

Определение параметров межзвездного поглощения света по данным каталога Hipparcos

Ф.А.Амосов, В.В.Витязев, А.С.Цветков

13 ноября 2015 г.

Аннотация

На основе данных космической миссии Hipparcos для близких звезд (до 500 пк) получена карта значений градиента покраснения в направлениях, определяемых центрами 3888 равновеликих площадок, построенных методом HealPix. Отсутствующие в каталоге Hipparcos данные о классе светимости 49542 звезд северного экваториального полушария были восстановлены с помощью линейного бинарного классификатора, построенного на обучающем множестве 39807 звезд спектральных классов III и V. Для III класса светимости точность классификатора равна 95%, полнота — 89%. Для V класса светимости соответствующие характеристики равны 91% и 96%. Для 98827 звезд каталога HIPPARCOS определены покраснения звезд по показателю цвета B-V. Произведено сравнение наших карт с аналогичными картами, по данным каталога 2MASS в ближней инфракрасной области.

1 Введение

Как известно [1], пылевые облака в Галактике приводят к общему ослаблению света звезд (экстинкции) и к покраснению звезд, то есть к увеличению их показателей цветов. Распределения покраснения звезд по небесной сфере может служить свидетельством о существовании межзвездной пыли в данном направлении.

Успех выявления пылевых облаков в двумерном и трехмерном варианте определяется имеющимися в нашем распоряжении массовыми каталогами звезд, содержащих информацию о координатах и параллаксах звезд, а также спектральные и фотометрические данные - спектральный класс, класс светимости и звездные величины в различных фотометрических полосах. За последние годы было осуществлено два грандиозных проекта, результаты которых позволили получить существенно продвижение в задаче об обнаружении облаков межзвездной пыли в Галактике. Первый проект — это миссия HIPPARCOS [3], позволившая получить на миллисекундном уровне точности значения параллаксов приблизительно для 100 000 звезд. Выполнение второго проекта — 2 MASS [4] позволило получить точную фотометрию примерно 470 миллионов звезд в ближней инфракрасной области (в полосах J — 1.2 μm , H — 1.6 μm и Ks — 2.2 μm). В работе [5] эти данные были использованы для исследования покраснения $E(J - K_s)$ 70 млн. звезд и построения трехмерной карты покраснения звезд в радиусе 1600 пк от Солнца. На основе этой работы была создана трехмерная карта межзвездного поглощения в ближайшем килопарсеке [6].

К сожалению, в каталоге Hipparcos спектральные классы звезд известны в основном только для отрицательных склонений. В настоящей работе поставлена задача устранить этот недостаток путем определения классов светимости звезд северного полушария с помощью метода машинного обучения [10], [11]. Теория обучения машин (machine learning, машинное обучение) — это самостоятельная математическая дисциплина, находящаяся на стыке прикладной статистики, численных методов оптимизации и дискретного анализа. В настоящее время технология машинного обучения нашла широкое применение в различных задачах обработки данных, в том числе и в астрономии [13]. Наша работа основана на построении бинарного классификатора, результатом работы которого является назначение звезде с известной абсолютной звездной величине класса светимости.

2 Исходные данные

В 2007 году вышла новая редакция астрометрических данных каталога Hipparcos [15] — каталог HIPNEWCAT (HIPparcos NEW astrometric CATalog). Утверждается, что точность положений, параллаксов и собственных движений всех звезд, ярче $H_P = 8$, улучшена в 4 раза, а для всех остальных звезд более, чем в 2 раза. Уменьшена взаимная корреляция параметров иногда в 10 раз. Именно эта версия использовалась в работе в качестве источника астрометрических данных.

В 2012 году вышел в свет каталог XHIP (An Extended Hipparcos Compilation) [17], в котором помимо сведений, содержащихся в HIPNEWCAT, собрана дополнительная богатая информация. Для 116 096 звезд приведена спектральная классификация, 46 392 звезд снабжены информацией о лучевых скоростях. В этой базе данных приведены результаты фотометрии в полосах I, J, K, заимствованной из каталога 2MASS. Для 114 315 звезд указаны спектральные классы, при этом для 73 842 звезд приведены классы светимости.

Каталог Hipparcos для большинства звезд содержит информацию о спектральном типе, полученную из наземных наблюдений. Основной источник — Мичиганский каталог [19] и несколько других каталогов. Однако, информация о спектральных классах приведена только для звезд южного экваториального полушария.

В нашей работе использовалась следующая информация:

- положения звезд (каталог Hipnewcat)

Таблица 1: Структура обучающего множества

класс светимости	$B - V < 0.6$	$B - V \geq 0.6$	всего
III	1947	16681	18628
V	15549	5630	21179
III & V	17496	22311	39807

- параллаксы звезд (каталог Hipnewcat)
- фотометрия (V_{mag} - каталог Hipparcos)
- класс светимости (каталог Hipparcos)

Дополнительные сведения об используемых данных можно найти в работах [7]- [9].

3 Способ получения классов светимости

Как мы увидели в обзоре данных каталога Hipparcos, практически у всех звезд северного экваториального полушария отсутствует класс светимости. Для нас его наличие чрезвычайно важно, ввиду того, что мы на основе класса светимости и спектрального класса рассчитываем истинное значение $B - V$ для звезд $(B - V)_{int}$. Тем самым, отсутствие класса светимости у половины звезд делает невозможным проведение наших расчетов для всего северного экваториального полушария.

Для исправления этого недостатка используем метод машинного обучения. Натренируем классификатор, который будет определять класс светимости для звезды по двум факторам — ее показателю цвета $B - V$ и ее абсолютной звездной величине M_V . Этих факторов должно быть достаточно, т.к. классы светимости теоретически разделимы на диаграмме Герцшпрунга-Рессела.

В каталоге HIPPARCOS доля звезд, которые не относятся ни к III, ни к V классам светимости мала (16,3%, 8058 из 49285). Поэтому, мы упростим задачу — обучим линейный бинарный классификатор, который будет предсказывать III или V класс. Сделаем это с помощью метода опорных векторов [11]. В качестве обучающего множества возьмем все звезды, у которых присутствует класс светимости III и V. Таких звезд 39807. Распределение этих звезд по классам светимости показано в таблице 1:

Диаграмма Герцшпрунга-Рессела, соответствующая нашему обучающему множеству, показана на рис. 1.

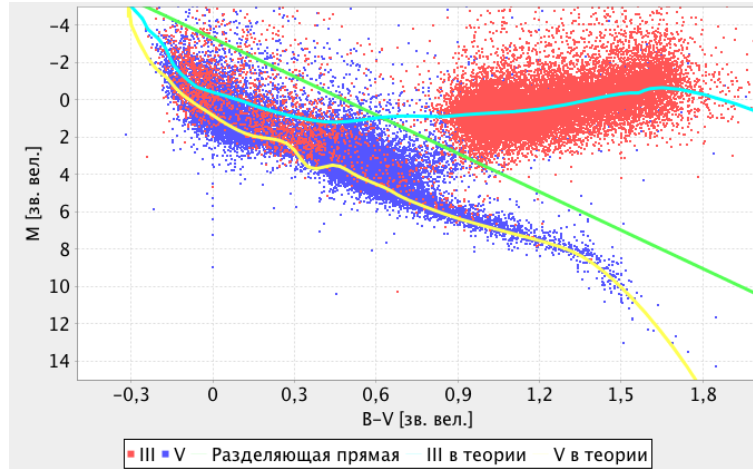


Рис. 1: Диаграмма Герцшпрунга-Рессела обучающего множества звезд (табл. 1). Разделяющая линия — зеленый цвет, теоретические кривые для III и V классов светимости показаны голубым и желтыми цветами соответственно.

Результатом обучения классификатора является разделяющая классы прямая, уравнение которой имеет следующий вид:

$$F(B - V, M) = -3.0752 \cdot (B - V) + 0.4485 \cdot M + 1.4793 = 0, \quad (1)$$

где $B - V$ — показатель цвета, M — абсолютная звездная величина.

Мы видим, что при показателе цвета $B - V \geq 0.6$ классификатор работает практически идеально, но при $B - V < 0.6$ и $F(B - V, M) > 0$ он всем звездам предсказывает V класс при большой доле звезд III класса в обучающем множестве (10.3%, 1758 из 17029). При $B - V < 0.6$ звезды III и V класса неразделимы, поэтому для этой половины можно принять другое решение — результат работы классификатора будет взвешенным средним III и V классов, то есть некоторым средним классом, $(B - V)_{int}$, у которого будет равен

$$(B - V)_{int} = w_1 \times (B - V)_{int}(III) + w_2 \times (B - V)_{int}(V). \quad (2)$$

В этой формуле веса w_1 и w_2 , ($w_1 + w_2 = 1$) логично взять в соответствии с априорной вероятностью классов в этой области (0.1 и 0.9).

В таблице двумерной спектральной классификации ([1]) у спектрального типа G2 показатель цвета при III классе 0.733, при V — 0.630. Это максимальная разница между III и V классом в интересующем нас диапазоне. Класс, соответствующий средневзвешенному решению будет иметь показатель цвета

$$(B - V)_{int} = 0.1 \cdot 0.733 + 0.9 \cdot 0.630 = 0.64, \quad (3)$$

то есть, даже в наихудшем случае отличие от V класса минимальное — 0.01^m — гораздо ниже уровня ошибок показателя цвета в каталоге HIPPARCOS. Поэтому в дальнейшем мы всегда будем использовать решение классификатора в этой области по V классу.

С учетом этого соглашения проведем оценку качества работы классификатора с помощью процедуры 10-fold кросс-валидации. Полученные оценки параметров классификатора (точность, полнота и F1-мера) приведены в следующей таблице:

Решение классификатора →	III	V	Класс	Точность	Полнота	F1-мера
III	16636	1992	III	95%	89%	92%
V	783	20396	V	91%	96%	93%

Как видим, классификатор имеет приемлемое качество, и с его помощью мы определили классы светимости 49542 звезд северного экваториального полушария. В этом каталоге содержится следующая информация:

- номер звезды в каталоге HIPPARCOS;
- галактические долготы и широты звезд, град;
- звездная величина в полосе V;
- спектральный класс;
- класс светимости.

4 Покраснение

Межзвездное поглощение может быть описано избытком цвета. Избыток цвета мы будем называть «покраснением». *Покраснение* звезды есть

$$E = E_{B-V} = (B - V)_{obs} - (B - V)_{int}, \quad (4)$$

где $(B - V)_{obs}$ — ее видимый показатель цвета звезды (с учетом межзвездного поглощения), а $(B - V)_{int}$ — теоретический показатель цвета звезды (без учета межзвездного поглощения). Значение $(B - V)_{obs}$ мы можем получить на основе данных фотометрии звезды из каталога. Значение $(B - V)_{int}$ мы получим воспользовавшись приведенной в [1] двумерной таблицей «спектральный класс, класс светимости — показатель цвета». То есть, для получения $(B - V)_{int}$ звезды нам нужно знать ее спектральный класс и класс светимости.

5 Карты градиентов покраснения звезд

В дальнейшем мы будем обозначать площадки, соответствующие «пикселям» разбиения Healpix как $\{P_i\}_{i=0}^{N_{pix}-1}$, а конусы, высекаемые соответствующими пикселями, — через $\{C_i\}_{i=0}^{N_{pix}-1}$ (У нас $N_{pix} = 3888$).

Такое разбиение позволит нам

1. рассмотреть ход покраснения в каждом конусе как одномерную функцию $E(r)$. Это корректно, ввиду того, что конусы достаточно узкие;
2. сделать наши результаты «независимыми», т.к. конусы не пересекаются;
3. поместить в каждый конус примерно одинаковое число звезд, чтобы избежать недостатка звезд в некоторых конусах.

Будем моделировать покраснение звезд вдоль луча зрения линейной функцией $E(r) = kr$. Свободный член можно считать нулевым, т.к. у ближайших звезд покраснения не должно быть ввиду очень малого объема пыли на луче зрения. Параметр k модели, соответствующей каждому конусу C_i , находится с помощью метода наименьших квадратов. При этом мы проводили отбор данных, так как метод наименьших квадратов очень чувствителен к выбросам, т.е. большие ошибки значений покраснения и/или параллакса могут очень сильно испортить результат. Для устранения этого недостатка после построения тренда по всем звездам в конусе, мы выбрасывали те звезды, которые давали самые большие отклонение от

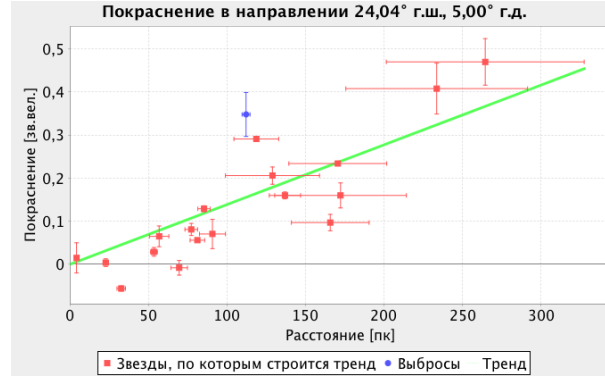


Рис. 2: Ход покраснения звезд по расстояниям в направлении $l = 5^\circ.00, b = 24.04^\circ$. По горизонтали расстояния в пк, по вертикали — покраснение (зв. вел.). При вычислении градиента k отброшена одна точка (синий цвет). Данные из каталога Hipparcos.

найденного тренда. Затем, новое значение тренда определялось уже только по оставшимся звездам. После выброса 10% самых плохих звезд процесс выбросов заканчивался.

Очевидно, что сильные выбросы дают далекие звезды, для которых параллакс определен с большой ошибкой. Отбрасывание таких звезд в нашем методе вычисления тренда приводит к уменьшению объема пространства, по которому проводится данное исследование. Легко показать, что доля оставшихся звезд, по которым были построены тренды, до 300 пк составляет 71%, до 400 пк — 83%, до 500 пк — 89%. Отсюда можно заключить, что подавляющее число звезд, для которых в нашей работе построены градиенты покраснения, находятся на гелиоцентрических расстояниях до 500 пк.

5.0.1 Выбросы, обусловленные большой ошибкой покраснения

Выбросы могут давать и звезды с большими ошибками в покраснении. Покраснение есть $E = (B - V)_{obs} - (B - V)_{int}$, поэтому его ошибка $\sigma_E = \sqrt{\sigma_{(B-V)_{obs}}^2 + \sigma_{(B-V)_{int}}^2}$. Ошибки $\sigma_{(B-V)_{obs}}$ приведены в каталоге Hipparcos. Можно показать, что ошибки более 0.05^m характерны только для звезд слабее 11^m . Но доля таких звезд менее 3%, поэтому ошибки $(B - V)_{obs}$ можно считать приемлемыми. К сожалению, ошибки $\sigma_{(B-V)_{int}}$ не представлены в работах [1], [2]. Значения в соответствующих таблицах $(B - V)_{int}$ указаны с точностью до двух знаков после запятой, поэтому можно полагать, что $\sigma_{(B-V)_{int}}$ имеет порядок 0.01^m . Для подтверждения данного уровня значений $\sigma_{(B-V)_{int}}$ можно рассчитать разницу значений $(B - V)_{int}$, приведенных в двух работах [1], [2], и увидеть, что значения показателя цвета существенно различаются только у поздних спектральных классов (от M0). Доля звезд спектрального класса M в каталоге Hipparcos менее 4%, поэтому можно принять значение $\sigma_{(B-V)_{int}} = 0.01^m$ для всех звезд.

Оценки уровня ошибок $(B - V)_{obs}$ и $(B - V)_{int}$ позволяют говорить о том, что уровень ошибок покраснения, полученного по данным каталога Hipparcos, составляет 0.03^m , что является приемлемым значением. Тем не менее, некоторые звезды имеют больше ошибки в покраснении и могут быть выброшены в нашем методе вычисления тренда.

5.1 Распределение коэффициента k по небесной сфере

Тем самым, ход покраснения в конусе C_i мы описываем одним числом k_i — скоростью роста покраснения в этом конусе. Она, как мы ранее выясняли, должна коррелировать с наличием пыли. Поэтому, составив карту распределения коэффициента k , мы составим двумерную панораму распределения пыли в окрестности Солнца с радиусом до 500 пк..

На рисунке 3 показана карта распределения градиентов покраснения звезд для звезд обучающего множества. Черным цветом закрашены те площадки, для которых звезды обучающего множества не позволяют вычислить покраснение, так как для них не известны классы светимости. Число таких площадок равно 1678.

Как было сказано выше, определение спектрального класса для остальных звезд, не входящих в обучающее множество, нами производилось с помощью двумерного линейного классификатора. Карта градиентов покраснения, построенная с помощью классификатора показана на рис.3 справа. Сравнение левой и правой карт на рис.3 показывает, что примененный нами метод машинного обучения для сортировки звезд по III и V классам светимости позволил получить информацию о межзвездном покраснении практически для всей небесной сферы.

5.2 Сравнение с результатами, полученными с помощью 2MASS

На рисунках 4 показаны распределения по небесной сфере показателей цвета $J - H$ и $H - K_s$ звезд по данным каталога 2MASS [4]. Мы видим, что показатели цвета в ближней инфракрасной области ($J - 1.2 \mu m$, $H - 1.6 \mu m$ и $K_s - 2.2 \mu m$), показывают те же области, которые мы получили по покраснению $E(B - V)$ в оптическом диапазоне. Строго говоря, наши результаты можно сравнивать только с показателями цветов, исправленными за стандартные значения, то есть с

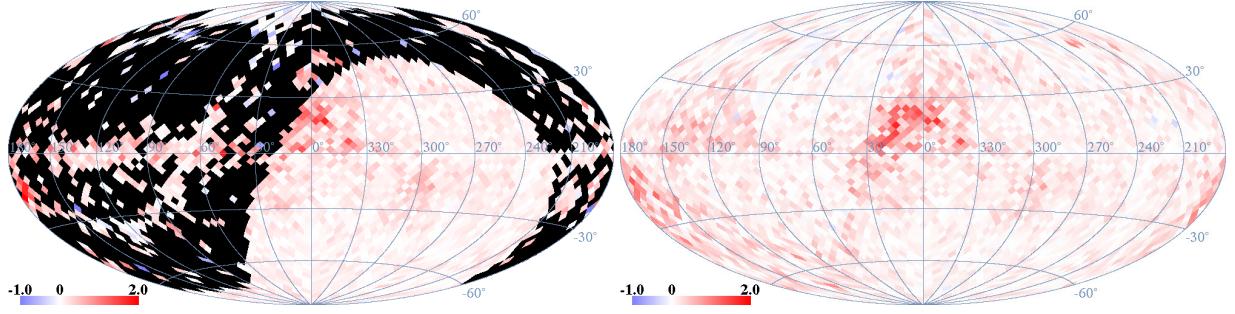


Рис. 3: Слева: распределение градиента покраснения, построенного по звездам, для которых можно определить покраснение только по данным каталога Hipparcos. Черным цветом отмечены 1678 пикселей, в которых нет даже 5-ти звезд. Справа: Распределение градиента покраснения, построенного по звездам, для которых можно определить значение покраснения (94199 звезд). Границы цветовой шкалы указаны в зв. вел. на кпк. Границы цветовой шкалы указаны в зв. вел. на кпк.

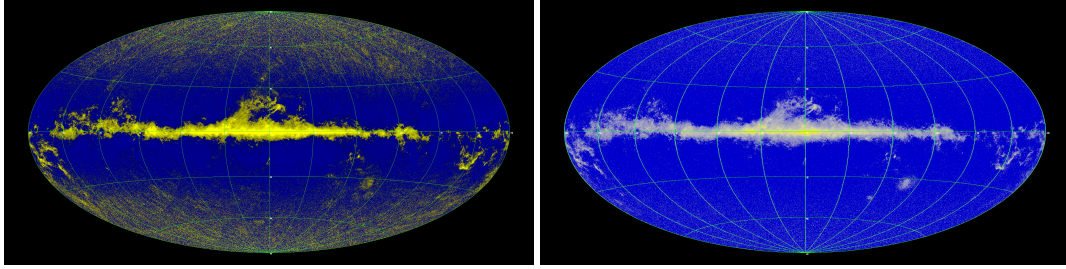


Рис. 4: Распределение по небесной сфере показателей цвета $J - H$ и $H - K_s$ (2MASS)

величинами

$$E = E_{J-H} = (J - H)_{obs} - (J - H)_{int}, \quad E = E_{H-K_s} = (H - K_s)_{obs} - (H - K_s)_{int}. \quad (5)$$

Однако, по данным [1], [18]) поправки $(J - H)_{int}$ и $H - K_s)_{int}$ слабо зависят как от спектра звезд, так и от спектрального класса, то есть практически одинаковы для всех звезд. Кроме того, они существенно меньше поправок $(B - V)_{int}$. Это обстоятельство позволяет считать, что карты цветов в ближней инфракрасной области рис.4 показывают детали, которые напрямую обуславливаются именно межзвездным поглощением.

6 Заключение

Перечислим основные результаты, полученные в работе.

- Создан бинарный классификатор, позволяющий звезде с известной абсолютной звездной величиной назначить класс светимости III или V. Для III класса светимости точность классификатора равна 95%, полнота — 89%. Для V класса светимости соответствующие характеристики равны 91% и 96%.
- Для 49542 звезд каталога HIPPARCOS с помощью бинарного классификатора определены классы светимости. Электронная версия этого каталога доступна по адресу github.com/amosov-f/dust-detector/tree/R-1.0/dust/src/test/resources/predict-lumin.txt
- Для 94199 звезд каталога HIPPARCOS определены покраснения звезд по показателю цвета $B - V$.
- Для звезд с расстояниями до 500 пк получена карта значений градиента покраснения в направлениях, определяемых центрами 3888 равновеликих площадок, построенных методом HealPix. Электронная версия этого каталога доступна по адресу github.com/amosov-f/dust-detector/tree/R-1.0/dust/src/test/resources/table-k.txt
- Определена статистическая надежность результатов для каждой площадки.
- Произведено сравнение карт покраснения, полученных нами в оптическом диапазоне, с аналогичными результатами, полученными в ближней инфракрасной области по данным каталога 2MASS. Практическое совпадение этих карт свидетельствует о надежной работе использованного нами бинарного классификатора.

Список литературы

- [1] Бинни, Меррифилд (J. Binney and M. Merrifield), *Galactic Astronomy* (Princeton: Princeton Univ. Press, 1998).
- [2] В. Страйжис, Многоцветная фотометрия звезд. Вильнюс, 1977.
- [3] Perryman M.A.C., Lindegren L., Kovalevsky J., Hog E., Bastian U., Bernacca P.L., Creze M., Donati F., Grenon M., Grewing M., van Leeuwen F., van der Marel H., Mignard F., Murray C.A., Le Poole R.S., Schrijver H., Turon C., Arenou F., Froeschle M., Petersen C.S., "The Hipparcos Catalogue"(1997A&A...323L..49P)
- [4] Скрутски и др., (Skrutskie, M. F., and 30 colleagues), The Two Micron All Sky Survey (2MASS). The Astronomical Journal 131, 1163-1183, (2006).
- [5] Г. А. Гончаров, Трехмерная карта покраснения звезд по фотометрии 2MASS: метод и первые результаты, ПАЖ, том 36, №8, с. 615–627, 2010.
- [6] Г. А. Гончаров, Трехмерная карта межзвездного поглощения в ближайшем килопарсеке, ПАЖ, том 38, №2, с. 108-121, 2012.
- [7] Цветков А.С. Руководство по практической работе с каталогом Hipparcos, СПб.: СПбГУ, (2005).
- [8] Цветков А.С. Руководство по работе с каталогом TYCHO-2, СПб.: СПбГУ, (2006).
- [9] А. С. Цветков, А. В. Попов, А. А. Смирнов. Неточности в спектральной классификации звезд каталога Tycho-2, ПАЖ, т.34, N1, с.1-11, (2008)
- [10] Hastie, T.; Tibshirani, R. & Friedman, J. (2001), *The Elements of Statistical Learning* , Springer New York Inc. , New York, NY, USA .
- [11] V.N.Vapnik *The Nature of Statistical Learning Theory* 1995.
- [12] Gorski et al. *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere* 2005.
- [13] Zeljko Ivezic, Andrew J. Connolly, *Statistics, Data Mining, and Machine Learning in Astronomy*. Princeton University Press. 533 p., 2014
- [14] Hog E., Baessgen G., Bastian U., Egret D., Fabricius C., Grossmann V., Halbwachs J.L., Makarov V.V., Perryman M.A.C., Schwendiek P., Wagner K., Wicenec A., "The Tycho Catalogue"*Astron& Astrophys*, 323L,57H. 1997.
- [15] Hipparcos, the New Reduction of the Raw Data van Leeuwen F.,*Astron. Astrophys.* 474, 653 (2007),
- [16] Wright et al. *Tycho-2 Spectral Type Catalog* 2003: The Astronomical Journal
- [17] Anderson, E.; Francis, Ch. XHIP: An extended hipparcos compilation, *Astronomy Letters*, Volume 38, Issue 5, pp.331-346, 2012.
- [18] V. Straižys and Romualda Lazauskaite. Intrinsic Color Indices and Luminosity Sequences of Stars in the 2MASS Two-color Diagram , *Baltic Astronomy*, vol. 18, 19–31, 2009.
- [19] Houk N. et al. Michigan Catalog of Two-dimensional Spectral Types for HD Stars. Vol. 1-5. Ann. Arbor: Univ. Michigan Dept. Astron. 1978, 1982, 1975, 1988, 1999.