Федеральное государственное автономное образовательное учреждение высшего образования «Московский физико-технический институт (национальный исследовательский университет)» Физтех-школа физики и исследований им. Ландау Кафедра проблем физики и астрофизики

Направление подготовки / специальность: 03.03.01 Прикладные математика и физика Направленность (профиль) подготовки: Современные проблемы физики и энергетики

СТАТИСТИЧЕСКИЙ АНАЛИЗ СВОЙСТВ МЕЖЗВЕЗДНОГО ПОГЛОЩЕНИЯ В ГАЛАКТИКЕ

(бакалаврская работа)

Студент: Муха Виктория Андреевна

(подпись студента)

Научный руководитель: Боли Пол Эндрю , PhD (к.ф.-м.н.)

(подпись научного руководителя)

Консультант (при наличии):

(подпись консультанта)

Москва 2022

Содержание

1	Анн	ютация	2
2	Вве	дение	3
	2.1	Цели работы	6
3	Пос	тановка задачи	7
	3.1	Обзор 2MASS	7
	3.2	Обзор GAIA	8
	3.3	Построение теоретической модели	8
4	Пар	аметры межзвездного поглощения	14
	4.1	Теория	14
	4.2	Закон Карделли	15
5	Бай	есовский подход	18
	5.1	Теория	18
6	При	пложение вероятностной модели к поставленной задаче	20
7	Me	год Монте-Карло и Цепи Маркова	25
	7.1	Теория	25
	7.2	Использование пакета МСМС	26
8	Ана	лиз результатов и дальнейшая работа	29
	8.1	Проверка алгоритма на теоретических моделях	29

	8.2	Применение к наблюдаемым данным	32
	8.3	Дальнейшая работа	34
9	Зак	лючение	36
10	Бла	годарности	37
11	При	ложение А. Код	43
12	При	ложение В. Графики	44

Список иллюстраций

3.1	Кривые пропускания из обзора 2MASS	7
3.2	Кривые пропускания из обзора Gaia	8
6.3	Диаграмма Герцшпрунга-Рассела как априорное распределение параметров светимости и эффективной температуры	21
6.4	Коэффициенты полинома для болометрической поправки (6.23)	22
12.5	Проекции цепей Маркова для звезд разных спектральных клас- сов: О9 (a), B5 (b)	44
12.6	Проекции цепей Маркова для звезд разных спектральных клас- сов: A1 (a), A7 (b)	45
12.7	Проекции цепей Маркова для звезд разных спектральных клас- сов: F3 (a), F9 (b), G5 (c)	46
12.8	Проекции цепей Маркова для звезд разных спектральных клас- сов: K1 (a), K7 (b)	47
12.9	Проекции цепей Маркова для звезд разных спектральных клас- сов: М9	47
12.10	ОПример анализа области Ra ∈ (243.5; 246.9), Dec ∈ (−38.5; −36.3) и построение проекций цепей Маркова для объектов в ней	48
12.11	1Пример анализа области Ra ∈ (243.5; 246.9), Dec ∈ (−38.5; −36.3) и построение проекций цепей Маркова для объектов в ней	49
12.12	2Пример построения цепей Маркова (a) и их проекций (b) для отдельной звезды	50
12.13	3Проекции цепей Маркова, построенные при недостатке априор- ных знаний	50

Список таблиц

8.1	Параметры	синтетических	моделей							•			•						30	
-----	-----------	---------------	---------	--	--	--	--	--	--	---	--	--	---	--	--	--	--	--	----	--

1 Аннотация

В данной работе был проведен анализ данных отдельных звезд для получения оценки их собственных характеристик (эффективная температура, светимость) и параметров межзвездного поглощения (полное поглощение A_{λ} и соотношение полного к селективному поглощению R_V) вещества, по углу зрения до рассматриваемого объекта. Для реализации поставленной задачи был разработан алгоритм статистического анализа спектров звезд в ближнем инфракрасном и оптическом диапазонах. Для этого были использованы современные обзоры 2MASS [1] и GAIA (DR2 [2] и eDR3 [3]). На основе данных из выбранных каталогов было получено спектральное распределение энергии звезд, построенное не только с использованием знаний о параметрах самих объектов, но так же с учетом параметров межзвездной среды, что позволило точнее определить искомые величины.

2 Введение

Одной из основных характеристик Галактики считается межзвездная среда. Несмотря на то что межзвездный газ составляет малую долю от видимого вещества, он оказывает значительное влияние на наблюдения звезд и требует корректного учета при проведении исследований звездного неба. Межзвездная пыль также относится к компонентам межзвездной среды наравне с межзвездным газом и является существенной преградой для получения точных измерений звезд и звездных объектов, что делает её заслуживающим внимания и необходимым для исследования параметром Галактики [4]. Помимо поглощения поверхности пылинок являются также катализатором для молекулярных реакций с последующим образованием сложных молекул. Как следствие, знание о характеристиках межзвездной среды дает возможность тщательнее исследовать области звездообразований и описывать происходящие в них процессы[5].

Параметры межзвездной среды остаются одним из самых интересных аспектов для изучения, в первую очередь, из-за того что полная детализированная карта распределения пылевых частиц в Галактике до сих пор не была получена, несмотря на то что значения параметров поглощения непосредственно влияют на многие измерения и информация об их значениях могла бы существенно повлиять на точность этих измерений. В данной работе рассмотрены два параметра, влияющие на рассеяние и поглощения, - это селективный параметр поглощения, R_V , зависящий от состава и распределения пылинок по размеру, и параметр поглощения межзвездной среды, A_λ , который зависит от количества поглощенного вещества (в настоящей работе используются параметры, введенные согласно закону межзвездного поглощения [6]). Стоит отметить также, что в работе рассмотрены ближний ИФ и оптический диапазон, для которых основной причиной поглощения света является пылевая среда, изучение которой вызывает особый интерес, чем и

3

был обусловлен выбор таких обзоров звездного неба как 2MASS [1] и GAIA (DR2 [2] и eDR3 [3]).

На протяжении долгого времени были совершены многочисленные попытки построить подробные карты распределения пыли в Галактике [7]. Так, например, в 1998 г. в работе [8] был описан метод построения 2D карты звездного неба для длин волн 100-240 мкм. Самым существенным недостатком описанного в упомянутой работе подхода является отсутствие зависимости от расстояния до исследуемого объекта, что не позволяло учитывать через какие области межзвездной среды проходил свет. Именно поэтому дальнейшие труды были направлены на то, чтобы получить 3D карты межзвездной среды Галактики.

Во многих работах был использован закон поглощения, полученный в работе [6]. Однако в большинстве из них селективный параметр поглощения R_V был взят константой $R_V = 3.1$, считавшейся оправданным усреднением, справедливым для значительного количества случаев. Впоследствии, было показано, что данное приближения не является достаточно точным и параметр R_V может менять в существенном диапазоне [9].

В 2006 году была опубликована работа [10] на основе данных 2MASS, в которой подход для получение оценки параметров межзвездной среды заключался в сравнении показателей цвета выделенного класса звезд с теоретическими моделями. Использование обзора 2MASS было обусловлено меньшей зависимости света в инфракрасном диапазоне от изменений в значениях R_V .

Карта покраснения звездного неба с использованием фотометрии PanSTARRS1 (PS1)¹, охватывающая диапазон волн 400-1100 нм, была построена и приведена в работе [11], основной особенностью которой является возможность использования метода для объектов, находящихся на больших расстояниях

 $^{^{1}} https://outerspace.stsci.edu/display/PANSTARRS/$

для наблюдения (~ 4.5 kpc).

Каталог параметра поглощения более чем для 11 млн звездных объектов был представлен в работе [12] с использованием каталогов XSTPS-GAC², 2MASS³ и WISE⁴ и сравнения наблюдаемых спектров с теоретическими моделями с использованием результатов из предыдущей работы автора [13]

Следует упомянуть, что в настоящей работе исследование параметров поглощения проводилось с помощью байесовской статистики. Такой подход также использовался для получение вероятностных оценок параметра в работе [14] для построение 3D карты пылевых облаков на расстоянии от 10 пк до 100 пк.

Из наиболее современных работ, посвященных построению карт поглощения, необходимо отметить [15], предоставляющую обновленный пакет dustmaps (Python) [16], [17], содержащий в себе данные о параметре поглощения с использованием каталогов 2MASS, Gaia DR2 и PS1 и оценку класса звезды с расстоянием до нее (все данные находятся в открытом доступе⁵). Также особое внимание следует обратить на алгоритм, разработанный Chen et al. на основе данных из 2MASS, Gaia DR2 и описанный в [18], для показателей цвета $E(G - K_S)$, $E(G_{BP} - G_{RP})$ и $E(H - K_S)$.

 $^{^{2}}$ http://english.pmo.cas.cn/

 $^{^{3}} https://irsa.ipac.caltech.edu/Missions/2mass.html$

 $^{{}^{4}}https://www.nasa.gov/mission_pages/WISE/main/index.html$

 $^{^{5}}$ http://argonaut.skymaps.info/

2.1 Цели работы

Основной целью данной работы является разработка рабочего алгоритма для получения вероятностной оценки параметров поглощения (A_{λ} и R_V) и характеристик звезд (светимость L, эффективная температура T_{eff} , расстояние до нее r) на основе данных из обзоров 2MASS [1] и Gaia DR2 [2]. В дополнение к основной задаче также осуществляется оценка расстояния до звезд. Для решения поставленной задачи проводилось сравнение наблюдаемых спектров звезд с синтетическими моделями и на основе байесовской статистики были получены статистические оценки искомых параметров. Проверка корректности работы алгоритма была осуществлена с помощью построения теоретических спектров с известными искомыми параметрами и применения алгоритма к ним. Для наглядности полученные результаты были выведены в виде графиков вероятностного распределения параметров.

Благодаря правильно подобранным подробным априорным данным селективный параметр поглощения R_V также вошел в вектор варьирующихся параметров. Возможность данного допущения является ключевым пунктом работы, так как именно значение этого параметра представляет наибольший интерес для изучения. В работах [6], [9] было показано, что значение R_V непосредственно связанно с плотностями пылевых областей, что позволяет связать данный параметр с оценкой размеров гранул пыли. Знание о параметрах межзвездной пыли позволяет не только увеличить точность измерений благодаря учёту поглощения, но также проводить анализ областей звездного неба на молекулярный состав.

3 Постановка задачи

В работе использованы два обзора звездного неба 2MASS [1] и GAIA (DR2 [2] и eDR3 [3]). Подход разработанного алгоритма состоит в сравнении построенного синтетического спектра звезды с наблюдаемым. Экспериментальные звездные величины или потоки были взяты из выше упомянутых каталогов, однако для построения корректной теоретической модели необходимы дополнительные для учета параметры, некоторые из которых также были получены из вышеупомянутых обзоров.

3.1 Обзор 2MASS

Наблюдения всего звездного неба телескопом 2MASS с апертурой 1.3 м проводились в ближнем инфракрасном диапазоне 1235 - 2159 нм (рис. 3.1) в трех фильтрах J, H, K_s ,



Рис. 3.1: Кривые пропускания из обзора 2MASS

3.2 Обзор GAIA

Наблюдения GAIA проводились в оптическом диапазоне в трех фильтрах G, G_{BP}, G_{RP} . Обзор предоставляет информацию о более чем полутора миллиардов звезд. Фильтры охватывают диапазон волн от 300 нм до 1050 нм (рис. 3.2). Помимо фотометрии, каталог GAIA содержит информацию о значениях параллакса для наблюдаемых звезд, которая также использовалась в настоящей работе.



Рис. 3.2: Кривые пропускания из обзора Gaia

3.3 Построение теоретической модели

Неоспоримым преимуществом является возможность объединения этих двух обзоров и рассмотрение их корреляции с возможностью выделения объектов, наблюдаемых в обеих миссиях, что дает дополнительную информацию для более корректного анализа. Цель данной работы состоит в получении статистического распределения параметров звезд и характеристик межзвездной среды. Для этого проводились сравнения спектров наблюдаемых объектов, полученных из обзоров, и теоретических. Основой для построения синтетических моделей послужили закон Рэлея-Джинса (3.1):

$$F_{teor}(\lambda, T) = \frac{2ck_BT}{\lambda^4} \tag{3.1}$$

И функция Планка (3.2):

$$F_{teor}(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1}$$
(3.2)

Данные теоретические спектры были выбраны как логическое дополнение к выбранным обзорам звездного неба. Для данных из 2MASS, то есть для ближнего ИК, спектр звезды имеет схожее с законом Рэлея-Джинса поведение (частный случай функции Планка), корректное использование которого, вместо функции Планка, может значительно сократить затраты по времени для проводимых вычислений, данная модель не применима к более коротким длинам волн. Для оптического диапазона взято приближение функцией Планка, хорошо совпадающее с наблюдениями при высоких температурах. Функция Планка применяется и к ближнему ИК в случае недостаточной точности закона Рэлея-Джинса. Стоит отметить, что для данных синтетических моделей спектр зависит только от температуры и длины волны, и сами функции ведут себя монотонно, что не позволяет учесть потенциальные скачки в спектрах и выделенные спектральные линии поглощения. Это приводит к тому, что подобные теоретические модели не способны точно описать поведение наблюдаемого источника. Для более корректных приближений необходимо подключить к рассмотрению различные модели звездных атмосфер, в первую очередь потому, что спектр может зависеть также от давления, концентрации и электронной плотности, а также стоит учитывать, что температура неоднородно распределена по объему объекта, что в свою очередь влияет и на излучение, зависящее от температуры. Однако в данной работе приближение теоретическими однородными спектрами является достаточным для оценки искомых параметров и включение в анализ звездных атмосфер на данном этапе является неоправданным усложнением модели, хоть и является желаемой оптимизацией алгоритма [19].

Вычисление наблюдаемого потока требует различных подходов в зависимости от рассматриваемого обзора звездного неба. Каталоги Gaia предоставляют данные о количестве фотоэлектронов в секунду, попавших на детекторы в различных фильтрах. Обработка подобных единиц измерения и получения из них потоков описана в документации к миссии [2]. Однако более корректным подходом является расчет модельных спектров в единицах Gaia, что и было проделано в данной работе. Все вычисления были проведены в системе Vega. Переход к нужным величинам для теоретических потоков осуществляется с помощью домножения получаемой величины на фактор $\frac{\lambda}{hc}$ для получения спектрального распределения фотонов в единицах $\frac{photon}{s.mm.m^2}$. После этого проводится интегрирование по всем длинам волн с учетом кривой пропускания фильтра. Для корректного результата также требуется домножить полученный результат на площадь телескопа ($P_a = 0.7278m^2$). Подобная процедура необходима для использования и сравнения синтетической модели и результатов, полученных Gaia. Стоит отметить, что работа в единицах измерения телескопа значительно уменьшает ошибку вычислений и является предпочтительнее, чем перевод этих единиц в обычные потоки $(\frac{erg}{s \cdot nm \cdot nm^2})$. Поэтому финальная формула для вычисления модельного спектра принимает вид (3.3)

$$F(\lambda, T) = P_a \cdot \int_{\lambda} filt(\lambda) \cdot \frac{\lambda}{hc} F_{teor}(\lambda, T) \cdot \dots d\lambda, \qquad (3.3)$$

где остальные необходимые члены для вычисления потока будут расписаны далее.

Работа с данными, полученными из обзора 2MASS, существенно отличается. К сожалению, прямой подход с вычислениями в единицах измерения телескопа в данном случае невозможен, что затрудняет процесс анализа и увеличивает погрешности вычислений. Архив 2MASS предоставляет информацию о звездных величинах для каждого фильтра, которые необходимо перевести в потоки.

Формула для перехода к нужным единицам требует учета нулевой точки в системе Vega, что, в свою очередь, ведет к необходимости получения потока от Vega, значения для которого были получены из архива CALSPEC⁶, предоставляющего данные в открытом доступе. Благодаря информации о потоке в диапазоне длин волн λ [нм] \in (90, 299000) были получены нулевые точки для каждого фильтра (3.4)

$$ZP_{Vega} = \frac{\int_{\lambda} Flux(\lambda) \cdot filt(\lambda) \cdot d\lambda}{\int_{\lambda} filt(\lambda) d\lambda}$$
(3.4)

С учетом полученных значений стало возможным посчитать потоки для объектов, наблюдаемых 2MASS, из данных о звездных величинах (3.5)

$$F_{obj} = \int_{\lambda} Z P_{Vega} \cdot 10^{-0.4G_{star}} \, d\lambda, \qquad (3.5)$$

где G_{star} - звездная величина объекта из обзора.

Как видно, из-за вынужденных дополнительных вычислений результаты получены с большей погрешностью, так как теперь необходимо учитывать ошибку в расчете нулевой точки.

⁶www.stsci.edu by Bohlin, Gordon, Tremblay (2014)

Теоретический поток для случая 2MASS выглядит следующим образом (3.6)

$$F(\lambda, T) = \int_{\lambda} filt(\lambda) \cdot F_{teor}(\lambda, T) \cdot \dots d\lambda, \qquad (3.6)$$

Основной задачей данной работы является вероятностная оценка параметров поглощения межзвездной среды, что приводит к необходимости добавления членов в формулы 3.3 и 3.6, отвечающие за поглощение, введенных в работе [6]. Тогда формулы для вычисления синтетических моделей примут вид (3.7), (3.8):

для Gaia:

$$F(\lambda,T) = P_a \cdot \int_{\lambda} filt(\lambda) \cdot \frac{\lambda}{hc} \cdot F_{teor}(\lambda,T) \cdot (\frac{R}{D})^2 \cdot 10^{0.4 \cdot A_0(\lambda,A_\lambda,R_V)} d\lambda, \quad (3.7)$$

для 2MASS:

$$F(\lambda, T) = \int_{\lambda} filt(\lambda) \cdot F_{teor}(\lambda, T) \cdot (\frac{R}{D})^2 \cdot 10^{0.4 \cdot A_0(\lambda, A_\lambda, R_V)} d\lambda, \qquad (3.8)$$

В этих формулах также появляется фактор, отвечающий за отношение радиуса объекта к расстоянию до него, необходимый для корректного построения теоретической модели. Вводится также аналогичное (3.7) и (3.8) формулам выражение, но с заменой квадрата радиуса на светимость для удобства дальнейших вычислений (3.9), (3.10)

$$F(\lambda,T) = P_a \int_{\lambda} filt(\lambda) \cdot \frac{\lambda}{hc} \cdot F_{teor}(\lambda,T) \cdot (\frac{1}{D})^2 \cdot \frac{L}{4\pi\sigma_{SB}T^4} \cdot 10^{0.4 \cdot A_0(\lambda,A_\lambda,R_V)} d\lambda,$$
(3.9)

$$F(\lambda,T) = \int_{\lambda} filt(\lambda) \cdot F_{teor}(\lambda,T) \cdot (\frac{1}{D})^2 \cdot \frac{L}{4\pi\sigma_{SB}T^4} \cdot 10^{0.4 \cdot A_0(\lambda,A_\lambda,R_V)} d\lambda, \quad (3.10)$$

Наконец, после проведения подобных преобразований, полученные величины можно использовать для анализа. Сравнение синтетических спектров с наблюдаемыми проводится в процессе разработанного алгоритма с использованием байесовской статистики. Основная идея состоит в реализации теоретических моделей с незакрепленными параметрами и поиском наилучшего совпадения с наблюдаемым. В данном случае нефиксированными переменными остаются эффективная температура T_{eff} , расстояние D, радиус R (или светимость L) и параметры межзвездного поглощения, A_{λ} и R_V .

4 Параметры межзвездного поглощения

4.1 Теория

Межзвездная среда заполнена неоднородным веществом, поглощающим и рассеивающим свет, что приводит к тому, что детектируемые с Земли спектры звезд отличаются от испускаемых. Данное явление было обнаружено более столетия назад [20] и непрерывно изучалось с тех пор. Один из самых распространенных методом измерения параметра поглощение основан на идее сравнения спектров двух звезд, у одного из которых этот параметр известен (4.11) (метод "стандартных свечей"[5]).

$$A_{\lambda}[mag] = 2.5 \cdot \log_{10} \cdot \left(\frac{F_{\lambda}^0}{F_{\lambda}}\right),\tag{4.11}$$

где F_{λ}^0 - поток от объекта с известным параметром поглощения (значение которого становится нулевой точкой отсчета); F_{λ} - поток от изучаемого объекта

Однако одним из наибольших недостатков данного подхода является необходимость наличия звезды со схожими с исследуемым объектом характеристиками, но при этом спектр которой не подвержен значительному поглощению межзвездной среды, поэтому задача о нахождении параметра поглощения требовала построения более эффективного подхода. Решение данной проблемы было предложено Карделли, в работе [6] которого была показана возможность построить полином, корректно описывающий взаимосвязь между параметром поглощения и длиной волны. Данная параметризация (позже доработанная в [21]) стала удобным способом описания характеристик поглощения межзвездной среды.

4.2 Закон Карделли

В данной работе использован закон поглощения из работы [6], в которой описан процесс получения и связи между параметрами межзвездного поглощения. Наряду с параметром, описывающим поглощение межзвездной среды вводится также селективный параметр поглощения, заданный отношением обычного параметра поглощения к разности его значений на двух длинах волн (4.13). Закон межзвездного поглощения имеет вид (4.12)

$$A_{\lambda} = A_0(a(\lambda) + \frac{b(\lambda)}{R_V}), \qquad (4.12)$$

где $a(\lambda)$ и $b(\lambda)$ - коэффициенты, заданные полиномами, зависящих от длины волны λ . A_0 - монохроматическое поглощение - и R_V - отношение полного поглощения к селективному (4.13) - в свою очередь от λ не зависят.

$$R_V = \frac{A_\lambda}{E(B-V)},\tag{4.13}$$

где E(B-V) - разность поглощения в полосе B (на длине волны 440.5 нм) и V (на длине волны 547.0 нм)

Выбор данной модели закона поглощения также обусловлен тем, что она создавалась именно для ближнего ИК и оптического диапазонов (также и для ультрафиолетового диапазона, но он в данной работе не рассматривается). Закон [6] построен на экспериментальных данных (в фильтрах: U, B, V, R, I, J, H, K, L).

Предложенная Карделли модель описывает поведение параметра поглощения, зависящего только от длины волны в диапазоне 125 нм - 3500 нм и параметра R_V . Основным преимуществом данного подхода является возможность корректно построить кривую поглощения при знании лишь одного параметра R_V , который, в свою очередь, можно определить, основываясь на дополнительных исследованиях. Благодаря построению такой параметризации стало возможно определить среднее значение параметра поглощения в Галактике для фиксированного $R_V = 3.1$, а также удалось отследить зависимость значения R_V и распределения плотности пылевых облаков на звездном небе, что привело к прямому использованию селективного параметра как характеристики размеров гранул пыли. Данный вывод дал существенный вклад в описание размеров и плотности пыли в изучаемых областях, так как экспериментально было обнаружено, что большие значения соответствуют более плотным областям пыли [21]. В работе Карделли было также подтверждено отсутствие зависимости закона поглощения R_V от λ на больших длинах волн (>700 нм)

Параметризация была выведена для среднего значения отношения параметра $A_{\lambda} \kappa A_{V}$, параметру в полосе V (выбор подобной нормализации был, в основном, обусловлен наличием значительного количества подробных данных именно для этой полосы) и имеет простой вид (4.14)

$$\langle A(\lambda)/A(V)\rangle = a(\lambda) + \frac{b(\lambda)}{R_V},$$
(4.14)

где $a(\lambda)$ и $b(\lambda)$ - коэффициенты, заданные полиномами, зависящими от длины волны λ , которые для ближнего инфракрасного и оптического диапазонов имеют вид (4.15)-(4.16)

$$300nm^{-1} \le x \le 1100nm^{-1}$$
:
 $a(x) = 0.574x^{1.61}$
 $b(x) = -0.527x^{1.61}$ (4.15)

$$\begin{split} 1100nm^{-1} &\leq x \leq 3300nm^{-1}, y = (x - 1.82): \\ a(x) &= 1 + 0.17699y - 0.50447y^2 - 0.02427y^3 + 0.72085y^4 \\ &\quad + 0.01979y^5 - 0.77530y^6 + 0.32999y^7 \\ b(x) &= 1.41338y + 2.28305y^2 + 1.07233y^3 - 5.38434y^4 \\ &\quad - 0.62251y^5 + 5.30260y^6 - 2.09002y^7 \end{split} \tag{4.16}$$

5 Байесовский подход

5.1 Теория

Для анализа и сравнения построенных модельных спектров с наблюдаемыми использовалась байесовская статистика для получения вероятностного распределения искомых параметров [22].

Согласно теореме Байеса вероятность реализации события A при известной дополнительной информации B(P(A|B)) будет равна отношению произведения вероятности события B при условии события A(P(B|A)) и вероятности события P(A) без наложения дополнительных условий к вероятности события P(B) без наличия добавочных условий (5.17)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)},$$
(5.17)

В данном случае P(B|A) приобретает смысл правдоподобности события, а P(A) и P(B) - априорных знаний о событиях. Их соотношение в результате дает апостериорное распределение вероятности события. Однако помимо этих членов в формуле также присутствует значение вероятности, отвечающее за априорные знания об исследуемом объекте P(B), что в реальности является неизвестной переменной. Поэтому введенное соотношение принимает вид (5.18)

$$P(A|B) \sim P(B|A)P(A), \tag{5.18}$$

Подобное приближение не является критичным, наоборот, преимущество

байесовской статистики заключается именно в возможности исключить прямое использование неизвестных переменных, представляющих небольшой для анализа интерес, и учесть их впоследствии в качестве константы.

Теперь становится возможным поиск необходимых параметров в процессе исследования (5.18). Если переходить от описания вероятности событий к описанию модели со свободными параметрами, то становится ясно, что сформулированный подобным образом закон позволяет проводить анализ различных моделей с варьирующимися параметрами для нахождения наилучшего совпадения, если задавать условия в виде минимизации χ^2 или максимизации правдоподобности $L(\overline{X}|\Theta)$. В данной работе применен второй подход с использованием модели нормального распределения для функции правдоподобия (6.25), тогда итоговая формула имеет вид (5.19)

$$P_{likelihood}(\overline{X}|\Theta)) = \prod_{i=0}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(Model_i(X_i|\Theta)) - func_i)^2}{2\sigma_i^2}\right), \quad (5.19)$$

где σ - ошибка данных func; Model $(\overline{X}|\Theta)$ - это модель со свободными параметрами, а func - это данные.

6 Приложение вероятностной модели к поставленной задаче

Описанный выше байесовский вероятностный подход при приложении к поставленной задаче приобретает вид (6.20)

$$P_{post}(dist, L, T_{eff}, A_{\lambda}, R_{V}|flux) \sim P_{likelihood}(flux|dist, L, T_{eff}, A_{\lambda}, R_{V}) \cdot P_{prior}(L, T_{eff}, dist|HRD, plx, plxerr, rlen), (6.20)$$

где априорная вероятность для значений температуры, светимости и расстояния заданы с помощью данных, полученных из диаграммы Герцшпрунга-Рассела (*HRD*) и данных из базы данных о расстоянии, созданной на основе данных из Gaia [23], [24]

Очевидно, что при данном задании априорной вероятности условия на параметры светимости и температуры не зависят от задания условий на расстояние, поэтому их можно разделить как независимые переменные, тогда формула (6.20) сводится к (6.21)

$$P_{post}(dist, L, T_{eff}, A_{\lambda}, R_{V}|flux) \sim P_{likelihood}(flux|dist, L, T_{eff}, A_{\lambda}, R_{V}) \cdot P_{prior}(dist|plx, plxerr, rlen) \cdot P_{prior}(L, T_{eff}|HRD), \quad (6.21)$$

Метод получения априорной вероятности для светимости и эффективной температуры построен на основе данных из диаграммы Герцшпрунга-Рассела. Благодаря тому что обзор Gaia содержит данные более чем о 1,5 миллиардах звездных объектов становится возможным построить диаграмму



Рис. 6.3: Диаграмма Герцшпрунга-Рассела как априорное распределение параметров светимости и эффективной температуры

распределения звезд по параметрам светимости и эффективной температуры [25] исключительно по экспериментальным данным⁷ рис. 6.3

На рис. 6.3 изображено распределение звезд по светимости и эффективной температуре в логарифмическом масштабе. Плотность распределения отображена на графической шкале. Стоит отметить, что изначальными данными для построения диаграммы являлись звездные величины M и эффективная температура T_{eff} . Для корректного перехода от звездных величин к светимости был применен подход, подробно описанный в документации⁸ обзора [2]. Зависимость светимости от звездной величины в таком случае принимает вид (6.22)

$$-2.5 \cdot L = M_G + BC_G(T_{eff}) - M_{bol}, \tag{6.22}$$

⁷https://vlas.dev/post/gaia-dr2-hrd/

 $^{{\}it 8'https://gea.esac.esa.int/archive/documentation/GDR2/Data analysis}$

где L - светимость в единицах солнечной светимости; $M_{bol} = 4.74mag$ абсолютная болометрическая величина, фиксирующая значение светимости Солнца; $BC_G(T_{eff})$ - болометрическая поправка, зависящая только от температуры и заданная в виде полинома (6.23)

$$BC_G(T_{eff}) = \sum_{i=0}^{4} a_i \cdot (T_{eff} - T_{eff}(Sun))^i, \qquad (6.23)$$

где a_i - коэффициенты, приведенные в таблице 6.4 посчитанные для значений болометрической поправки и ошибки к ней [26], предоставленные к общему доступу в документации⁹ Gaia DR2.

-	N	
	$BC_{ m G}$	$\sigma(BC_{\rm G})$
	4000-8	8000 K
a_0	$6.000 imes10^{-02}$	$2.634 imes10^{-02}$
a_1	$6.731 imes10^{-05}$	$2.438 imes10^{-05}$
a_2	$-6.647 imes10^{-08}$	$-1.129 imes10^{-09}$
a_3	$2.859 imes10^{-11}$	$-6.722 imes10^{-12}$
a_4	$-7.197 imes10^{-15}$	$1.635 imes10^{-15}$
	3300-4	4000 K
a_0	$1.749 imes10^{+00}$	$-2.487 imes 10^{+00}$
a_1	$1.977 imes10^{-03}$	$-1.876 imes10^{-03}$
a_2	$3.737 imes10^{-07}$	$2.128 imes10^{-07}$
a_3	$-8.966 imes10^{-11}$	$3.807 imes10^{-10}$
a_4	$-4.183 imes10^{-14}$	$6.570 imes10^{-14}$

Рис. 6.4: Коэффициенты полинома для болометрической поправки (6.23)

Для определения априорной вероятности распределения параметра расстояния были рассмотрены работы [23], [24], [27], в которых расписан метод получения вероятностной оценки расстояния по данным параллакса и ошибки параллакса из Gaia. Выбор подхода, использованного в данной работе, был обусловлен, в первую очередь, согласованностью с номером рассматриваемого обзора (Gaia DR2) и его обработкой [23], но также поиском оптимального решения, не несущего излишней точности и впоследствии чрезмерных

 $^{^9\ &#}x27;https://gea.esac.esa.int/archive/documentation/GDR2/Data analysis$

затрат по времени. В работе [23] приведен алгоритм статистической оценки параметра расстояния с использованием априорного распределения в виде экспоненциально убывающей функции и функции правдоподобия нормального распределения (6.24)

$$P\left(dist \mid plx, \sigma_{plx}, L_{\rm sph}(l, b)\right)$$

$$= \begin{cases} \frac{1}{Norm} dist^{2} \exp\left[-\frac{dist}{L_{\rm sph}(l, b)} - \frac{1}{2\sigma_{plx}^{2}} \left(plx - \varpi_{\rm zp} - \frac{1}{dist}\right)^{2}\right] & \text{если } dist > 0\\ 0 & \text{иначе.} \end{cases}$$

$$(6.24)$$

где dist - искомое расстояние; L(l,b) - это порядок длины, зависящий от широты и долготы в галактической системе координат; plx - это параллакс; σ_{plx} - ошибка параллакса; ϖ_{zp} - нулевая точка параллакса в системе Gaia

Так выглядит апостериорное распределение параметра расстояния для данных параллакса, взятых из второго релиза Gaia. В данной работе полученная апостериорная вероятность (6.24) используется в качестве априорных знаний.

Возвращаясь к общей задаче, перед окончательным переходом к получению апостериорной оценки искомых параметров, выбранных в данной работе, необходимо ввести функцию правдоподобия. Было принято решение о рассмотрении модели с нормальным распределением (6.25), когда расхождение теоретической модели с наблюдаемыми данными будет приводить к экспоненциальному затуханию.

23

$$P_{likelihood}(flux|dist, L, T_{eff}, A_{\lambda}, R_{V}) =$$

$$= \prod_{i=0}^{N} \frac{1}{\sqrt{2\pi\sigma_{flux_{i}}^{2}}} \exp\left(\frac{-(Flux_{model}^{i} - Flux_{obj}^{i})^{2}}{2\sigma_{Flux_{i}}^{2}}\right)$$
(6.25)

Стоит также отметить, что из-за отсутствия априорных знаний о параметрах межзвездного поглощения единственные ограничения, которые были введены для этих переменных, - это верхние и нижние границы значений параметров. За априорную оценку было принято решение брать неинформативное априорное распределение (6.26)

$$P(\theta) \sim \frac{1}{\theta},\tag{6.26}$$

Тогда после введения всех нужных величин, наконец, стало возможным ввести само апостериорное распределение для свободных параметров (6.27)

$$P_{post}(dist, L, T_{eff}, A_{\lambda}, R_{V}|flux) = \frac{1}{Norm} P_{prior}(L, T_{eff}|HRD) \cdot P_{prior}(dist|plx, plxerr, rlen) \cdot P_{likelihood}(flux|dist, L, T_{eff}, A_{\lambda}, R_{V}) \cdot P_{uninf.prior}(A_{\lambda}, R_{V})$$
(6.27)

7 Метод Монте-Карло и Цепи Маркова

Приведенное выше описание вероятностей, используемое в теории, основанной на законе Байеса, требует корректного и оптимального применения для получения наглядных результатов. Для этого был привлечен дополнительный модуль, в котором байесовский подход применяется в методе Монте-Карло с привлечением теории цепей Маркова. Подобный выбор в применении этого подхода был сделан по большей части из-за значительных преимуществ данного алгоритма (MCMC) для нахождения апостериорного распределения при большом количестве свободных параметров. MCMC позволяет проводить анализ в многомерном пространстве параметров за допустимое (оптимальное) количество времени.

7.1 Теория

Метод Монте-Карло с использованием марковских цепей состоит из простой итерационной схемы, позволяющей получить выборку из непрерывного или сложного распределения. В данной работе выборка необходима из апостериорного распределения. В формуле (5.17) было показано, что апостериорная вероятность зависит от априорных знаний об объекте исследования без дополнительных условий P(B) - фактор, который невозможно или неоправданно затратно высчитывать для решения задачи. Поэтому необходимо использовать другой подход, в котором бы нормализующим фактором можно было бы пренебречь или выразить через константу. Основным преимуществом использования цепей Маркова является тот факт, что вероятность оказаться в каком-либо состоянии зависит лишь от состояния, в котором система находится на данный момент, а не от всех предыдущих шагов (7.28)

$$P(A_{i+1} = p_{i+1} | A_i = p_i, A_{i-1} = p_{i-1}, A_{i-2} = p_{i-2} \dots)$$

= $P(A_{i+1} = p_{i+1} | A_i = p_i),$ (7.28)

где p_i - состояние, в котором система A может оказаться.

Самым простым примером применения МСМС является процесс, состоящий из пяти простых пунктов:

- 1. Выбор случайного начального состояния (№1)
- 2. Выбор случайного состояния на следующем шаге (№2)
- 3. Проверка вероятности оказаться в состоянии, выбранном на Шаге 2
- Сравнение вероятности находится в состоянии №2 с вероятностью находиться в состоянии №1
- 5. Если вероятность оказаться в состоянии №2 больше, то продолжить цепь, начиная с Шага 2, иначе остаться на месте и повторить с Шага 2

7.2 Использование пакета МСМС

Для получения результатов вероятностного распределения значений параметров был использован пакет [28] *етсее*, использующий метод MCMC, совместно с байесовской статистикой.

Алгоритм работы пакета *emcee* [28] заключается в одновременном построении сразу большого количества выборок, основываясь на апостериорной вероятности. Число выборок задается с помощью вспомогательных элементов (walkers), прописанных в структуре кода, основной функцией которых является исследование пространства параметров. Благодаря использованию цепей Маркова рассмотрение всевозможных значений параметров стало оптимальной для реализации задачей, вследствие того что следующий рассматриваемый вектор значений параметров зависит только от состояния системы, в котором она находится, без влияния того, как она в этом состоянии оказалась. Такой подход оправдывает приближение, сделанное в (5.18).

Для реализации алгоритма с использованием *етсее* согласно байесовской статистике также необходимо введение дополнительных функций для расчета: функции правдоподобности системы и априорных вероятностей, имеющих вид функций из (6.27) и задающих апостериорное распределение, выборка из которого и выполняется с помощью цепей Маркова описанным выше способом. Однако в данном случае для удобства вычисления проводятся в логарифмических единицах (тогда функция правдоподобности принимает вид (7.29))

$$log(P_{likelihood}(flux|dist, L, T_{eff}, A_{\lambda}, R_{V})) =$$

= (-0.5) $\sum_{i=0}^{N} (log(2\pi\sigma_{flux_{i}}^{2}) + (\frac{-(Flux_{model}^{i} - Flux_{obj}^{i})^{2}}{2\sigma_{Flux_{i}}^{2}}))$ (7.29)

После введения всех вспомогательных функций итерационный процесс алгоритма можно свести к нескольким простым шагам:

- 1. Выбор вектора начального распределения параметров (также определяется разброс величин от начального распределения, значение которых становятся начальными для каждого элемента, исследующего пространство параметров (walkers))
- Нахождения априорной вероятности с набором значений переменных из №1

- 3. Вычисление функции правдоподобности со значениями из №1
- 4. Определение апостериорной вероятности для векторов параметров из №1
- 5. Выбор следующего вектора начального состояния
- 6. Вычисление пунктов №2, 3, 4 с новыми данными из №5
- 7. Если вероятность оказаться в состоянии №5 больше, то цепь продолжается с этого вектора значений, иначе возвращается к состояние из №1 и ищется другое новое состояние №5

Видно, что в таком случае следующий шаг действительно не зависит от чего-то кроме состояния системы в момент рассмотрения. К тому же становится ясным, что количество walkers, задающих количество одновременно рассматриваемых начальных векторов значений, которые продвигаются по пространству параметров, напрямую влияет на время схождения цепей к значениям параметров близких к истине (согласно [28] вклад количества walkers значительнее, чем вклад количества итераций).

Наконец, после использования алгоритма *етсее* становится возможным реализация основной цели данной работы - анализ полученных результатов статистического распределения искомых параметров.

8 Анализ результатов и дальнейшая работа

Все упомянутые в этой главе графики можно найти в Приложении В в секции Графики. В качестве визуализация полученных значений были построены контуры областей с использованием модуля corner.py [29], являющиеся проекциями цепей Маркова в пространстве параметров и иллюстрирующие вероятностное распределение значений изучаемых параметров. Кроме того, на графиках такого вида стало возможно проследить за относительным распределением параметров и выявить возможные связи между ними. Помимо контуров также были построены гистограммы для каждого параметра, иногда демонстрирующие наличие пиков в распределении.

8.1 Проверка алгоритма на теоретических моделях

Проверка корректности работы алгоритма проводилась с помощью синтетических моделей, построенных аналогично теоретическим (3.9) и (3.10), но с зафиксированными значениями параметров. Значения параметров были взяты из Приложения А работы [30] и выбраны таким образом, чтобы спектры соответствовали звездам главной последовательности для разных спектральных классов. Для анализа были выбраны звезды классов: *О*9, *B*5, *A*1, *A*7, *F*3, *F*9, *G*5, *K*1, *K*7, *M*3, *M*9 с параметрами из табл. 8.1

В отличие от расчета вероятностного распределения для реальных объектов с правильно выбранным распределением расстояния до него, для теоретических моделей подобный подход невозможен, из-за чего было принято решение выбрать равномерное распределение для априорных знаний о расстоянии. Данное приближение существенно влияет на точность определение параметра расстояния. что приводит к сдвигу в значениях и других величин, однако даже так результаты близки к заданным изначально значениям.

Спектральный	Радиус,	Эффективная	Расстояние,	$A_V,$	$R_V,$
класс	Rsun	температура, К	pc	mag	mag
O9	6.32	32961.0	10.0	2	5.1
B5	3.55	14800.0	10.0	2	5.1
A1	2.37	9480.0	10.0	2	5.1
A7	1.95	7840.0	10.0	2	5.1
F3	1.58	6770.0	10.0	2	5.1
F9	1.09	6010.0	10.0	2	5.1
G5	0.92	5570.0	10.0	2	5.1
K1	0.78	5081.0	10.0	2	5.1
K7	0.55	4064.0	10.0	2	5.1
M3	0.23	3320.0	10.0	2	5.1
M9	0.10	2360.0	10.0	2	5.1

Таблица 8.1 Параметры синтетических моделей

На рис. 12.5 - 12.9 четко видно соотношения распределений найденных величин с заданными параметрами. Гистограммы, наглядно демонстрирующие пики распределений, также предоставляют информацию о наиболее вероятном значении и интервале ошибки. Для анализа распределения и получения конкретных величин параметров были посчитаны квантили (процентали) с вероятностными значениями [0.16, 0.5, 0.84], которыми фиксируются нижняя и верхняя границы, а также среднее значение. В данном случае распределение не является нормальным, поэтому выбор равновесных нижних и верхних граней не несет в себе дополнительной точности.

На рис. 12.5 для синтетических моделей классов *O*9, *B*5 хорошо видно, что для сильно горячих звезд алгоритм дает некорректные результаты. Это легко объясняется тем, что в данной работе в качестве априорных данных была взята диаграмма Герцшпрунга-Рассела (*HRD*), построенная на основе данных из обзора Gaia (рис. 6.3). Как видно, на диаграмме практически отсутствуют горячие звезды (температура не превышает 10000 K), что приводит к необходимости подключения других каталогов для учета звезд классов *O*, *B*.

Для классов звезд A1, A7 результаты (рис. 12.6) уже входят в пределы выбранных интервалов ошибок. Как было упомянуто выше, простой выбор априорных знаний о расстоянии дает неточности в определении этого параметра. При сравнении сочетаний параметров видно, что вклад некоторых комбинаций, согласно *HRD*, имеет большую вероятность. Так, например, на графике для звезды A7 (рис. 12.6), видно, что звезда с данной температурой более вероятно будет иметь меньшую светимость, но находиться ближе по расстоянию. При задании более точных априорных данных для параметра расстояния вклад априорных данных из *HRD* имел бы не такой большой вес и оценка всех параметров была бы корректнее. Аналогичные проблемы возникают и для звезд классов *F3*, *F9*, *G5*, *K1*, *K7*, *M3*, однако из рис. 12.7 - 12.8 видно, что для данного диапазона температур совпадения полученных величин с истинными наиболее точные, что логично объясняется существенным количеством начальных данных (*HRD*) именно для таких звезд.

Звезды с очень низкими температурами плохо разрешаются алгоритмом по той же причине, что и сильно горячие (проверка проводилась для звезды класса *M9* рис. 12.9). Температуры ниже 3000 К не учитываются на диаграмме Герцшпрунга-Рассела, что приводит к слабым априорным данных и таким результатам.

Стоит отметить, что несмотря на наличие сильной зависимости от данных из *HRD* при рассмотрении синтетических моделей, данная связь хоть и присутствует в работе с реальными объектами, но имеет меньший вклад, так

31

как для наблюдаемых объектов используются известные данные о расстоянии. Благодаря ненулевой вероятности значений параметров вне заполненных областей на диаграмме Г-Р, с помощью алгоритма возможно рассмотрение звезд неучтенных на *HRD* спектральных классов.

8.2 Применение к наблюдаемым данным

После проверки на синтетических моделях алгоритм был задействован в анализе реальных звезд. Пример запроса (ADQL query) на получение и объединение данных из обзоров Gaia DR2 и 2MASS приведен в Приложении А. Приведенный пример задает координаты области звездообразования Lupus V, объекты которого находятся на расстоянии ≤ 200 pc. Графики, полученные для данной области, приведены в Приложении В (рис. 12.10-12.11). Синими линиями на графиках обозначены значения параметров, полученных для расстояния из [23], а для остальных параметров из обзора Gaia DR2 [2].

Для иллюстрации вывода результатов были построены контуры статистического распределения параметров для одного точечного объекта (рис. 12.12 (b)), представляющие собой проекции цепей Маркова, и вывод их самих для наглядности алгоритма построения (рис. 12.12 (a)). Таким образом схождение цепей Маркова к какому-либо значению может считаться схождением к самому вероятностному распределению свободных параметров. Как видно на гистограммах, почти все распределения параметров имеют ярко выраженные пики, что делает возможным вывод о наиболее вероятных значениях искомых параметров.

Тем не менее, анализ отдельного объекта не несет в себе полной картины статистического распределения параметров межзвездной среды. Именно поэтому основной целью данной работы является рассмотрение областей звездного неба с последующим выводом о свойствах межзвездной среды. Имеет смысл рассматривать звезды из одной области, находящиеся на приблизительно одном расстоянии для объективного и оптимального на данном этапе исследования выбранной области. Именно поэтому были построены графики, иллюстрирующие проведенный анализ данных о звездах из одной области (рис. 12.10-12.11)

В основном, распределение параметров межзвездной среды для одной области и рассматриваемых звезд, находящихся на приблизительно одинаковом расстоянии, схоже по структуре друг с другом. Однако любые отклонения в значениях для подобных объектов легко интерпретируется как изменение свойств среды по данному направлению. Именно такой подход в оценке характеристик межзвездной среды необходимо провести для корректного определения параметров межзвездного поглощения. Поэтому помимо построения статистического распределения свободных параметров требуется также проведение сравнительного анализа, для предоставления выводов о структуре и характеристиках межзвездной среды.

Не все объекты возможно корректно проанализировать разработанным методом. Неверные или недостаточные априорные данные могут сильно ухудшить работу алгоритма. Несмотря на то что обзор Gaia содержит в себе информацию об огромной количестве объектов, в том числе и об их параллаксе, определение которого является одной из основных целей этой миссии, эти значения определены не всегда, что сразу же лишает возможности использовать точные данные для этого параметра в априорной вероятности. Это приводит к значительному увеличению неопределенности в вычислениях. Еще одной причиной для слабых результатов может являться наличие объектов в обзоре Gaia, не являющихся звездами (например, галактики), но рассмотренные алгоритмом как таковые.

Описанные выше сложности могут привести к некорректному построению

33

вероятностного распределения, и марковские цепи или не сходятся, или сходятся к начальному распределению, что не несет в себе никаких дополнительных знаний об объекте и окружающей его среде. Один из таких примеров изображен на рис. 12.13, на котором видно отсутствие выделенных пиков в распределении, что означает расхождение цепей Маркова в пространстве параметров. Такие случаи требуют дополнительного анализа результатов для правильной интерпретации полученных значений параметров поглощения межзвездной среды, особенная необходимость в котором присутствует при изучении отдельных областей звездного неба для верного понимания усредненной структуры межзвездного вещества в них.

8.3 Дальнейшая работа

Анализ полученных результатов позволил увидеть и выделить пункты, которые существенно могут улучшить работу алгоритма. В первую очередь, в планах развития данной работы стоит выделить оптимизацию кода для создания интерфейса, позволяющего использовать его для получения значений параметров любого интересующего объекта с известными координатами. Помимо технической составляющей планируется подключение других обзоров звездного неба (PanSTARRS1, TYCHO и т.д.) для покрытия большего диапазона длин волн и увеличения количества информации об объектах. К тому же анализ поведения селективного параметра поглощения на разных длинах волн позволяет изучать различные характеристики межзвездного вещества, что может стать дополнительным направлением исследования. Как было замечено выше, на используемую диаграмму Г-Р необходимо внести данные о звездах класса О и ранние подклассы В, а также холодные звезды класса М. Частью планируемых улучшений алгоритма может стать возможность выбора спектра теоретической модели, то есть кроме использованной в настоящей работе функции Планка, в работу могут быть включены функции различных звездных атмосфер.

9 Заключение

Данная работа была посвящена разработке алгоритма, позволяющего получить вероятностную оценку на параметры поглощения в Галактике (A_{λ}, R_V) . Дополнительным аспектом работы стало исследование характеристик звезд (светимость L, эффективная температура T_{eff}) и оценка расстояния до них D.

В качестве проверки корректности работы алгоритма были построены синтетические модели звезд разных классов, анализ которых позволил увидеть интервалы параметров, для которых разработанный подход применим (классы: F, G, K, ранние подклассы M), а какие требует дополнительных априорных данных (классы: O, ранние подклассы B, поздние подклассы M).

При изучении реальных точечных объектов с помощью созданного алгоритма благодаря данным о параллаксе из обзора Gaia DR2 удалось также получить оценку на расстояние до рассматриваемых объектов с бо́льшей точностью, чем это было возможно для теоретических моделей. После применения разработанного метода к оценке параметров отдельных областей звездного неба был сделан вывод о необходимости дополнительного анализа полученных результатов для учета особенностей рассматриваемых объектов в ней. Конкретные заключения о характеристиках межзвездной среды в отдельных областях требуют усредненного анализа по звездам в ней для корректной интерпретации полученных результатов.

В настоящей работе приведен метод исследования параметров межзвездной среды и характеристик звезд с помощью алгоритма, написанного на языке Python. Дальнейшая оптимизация данного кода сделает возможным использование алгоритма в общем доступе, а также с помощью него станет возможным исследование областей звездообразования как часть крупномасштабной задачи об изучении молодых звездных объектов в Галактике.

36

10 Благодарности

Я хотела бы выразить благодарность своему научному руководителю. Полу Боли, за помощь в выборе направления работы и за поддержку на протяжении всего процесса решения поставленной задачи, а также за полученные знания и возможность развиваться в интересующей меня области астрофизики. Мне хотелось бы поблагодарить всю группу, занимающуюся наблюдательной астрофизикой в Лаборатории фундаментальных и прикладных исследований релятивистских объектов Вселенной МФТИ, за советы и комментарии при подготовке к защите дипломной работы. В настоящей работе были использованы данные из обзора Gaia, миссии Европейского Космического Areнства (ESA) (https://www.cosmos.esa.int/gaia), данные были обработаны Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Финансирование DPAC обеспечивалась национальными учреждениями, в частности учреждениями, участвующими в многостороннем соглашении по миссии Gaia. В работе также были использованы данные миссии 2MASS (Two Micron All Sky Survey), которая является совместным проектом Университета Массачусетса и Калифорнийского технологического института, финансируемые Национальным управлением по аэронавтике и исследованию космического пространства (NASA) и Национальным научным фондом. Работа выполнена при поддержке гранта PHΦ № 18-72-10132

Список литературы

- M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. V. Dyk, and S. Wheelock, "The two micron all sky survey (2mass)," *The Astronomical Journal*, vol. 131, no. 2, pp. 1163–1183, 2006.
- [2] Gaia Collaboration et al., "Gaia data release 2 summary of the contents and survey properties," Astronomy and Astrophysics, vol. 616, p. A1, 2018.
- [3] Gaia Collaboration et al., "Gaia early data release 3. summary of the contents and survey properties," *Astronomy and Astrophysics*, vol. 649, p. A1, 2021.
- [4] A. V. Zasov and K. A. Postnov, "Obshchaya astrofizika (general astrophysics)," *Fryazino: Vek 2*, pp. 91–133, 2011.
- [5] B. T. Draine, *Physics of the interstellar and intergalactic medium*. Princeton University Press, 2011.
- [6] J. A. Cardelli, G. C. Clayton, and J. S. Mathis, "The Relationship between Infrared, Optical, and Ultraviolet Extinction," *The Astrophysical Journal*, vol. 345, p. 245, 1989.
- M. Richard James Hanson, Mapping 3D extinction and structures in the Milky Way. Doctor of Natural Sciences dissertation, Ruperto-Carola-University of Heidelberg, 2017.
- [8] D. J. Schlegel, D. P. Finkbeiner, and M. Davis, "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds," *The Astronomical Journal*, vol. 500, no. 2, pp. 525–553, 1998.

- [9] J. A. Cardelli, G. C. Clayton, and J. S. Mathis, "The Determination of Ultraviolet Extinction from the Optical and Near-Infrared,", vol. 329, p. L33, 1988.
- [10] D. J. Marshall, A. C. Robin, C. Reylé, M. Schultheis, and S. Picaud, "Modelling the Galactic interstellar extinction distribution in three dimensions," *Astronomy and Astrophysics*, vol. 453, no. 2, pp. 635–651, 2006.
- [11] E. F. Schlafly, G. Green, D. P. Finkbeiner, M. Jurić, H.-W. Rix, N. F. Martin, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R.-P. Kudritzki, E. A. Magnier, N. Metcalfe, J. S. Morgan, P. A. Price, C. W. Stubbs, J. L. Tonry, R. J. Wainscoat, and C. Waters, "A MAP OF DUST REDDENING TO 4.5 kpc FROM pan-STARRS1," *The Astrophysical Journal*, vol. 789, no. 1, p. 15, 2014.
- [12] B.-Q. Chen, X.-W. Liu, H.-B. Yuan, H.-H. Zhang, M. Schultheis, B.-W. Jiang, Y. Huang, M.-S. Xiang, H.-B. Zhao, J.-S. Yao, and H. Lu, "A three-dimensional extinction map of the Galactic anticentre from multiband photometry," *Monthly Notices of the Royal Astronomical Society*, vol. 443, no. 2, pp. 1192–1210, 2014.
- [13] B. Q. Chen, M. Schultheis, B. W. Jiang, O. A. Gonzalez, A. C. Robin, M. Rejkuba, and D. Minniti, "Three-dimensional interstellar extinction map toward the Galactic bulge," *Astronomy and Astrophysics*, vol. 550, p. A42, 2013.
- [14] R. Lallement, J. L. Vergely, B. Valette, L. Puspitarini, L. Eyer, and L. Casagrande, "3D maps of the local ISM from inversion of individual color excess measurements," *Astronomy and Astrophysics*, vol. 561, p. A91, 2014.
- [15] G. M. Green, E. Schlaffy, C. Zucker, J. S. Speagle, and D. Finkbeiner, "A 3D Dust Map Based on Gaia, Pan-STARRS 1, and 2MASS," *The Astrophysical Journal*, vol. 887, no. 1, p. 93, 2019.

- [16] G. M. Green, E. F. Schlafly, D. P. Finkbeiner, H.-W. Rix, N. Martin, W. Burgett, P. W. Draper, H. Flewelling, K. Hodapp, N. Kaiser, R. P. Kudritzki, E. Magnier, N. Metcalfe, P. Price, J. Tonry, and R. Wainscoat, "A three-dimensional map of Milky Way dust," *The Astrophysical Journal*, vol. 810, no. 1, p. 25, 2015.
- [17] G. M. Green, E. F. Schlafly, D. Finkbeiner, H.-W. Rix, N. Martin, W. Burgett, P. W. Draper, H. Flewelling, K. Hodapp, N. Kaiser, R.-P. Kudritzki, E. A. Magnier, N. Metcalfe, J. L. Tonry, R. Wainscoat, and C. Waters, "Galactic reddening in 3d from stellar photometry - an improved map," *Monthly Notices of the Royal Astronomical Society*, vol. 478, no. 1, pp. 651–666, 2018.
- [18] B. Q. Chen, Y. Huang, H. B. Yuan, C. Wang, D. W. Fan, M. S. Xiang, H. W. Zhang, Z. J. Tian, and X. W. Liu, "Three-dimensional interstellar dust reddening maps of the Galactic plane," *Monthly Notices of the Royal Astronomical Society*, vol. 483, no. 4, pp. 4277–4289, 2019.
- [19] Н.А. Сахибуллин, "Теоретическая астрофизика. Звездные атмосферы часть 1," Учебное пособие, 2015.
- [20] E. E. Barnard, "On a nebulous groundwork in the constellation Taurus.," The Astrophysical Journal, vol. 25, pp. 218–225, 1907.
- [21] E. L. Fitzpatrick, "Correcting for the Effects of Interstellar Extinction," The Publications of the Astronomical Society of the Pacific, vol. 111, no. 755, pp. 63–75, 1999.
- [22] C. A. L. Bailer-Jones, Practical Bayesian Inference: A Primer for Physical Scientists. Cambridge University Press, 2017.
- [23] C. A. L. Bailer-Jones, J. Rybizki, M. Fouesneau, G. Mantelet, and R. Andrae,
 "Estimating distance from parallaxes. IV. distances to 1.33 billion stars in gaia
 data release 2," *The Astronomical Journal*, vol. 156, no. 2, p. 58, 2018.

- [24] C. A. L. Bailer-Jones, J. Rybizki, M. Fouesneau, M. Demleitner, and R. Andrae, "Estimating distances from parallaxes. v. geometric and photogeometric distances to 1.47 billion stars in gaia early data release 3," *The Astronomical Journal*, vol. 161, no. 3, p. 147, 2021.
- [25] Gaia Collaboration et al., "Gaia data release 2 observational hertzsprungrussell diagrams," Astronomy and Astrophysics, vol. 616, no. A10, 2018a.
- [26] R. Andrae, M. Fouesneau, O. Creevey, C. Ordenovic, N. Mary, A. Burlacu,
 L. Chaoul, A. Jean-Antoine-Piccolo, G. Kordopatis, A. Korn, Y. Lebreton,
 C. Panem, B. Pichon, F. Thévenin, G. Walmsley, and C. A. L. Bailer-Jones,
 "Gaia data release 2: first stellar parameters from apsis," Astronomy and Astrophysics, vol. 616, p. 29, 2018.
- [27] C. A. L. Bailer-Jones, "Estimating distances from parallaxes," *Publications of the Astronomical Society of the Pacific*, vol. 127, no. 956, pp. 994–1009, 2015.
- [28] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The mcmc hammer," PASP, vol. 125, pp. 306–312, 2013.
- [29] D. Foreman-Mackey, "corner.py: Scatterplot matrices in python," The Journal of Open Source Software, vol. 1, p. 24, jun 2016.
- [30] R. van Boekel, B. Benneke, K. Heng, R. Hu, N. Madhusudhan, S. Quanz, Y. Bétrémieux, J. Bouwman, G. Chen, L. Decin, R. de Kok, A. Glauser, M. Güdel, P. Hauschildt, T. Henning, S. Jeffers, S. Jin, L. Kaltenegger, F. Kerschbaum, O. Krause, H. Lammer, A. Luntzer, M. Meyer, Y. Miguel, C. Mordasini, R. Ottensamer, T. Rank-Lueftinger, A. Reiners, T. Reinhold, H. M. Schmid, I. Snellen, D. Stam, Z. Sun, and B. Vandenbussche, "The Exoplanet Characterization Observatory (EChO): performance model EclipseSim and applications," in *Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave* (M. C. Clampin, G. G. Fazio, H. A.

MacEwen, and J. Oschmann, Jacobus M., eds.), vol. 8442 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 84421F, 2012.

11 Приложение А. Код

SELECT

gaia.source_id, gaia.ra, gaia.dec, gaia.parallax, gaia.parallax_error,

- gaia.phot_g_mean_flux, gaia.phot_g_mean_flux_error, gaia.phot_bp_mean_flux,
- gaia.phot_bp_mean_flux_error, gaia.phot_rp_mean_flux,
- gaia.phot_rp_mean_flux_error,
- gaia.teff_val, gaia.teff_percentile_upper, gaia.radius_val,
- gaia.radius_percentile_upper, gaia.lum_val, gaia.lum_percentile_upper, dis.source_id,
- best.best_neighbour_multiplicity, best.number_of_mates, tm.ra, tm.dec, tm.j_m,
- tm.j_msigcom, tm.h_m, tm.h_msigcom, tm.ks_m, tm.ks_msigcom, dis.r_est,
- dis.r_len, dis.r_lo, dis.r_hi
- FROM gaiadr2.gaia_source AS gaia
- JOIN external.gaiadr2_geometric_distance as dis
- $ON dis.source_id = gaia.source_id$
- JOIN gaiadr2.tmass_best_neighbour AS best
- ON gaia.source_id = best.source_id
- JOIN gaiadr1.tmass_original_valid as tm
- $ON best.tmass_oid = tm.tmass_oid$
- WHERE gaia.dec > -38.5 AND gaia.dec < -36.3 AND gaia.ra > 243.5 AND
- gaia.ra<246.9

12 Приложение В. Графики

Проверка работоспособности алгоритма на синтетических моделях. Синими линиями показаны истинные значения.



Рис. 12.5: Проекции цепей Маркова для звезд разных спектральных классов: O9 (a), B5 (b)



Рис. 12.6: Проекции цепей Маркова для звезд разных спектральных классов: A1 (a), A7 (b)



Рис. 12.7: Проекции цепей Маркова для звезд разных спектральных классов: F3 (a), F9 (b), G5 (c)



Рис. 12.8: Проекции цепей Маркова для звезд разных спектральных классов: K1 (a), K7 (b)



Рис. 12.9: Проекции цепей Маркова для звезд разных спектральных классов: M9



Рис. 12.10: Пример анализа области Ra $\in (243.5; 246.9),$ Dec $\in (-38.5; -36.3)$ и построение проекций цепей Маркова для объектов в ней48



Рис. 12.11: Пример анализа области Ra \in (243.5; 246.9), Dec \in (-38.5; -36.3) и построение проекций цепей Маркова для объектов в ней



Рис. 12.12: Пример построения цепей Маркова (a) и их проекций (b) для отдельной звезды



Рис. 12.13: Проекции цепей Маркова, построенные при недостатке априорных знаний