}(r)$ всех оптических кандидатов в кружке радиуса r:

$$N_{\rm o}(r) = N_{\rm x} \left(p_{\rm c} \left[1 - \exp\left(\frac{-r^2}{2\sigma^2}\right) \right] + \rho \pi r^2 \right)$$

Вклад компаньонов

- $p_{\rm c}(F_{{\rm X},0.5-2})$ вероятность наличия в оптическом каталоге компаньона
- *р* плотность фоновых (ложных) источников

Вклад «ложных» источников

рентгеновского источника (зависит от потока рентгеновского источника) • $\sigma(\mathscr{L})$ — стандартное отклонение нормального распределения компаньонов



Дерево поиска компаньона

уровень 1 (pos + phot)

 $p_{\varnothing} \ge p'_{\varnothing}$

(компаньон отсутствует)

уровень 2 (pos) **уровень 2** (pos + phot)

• $p_{\text{match}}(o_{i})$ — вероятность заданного оптического источника быть компаньоном для заданного рентгеновского

• p_{\emptyset} — вероятность отсутствия компаньона для рентгеновского источника с заданным потоком *в заданном оптическом поле*



Модель

Ограничения модели

- ИСТОЧНИКОВ
- Распределение оптических источников пуассоновское с плотностью объектов ρ

$$N_{\rm o}(r) = N_{\rm x} \left(p_{\rm c} \left[1 - \exp\left(\frac{-r^2}{2\sigma^2}\right) \right] + \rho \pi r^2 \right)$$

• Функция ошибки на положение рентгеновских источников описывается двумерным нормальным распределением

• Ошибка локализации оптических источников пренебрежимо мала по сравнению с ошибками локализации рентгеновских

Функция правдоподобия L

Определена в интервале по $\mathscr L$ и эффективной звездной величине

Среднее количества кандидатов в кольце *i*, ограниченном окружностями с радиусами r_i , r_{i+1} :

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

Функция п
$$L = \sum_{i}^{L}$$



Модель

Эффективная звездная величина m_{eff}

• Единый фотометрический параметр вместо нескольких порогов в разных фильтрах

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

где c_g, c_r, c_z — выбранные константы, величины в фильтрах g, r, z заданы в АВ-системе



~ 22.0 21.5 ι_{20.0} Ζ

Модель

Эффективная звездная величина m_{eff}

• Единый фотометрический параметр вместо нескольких порогов в разных фильтрах

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

где c_g, c_r, c_z — выбранные константы, величины в фильтрах g, r, z заданы в АВ-системе



- 22.0 21.5 ι_{20.0} Ζ

Модель Распределение оптических кандидатов по *m*_{eff}



Распределения всех оптических объектов, найденных внутри радиуса 2σ от рентгеновских источников СРГ/еРОЗИТА

Модель Распределение оптических кандидатов по *m*_{eff}



Распределения всех оптических объектов, найденных внутри радиуса 2σ от рентгеновских источников СРГ/еРОЗИТА

Модель

Распределение оптических кандидатов по угловому расстоянию

Распределения кандидатов по угловому расстоянию до рентгеновского источника

- Отличаются в разных диапазонах по правдоподобию детектирования $\mathscr L$
- Построены с разными пороговыми значениями $m_{\rm eff}$
- Получаем набор параметров $\sigma, p_{\rm c}, \rho$

100

75

50

25

 $^{\circ}$

$$N_{\rm o}(r) = N_{\rm x} \left(p_{\rm c} \right)$$







Модель Ошибка локализации $\sigma(\mathscr{L})$

• Вычисленные нами ошибки локализации σ рентгеновских источников в среднем выше ошибок σ_{det} , измеренных алгоритмом детектирования источников



Зависимость ошибки локализации рентгеновских источников LH от правдоподобия их детектирования

Модель Вероятность наличия компаньона $p_{\rm c}(F_{{\rm X}.0.5-2})$

- Время экспозиции СРГ/еРОЗИТА слабо менялось внутри поля LH, поэтому для пересчета правдоподобия детектирования $\mathscr L$ в рентгеновский поток: $\lg(F_{X,0.5-2}) = 0.66 \times \lg(\mathscr{L}) - 15.09$
- Ошибки на параметр $p_{\rm c}$ не являются гауссовыми, поэтому мы преобразовали $p_{\rm c}$ из интервала [0,1]на всю числовую ось при помощи логистического преобразования:

$$logit(p_c) = ln\left(\frac{p_c}{1 - p_c}\right)$$

• Обратное логистическое преобразование:

$$p_{\rm c} = \frac{\exp[logit(p_{\rm c})]}{1 + logit(p_{\rm c})}$$
 0.4

2

logit(p_c)

 $p_{\rm c}$



Модель Плотность ложных источников $ho(\mathscr{L})$

- С увеличением правдоподобия детектирования рентгеновских источников от 6 до 1000 плотность оптических объектов поля падает на порядок
- Более яркие рентгеновские источники, в среднем, имеют более яркие оптические компаньоны



Значения плотности ложных источников ρ в разных интервалах по \mathscr{L} для $p_{\lim} = 0.85$.

Модель Плотность ложных источников $ho(\mathscr{L})$

- С увеличением правдоподобия детектирования рентгеновских источников от 6 до 1000 плотность оптических объектов поля падает на порядок
- Более яркие рентгеновские источники, в среднем, имеют более яркие оптические компаньоны



Значения плотности ложных источников ρ в разных интервалах по \mathscr{L} для $p_{\rm lim}=0.85$.

Модель Семейство моделей для разных значений $p_{ m lim}$



Справа: Вероятность наличия у источника СРГ/еРОЗИТА оптического компаньона как функция от рентгеновского потока — $p_{\rm c}(F_{{
m X},0.5-2})$. Показаны кривые для различных значений выбора максимального квантиля распределения оптических компаньонов p_{\lim} .

p_{lim} •••• 0.995 0.99--- 0.95 0.90.85 0.8 0.75 0.7 0.650.60.55 0.5



Модель Семейство моделей для разных значений $p_{\rm lim}$



Слева: Значения плотности ложных источников ho в разных интервалах по \mathscr{L} для $p_{\mathrm{lim}}=0.85$. Справа: Зависимости плотности ложных источников от правдоподобия рентгеновского детектирования для различных значений выбора максимального квантиля распределения оптических компаньонов $p_{
m lim}$. Горизонтальной сплошной линией отмечены значения плотности по всем источникам DESI LIS.

Модель Семейство моделей для разных значений $p_{\rm lim}$



Зависимость ошибки локализации объекта от правдоподобия рентгеновского детектирования — $\sigma(\mathscr{L})$. Показаны кривые, измеренные при различных значениях максимального квантиля распределения оптических компаньонов p_{\lim}_{28}

Модель

Вероятность компаньона p_{match}

поле с плотностью источников ρ_{i}^{eff} , является компаньоном, определяется следующим выражением:

 $p_{\text{match}} = \frac{1}{p_{\text{c,i}}^{\text{eff}} \exp}$

Значения $p_{\rm c,i}^{\rm eff}$, $\rho_{\rm i}^{\rm eff}$ вычислены при помощи позиционно-фотометрической модели. 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 $m_{\rm eff}$): $p_{\rm c}^{\rm pos}$, $\rho^{\rm pos}$.

Вероятность, что оптический кандидат і, находящийся на расстоянии $r_{
m i}$ от рентгеновского источника (с параметрами σ и $p_{
m c,i}^{
m eff}$), в оптическом

$$\frac{\rho_{\mathrm{c,i}}^{\mathrm{eff}} \exp\left(\frac{-r_{\mathrm{i}}^{2}}{2\sigma^{2}}\right)}{p\left(\frac{-r_{\mathrm{i}}^{2}}{2\sigma^{2}}\right) + 2\pi\rho_{\mathrm{i}}^{\mathrm{eff}}\sigma^{2}}$$

Модель

Вероятность отсутствия компаньона $p_{oldsymbol{\emptyset}}$

$$p_{\emptyset} = \frac{1}{1 + \frac{1}{2\pi\sigma} \sum_{i=1}^{n_{o}} \exp\left(\frac{-r_{i}^{2}}{2\sigma^{2}}\right) \frac{p_{c,i}'}{\rho_{i}'}},$$

где $n_{\rm o}$ — количество оптических кандидатов DESI LIS в окрестности источника еРОЗИТА, $p'_{\rm c,i} = p_{\rm c}^{\rm pos} - \overline{p_{\rm c,i}^{\rm eff}}$, $\rho'_{\rm i} = \rho^{\rm pos} - \overline{\rho_{\rm i}^{\rm eff}}$. Параметры $\overline{p_{\rm c,i}^{\rm eff}}$, $\overline{\rho_{\rm i}^{\rm eff}}$ также как и $p_{\rm c,i}^{\rm eff}$, $\rho_{\rm i}^{\rm eff}$ получены с помощью модели, $m_{\rm eff}$ которой имеет наименьшее отличие от $m_{\rm eff}$ кандидата, но при этом $m_{\rm eff}$ выбранной модели должен быть *меньше* $m_{\rm eff}$ кандидата.

Введение Данные Иодель Результаты Выводы

Введение Данные Модель Результаты

Выводы

Результаты Калибровочная зависимость ошибки локализации



- Зависимость ошибки локализации рентгеновского источника от его вероятности детектирования была измерена нами при поиске наилучших параметров модели отождествления в интервалах по *Э* на основе оптических данных DESI LIS в окрестности рентгеновских источников из области Дыры Локмана
- Была найдена калибровочная зависимость между ошибкой локализации точечных рентгеновских источников на основе данных об их оптических компаньонах

$$\sigma_{\text{corr},1} = 1.11 \sqrt{\sigma_{\text{det}}^2 + 0.68^2}$$
 $\sigma_{\text{corr},2} = 0.87 \sqrt{\sigma_{\text{det}}^{2.53} + 1.12^2}$

 $\sigma_{\rm corr, eFEDS} = 1.15 \sqrt{\sigma_{\rm det}^2 + 0.7^2}$ (Brunner et al., 2021)





Вероятность наличия компаньона $p_{\rm c}(F_{{\rm X}.0.5-2})$

- Глубина обзора влияет на поведение рс
- На рентгеновском потоке $F_{{
 m X},0.5-2} pprox ~ 3 \cdot 10^{-15}$ эрг/с/см² (соответствует пороговой чувствительности в области Дыры Локмана) вероятность иметь компаньон $p_{\rm c}$ составляет $\approx 74\%$ и $\approx 45\%$ для обзора DESI LIS и SDSS, соответственно
- Для $F_{{
 m X},0.5-2}pprox~10^{-14}$ эрг/с/см 2 , (что соответствует ожидаемой пороговой чувствительности четырехлетнего обзора в экваториальной плоскости) значение $p_{
 m c}$ для DESI LIS и SDSS практически сравниваются





F_{x, 0.5 - 2}



источников СРГ/еРОЗИТА

Точность отождествления:

 $Precision_{X} =$

где $N_{\rm X}$ — общее количество всех рентгеновских источников в выборке, $\hat{N}_{c+h}^{\star} = \hat{N}_{c}^{\star} + \hat{N}_{h}^{\star} - количество рентгеновских$ источников, у которых оптический компаньон найден правильно (\hat{N}_{c}^{\star}) или правильно указан факт отсутствия компаньона в выбранном фотометрическом обзоре ($\hat{N}_{\rm h}^{\star}$)

$$=\frac{\hat{N}_{\rm c+h}^{\star}}{N_{\rm X}},$$



источников СРГ/еРОЗИТА

	Источники с компаньоном (с)	Источники без компаньона (h)
Полнота (Recall)	Recall _c $= \frac{\hat{N}_{c}^{\star}}{N_{c}}$, где N_{c} — полное число рентгеновских источников с оптическими компаньонами в выборке.	$\begin{aligned} \text{Recall}_{h} &= \frac{\hat{N}_{h}^{\star}}{N_{h}}, \\ \text{где } N_{h} &- \text{полное число рентгеновских} \\ \text{источников без оптических компаньонов в} \\ \text{выборке.} \end{aligned}$
Точность (Precision)	Precision _c = $\frac{\hat{N}_{c}^{\star}}{\hat{N}_{c}}$, где \hat{N}_{c} – число рентгеновских источников, отобранных алгоритмом отождествления как источники с компаньонами.	Precision _h = $\frac{\hat{N}_{h}^{\star}}{\hat{N}_{h}}$, где \hat{N}_{h} — число рентгеновских источников отобранных алгоритмом отождествления как источники без компаньонов.



Привлечение фотометрической информации

Тестовый каталог на основе данных 4XMM, CSC 2.0



- Введение фотометрической информации позволяет сократить количество ошибок при поиске компаньонов
- Для DESI LIS это улучшение составляет \approx 13%
- Далее используем только pos+phot

компаньон присутствует в поле

компаньон присутствует в поле и найден







источников СРГ/еРОЗИТА



- Точность отождествления всей выборки рентгеновских объектов из области Дыры Локмана потоком выше $3 \cdot 10^{-15}$ эрг/с/см² составляет 78%
- С потоком выше 10^{-14} эрг/с/см² 93% (левая панель, сплошная линия)

 N_X — общее количество всех рентгеновских источников в выборке • $\hat{N}_{c+h}^{\star} = \hat{N}_{c}^{\star} + \hat{N}_{h}^{\star} -$ количество рентгеновских источников, у которых оптический компаньон найден правильно ($\hat{N}_{\mathrm{c}}^{\star}$) или правильно указан факт отсутствия компаньона в выбранном фотометрическом обзоре ($\hat{N}^{\star}_{
m h}$)

Precision_X :



Результаты Точность и полнота отбора

для источников СРГ/еРОЗИТА с компаньоном и без

- Модель позволяет производить поиск компаньонов для рентгеновских источников в области Дыры Локмана с потоком выше 10^{-14} эрг/с/см² с точностью 94% и полнотой 94% (сплошная и штриховая линия соответственно, левый верхний график)
- 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 линия соответственно, левый нижний график)

Все значения приведены для $p_{\emptyset} = 0.95$ (соответствует вертикальной линии на графиках) и относительно оптического обзора DESI LIS



Сплошная линия соответствует точности для источников с компаньоном (Precision, верхние панели) и для источников без компаньона (Precision_h, нижние панели). Штриховая линия соответствует полноте отбора для для источников с компаньоном (Recall_c, верхние панели) и для источников без компаньона (Recall_h, нижние панели). Метрики были вычислены относительно обзора DESI LIS для двух различных порогов по рентгеновскому потоку.

40



Введение Данные Модель

Результаты Выводы

Введение Данные Модель Результаты Выводы



42

- ошибка локализации рентгеновского источника $\sigma_{
 m det}$ (полученная при помощи стандартного пайплайна детектирования точечных рентгеновских источников eSASS) может быть уточнена следующим образом: $\sigma_{\rm corr} = 0.87 \sqrt{\sigma_{
 m det}^{2.53} + 1.12^2}$.
- Измерена зависимость вероятности наличия оптического компаньона от рентгеновского потока источника $p_{
 m c}(F_{{
 m X},0.5-2})$ для фотометрических обзоров различной глубины (DESI LIS и SDSS). На рентгеновском потоке $3 \cdot 10^{-15}$ эрг/с/см² (чувствительность обзора еРОЗИТА в области Дыры Локмана) значение $p_{
 m c}$ составляет pprox74 % для DESI LIS и pprox45 % для SDSS. На $F_{
 m X,0.5-2}$ = 10^{-14} эрг/с/см 2 (соответствует чувствительности четырехлетнего обзора СРГ/еРОЗИТА в экваториальной плоскости) — \approx 82% и \approx 75% для DESI LIS и SDSS соответственно.
- 3 · 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 $3 \cdot 10^{-15}$ эрг/с/см² составляет 78%. С потоком выше ро (вероятность отсутствия компаньона в каталоге DESI LIS для заданного рентгеновского источника и его оптического окружения).
- потоком выше $3 \cdot 10^{-15}$ эрг/с/см² при точности 77% достигается полнота 86%. Значения приведены для $p_{\emptyset} = 0.95$.
- Для рентгеновских источников с потоком выше 10^{-14} эрг/с/см² при точности 77% достигается полнота 74%. Для источников с потоком выше $3 \cdot 10^{-15}$ эрг/с/см² при точности 87% достигается полнота 41%. Значения приведены для $p_{\emptyset} = 0.95$.

• По данным об оптических источниках (DESI LIS) в полях рентгеновских объектов, была измерена зависимость ошибки локализации рентгеновских источников СРГ/еРОЗИТА в области Lockman Hole от правдоподобия их детектирования $\sigma(\mathscr{L})$. Мы нашли, что средняя

Построена модель оптического отождествления рентгеновских источников СРГ/еРОЗИТА в поле Lockman Hole, 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 $F_{{
m X},0.5-2}$ > 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 варьируется в зависимости от выбора параметра

• 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

• Модель отождествления позволяет производить отбор рентгеновских источников без оптического компаньона в обзоре DESI LIS.