

Вейвлет-анализ**звездной плотности астрометрических каталогов**¹

1. Введение. Обнаружение звездных скоплений и их исследование является одной из основных задач галактической астрономии, поскольку распределение скоплений в пространстве, их кинематические характеристики и звездный состав отражают общие свойства диска Галактики. По современным оценкам в Галактике содержится около 100 000 скоплений, из которых в настоящее время выявлено около 2%. Успех поиска скоплений определяется двумя факторами. Во-первых, необходимо иметь звездные обзоры, содержащие данные о миллионах звезд. Во-вторых, необходимы специальные математические методы, позволяющие перерабатывать огромные массивы данных в разумные сроки. В последнее время наблюдается взрывообразный процесс создания массовых астрометрических каталогов. Отметим здесь каталог ТУСНО-2 [1], содержащего более 2.5 миллионов звезд. Каталог полон на 99% до $V=11$ и на 95% до $V=11.5$. Положения звезд были получены с помощью переработки измерений, сделанных в рамках проекта ТУСНО на борту искусственного спутника HIPPARCOS, а в качестве первой эпохи для вывода собственных движений звезд было использовано около 140 наземных астрометрических каталогов. Временной базис приблизительно в 100 лет обеспечил высокую точность собственных движений звезд (1-3 мсд в год), а повторная обработка наблюдений, выполненных в космосе, дала точность координат около 10 мсд для $V < 9$ и 100 мсд для $V < 12$.

Использование современных технологий позволило наземной астрометрии создать каталоги, содержащие десятки и сотни миллионов звезд. Хорошим примером является проект UCAC (USNO CCD Astrometric Catalog). Предварительная версия этого каталога UCAC2 [2] содержит 48 миллионов звезд от южного полюса небесной сферы до склонения $+50$ градусов. Координаты звезд измерены на 8-дюймовом рефракторе в Чили и во Флагстаффе (США). Для звезд ярче 12.5 звездной величины собственные движения были получены с привлечение наблюдений на спутнике HIPPARCOS по методике ТУСНО-2, для более слабых звезд использовались измерения пластинок, снятых по программам Lick Northern Proper Motions (NPM) и Southern Proper Motions (SPM). Позиционная точность колеблется от 20 до 70 мсд в зависимости от блеска звезд; ошибки собственных движений звезд составляют 1-7 мсд в год и также зависят от звездной величины. После завершения этот каталог охватит все небо и будет содержать 80 миллионов звезд от $V=7.5$ до $V=16.5$. Прогнозируемая точность координат звезд 15-70 мсд, а точность собственных движений звезд будет находиться в интервале от 1 до 5 мсд в год.

В настоящее время своеобразным рекордсменом стал каталог USNO-B1.0, в котором содержится информация о 1 042 618 261 звезде [3]. Этот каталог представляет собой полный список звезд до $V=21$, построенный с помощью оцифровки пластинок нескольких фотографических обзоров неба. В каталоге B1.0 приведены прямые восхождения, склонения, собственные движения и оценки блеска звезд. Позиционная точность на современную эпоху оценивается величиной 200 мсд. Точность определения собственных движений звезд не приводится.

¹Работа поддержана грантом РФФИ 07-02-92114-ГФЕН-а и грантом Президента РФ для поддержки ведущих научных школ (НШ-1323.2008.2).

Для удобства использования столь обширных данных в Военно-морской обсерватории США создана астрометрическая база данных NOMAD (Naval Observatory Merged Astrometric Dataset) [4]. Для более чем 1 миллиарда звезд в ней приведены наиболее точные координаты, собственные движения и многополосная фотометрия, включая ближнюю инфракрасную область, заимствованную из проекта 2MASS. Средняя плотность каталога достигает несколько десятков тысяч звезд на кв. градус.

Массовые каталоги звезд, ставшие доступными астрономической общественности в последние годы, позволяют поставить целый ряд новых задач, решение которых ранее было не доступным. Однако, работа с такими огромными массивами данных сопряжена со значительными вычислительными трудностями. Отметим здесь, что авторы каталога NOMAD не представили пользователям никаких программных продуктов, способствующих работе с этой базой данных.

Настоящая статья ставит перед собой две цели. Первая из них – это разработка необходимых программных средств доступа к массиву данных NOMAD. Второй целью является поиск рассеянных скоплений в Галактике. Для выявления неоднородностей в распределении звезд на небесной сфере будет применяться техника вейвлет-преобразования. В наших предыдущих исследованиях [5], проводимых по данным каталога TUCHO-2, содержащего около 2,5 млн. звезд, была показана высокая эффективность применения вейвлет-анализа для выделения скоплений звезд, а также для оценки вероятности присутствия в областях повышенных концентраций звезд физических скоплений. Усовершенствованное программное обеспечение и полученные результаты в будущем смогут быть использованы для обработки данных для 1 млрд. звезд, ожидаемых в результате выполнения миссии GAIA.

2. Каталог NOMAD. Каталог NOMAD является компиляцией нескольких астрометрических каталогов: Hipparcos [6], UCAC2 [2], Tycho-2 [1] и USNO-B [3], кроме этого для примерно половины всех звезд в каталог включены данные о фотометрии в полосах J, H, K, взятые из проекта 2MASS [7]. Около 340 млн. звезд имеют данные о собственных движениях. Скажем только, они очень разнородные и имеют низкую точность. В данной работе мы не будем их рассматривать. Конечно же, в каталоге такой мощности отсутствуют параллаксы.

Для звезд каталога NOMAD приведены данные о шестиполосной фотометрии в диапазонах B, V, R, J, H, K. Однако, далеко не для всех звезды имеются все шесть звездных величин одновременно, а точность значений B, V и, особенно, R оставляет желать лучшего. В таблице 1 представлены данные о наличии фотометрии. Следует отметить, что наличие данных в полосах B, V, R носит существенно зонный характер. Существуют площадки небесной сферы, в которых данных много, и наоборот, существуют площадки, в которых данных почти нет, либо данные носят выборочный характер. Данные о звездных величинах J, H, K (заимствованные из каталога 2MASS) значительно менее подвержены наблюдательной селекции.

Распределение *звездной плотности* показано на рис. 1. Как мы видим, каталог охватывает весьма обширные области галактики, поскольку хорошо виден не только галактический диск, но и центральный балдж. Звездная плотность в отдельных направлениях превосходит 1 млн. на кв. градус! Каталог содержит даже звезды некоторых ярких галактик (Рис.2)!

Основная информация каталога – информация о звездных величинах и звездной плотности в различных фотометрических полосах. Такая информационная структура каталога позволяет решать не только задачи малого поля (звездные карты отдельных

Таблица 1: Число звезд каталога NOMAD с фотометрическими данными.

Полоса	Есть данные	Нет данных
B	912 млн.	205 млн.
V	267 млн.	850 млн.
R	1 млрд.	117 млн.
J, H, K	471 млн.	647 млн.

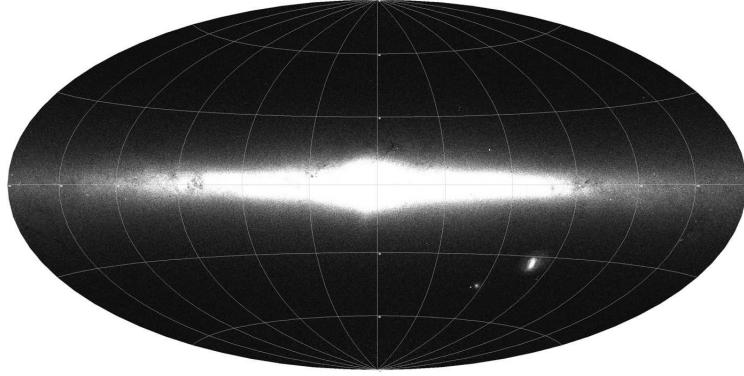


Рис. 1: *Распределение звездной плотности каталога NOMAD.*

областей, например, для идентификации объектов, поиска малых планет и т.п.), но и глобальные задачи, когда важна информация всего каталога как единого целого. Одной из таких задач может быть задача поиска неизвестных рассеянных (или даже шаровых) звездных скоплений и иных аномалий в концентрации звезд.

3. Одномерный вейвлет-анализ. Развитие техники вейвлет-анализа дает мощный аппарат для решения подобного рода задач. Существует большое разнообразие вейвлет-техник: вейвлеты на плоскости, используемые при анализе фотопластинок и ПЗС-изображений, и даже специальные вейвлеты на сфере [8], применяющиеся при анализе фонового микроволнового излучения. Вейвлеты на плоскости применялись нами ранее [5] при анализе звезд каталога Tycho-2. Однако неоднородности, которые представляют для нас интерес в настоящей статье, имеют малый размер (не более одного-двух десятков угловых минут), поэтому мы предложим более простой способ проведения обзора всей небесной сферы с помощью одномерного вейвлет-анализа [9] с использованием МНАТ-функции. Как оказалось, одномерный анализ позволяет существенно ускорить вычисления, что очень важно при работе с большими каталогами. Предположим, что у нас имеется одномерный и равномерный ряд какой-либо величины, например, звездной плотности с постоянным шагом по галактической долготе для выбранного значения галактической широты. С помощью техники вейвлет-анализа мы можем выделить неоднородности разных масштабов в этом ряду. Приведем основные формулы, решающую данную задачу.

Пусть x_k – ряд исходных значений, Δt – шаг выборки, N – число точек ряда:

$$x_k = x(t_k), \quad t_k = \Delta t \cdot k, \quad k = 1, 2, \dots, N. \quad (1)$$

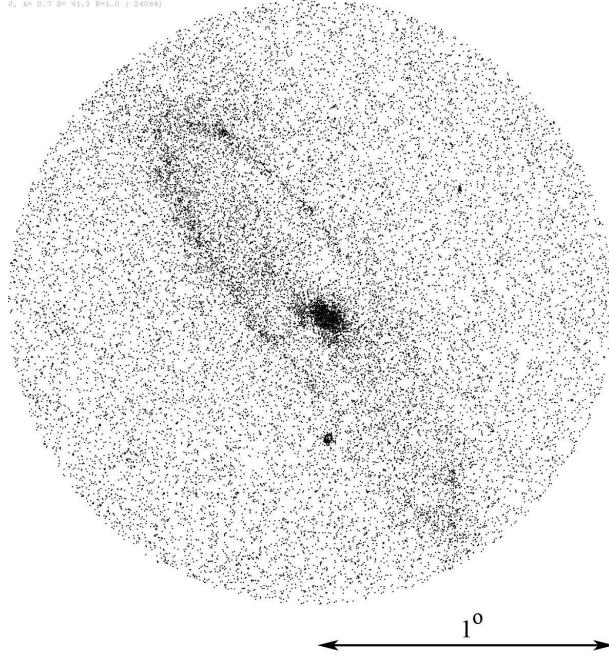


Рис. 2: Карта распределения звезд каталога NOMAD в окрестности галактики M31.

В качестве первого этапа обработки найдем среднее значение ряда

$$\langle x \rangle = \frac{1}{N} \sum_{k=1}^N x_k \quad (2)$$

и центрируем ряд

$$x_k^{\circ} = x_k - \langle x \rangle, \quad k = 1, 2, \dots, N. \quad (3)$$

Оценим также дисперсию ряда

$$\sigma_0^2 = \frac{1}{N-1} \sum_{k=1}^N (x_k^{\circ})^2, \quad (4)$$

которая будет использована в критерии значимости (13).

В качестве вейвлет-функции в нашей задаче мы будем использовать так называемый МНАТ-вейвлет (Mexican HAT, «Мексиканская шляпа»):

$$\psi(t) = (1 - t^2)e^{-t^2/2}. \quad (5)$$

В случае МНАТ-вейвлета дискретное вейвлет преобразование описывается следующими выражениями [9]:

$$W(a_i, b_j) = \frac{1}{n(a_i, b_j)} \sum_{k=1}^N x_k^{\circ} \psi\left(\frac{t_k - b_j}{a_i}\right), \quad (6)$$

$$n(a_i, b_j) = \sum_{k=1}^N \exp \left(-\frac{1}{2} \left(\frac{t_k - b_j}{a_i} \right)^2 \right). \quad (7)$$

Здесь: $a_i = i\Delta a$ – масштаб вейвлета; $b_j = j\Delta b$ – сдвиг вейвлета;

Для практической реализации мы должны договориться о дискретизации аргументов. В нашем случае удобно принять шаг масштаба и сдвига равными друг другу и равными шагу Δt :

$$\Delta a = \Delta b = \Delta t. \quad (8)$$

Как мы увидим, это позволит сильно упростить формулы, применяемые в вычислениях. Функция (5) сильно локализована в пространственной области. Можно показать, что

$$\int_{-L}^{+L} (1 - t^2) e^{-t^2/2} dt < 0.01 \quad (9)$$

для $L \approx 3.6$, что позволяет значительно уменьшить число слагаемых в формулах (6) и (7). Окончательно, принимая во внимание (8), мы можем записать

$$W_{i,j} = \frac{1}{n_{i,j}} \sum_{k=j-j^*}^{k=j+j^*} x_k \circ \psi \left(\frac{k-j}{i} \right), \quad (10)$$

$$n_{i,j} = \sum_{k=j-j^*}^{k=j+j^*} \exp \left(-\frac{1}{2} \left(\frac{k-j}{i} \right)^2 \right), \quad (11)$$

в этих выражениях граница индексов суммирования будет записываться как $j^* = [3.6 i]$.

В вейвлет-анализе существует еще одна проблема – краевые эффекты, которые ограничивают допустимые значения индекса j в формулах (10), (11), так как произойдет выход за границу массива x_k . В случае анализа звездной плотности вдоль круга заданной широты эти эффекты отсутствуют в силу замкнутости на себя ряда x_k , а именно всегда можно считать $x_{k+N} = x_k$ и $x_{k-N} = x_k$, где N – общее число элементов массива.

Формулы (10) и (11) имеют еще один плюс – они, насколько это возможно, оперируют с целыми числами, что положительно влияет на производительность вычислений.

В качестве результата вейвлет-анализа обычно рассматривают не сами элементы массива $W_{i,j}$, а квадраты их модулей. В нашем случае оказалось более целесообразным использовать не квадрат модуля, а величину

$$S_{i,j} = W_{i,j} \cdot |W_{i,j}|. \quad (12)$$

Положительное значение величины $S_{i,j}$ свидетельствует о повышенной концентрации звезд по сравнению с окружающими областями, а отрицательное – о пониженной. Массив значений $S_{i,j}$ (или $|S_{i,j}|$) в вейвлет-анализе часто называют *скалограммой*.

К полученному массиву значений $S_{i,j}$ нужно применить еще критерий выделения сигнала из шума. Поскольку для чисто шумового сигнала значения квадратов вейвлет-преобразования (6) представляют собой случайные величины, распределенные по закону хи-квадрат с одной степенью свободы, можно показать, что значение $|S_{i,j}|$ является статистически значимым с вероятностью p в случае, если

$$|S_{i,j}| \geq \sigma_0^2 Z_{i,j} T_p, \quad (13)$$

где порог обнаружения сигнала $Z_{i,j} T_p$ определяется следующими величинами:

$$Z_{i,j} = \frac{1}{n^2(a_i, b_j)} \sum_{k=1}^N \psi^2 \left(\frac{t_k - b_j}{a_i} \right), \quad (14)$$

$$p = \operatorname{erf} \left(\sqrt{\frac{T_p}{2}} \right), \quad (15)$$

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt. \quad (16)$$

3. Пример одномерного вейвлет-анализа. Для начала проведем вейвлет-анализ модельного ряда x_k , содержащего две концентрации разного масштаба (рис. 3). Мы видим, что на вейвлет-изображении «скопление» меньшего размера находится в области меньших масштабов, максимальное значение соответствует 20° . При резкой границе неоднородностей возникают "дифракционные" лепестки с отрицательной плотностью.

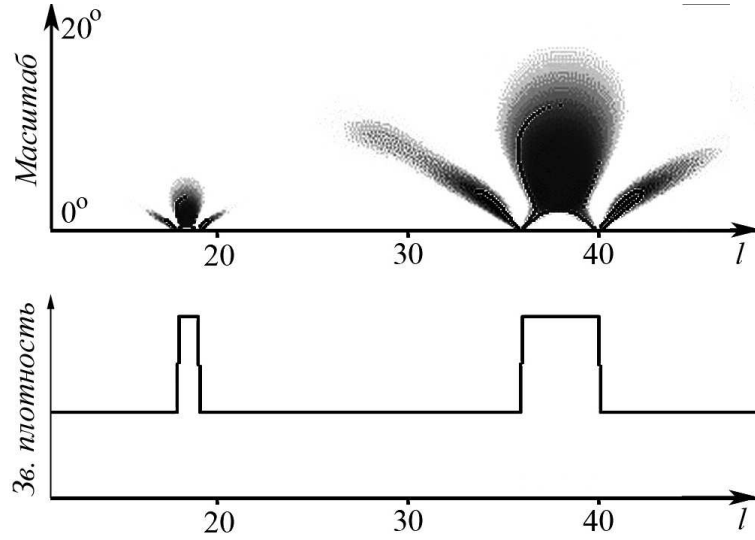


Рис. 3: Вейвлет-преобразование (верхний рисунок) тестового ряда (нижний рисунок)

Рассмотрим теперь пример вейвлет-анализа реальной области небесной сферы, расположенной вдоль малого круга небесной сферы, ограниченной галактическими широтами $b = +32^\circ \pm 0.5^\circ$. В качестве ряда данных x_k возьмем средние значения звездной плотности с шагом $\Delta t = 10'$. В силу принятого нами соотношения (8) это же значение будут иметь шаг сдвига Δb и шаг масштаба Δa . Осталось определиться с максимальным размером масштабов искомых неоднородностей плотности. Хотя теоретически возможен поиск неоднородностей с угловым размером вплоть до 180° , но в нашем случае это не имеет никакого смысла. Угловой размер рассеянных звездных скоплений в большинстве не превышает $d = 1^\circ$, поэтому мы можем ограничиться диапазоном индекса $i \leq d/\Delta a$. Рис. 4 представляет фрагмент вейвлет-анализа звездной плотности в области галактических долгот $-170^\circ \div -120^\circ$. В районе $l = -153^\circ$ вейвлет-анализ показывает

отчетливую концентрацию, вызванную звездным скоплением М 44 (Ясли). Конечно, в данном случае эта концентрация заметна непосредственно на графике звездной плотности (нижняя часть рис. 4), но, как мы увидим далее, метод вейвлет-анализа позволяет обнаруживать аномалии концентраций и в менее очевидных случаях.

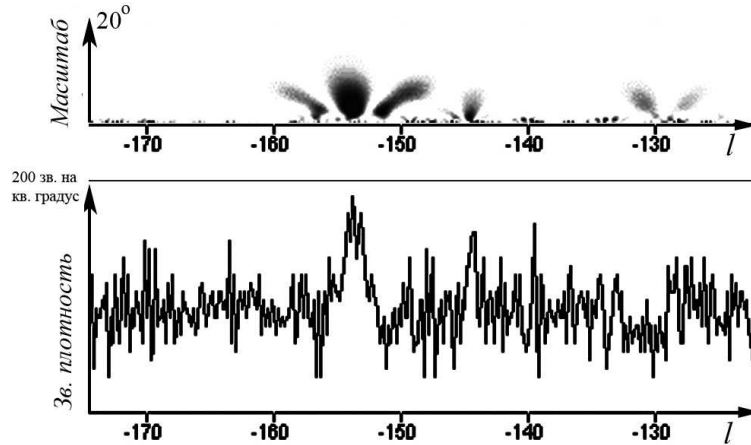


Рис. 4: Пример вейвлет-преобразования в области звездного скопления Ясли

3. Вейвлет-обзор всей небесной сферы.

Для широкомасштабного поиска локальных неоднородностей звездной плотности был предложен следующий алгоритм:

- Задавался некоторый критерий отбора звезд, например, по принадлежности звезды заданному диапазону звездной величины в определенной полосе.
- Выполнялось деление небесной сферы на узкие (обычно 1° или 0.5°) зоны по галактической широте, которые затем разбивались на площадки с заданным шагом по долготе (от $5'$ до $10'$).
- Для каждой площадки определялась звездная плотность, таким образом строился ряд данных x_k для каждой зоны.
- Полученные ряды звездной плотности по долготе для всех значений галактической широты обрабатывались с помощью техники вейвлет-анализа.

Далее возникает вопрос о визуализации полученных массивов как единого целого. Построение трехмерного изображения всей небесной сферы не только проблематично, но и не будет обладать большой наглядностью. Для того чтобы остаться в рамках двух измерений, преобразуем скалограмму $S_{i,j}$ в одномерный массив G_j , называемый *скейлограммой*:

$$G_j = \frac{1}{i_{max} - i_{min} + 1} \sum_{i=i_{min}}^{i_{max}} S_{i,j}. \quad (17)$$

Задавая разные значения пределов суммирования i_{min} и i_{max} , можно сделать так, чтобы скейлограммы отражали более или менее крупные детали. Например, суммируя от

$i_{min} = 1$ до $i_{max} = 6$ при $\Delta a = 10'$, мы получим в скейлограмме следы от деталей, размер которых заведомо меньше 1° . Напротив, задавая $i_{min} > 6$, мы переходим к изучению структур размером больше 1° . Таким образом скейлограмма хотя и преобразует одномерный массив x_k в одномерный же G_j , однако варьируя значениями i_{min} и i_{max} мы можем фильтровать масштаб видимых в скейлограмме деталей.

Полученные для каждого круга галактической широты скейлограммы можно отображать в какой-либо проекции небесной сферы, например в широко распространенной проекции Айтоффа [6]. Такие изображения мы назвали *вейвлет-обзорами* небесной сферы. При значении $\Delta t = 10'$ и менее такие вейвлет-обзоры представляют собой весьма большие изображения (3000×1500 точек и больше) и приведены в интернете [10]. В данной работе мы показываем лишь фрагмент одного изображения (рис. 5), построенного для выборки звезд, имеющих звездную величину $J = 14^m \div 16^m$. Выделение сигнала из шума произведено с заданной вероятностью $p = 0.99$. На полученный вейвлет-обзор мы нанесли объекты каталога Revised NGC [11].

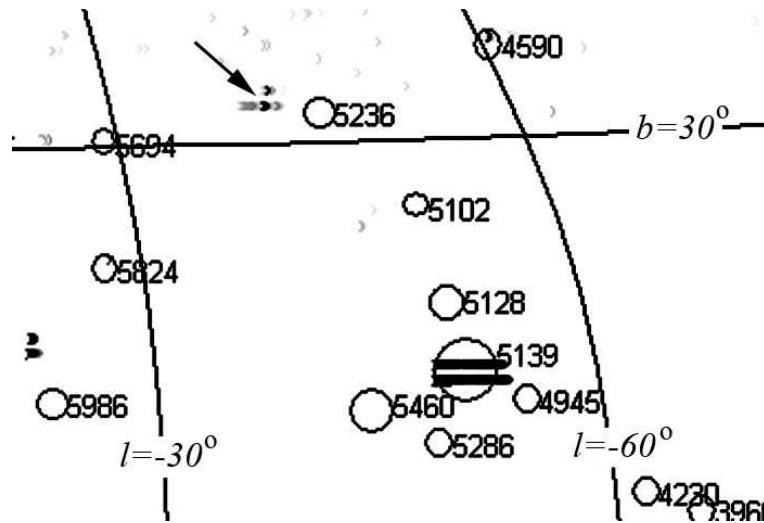


Рис. 5: Фрагмент вейвлет-обзора небесной сферы с нанесенными объектами каталога NGC

Анализ данного изображения показывает, что мы должны рассматривать три класса объектов:

- Известные объекты, выделяемые методом вейвлет-анализа;
- Известные объекты, не обнаруживаемые методом вейвлет-анализа;
- Неизвестные объекты, выделяемые методом вейвлет-анализа.

Если с первыми объектами все понятно, то уже вторые нуждаются в объяснении. Действительно, звездные скопления видны не во всех диапазонах звездных величин. Например, такое звездное скопление как Плеяды (M 45) не заметно в области слабых звезд, зато прекрасно выделяется вейвлет-анализом при переходе к звездам с блеском $8^m - 10^m$.

Наибольший интерес для нас представляет третий класс объектов: области повышенной концентрации звезд, которые не отмечены в каталогах шаровых [12] и рассеянных [13] звездных скоплений. Можно надеяться на открытие новых, ранее неизвестных рассеянных скоплений в ИК-диапазоне при использовании звездных величин J, K, H или R. Похоже, это имеет место в области с координатами $l = -47^\circ$, $b = +32^\circ$, которая показана на рис. 6, хотя этот факт нуждается в уточнении.

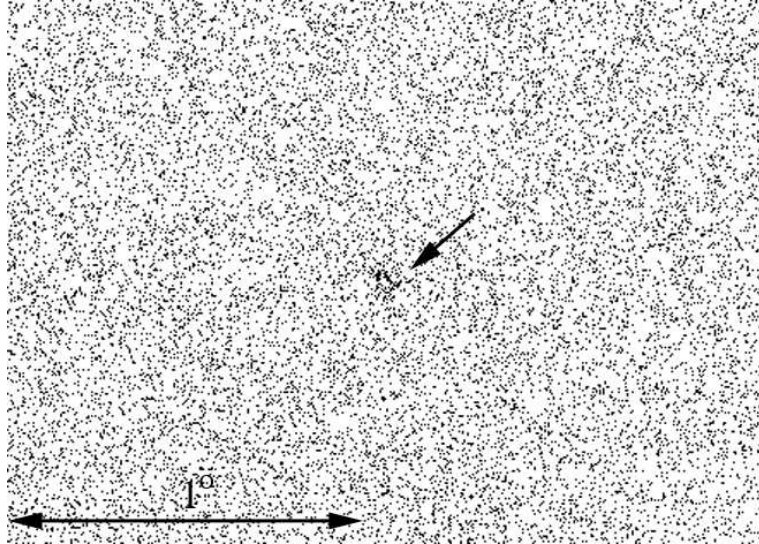


Рис. 6: Вероятно найденное слабое рассеянное скопление

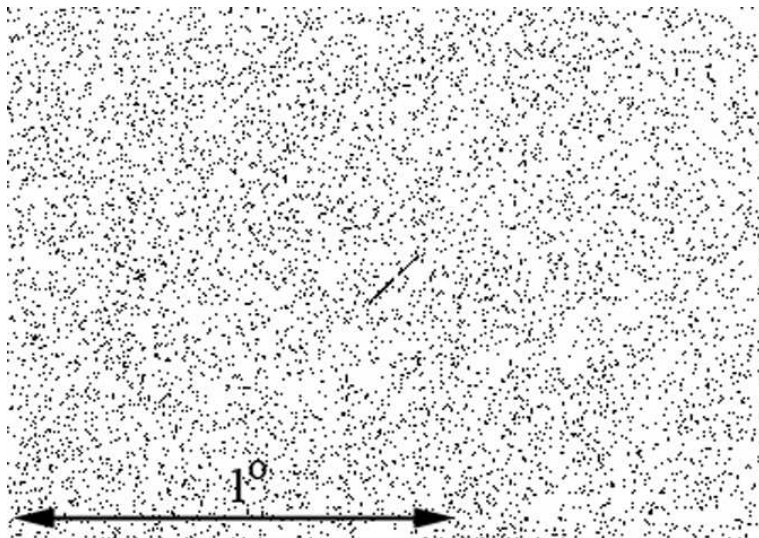


Рис. 7: Объект неизвестной природы

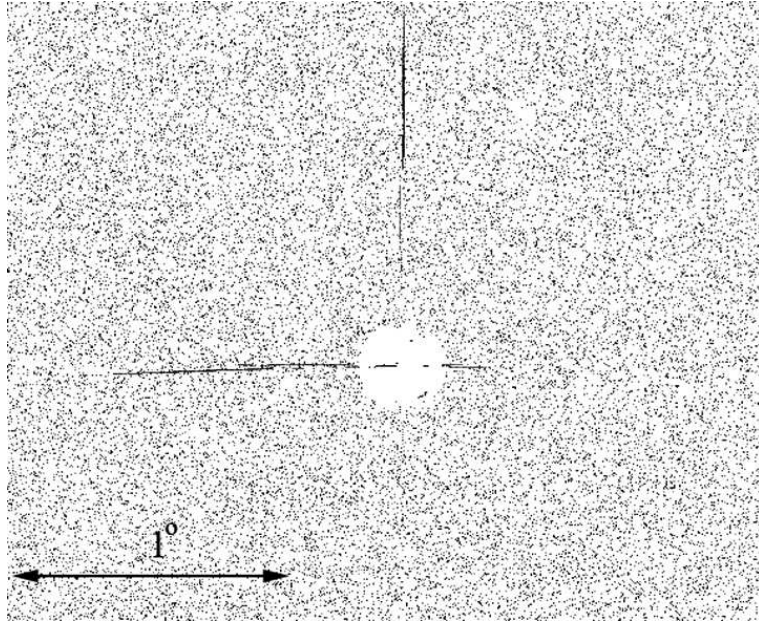


Рис. 8: Артефакт каталога - дифракционные лучи

Иногда мы встречаемся с объектами, природу которых еще надо выяснять. Например, в области с координатами $l = -57$, $b = 40.5$ мы видим странную полосу, которая в принципе может быть галактикой, видимой с ребра (рис. 7), но может быть и чем-то иным.

Однако значительная часть обнаруженных пиков звездной плотности является образованиями совсем другого рода, не связанными с реально существующими объектами. Например, рассмотрим четко выделенный объект на рис. 5 расположенный рядом с NGC 5236. На рис. 8 мы видим, что из-за автоматического построения каталога в него по ошибке попали дифракционные лучи яркой звезды, и эти лучи записаны в каталоге как звезды! Следует сказать, что в полосе R такие артефакты нами были выявлены ранее, но мы не ожидали что и в каталоге 2MASS мы встретимся с подобными явлениями (хотя и в меньшем количестве).

4. Заключение. Данная работа является первым шагом в области инфракрасных обзоров больших каталогов. На следующей стадии будет построен каталог обнаруженных объектов и их классификация. Применение техники инфракрасного анализа может помочь в решении и чисто астрометрической задачи – улучшения каталога, избавлении его от ошибочных данных.

Список литературы

- [1] *Tycho-2*, Hog, E. et al., *Astron. Astrophys.*, 355, L27 UCAC2 Zacharias, N. et. (2000)
- [2] *UCAC2*, Zacharias, N. et al., *Astron. J.*, 127, 3043 (2004)

- [3] *USNO-B*, Monet, D. G. et al., *Astron. J.*, 125, 984 (2003)
- [4] Naval Observatory Merged Astrometric Dataset, <http://www.navy.mil/nomad> (2002).
- [5] *Kazakevich E. , Vityazev V. , Orlov V. Tycho-2: Search for Stellar Groups using Wavelet Transform, ASP Conf. Ser. Order and chaos in stellar and planetary systems, V. 316, 2004,p.243-246.*
- [6] *ESA*, The Hipparcos and Tycho Catalogues 1997, ESA SP-1200, (1997).
- [7] *2MASS*, <http://www.ipac.caltech.edu/2mass/releases/allsky/>
- [8] *Мак Эвен и др.* (J. D. McEwen, M. P. Hobson, D. J. Mortlock, A. N. Lasenby), *IEEE Transactions on Signal Processing* (2005)
- [9] *В.В.Витязев*, Вейвлет-анализ временных рядов, СПбГУ (2001).
- [10] Сайт Астрономического института СПбГУ (<http://astro.pu.ru>)
- [11] *Sulentic, J. W., Tifft, W. G.*, Revised New General Catalogue of Nonstellar Astronomical Objects, Tuscon (1977).
- [12] *Harris W.E.*, A catalog of parameters for globular clusters on the Milky Way, *Astron. J.* 112, 1487 (1996)
- [13] *Dias W.S., Alessi B.S., Moitinho A., Lepine J.R.D.* New catalog of optically visible open clusters and candidates (V2.8), *Astron. Astrophys.*, 389, 871 (2002)

В и т я з е в В. В., Ц в е т к о в А. С. Вейвлет-анализ звездной плотности астрометрических каталогов

Вестн. С.-Петербург. ун.-та. Сер.1. 2009. Вып.____ (©____). С. 00 — 00.

Мы представляем быстрый и эффективный алгоритм поиска скоплений звезд и других неоднородностей звездной плотности массовых астрометрических каталогов. Наш подход основан на использовании одномерного вейвлет-преобразования, достаточного для исследования неоднородностей малых (до 1 градуса) масштабов. В каталоге NOMAD, содержащем более 1 миллиарда звезд, с помощью этого метода обнаружены известные и, по-видимому, новые звездные скопления в различных фотометрических полосах. Найдены также многочисленные артефакты каталога. Предлагаемая техника может быть использована при анализе будущих массовых каталогов, в частности, данных, которые будут получены космическим аппаратом GAIA.

КЛЮЧЕВЫЕ СЛОВА: вейвлеты, звездные скопления, астрометрические каталоги

V. V. Vityazev, A. S. Tsvetkov. Wavelet analysis of stellar density in huge astrometric catalogues

The fast and effective wavelet technique for searching the heterogeneities of stellar density in large astrometric catalogues is presented. Our approach is based on one dimension wavelet transform which turned out to be sufficient for detecting small scale (up to 1 degree) details in stellar distribution on the celestial sphere. Among one billion NOMAD stars the known and unknown globular and open clusters have been detected in various photometric bands. A lot of artifacts in NOMAD data were found in addition. This technique can be used for catalogues of the nearest future including the products of GAIA mission.

KEY WORDS: wavelets, stellar clusters, astrometric catalogues