

ПАКЕТ ПРИКЛАДНЫХ ПРОГРАММ ДЛЯ АНАЛИЗА ФРАКТАЛЬНЫХ СВОЙСТВ МНОЖЕСТВ НАГРУЖЕННЫХ ТОЧЕК

Аннотация

В статье описываются возможности и основные особенности разработанного автором программного пакета по проведению анализа фрактальных свойств множеств изолированных точек. Пакет включает в себя возможности загрузки готовых множеств и моделирования стандартных фрактальных и однородных распределений, а также средства для анализа их фрактальных свойств (корреляционную функцию, функцию распределения). Программа может быть полезна не только для проведения научных исследований, но и в учебном процессе.

Ключевые слова: фракталы, компьютерное моделирование, фрактальные модели.

1. ВВЕДЕНИЕ

С момента выхода книги Мандельброта [1] фрактальная геометрия нашла свое применение во многих приложениях в различных областях знания. Идея данного пакета пришла из космологии, где актуален вопрос о том, распределены ли галактики во Вселенной однородно или фрактально. В этой модели Вселенная рассматривается как множество изолированных точек – галактик. Однако интерес при проведении анализа представляют не только пространственные координаты галактик (непосредственно «отвечающие» за фрактальные свойства), но и другие параметры – яркость, морфологический тип и т. п., – которые используются для фильтрации объектов (построения выборок) перед проведением анализа. Например, могут представлять интерес фрактальные свой-

ства не всех галактик, а только определенного типа. Кроме того, так как мы можем видеть с Земли только часть галактик – чем дальше галактика, тем ярче она должна быть, чтобы ее можно было зарегистрировать – интерес представляет степень искажения фрактальных свойств из-за наблюдательной селекции.

Кроме анализа естественных распределений точек, важно исследование модельных фрактальных и однородных множеств точек – как для сравнения результатов моделирования с реальностью, так и для проверки самих методов расчета фрактальных характеристик. В связи с этим представляет интерес создание программного пакета, который бы позволил проводить моделирование и анализ фрактальных свойств множеств изолированных точек, снабженных дополнительными характеристиками – весами. Хотя описываемый пакет и берет свое начало в космологии, его применение возможно и к

другим подобным множествам – молекул в газе и т. п. (или даже в экономике – множества людей, предприятий и т. д.). Примеры применения фрактального подхода в физике и астрономии имеются в [2].

Представляемый автором программный пакет SWP¹ содержит средства для создания искусственных (модельных) множеств точек, загрузки внешних множеств (естественных каталогов и т.п.), их преобразования, построения выборок и анализа фрактальных характеристик (в первую очередь, корреляционной функции). Программа SWP представляет собой интерпретатор языка сценариев.

2. ПОСТАНОВКА ЗАДАЧИ

Поставленная задача включает в себя разработку программного пакета, позволяющего проводить исследования ряда фрактальных свойств множеств изолированных нагруженных точек. В настоящее время существует множество программ по генерации фрактальных множеств, рассчитанных главным образом на визуализацию или 3D-моделирование фрактальных изображений, а исследователи, применяющие численные методы фрактальной геометрии, как правило, не публикуют создаваемые ими программы. В данном случае требуется создание пакета, который позволил бы объединить в себе возможности моделирования, исследования свойств моделей и естественных множеств и визуализацию результатов. Пакет также должен удовлетворять ряду дополнительных требований:

– в программе должна быть предусмотрена как возможность загружать готовые (полученные сторонними средствами) множества точек (например, каталоги галактик), так и средства моделировать базовые стандартные фрактальные распределения разной размерности (для сравнения поведения свойств с реальными распределениями и проверки методов исследования);

– программа должна уметь работать с большими множествами (например, для исследования модельной выборки в узком те-

лесном угле с ограничением по дополнительному параметру в 10^5 точек исходное множество может иметь размер $\sim 2 \cdot 10^8$ точек);

– должны быть реализованы средства для массовой обработки различных выборок (например, сетку моделей разной глубины или исследование выборок реального множества с разным значением ограничения по параметру); обработка должна быть максимально автоматизирована, требуется возможность пакетной обработки выборок;

– необходимо сокращение числа вычислений за счет повторного использования ранее полученных результатов как при возникновении одинакового этапа при похожих задачах, так и при выполнении вычислений с измененными параметрами (когда изменение параметров приводит к расхождению лишь на каком-то этапе расчета);

– итогом работы программы должны быть как числовые результаты (например, величина фрактальной размерности), так и графики (например, корреляционной функции).

Исходя из требований автоматизации обработки выборок, а также из-за многоэтапности процесса анализа каждой выборки – создать множество, взять выборку, провести анализ, построить график – автором было принято решение реализовать программу в виде интерпретатора языка сценариев.

3. ОСОБЕННОСТИ РЕАЛИЗАЦИИ

Язык сценариев разработан специальным образом для удобства выполнения требуемых операций: основными типами данных этого языка являются нагруженные точки и их множества. Координаты галактик в анализируемой выборке могут задаваться не только в виде декартовых координат точки, но и, например, с помощью сферических координат и красного смещения, которое есть функция расстояния. Поэтому понятие «точка» и «вес» в программе понимаются очень широко: речь идет о типе данных, который содержит в себе набор полей – значений, доступ к которым возможен (для удобства пользователя) по их имени.

¹ SWP можно расшифровать как Statistic for Weighted Points или в традициях GNU рекурсивно, как SWP is for Weighted Points.

Более того, так как значения некоторых полей могут быть вычисляемыми через значения других полей (сферические координаты через декартовы, расстояние через красное смещение и т. п.), то поля точки бывают двух видов – постоянные (атрибуты) и вычисляемые (свойства). Свойства представляют собой функции, аргументом которых является точка, возвращающие числовое значение, которое вычисляется каждый раз при обращении к соответствующему полю. Таким образом, при изменении значения атрибутов точки значения свойств будут меняться соответственно.

Тип данных «множество точек» используется в программе шире, чем собственно множество точек в пространстве. Доступ к «весам» точки по имени поля позволяет создавать множества произвольных точек. Например, во множество можно записать значения табулированной функции. Так и записывается результат вычисления корреляционной функции, которая строится по точкам.

Программа содержит инструментарий для загрузки исследуемых множеств в текстовом формате CSV.

4. ОСНОВНЫЕ ВОЗМОЖНОСТИ ПО МОДЕЛИРОВАНИЮ И ОБРАБОТКЕ МНОЖЕСТВ

Моделирование множеств. В программе реализована возможность моделировать случайное однородное распределение точек (в R^n) – каждая координата точки есть случайное число в интервале $(-1, +1)$, а также алгоритм для создания случайного фрактального распределения в R^n с любой наперед заданной фрактальной размерностью $D < n$. Для генерации случайных множеств используется датчик псевдослучайных чисел Mersenne Twister [3].

Для создания фрактального распределения заданной размерности используется алгоритм генерации обобщенного на случай R^n случайного канторова множества (также называемого «пылью Кантора»). Процесс создания множества следующий. Берется гиперкуб

$$\underbrace{[0, 1] \times [0, 1] \times \dots \times [0, 1]}_n,$$

который считается нулевым поколением генерации, и делится на M^n равных кубиков (каждое ребро делится на M равных отрезков, число M выбирается заранее). Для каждого кубика принимается решение, «выживет» он или нет в соответствии с выбираемой вероятностью «выживания» p . Множество выживших кубиков объявляется следующим поколением. Для каждого кубика нового поколения алгоритм перезапускается до тех пор, пока не будет достигнуто последнее поколение. Хотя канторовым множеством является предел при устремлении номера поколения в бесконечность, содержащий континуум точек, в реальных моделях приходится выбирать поколение, на котором следует остановить итерации. Пакет позволяет выбрать последнее поколение по его номеру или по достижении нужного конечного количества кубиков (точек). В каждом кубике последнего поколения выбирается случайная точка – каждая координата точки есть случайное число, равномерно распределенное на ребре кубика. Множество этих выбранных точек и является результатом генерации. Теоретическая оценка размерности такого множества рассчитывается по формуле

$$D = \log_M (pM^n)$$

(реально пользователь задает размерность D и число делений M , а программа вычисляет вероятность выживания p). О генерации подобных множеств и исследовании их фрактальных свойств написано в [4].

Снабжение случайными весами. В программе реализован алгоритм Фон Неймана для генерации случайных чисел, распределенных по указанному дифференциальному закону $f(x)$ на отрезке $[a, b]$, для придания модельных весов (например, яркости галактик). Алгоритм состоит в генерации пары случайных чисел $x_0 \in [a, b]$ и $y_0 \in [0, \max_{[a,b]} f]$. Если $y_0 < f(x_0)$, то число x_0 принимается за результат генерации, в противном случае генерируется новая пара чисел. Известно, что числа x_0 распределены как $f(x)$. Так как в SWP реализованы средства построения функции распределения

значения выбранного поля у всех точек множества, результат можно проверить экспериментально.

Выделение подвыборок. С помощью программы можно выделить подвыборку из множества, как смоделированного, так и загруженного, по любому из весов, в том числе вычисляемому (свойству). Всякая получаемая подвыборка есть новое множество. Особо обрабатывается выделение подвыборок по пространственным границам. Также возможно преобразование множества (преобразование всех точек множества), особый интерес среди которых представляют преобразования пространства (гомотетия, сдвиг). О границах и их поведении при данных преобразованиях написано ниже.

Корреляционные функции. Для исследования фрактальных множеств изолированных точек используются корреляционные функции, характеризующие свойства сгущения (взаимного расположения) точек – зависимость функции количества точек на данном расстоянии от расстояния (масштаба). Подробно теория корреляционных функций и алгоритмы их вычисления описаны в [4, 5]. Следуя [4], различают полную и редуцированную корреляционные функции, оба варианта реализованы в пакете. Полная корреляционная функция есть характеристика плотности множества на данном масштабе, в то время как редуцированная – отклонение плотности на данном масштабе от средней плотности во всем множестве.

Примером полной корреляционной функции является условная плотность Γ (точнее, – концентрация, так как речь идет о множествах изолированных точек). Она определяется формулой

$$\Gamma(r) = \left\langle \frac{|T \cap B(P, r)|}{|B(P, r)|} \right\rangle_{P \in T: B(P, r) \subset T_E},$$

где T – исходное множество точек, $B(P, r)$ – шар с центром в точке P и радиусом r , $|B(P, r)|$ – его объем, в числителе стоит количество точек множества T , попавших в шар радиуса r вокруг пробной точки P , в знаменателе – объем этого шара, дробь (концентрация точек в шаре радиуса r вокруг

пробной точки P) усредняется по тем точкам множества P , шары вокруг которых целиком лежат внутри пространственных границ исходного множества T_E . Таким образом, условная концентрация на масштабе r есть средняя концентрация точек в шарах радиуса r , центры которых располагаются в точках множества (отсюда термин «условная»), причем в расчет берутся только те шары, которые целиком лежат в пространственных границах множества. Формально условную плотность можно рассчитать для $r_{\min} < r < R_{\max}$, где r_{\min} – наименьшее расстояние между двумя точками, а R_{\max} – радиус наибольшего «пригодного» шара. Для фрактальных и однородных ($D = n$) множеств

$$\Gamma(r) = Ar^{D-n},$$

хотя, как показывает численный эксперимент, проводимый с помощью пакета, это не выполняется на всем участке $r_{\min} < r < R_{\max}$.

Редуцированная корреляционная функция по Гамильтону рассчитывается при помощи генерации однородного множества сравнения в тех же границах T_E , что и исследуемое множество (в программе предусмотрена соответствующая возможность). Оценка выполняется по формуле

$$\xi(r) = \frac{N_r N_d}{(N_r - 1)(N_d - 1)} \frac{DD(r)RR(r)}{DR(r)^2} - 1,$$

где N_d – количество точек в исходном множестве, N_r – количество точек во множестве сравнения, $DD(r)$ – количество пар точек исходного множества, лежащих на расстоянии r друг от друга, $RR(r)$ – аналогичное количество пар точек множества сравнения (эти пары учитываются дважды), $DR(r)$ – количество пар «точка исходного множества» – «точка множества сравнения», лежащие на расстоянии r друг от друга. При расчете для каждого конечного промежутка $[r_i, r_i + \Delta r]$, где $r_i = r_{\min} + i\Delta r$ вычисляется количество пар точек, расстояние между которыми попадает в этот промежуток (ширина промежутков $\Delta r = \frac{r_{\max} - r_{\min}}{N}$, где число N выбирается пользователем. Для однородного множества $\xi(r) \equiv 0$, для фрак-

тального $\xi(r) \equiv Ar^{D-n}$. Вычисление корреляционной функции возможно для r от r_{\min} до максимального расстояния между точками r_{\max} .

5. АППРОКСИМАЦИЯ ПРЯМОЙ

При исследовании корреляционных функций встает задача выделения прямолинейного участка функции. Как уже отмечалось ранее, теоретически ожидается, что корреляционная функция и условная плотность удовлетворяют соотношению

$$f(r) = Ar^{D-n}, \quad (1)$$

или (в логарифмических координатах)

$$\log f(\log r) = (D-n) \log r + b, \quad (2)$$

где n – размерность пространства, а D – оценка фрактальной размерности. На практике корреляционная функция может быть рассчитана для $r_{\min} < r < r_{\max}$ (редуцированная) или $r_{\min} < r < R_{\max}$ (полная) – как описано выше. Однако по ряду причин на всем этом промежутке она даже приближенно не удовлетворяет соотношению (1). Наиболее существенные причины отклонения – это, во-первых, то, что на концах промежутка функция определяется ненадежно из-за недостаточного количества пар точек, по которым она вычисляется. Во-вторых, что существуют множества, которые показывают различную фрактальную размерность на разных масштабах.

В связи с этим встает задача нахождения участков, где функция является прямолинейной, а потом аппроксимировать их методом наименьших квадратов для нахождения фрактальной размерности. Конечно, такой участок можно подбирать вручную, но при обработке достаточно большой сетки моделей это становится нереальным, в связи с чем возникает задача автоматизировать этот процесс. Чтобы снять эту проблему, в пакет включено средство для автоматического нахождения участков прямолинейности табулированной функции.

Участок табулированной функции f , заданной на точках x_0, x_1, \dots, x_n , считается прямолинейным, если

$$\sqrt{\frac{1}{n} \sum_{i=1}^{n-1} \left(2 \frac{f(x_i) - l(x_i)}{x_{i+1} - x_{i-1}} \right)^2} < s_0,$$

где l – искомая аппроксимирующая прямая, построенная на отрезке $[x_0, x_n]$ методом наименьших квадратов, s_0 – выбираемый «критерий прямолинейности». Для рассматриваемых случаев (условной плотности и корреляционной функции) этот критерий показывает себя достаточно не зависимым от количества точек во множестве и количества точек табулирования функции (автор брал 1.0 для условной плотности и 3.0 для корреляционной функции, так как, видимо, последняя флуктуирует больше).

После того как участки прямолинейности выделены, программа позволяет взять требуемое число самых длинных из них и посмотреть на параметры аппроксимации – коэффициенты прямой и участок, на котором эта прямая проведена, – чтобы получить оценку фрактальной размерности и масштаба, на котором множество имеет данную размерность.

6. ПРОСТРАНСТВЕННЫЕ ГРАНИЦЫ МНОЖЕСТВ

Важной задачей при создании пакета был учет пространственных границ множества. Такие операции, как, например, построение функции условной плотности требуют знания расстояния до границы от каждой точки. Кроме того, при проведении пространственного преобразования, например, параллельного переноса, граница должна меняться соответственно. Каждое множество снабжается границей, определяющей часть пространства, в котором лежит множество. Граница множества есть совокупность элементарных границ, таких, как, например, сфера. Частью пространства, в котором лежит множество, может являться по выбору пользователя внутренность или «внешность» сферы. Точка лежит в границах множества, если она принадлежит части пространства, выделенной каждой из элементарных границ (берется пересечение частей пространства); расстояние от точки до

границы есть минимум среди расстояний до каждой из них. Пространственное преобразование множества производится не только над каждой точкой множества – каждая элементарная граница подвергается такому же преобразованию.

В настоящее время реализовано лишь несколько элементарных границ – прямоугольный параллелепипед, сфера или сферическая оболочка (в R^n), ограничение по угловым сферическим координатам $0 \leq \alpha < 2\pi$, $-\pi/2 \leq \delta \leq \pi/2$ (в R^3), ограничение по полярному углу (в R^2). Реализованные в SWP преобразования параллельного переноса и гомотетии корректно преобразуют сферические и прямоугольные границы, проектирование или сечение вдоль одной из осей координат также корректно преобразует параллелепипед в прямоугольник (реализованы параллелепипеды с ребрами параллельными осям), сферу в окружность, ограничение по длине в ограничение по полярному углу.

7. ОСОБЕННОСТИ СОХРАНЕНИЯ РЕЗУЛЬТАТОВ

Множества точек могут быть очень большими. Например, для выделения сферической подвыборки в $2 \cdot 10^5$ точек (галактик, видимых с Земли) может потребоваться стартовать с исходного фрактального множества в $3 \cdot 10^8$ точек (> 7 Гб). Таким образом, модельные множества могут не уместиться в оперативной памяти. В то же время, все операции, связанные с созданием, преобразованием и фильтрацией множеств, требуют лишь последовательного доступа к их элементам (точкам). Операции, которые требуют произвольный доступ к элементам – такие, как вычисление корреляционной функции, – имеют гораздо большую временную сложность (n^2), а значит, будут применяться лишь к множествам меньшего размера¹.

¹ Автор в первую очередь оптимизировал код именно этих алгоритмов. Так вычисление фрактальной размерности при помощи условной плотности 100000-точечного множества занимает менее 20 минут на процессоре AMD Athlon с частотой 3.0 ГГц, а с помощью корреляционной функции Гамильтона (включая нахождение ее наклона и нуля) с использованием 300000-точечного однородного множества сравнения потребуется на порядок больше времени.

² <http://gnuplot.info>

В связи с этим программа сохраняет все создаваемые множества на постоянном носителе (диске). Каждому множеству автоматически присваивается уникальное имя файла. Это имя генерируется, исходя из способа создания этого множества – параметров генерации или исходного множества, способа и параметров выделения выборки и т. п. Созданные одинаковым способом множества будут иметь одинаковое имя.

Это обстоятельство позволило реализовать в пакете защиту от двойного пересчета: при повторном запуске того же сценария или при создании того же множества в том же или в другом сценарии оно не будет создаваться вторично. Программа проверяет наличие на диске этого файла, и если файл уже существует, то повторного вычисления (генерации множества или вычисления корреляционной функции и т. п.) не производится.

Одним из наиболее важных результатов, выдаваемых программой (помимо возможности вывода результатов расчета на терминал), являются графики. Сама программа графиков не строит, но язык содержит средства для автоматической генерации скриптов Gnuplot² и вызова Gnuplot'a (должен быть установлен отдельно) для создания графиков в формате EPS, который можно использовать для вставки в документ LaTeX. Для создания графиков не обязательно знать язык Gnuplot, средств языка пакета достаточно для формирования графиков результирующих функций, множеств, построения на них дополнительных элементов и подписей и т. п.

8. ПРИМЕР ИСПОЛЬЗОВАНИЯ

На рис. 1–3 приведены примеры работы программы по обработке случайного плоского канторова множества с наперед заданной фрактальной размерностью, изображе-

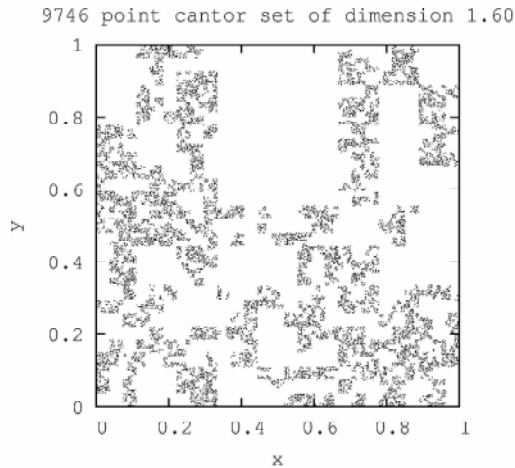


Рис. 1

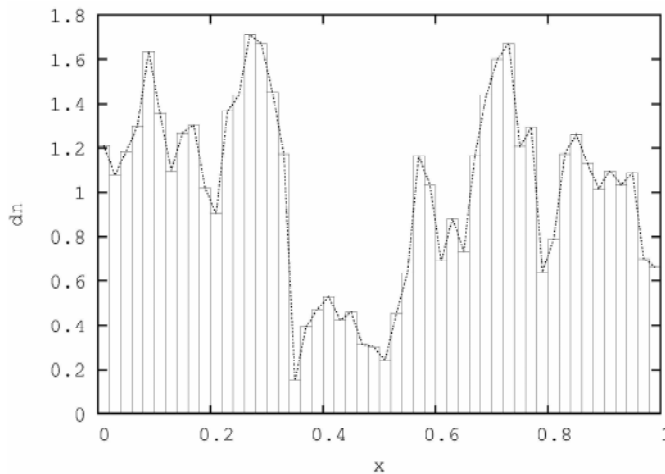


Рис. 2

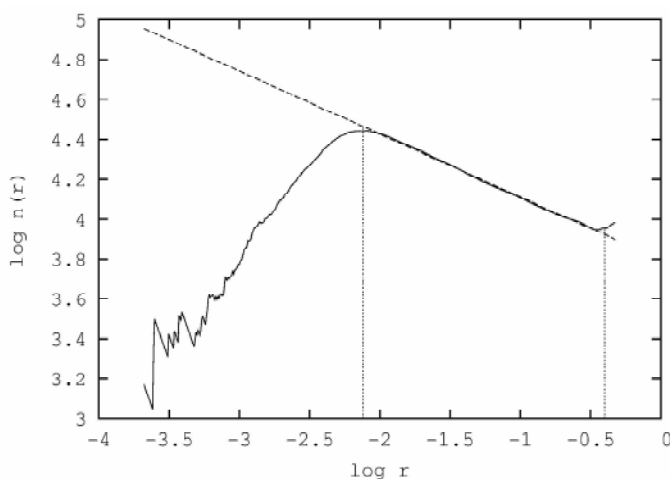


Рис. 3

ние которого показано на рис. 1. Построена плотность вероятности распределения точек по значению координаты x , ее график показан на рис. 2. На рис. 3. показан график функции условной плотности в логарифмических координатах. Последняя аппроксимируется прямой, берется одна самая длинная прямая. На печать программа вывела значение ее углового коэффициента $k = -0.315$, что соответствует оценке фрактальной размерности $D = k + 2 = 1.69$. На рис. 3. показан график корреляционной функции и аппроксимирующей прямой с границами отрезков аппроксимации.

Пакет и его исходный код распространяются свободно на условиях GNU General Public License v3.0 и доступны на сайте <http://swpproject.sourceforge.net>. Там же доступна документация, полное описание возможностей языка. На сайте можно найти скрипт, с помощью которого получены рисунки, приведенные в статье. Пакет написан на языке C++, содержит около 22000 строк и является кроссплатформенным (автором проверена работоспособность на Linux и Windows).

Некоторые полученные с помощью программы астрономические результаты изложены в [6]. Программа может быть полезна не только для проведения научных исследований, но и в учебном процессе, так как с ее помощью можно иллюстрировать особенности фрактальной геометрии. Студенты могут получить практику в моделировании случайных фрактальных множеств различной размерности, построении корреляционных функций и нахождении их характеристик, «почувствовать» различие между однородностью и фрактальностью и т. п.

Литература

1. Мандельброт Б. Фрактальная геометрия природы. М.: Институт компьютерных исследований, 2002.
2. Gabrieli A., Sylos Labini F., Joyce M., Pietronero L. Statistical physics for cosmic structures. Berlin–Heidelberg–New York–Barcelona–Hong Kong–London–Milan–Paris–Tokyo: Springer, 2004.
3. Matsumoto M., Nishimura T. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator // ACM Transactions on Modeling and Computer Simulation. 1998. Vol. 8. I. 1. P. 3–30.
4. Addison Paul S. Fractals and chaos: an illustrated course. Bristol, Philadelphia: Institute of Physics Publishing, 1997.
5. Sylos Labini F.; Vasilyev N.L. Extension and estimation of correlations in cold dark matter models // Astronomy & Astrophysics. 2008. Vol. 477. I. 2. P. 381–395.
6. Ловягин Н.Ю. Статистические свойства пространственного распределения галактик // Бюлл. спец. астрофиз. обсерв. 2009. Т. 64. Вып. 3. С. 223–235.

Abstract

In the paper capabilities and main features of developed by author program package for analysis of fractal properties of sets of isolated points are described. Package includes capabilities to load ready sets and to model standard fractal and uniform distributions, and tools to analysis of its fractal properties (correlation function, distribution function). Program might be useful not only for scientific research but in studying.

Keywords: fractals, computer modelling, fractal models.

*Ловягин Никита Юрьевич,
ассистент кафедры информатики
математико-механического
факультета СПбГУ,
lovuyagin@mail.com*



Наши авторы, 2011.
Our authors, 2011.