

### 1. Цели и задачи

Имеется случайная выборка  $N$  однотипных объектов (звёзды, астероиды, экзопланеты и т.п.), в которой каждый характеризуется  $n$ -мерным вектором параметров  $x_i$ .

Распределение объектов можно охарактеризовать функцией плотности  $f(x)$ . Оно может быть более-менее однородно, и тогда  $f(x)$  имеет гладкий вид без пиков, провалов и т.п. Но могут быть и неоднородности в виде локальных концентраций и разрежений вокруг каких-то значений  $x$ , и тогда  $f(x)$  будет демонстрировать мелкомасштабные структуры.

Однако флуктуации плотности могут возникать и сами по себе, вследствие естественных неоднородностей случайной выборки. С ростом числа объектов такие неоднородности постепенно выглаживаются по обычному статистическому закону  $N^{-1/2}$ , однако для всякого  $N$  существует достаточно мелкий масштаб, где эти спонтанные флуктуации остаются большими. Это явление называется дробовой шум, или шум малой выборки.

Говоря неформально, наша задача — научиться отделять реальные флуктуации плотности, отражающих структуру функции  $f(x)$ , от фиктивных, вызванных только дробовым шумом выборки. Такая задача представляет собой смесь кластерного анализа и статистической оценки плотности (density estimation).

### 3. Некоторые результаты

#### 3.1 Экзопланеты: одномерные распределения ряда параметров

- Статистически значимое обнаружение нового семейства экзопланет-гигантов, представленное локальным уплотнением около орбитального периода 450 сут., предположительно связано с эффектом ледяной аккумуляции в протопланетном диске.
- Подтверждение (правда, не вполне статистически уверенное) наличия «долины периодов» в распределении радиусов экзопланет, около значения 2 радиуса Земли (Fulton et al. 2018, тоже связано с физикой планетообразования).

#### 3.2 Поиск семейств астероидов

- Исследовались двумерные распределения попарно взятых собственных элементов ( $a$ ,  $e$ ,  $\sin i$ ). То есть, распределения на двумерных проекциях этого трёхмерного пространства.
- Выделено 44 астероидных семейства с подтверждённой статистической значимостью. Из них большинство известные (по методу HCM), хотя есть как минимум одно семейство, нигде ранее не упомянутое (астероид с наименьшим номером 2353).
- Заметно мешает полноценному анализу эффект перекрытия семейств на двумерных проекциях, из-за чего многие известные семейства проверить не удалось. Эту проблему может решить трёхмерный анализ, который будет проведён в ближайшем будущем.

#### 3.3 Высокоскоростные звёздные потоки в окрестности Солнца по данным GAIA DR2

- Исследовались распределения звёзд солнечной окрестности (200 пк) на двумерных проекциях пространства скоростей ( $U$ ,  $V$ ,  $W$ ), по каталогу GAIA DR2.
- Обнаружено около 15 неперекрывающихся статистически значимых групп, почти все они новые и при этом являются высокоскоростными.
- Эффект наложения групп в двумерных проекциях в этой задаче ещё сильнее, так как семейства расположены плотнее. Эта выборка также требует полноценного трёхмерного анализа.

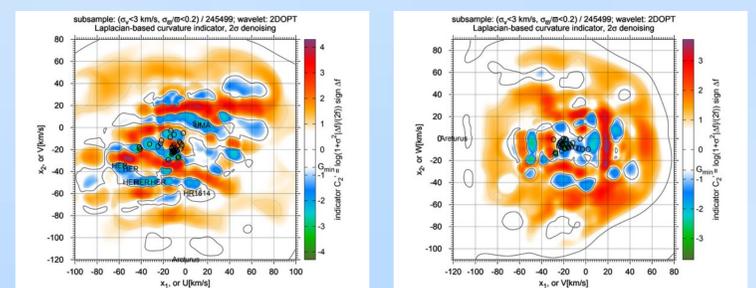
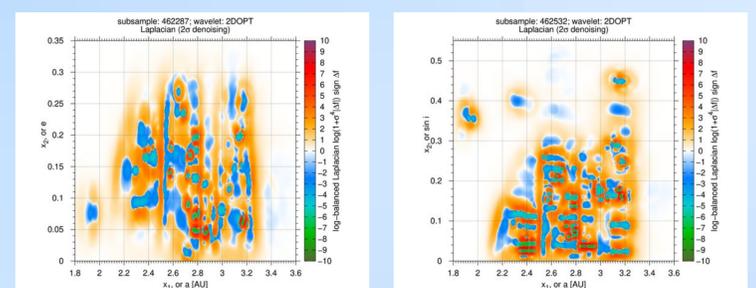
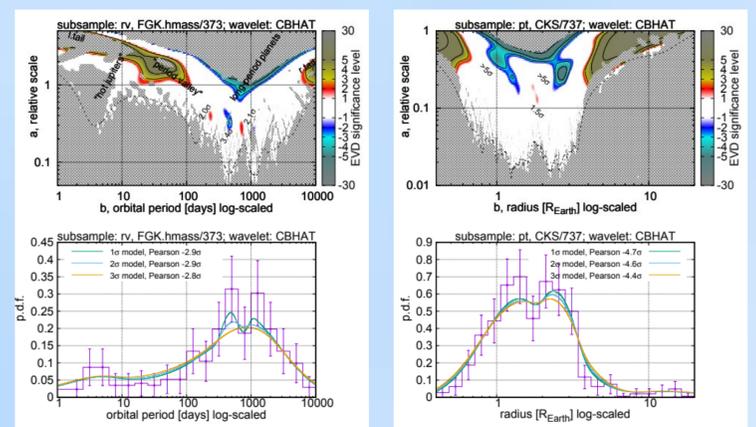
### 2. Метод

Для решения задачи мы используем метод вейвлет-анализа, который осуществляет разложение функции  $f(x)$  на структуры разных масштабов, но притом более гибко, чем Фурье-анализ.

Заметим, что это не тот классический вейвлет-анализ, который используется при анализе временных рядов или изображений. Там искомая функция определяется явно, через наблюдаемые значения  $f_i=f(x_i)$  в заданных (детерминированных) узлах  $x_i$ . Здесь же  $f(x)$  определяется неявно, через локальную плотность случайных  $x_i$  в окрестности заданной точки. Это отличие существенно меняет всю математику метода.

Для доведения алгоритма до рабочего состояния потребовалось решить ряд проблем математического характера, решения которых ранее не было известно:

- оценка вероятностных уровней шума в вейвлет-преобразовании с помощью обобщенного методов Райса (теория экстремальных значений случайных полей);
- определение области гауссовости вейвлет преобразования, внутри которой возможно проводить поиск структур;
- определение оптимальных вейвлетов, позволяющих минимизировать получаемый шум, а также расширить область гауссовости;
- алгоритм построен для размерности размерностей  $n=1,2,3$  (некоторые его детали существенно зависят от  $n$ ).



### Ссылки

1. C++ код алгоритма в свободном доступе: <https://sourceforge.net/projects/waveletstat/>
2. Одномерный алгоритм анализа: Baluev R.V. 2018, Astron. & Comput., 23, 151
3. Применение одномерного алгоритма к экзопланетам: Baluev R.V. & Shaidulin V.Sh. 2018, Astrophys. & Space Sci., 363, 192
4. Двумерный алгоритм анализа: Baluev R.V., Rodionov E.I., Shaidulin V.Sh. 2019, preprint 1903.10167 (Appl. & Comput. Harmonic Anal. submitted)
5. Применение двумерного алгоритма к астероидам: Baluev R.V. & Rodionov E.I. 2020, Celest. Mech. Dyn. Astron., 132, 34
6. Применение двумерного алгоритма к звёздному населению: Baluev R.V., Shaidulin V.Sh., Veselova A.V. 2020 Acta Astronomica, 70, 141