**Липецкий государственный технический университет**

Кафедра прикладной математики

Специальность 230401.65 Прикладная математика

*ПОЯСНИТЕЛЬНАЯ ЗАПИСКА   
К ДИПЛОМНОЙ РАБОТЕ*

Синтез самоподобного трафика

с использованием модели FARIMA

Студент     ( гр.ПМ-08-1Толкунов Н. С. )

Руководитель работы     ( асс. Жбанов С.А. )

Консультанты

нормоконтроль  
программного обеспечения     ( к.т.н., Федоркова Г.О. )

нормоконтроль  
оформления работы     ( к.т.н., Федоркова Г.О. )

Рецензент     ( )

Работа рассмотрена кафедрой и допущена к защите в ГАК  07.06.2013

Заведующий кафедрой     ( д.т.н., проф. Погодаев А.К. )

Липецк – 2013г.

ЛИПЕЦКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

|  |  |
| --- | --- |
| **Факультет** ФАИ gh | **Зав. кафедрой**  Погодаев А.К. в |
| **Кафедра** ПМ ап | “ 29 ” марта 2013 г. |

**ЗАДАНИЕ КАФЕДРЫ**

|  |  |
| --- | --- |
| **Студенту** Толкунову Николаю Сергеевичу | **Группы** ПМ-08-1 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 1. | **Тема** | Синтез самоподобного трафика с использованием модели FARIMA | | | | |
|  |  | | | | | |
|  |  | | | | | |
| 2. | **Исходные данные** | | Данные о количестве попаданий звонящих в колл-центр | | | |
|  | «Ростелеком» в очередь, количество передаваемых пакетов в сетях компании Bellcore. | | | | | |
| 3. | **Используемые методы и модели** | | | |  | |
|  | В работе используются методы и модели анализа временных рядов, численные | | | | | |
|  | методы синтеза сетевого трафика. | | | | | |
|  |  | | | | | |
|  |  | | | | | |
|  |  | | | | | |
|  |  | | | | | |
|  |  | | | | | |
| 4. | **Ожидаемые результаты** | | |  | | |
|  | Методы оценки параметров модели FARIMA и синтеза самоподобного трафика с ее | | | | | |
|  | помощью; аппробация модели и методов на исходных данных; программный продукт | | | | | |
|  | для оценки параметров модели FARIMA и синтеза самоподобного трафика. | | | | | |
|  |  | | | | | |
|  |  | | | | | |
|  |  | | | | | |
|  |  | | | | | |
| 5. | **Срок сдачи работы руководителю** | | | | | 07.06.2013 |
|  |  | | | | | |
|  |  | | | | | |
| 6. | **Дата выдачи** **задания** | | | | | 29.03.2013 |
|  |  | | | | | |
|  |  | | | | | |
| 7. | **Руководитель работы** | | | | | / Жбанов С.А. / |
|  |  | | | | |  |
|  |  | | | | | |
| 8. | **Задание принял к исполнению студент** | | | | | / Толкунов Н.С. / |
|  |  | | | | | |

**Аннотация**

С. 53 . Ил. 29 . Табл. 1. Литература: 15 назв.

FARIMA, ВРЕМЕННЫЕ РЯДЫ, СЛУЧАЙНЫЕ ПРОЦЕССЫ, САМОПОДОБИЕ

В данной работе рассматривается задача моделирования самоподобного трафика, дается обзор методов ее решения, описывается модель FARIMA и алгоритмы ее настройки. Приводятся результаты моделирования трафика по реальным данным.

ГРАФИЧЕСКАЯ ЧАСТЬ

Плакат 1. Постановка задачи

Плакаты 2, 3. Показатель Хёрста

Плакаты 4, 5. Самоподобные процессы

Плакат 6. Обзор методов моделирования

Плакаты 7, 8, 9. Модель FARIMA

Плакат 10. Интерфейс приложения

Плакаты 11, 12. Результаты моделирования

Плакат 13. Результаты работы

Всего плакатов……………………………………………………………13

**Оглавление**

[1 Самоподобный трафик. Методы моделирования 8](#_Toc358844710)

[1.1 Определение самоподобного процесса 8](#_Toc358844711)

[1.2 Обзор методов моделирования самоподобного телетрафика 13](#_Toc358844712)

[1.3 Постановка задач исследования 19](#_Toc358844713)

[2 Построение математической модели самоподобного телетрафика 20](#_Toc358844714)

[2.1 Оценка показателя Хёрста 20](#_Toc358844715)

[2.2 Оператор дробной разности 29](#_Toc358844716)

[2.3 Модель FARIMA 30](#_Toc358844717)

[3 Программная реализация синтеза самоподобного трафика 34](#_Toc358844718)

[3.1 Описание программы 34](#_Toc358844719)

[3.1.1 Общие сведения 34](#_Toc358844720)

[3.1.2 Функциональное назначение 34](#_Toc358844721)

[3.1.3 Описание логической структуры программы 34](#_Toc358844722)

[3.1.4 Используемые технические средства 36](#_Toc358844723)

[3.1.5 Установка и удаление программы 36](#_Toc358844724)

[3.1.6 Вызов и загрузка 37](#_Toc358844725)

[3.1.7 Входные данные 37](#_Toc358844726)

[3.1.8 Выходные данные 38](#_Toc358844727)

[3.2 Руководство оператора 38](#_Toc358844728)

[3.2.1 Назначение программы 38](#_Toc358844729)

[3.2.2 Условия выполнения программы 39](#_Toc358844730)

[3.2.3 Выполнение программы 39](#_Toc358844731)

[4 Моделирование трафика 44](#_Toc358844732)

[Заключение 53](#_Toc358844733)

[Список источников 54](#_Toc358844734)

**Введение**

При проектировании, запуске и эксплуатации информационных телекоммуникационных сетей одной из основных проблем является задача обеспечения качества обслуживания (заданных уровней задержек, потерь и пр.) при обработке потока данных – трафика, являющегося следствием информационного обмена между системами.

До недавнего времени теоретическую базу для проектирования систем распределения информации обеспечивала теория телетрафика. Данная теория хорошо описывает процессы, происходящие в таких системах распределения информации, как телефонные сети, построенных по принципу коммутации каналов. Наиболее распространенной моделью потока вызовов (данных) в теории телетрафика является простейший поток (стационарный ординарный поток без последействия), также называемый стационарным пуассоновским потоком [1].

Настоящий период бурного развития высоких технологий привел к появлению и повсеместному распространению сетей с пакетной передачей данных, которые постепенно стали вытеснять системы с коммутацией каналов, но, по-прежнему, они проектировались на основе общих положений теории телетрафика.

В 1993 году группа американских исследователей опубликовали результаты своей новой работы, которая в корне изменила существующие представления о процессах, происходящих в телекоммуникационных сетях с коммутацией пакетов. Эти исследователи изучили трафик в информационной сети корпорации Bellcore и обнаружили, что потоки в ней нельзя аппроксимировать простейшими и, как следствие, они уже имеют совершенно иную структуру, чем принято в классической теории телетрафика. В частности, было установлено, что трафик такой сети обладает так называемым свойством «самоподобия», т.е. выглядит качественно одинаково при почти любых масштабах временной оси, имеет память (последействие), а также характеризуется высокой пачечностью [2].

В результате теоретический расчет параметров системы распределения информации, предназначенной для обработки такого трафика, по классическим формулам дает некорректные и неоправданно оптимистические результаты. Более того, привычные алгоритмы обработки трафика, созданные для работы с простейшими потоками, оказываются недостаточно эффективными для потоков с самоподобием.

В вопросе моделирования самоподобного телетрафика остается множество нерешенных задач:

* фактически отсутствует строгая теоретическая база, которая пришла бы на смену классической теории массового обслуживания при проектировании современных систем распределения информации с самоподобным трафиком;
* нет единой общепризнанной модели самоподобного трафика;
* не существует достоверной и признанной методики расчета параметров и показателей качества систем распределения информации при влиянии эффекта самоподобия;
* отсутствуют алгоритмы и механизмы, обеспечивающие качество обслуживания в условиях самоподобного трафика.

**1 Самоподобный трафик. Методы моделирования**

1.1 Определение самоподобного процесса

Рассмотрим случайный процесс как дискретную последовательность случайных величин:



где  – случайная величина с заданным законом распределения.

Будем предполагать, что все рассматриваемые случайные процессы имеют ограниченную ковариацию  и следовательно дисперсию .

Поведение процесса  при агрегировании имеет принципиальное значение, поскольку некоторые механизмы в теории систем распределения информации, например, распределение пропускной способности, часто рассматриваются применительно к агрегированным процесса [2].

Случайный процесс будет обладать свойством самоподобия, если агрегированный процесс *m*-го порядка



будет иметь корреляционную функцию  совпадающую с корреляционной функцией исходного случайного процесса  для любых[3]. При выполнении данного условия можно утверждать, что дисперсия агрегированного процесса  убывает согласно выражению



т.е. вариация агрегированных процессов – средних выборок – уменьшается медленнее, чем величина, обратная размеру выборки.

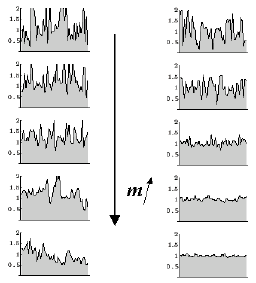


Рисунок 1. Временные реализации реального сетевого (самоподобного) трафика (слева) и традиционной несамоподобной (Пуассоновской) модели телетрафика (справа) при различных масштабах временной оси. Сверху вниз масштаб временной оси укрупняется

В результате в самоподобных процессах имеет место явления долгосрочной зависимости, которое приводит к расходимости корреляционных функций процесса:



Наконец энергетический спектр самоподобных процессов описывается выражением



Важнейшим параметром, характеризующим «степень» самоподобия  случайного процесса, является параметр Хёрста (Hurst). Он является индикатором самоподобия процесса, а также свидетельствует о наличии у него таких свойств как персистентность/антиперсистентность и продолжительная память [4].

Следует заметить, что в случае  говорят о персистентном (поддерживающемся) поведении процесса, либо о том, что процесс обладает

длительной памятью. Другими словами, если в течение некоторого времени в прошлом наблюдались положительные приращения процесса, то есть происходило увеличение, то и впредь в среднем будет происходить увеличение. Иначе говоря, вероятность того, что процесс на  шаге отклоняется от среднего в том же направлении, что и на  шаге настолько велика, насколько параметр  близок к единице. Таким образом, персистентные стохастические процессы обнаруживают четко выраженные тенденции изменения при относительно малом «шуме».

В случае  говорят об антиперсистентности процесса. Здесь высокие значения процесса следуют за низкими, и наоборот. Другими словами, вероятность того, что на  шаге процесс отклоняется от среднего в противоположном направлении (по отношению к отклонению на шаге) настолько велика, насколько параметр  близок к нулю.

При *H* равном  отклонения процесса от среднего являются действительно случайными и не зависят от предыдущих значений.

Заметим, что именно свойство персистентности оправдывает применение для моделирования и предсказания самоподобных рядов AR (авторегрессионных) моделей. В частности, получили широкое распространение такие авторегрессионные модели, как ARMA (процесс скользящего среднего), ARIMA (интегральный процесс скользящего среднего) и FARIMA (фрактальный интегральный процесс скользящего среднего).

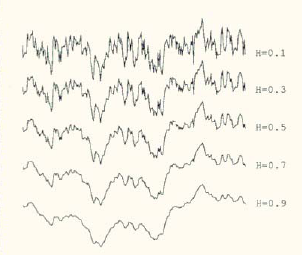


Рисунок 2. Реализация процесса с различным показателем Хёрста

Если для случайного процесса определить его среднее как выборочную дисперсию положить равной и определить изменчивость случайного процесса по формуле  то для большинства самоподобных процессов будет выполняться условие



или



где  – параметр Хёрста.

Величина , причем для процессов не обладающих свойством самоподобия , а для самоподобных процессов данный параметр обычно изменяется в пределах от 0,7 до 0,9. Параметр , который был введен ранее для характеристики асимптотических свойств самоподобного процесса, связан с параметром Хёрста соотношением



На практике параметр Хёрста можно определить путем построения графика отношения  в зависимости от  при разных и вычислить величину  как тангенс угла наклона полученной линии. Следует заметить, что полученное множество точек не будет строго линейным, поэтому их следует аппроксимировать линией, например, по методу наименьших квадратов. Данная методика определения параметра Хёрста получила название R/S-метод [3].

Впервые на самоподобие процессов, описывающих трафик в пакетных сетях, обратили внимание в начале 90-х годов прошлого века. В частности на основе результатов экспериментов анализа трафика в пакетной сети было установлено, что распределение числа пакетов в единицу времени очень хорошо описывается самоподобным случайным процессом с параметром Хёрста около 0,65-0,8 [4].

R/S-метод дает лишь приближенное значение показателя Хёрста, поэтому для его вычисления целесообразно пользоваться несколькими методиками и сравнения полученных результатов. Рассмотрим метод определения величины  на основе периодограммного анализа [5].

Для самоподобного случайного процесса вычисляется периодограмма по формуле



где  - длина временного ряда.

Учитывая, что самоподобность влияет на характер спектра при должен получаться график зависимости спектральной плотности вида

при 

Из последнего выражения следует, что множество случайных точек  будут располагаться линейно с коэффициентом наклона линии . На практике для вычисления оценки должны использоваться только нижние 10% частот, т.к. описанное выше поведение справедливо только для области частот, близких к нулю. Основным недостатком данного метода является большой объем вычислений при построении оценки показателя Хёрста.

1.2 Обзор методов моделирования самоподобного телетрафика

Методы моделирования сетевого трафика, концептуально можно разделить на два класса – аналитические и имитационные. Аналитическая модель – это совокупность математических выражений, формально описывающая моделируемый объект или процесс. Такие модели удобны для проведения теоретических исследований, однако, для большинства источников построение адекватной аналитической модели крайне затруднительно.

Имитационная модель – это набор алгоритмов, генерирующий некую последовательность, которая по своим характеристикам близка к реальной (экспериментально снятой с действующего объекта) последовательности. В качестве такой последовательности, например, может быть сетевой трафик. Использование имитационных моделей является зачастую более предпочтительным и удобным. В тоже время, как правило, имитационные модели имеют узкую специфику, и применение таких моделей требует значительной работы для адаптации модели под новые условия применения.

Возможны также комбинированные модели, сочетающие в себе аналитическую и алгоритмическую части [6]. На сегодняшний день разработано множество моделей, предназначенных для имитации фрактального трафика. Анализ доступных публикаций по моделированию самоподобного трафика позволяет выделить следующие модели:

*Фрактальное броуновское движение (Fractional Brown Motion - FBM).* В основе модели FBM лежит случайный процесс, начинающийся в начале координат с независимыми бесконечно малыми гауссовскими приращениями. FBM описывается аналитически. Также для генерации FBM широко используются алгоритмы случайного перемещения средней точки (RMD-алгоритм) и алгоритмы последовательного случайного сложения (SLA-алгоритм) [7].

*Фрактальный гауссовский шум (Fractional Gaussian Noise - FGN)* [8]. FGN – стационарный в широком смысле стохастический процесс с определенным параметрами (средним значением, дисперсией, Хёрста) и автокорреляционной функцией заданного вида. По сравнению с обычным гауссовым шумом, FGN имеет дополнительный параметр Хёрста, который количественно определяет степень фрактального масштабирования. Основная трудность использования FBM и FGN – подбор наилучших значений параметров для получения генерируемого трафика, близкого по свойствам к экспериментально снятым реализациям трафика.

*Хаотические отображения (Chaotic Map – СМАР)*. Такие модели являются достаточно распространенными и концептуально простыми, они используют меньшее число параметров, чем FGN и FBM, и их выбор имеет более наглядную трактовку.

*Модели на основе техники «динамического моделирования Маркова» (Dynamic Markov Modeling – DMM)*. Эти модели представляют собой автоматы с конечным числом состояний, изображаемые орграфами или диаграммами состояний модели. В процессе обучения модели, при получении очередного символа входного потока, происходит переход модели в следующее состояние и модификация частотных счетчиков, соответствующих вероятностям переходов. Выходом модели является набор вероятностей появления символов [9].

*Модели с использованием нечеткой логики*. Построение нечетких моделей, как правило, основано на настройке функций принадлежности по параметрам нечетких множеств, используемых в правилах, весов правил и на настройке операций. В [14] предложено использование нечеткой модели временного ряда нечетких тенденций, позволяющей эффективно моделировать и прогнозировать работу сложной технической системы.

*Нейросетевые модели*, которые позволяют решить задачу аппроксимации функций нескольких переменных по обучающей выборке путем погружения временного ряда в многомерное пространство.

*Авторегрессионные модели (Autoregressive Models – AR)* широко применяются для моделирования и предсказания благодаря свойству длительной памяти самоподобных процессов. В этих моделях текущее значение генерируемой величины рассчитывается как взвешенная сумма  предыдущих отсчетов плюс случайная переменная. Как разновидности таких моделей используются модели ARMA (процесс скользящего среднего), ARIMA (интегральный процесс скользящего среднего) и FARIMA (фрактальный интегральный процесс скользящего среднего). К достоинствам последней необходимо отнести возможность гибкого управления корреляционной структурой (кратковременной и долговременной зависимостями, произвольным распределением).

*Фрактальные точечные процессы (Fractal Point Process – FPP)* являются очень наглядными для моделирования самоподобного трафика. Простейший точечный процесс представляется на временной оси неубывающей ступенчатой функцией, моменты роста которой являются случайными. Существует много модификаций FPP, которые достаточно экономичны и вычислительно эффективны.

*ON/OFF-модели*. В этих моделях трафик рассматривается как комбинация источников, которые его генерируют. Каждый источник имеет следующую структуру. Некоторый период времени они могут генерировать пакеты информации (так называемые ON–периоды), при этом внутри одного периода пакеты приходят с одинаковыми интервалами между ними. После ON–периода следует OFF–период, когда источник не генерирует пакеты. Отметим, что ON/OFF – процесс классифицирован как чередующийся фрактальный процесс восстановления - одна из разновидностей FPP.

*Фрактальное движение Леви (Fractional Levi Motion – FLM)* относится к так называемым устойчивым процессам. В основе его моделирования лежат симметричные -устойчивые распределения, характеризуемые кроме показателя Хёрста, еще и показателем Леви. FLM можно рассматривать как некое обобщение FBM и эффективно использовать для моделирования интенсивности трафика или скорости передачи, имеющих теоретически бесконечную дисперсию.

*Мультифрактальные модели (Multifractional - MF)* удачно воспроизводят трафик, агрегированный от нескольких существенно отличающихся источников. Мультифрактальность трафика проявляется в изменении статистических свойств реализации трафика при изменении масштаба агрегирования [11]. Для описания таких свойств вводятся дополнительные масштабная функция и моментный коэффициент. В основе MF моделей лежат консервативные бинарные мультипликативные каскады.

*Вейвлет модели (Wavelet Models)* строятся на основе обратного дискретного вейвлет-преобразования, который состоит в формировании с помощью масштабных и вейвлет-коэффициентов дискретного временного ряда, используя функции детализации различного масштаба на основе прототипа полосовой вейвлет-функции и низкочастотной скейлинг-функции. Вейвлет-модели могут иметь различное количество параметров (три и более) и достаточно эффективны для моделирования самоподобного трафика. Очень близкие по свойствам к вейвлет-моделям являются модели на основе преобразования всплесков [12].

*Модели на основе классических систем массового обслуживания*. Как правило, такие модели удачно описывают трафик с пуассоновскими потоками. Однако, как показано в [13], такая модель как М/G/∞ способна создать приблизительно самоподобный трафик путем управления поведением «хвоста» произвольного распределения обслуживания пользователей, создавая тем самым долговременную зависимость.

Практически все рассмотренные модели хорошо подходят для моделирования самоподобного трафика данных в телекоммуникационных сетях с коммутацией пакетов. Все модели обладают такими необходимыми для качественного моделирования свойствами, как долговременная зависимость, масштабируемость, стационарность и т.д. Однако современные исследования экспериментально снятых реализаций трафика показывают, что характеристики трафика могут изменяться в самых широких пределах и зависеть от большого числа параметров и настроек реальных сетей, характеристик протоколов, передаваемой информации и поведения пользователей.

Кроме перечисленных выше, выявлены, например, такие характеристики трафика как наличие кратковременных зависимостей, нестационарность, мультифрактальность.

Общим недостатком используемых в настоящее время моделей сетевого трафика является их направленность на какую–либо конкретную разновидность трафика либо сети и отсутствие универсальности, хотя некоторые авторы претендуют на универсальность разработанных моделей. Кроме того, применение их на практике приводит к большому объему исследовательской работы, требуемой для адаптации (обучения) модели к параметрам сетевой конфигурации или параметра трафика. Все это значительно усложняет построение универсальной модели, из-за большого разнообразия, как самих источников, так и сетевых конфигураций, оказывающих влияние на их работу. Отметим также, что поскольку реальный трафик, как правило, не является строго самоподобным, простые модели, такие как, например, FBM, FGN, FPP, FLM и М/G/∞, не всегда могут адекватно описать реальный трафик.

Адекватность описания реального трафика достигается путем усложнения моделей, объединения нескольких моделей, введения дополнительных параметров. Естественно для более сложных моделей требуются большие вычислительные возможности или большее время для генерирования реализаций трафика. Это не является проблемой при проведении единичных экспериментов и исследований, когда время генерации очередного отсчета трафика или всей реализации ограничено «терпеливостью» исследователя. Но как только встает проблема применения некой модели для прогнозирования трафика с целью дальнейшего оптимального управления ресурсами сети, фактор реального времени накладывает жесткие ограничения на сложность модели при заданных вычислительных возможностях сетевых узлов и агентов управления.

Прогнозы рынка современных видов услуг и реальные статистические данные операторов свидетельствуют, в ряде случаев, о существенных расхождениях. В отличие от телефонии прогностические оценки трафика данных очень недостоверны. Это свидетельствует о недостаточном развитии теории прогнозирования мультисервисного (самоподобного) трафика. С дру-

гой стороны, именно на знании параметров трафика базируется основная часть работ по расчету сети.

1.3 Постановка задач исследования

Классические методы моделирования трафика – фрактальный Гауссовский шум, фрактальное Броуновское движение и пр. – как правило, описывают только долгосрочные зависимости, однако исследования показывают, что в реальном трафике присутствуют и долгосрочные и краткосрочные зависимости. FARIMA позволяет моделировать долгосрочные и краткосрочные зависимости. Кроме того, FARIMA-процесс после операции дробного дифференцирования сводится к модели ARMA, что упрощает работу с ней и позволяет использовать уже известные методы построения моделей.

Целью данной работы стала разработка математического и программного обеспечения для моделирования самоподобного трафика на основе модели FARIMA для последующего анализа системы массового обслуживания

Задачами исследования стали:

* критический обзор существующих методов для моделирования трафика, сравнительный анализ;
* исследование аппарата FARIMA-модели, ее нестационарного аналога;
* создание программного продукта для моделирования трафика;
* апробация полученной модели на реальных данных и сравнение результатов моделирования и реальных данных;

**2 Построение математической модели самоподобного телетрафика**

2.1 Оценка показателя Хёрста

Известно несколько методов оценки показателя Хёрста, характеризующего статистическую самоподобность случайных процессов и временных рядов. В основном это эвристические методы, основанные на анализе R/S-статистики и графика изменения дисперсии наблюдаемого процесса. Специфические свойства энергетического спектра статистически самоподобного случайного процесса (временного ряда) или его корреляционной функции используются в оценке Витла [14]. Перечисленные методы, при решении задач обработки процессов и полей (изображений) со статистически самоподобной (фрактальной) структурой используются преимущественно как диагностический инструмент для определения персистентности в данных и не учитывают неизбежные ошибки при их получении (шум в наблюдении) и возможную априорную неопределенность относительно несущественных, но неизвестных параметров в модели наблюдения. Следует отметить метод оценки показателя Хёрста, основанный приближенном представлении самоподобных случайных процессов в конечномерном базисе вейвлет-функций, приспособленном для анализа нестационарных процессов. Однако оценки, полученные этим методом, получаются смещенными.

Таким образом, проверка на практике самоподобности с использованием оценки показателя Хёрста является сложной задачей, а перечисленные методы ее решения непригодны для статистического анализа и недостаточно эффективны, если требуется более тонкий анализ данных искаженных различного рода помехами в реальных условиях, когда оперируют конечным наборам данных. Например, при распознавании двух или более областей двумерного объекта, с различными параметрами фрактальной размерности (показателями Хёрста) на фоне шумовой компоненты в виде белого гауссовского шума.

Ниже рассматривается статистическая постановка задачи совместного определения параметров специального класса случайных процессов (фрактального броуновского шума (ФБШ)), полученного взятием приращений фрактального броуновского движения (ФБД) [14], наблюдаемого на фоне белого гауссовского шума, а также подход к ее решению, основанный на использовании функции максимального правдоподобия. Исследуется потенциальная точность совместной оценки параметров ФБШ.

Предположим, что в течение времени наблюдается смесь:

, , ,

фрактального броуновского движения , где  – показатель Хёрста, а также белого гауссовского шума  со спектральной плотностью мощности  и нулевым математическим ожиданием.

Основные статистические свойства ФБД – процесса следующие:

* среднее значение ;
* дисперсия ;
* нормированная функция корреляции ;
* стационарные приращения ;

Процесс  характеризуется нестационарной функцией корреляции. Не снижая общности постановки задачи, будем полагать, что обработке подлежит приращение наблюдения

,

здесь  - приращение текущего момента времени.

Первое слагаемое в этом выражении ФБШ. Автокорреляционная функция приращения наблюдения стационарна и имеет вид

.

В результате требуется дать статистически оптимальные оценки параметров  и  по данным . В рамках имеющейся априорной и измерительной информации для оценивания искомых параметров будем опираться на метод максимального правдоподобия.

Воспользуемся спектральным представлением логарифма функции максимального правдоподобия, которое в предположении, что размер мелких деталей спектра мощности процесса  гораздо меньше характерного масштаба его изменения, имеет вид:



где  – спектр Фурье наблюдения ,

 – энергетический спектр наблюдения, в котором первое слагаемое – энергетический спектр ФБШ , а второе – двусторонняя спектральная плотность шума наблюдения.

В предположении, что спектр ФБШ  ограничен по полосе  и выполняется условие большого отношения сигнал/шум в наблюдении –  (), оценка параметра  может быть получена из уравнения максимального правдоподобия в явном виде независимо от значения параметра :

,

где - полоса частот.

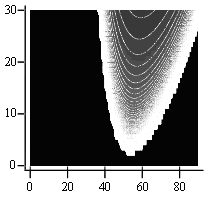
Следует иметь в виду, что оценка  может принимать отрицательные значения. В тоже время, так как  является спектральной амплитудой, то она должна быть неотрицательной величиной. В противном случае функция правдоподобия может быть неопределенна в области ее вещественных значений. Тогда за максимально правдоподобную оценку принимается значение



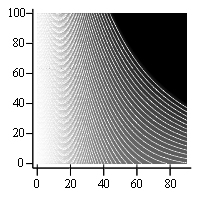
На рисунке 3*a*, где в градациях серого цвета показано изменение оценки  в плоскости , хорошо видна граница области, имеющей уровень черного цвета.

Из рисунка 3*б* видно, что на множестве возможных значений параметра оценка дисперсии  статистически самоподобного случайного процесса принимает минимальное значение в окрестности истинного его значения и практически не зависит от спектральной плотности шума  из области ее допустимых значений. Именно это свойство используется в одном из эвристических методов оценки параметра *H*. Для реализации статистически квазиоптимальной оценки параметра  необходимо найти минимум функции , при значении спектральной амплитуды . Вид этой функции в случае, когда оценка  получена по 256-ти значениям спектра наблюдаемого сигнала, а отношении сигнал/шум  показан на рисунке 4. Функция имеет минимальное значение при , которое принимается за максимально правдоподобную оценку истинного значения параметра . Функция  строилась с шагом дискретизации  и .

Рисунок 3. Изменение оценки  в плоскости 



*б*







*а*

Очевидно, что качество оценки параметра  зависит от величины  При малых выборках отсчетов спектра точность и надежность оценки  позволяют определить интервальные оценки. В таблице представлены данные оценки неизвестного математического ожидания случайной величины  заданной на множестве из ста независимых оценок  по выборочной средней  при помощи доверительного интервала с надежностью (доверительной вероятностью)  в случае, когда ее дисперсия неизвестна.

Границы доверительных интервалов оценок  показанные в таблице, рассчитаны для значения параметра  связанного как с антиперсистентным () так и с персистентным () характером ФБШ. Легко заметить, что при небольших значениях величины  проявляется заметное смещение оценки параметра к значению *H* равному  характерному для гауссовского шума. По мере увеличения отношения сигнал/шум смещение доверительного интервала относительно истинного значения параметра  уменьшается.







Рисунок 4. Вид функциипри

Границы доверительных интервалов оценок 

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Сигнал/Шум | | | |
|  |  |  |  |
|  | Верх.гр | 0,362 | 0,215 | 0,177 | 0,132 |
| Нижн. гр | 0,346 | 0,194 | 0,156 | 0,111 |
| Интервал | 0,016 | 0,021 | 0,021 | 0,021 |
|  | Верх.гр | 0,56 | 0,746 | 0,844 | 0,856 |
| Нижн. гр | 0,539 | 0,724 | 0,82 | 0,833 |
| Интервал | 0,021 | 0,022 | 0,024 | 0,023 |

В случае отсутствия априорной информации о параметре , что вполне вероятно в практике измерений, оценка  будет функцией неизвестной величины . В этом случае оценка параметра  может проводиться совместной минимизацией функции  по обоим аргументам.

Рассмотрим вариант оценки параметра , когда из уравнения максимального правдоподобия  при большом  и учете в разложении в ряд функции  элементов не выше второй степени по, находится оценка параметра :

.

В этом случае оценка  определяется как минимальное значение функции максимального правдоподобия  в плоскости  по обоим аргументам. На рисунке 5, в градациях серого цвета отображены ее значения и линии равного уровня. Минимальное значение функции (уровень черного) находится в окрестности истинных значений параметров  и  (для наглядного изображения функции на плоскости  значения аргументов на осях смещенные). Это свойство функции сохраняется для всех возможных значений параметра .

Таким образом, получена функция, позволяющая провести статистически оптимальный анализ данных в присутствии шума в наблюдении, в условиях ограниченного количества отсчетов выборки, а так же неизвестных параметров: дисперсии сигнала  и спектральной плотности шума .

Рассчитаем потенциальную точность совместной оценки параметров ФГШ, обращая матрицу Фишера их совместной оценки:

.

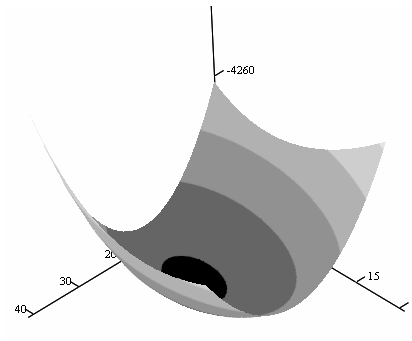








Рисунок 5. Значения  в плоскости 

Дисперсия ошибок оценки параметров  и  в зависимости от истинного значения  и спектральной плотности шума  изображены на рисунках 6а и 6б соответственно. На рисунке 6а видно, что дисперсия ошибок оценки параметра  возрастает с увеличением спектральной плотности шума и при этом зависит от истинного его значения.

*а б*

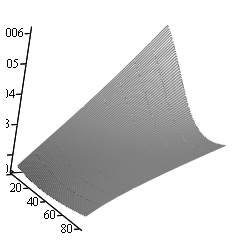
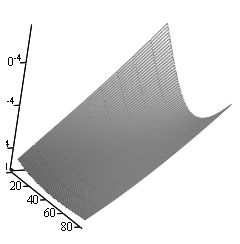










Рисунок 6. Дисперсия ошибок оценки параметров  и 

Дисперсия ошибок оценки параметров  и  в зависимости от истинного значения  и дисперсии фрактального броуновского шума (рисунки 7а, 7б) соответственно.

*а б*

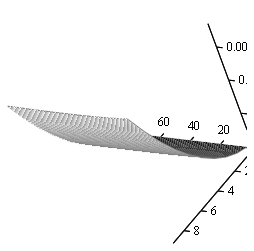
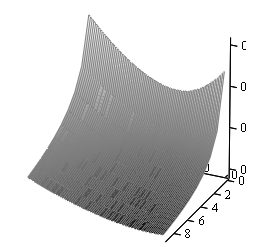










Рисунок 7. Дисперсия фрактального броуновского шума

Таким образом, получена потенциальная точность оценки показателя Хёрста ФГШ, неизвестных значений его дисперсии и спектральной плотности шума при их совместной оценки методом максимального правдоподобия. Проанализировано влияние уровня шума при условии ограниченного размера выборки наблюдения.

2.2 Оператор дробной разности

Понятие дробной разности, которое расширяет понятие разности целого порядка, используемой в модели ARIMA [4]. Процесс случайного блуждания ARIMA(0,1,0),который является дискретным аналогом броуновского движения, определяется как



где  – оператор взятия разности первого порядка,

 – лаговый оператор ().

Тогда, используя биномиальные коэффициенты, можно получить оператор взятия дробной разности



Биномиальный коэффициент для натуральных значений *d* и *k* можно вычислить по формуле



Используя гамма-функцию, которая является расширением факториала для комплексных чисел ( для натуральных *d*), получим формулу вычисления биномиального коэффициента для действительных *d* и *k*:



Т. о. оператор взятия дробной разности можно записать как



* 1. Модель FARIMA

Модель FARIMAявляется расширением модели ARIMA, позволяющим параметру dпринимать нецелые значения. Модель FARIMA для вре­менного ряда может быть представлена в виде



где 

 – лаговый оператор, т.е. 

 – дробно-разностный оператор [13],



 – белый шум, 

Условие  вытекает из стационарности и обратимости FARIMA-процесса, т.к. при  процесс FARIMAстационарен, а при обратим [1]. Временной ряд с любым *d* можно привести к условию  взятием нужного числа целых разностей. При процесс FARIMA(*p*, 0, *q*) представляет собой обычный про­цесс ARMA(*p*, *q*). Процесс FARIMA (0, *d*, 0), т.е. дробно-разностный шум (FDN) – простейший вид процесса FARIMA и может описы­вать только долгосрочные зависимости [13]. Параметр *d* показывает степень самоподобия процесса и , где *H* – показатель Хёрста. Процесс FARIMA(*p*, *d*, *q*)можно представить как процесс ARMA, в котором белый шум заменен дробно-разностным



где  – дробно-разностный шум [13].

Таким образом, реализация FARIMA-процесса сводится к 2-м шагам:

* Сгенерировать  (при );
* Сгенерировать ARMA-процесс с помощью полученного [13];

Генерация дробно-разностного шума с использованием гамма-функций – вычислительно сложная задача, поэтому был предложен рекурсивный алгоритм [14]:

Если разложить в ряд Тейлора , получим



умножив каждый член ряда на –1 получим



Теперь умножим на 



использовав свойство гамма-функции , получим







Тогда  эквивалентно







Процедура оценки параметров для известного временного ряда состоит в следующем:

* Получить из исходного ряда ряд с нулевым математическим ожиданием  ;
* Получить оценку параметра *d*для ряда, используя известные методы оценки параметра Хёрста [13];
* Взять дробную разность ряда 



На практике нам нужны значения ряда , где *M* должно быть достаточно велико. Их можно оценить, используя дополнительный обратный прогноз (backward prediction), основанный на модели авторегрессии [3];

* Используя методы оценки параметров ARMA модели, оценить параметры , и  ряда  [13];

**3 Программная реализация синтеза самоподобного трафика**

3.1 Описание программы

3.1.1 Общие сведения

Наименование: «Синтез самоподобного трафика».

Исполняемый файл: STS.exe.

Программа разработана в интегрированной среде Embarcadero C++ Builder XE, на языке программирования C++, предназначена для эксплуатации на компьютерах IBM PC и совместимых с ними.

3.1.2 Функциональное назначение

Программа «Синтез самоподобного трафика» предназначена для расчета параметров модели FARIMAпо введенным данным о наблюдениях самоподобного трафика, синтеза трафика по данной модели и предсказывания поведения самоподобного процесса. Разработанный программный продукт может применяться для расчета характеристик СМО в системах с самоподобным трафиком и для исследования свойств самоподобных процессов.

3.1.3 Описание логической структуры программы

Программный продукт состоит из пяти объектов форм и других стандартных объектов-элементов среды Builder C++.Пользовательские классы в программе не используются.

Объект Form2 содержит следующие методы:

* Button1Click – открытие окна «Параметры модели» (Form4) для ввода параметров ARMA составляющей модели FARIMA;
* Button2Click – оценка параметров модели FARIMA для значений временного ряда, хранящихся в файле, открытие окна «График» (Form3)с графиком исходного ряда, настроенной на него моделью и значениями ряда;
* CheckBox1Click – обработка нажатия на флажок «оценить». Если флажок выбран, поле ввода параметра *d* становится неактивным;
* UpDown1Click – обработка нажатия на кнопки увеличения/уменьшения количества параметров авторегрессии;
* UpDown2Click – обработка нажатия на кнопки увеличения/уменьшения количества параметров скользящего среднего;

Объект Form4содержит следующие методы:

* Button1Click – синтез самоподобного трафика по введенным параметрам и открытие окна «Синтез трафика» (Form5) с графиком и значениями сгенерированного временного ряда;

Объект Form4содержит следующие методы:

* Button1Click – синтез самоподобного трафика по введенным параметрам и открытие окна «Синтез трафика» (Form5) с графиком и значениями сгенерированного временного ряда;

Используемые функции:

* long double f\_diff (int n, long double \* ystore, long double \* pistore)

Функция производит дробное дифференцирование временного ряда. Параметры: количество элементов ряда, элементы ряда, коэффициенты для вычисления дифференциала;

* long double F (long double \* a, long double \* b, long double \* x, int lim=5)

Функция для вычисления в формуле логарифмической функции правдоподобия [15]. Параметры: массив параметров авторегрессии, массив параметров скользящего среднего, массив значений временного ряда, количество итераций;

* long double FDN (int n, long double a, long double \* ystore, long double \* pistore)

Функция для генерации дробно-разностного шума (fractional differencing noise). Параметры: число значений генерируемого ряда, значение белого шума для генерации одного значения FDN, массив для записи результатов, массив коэффициентов дифференцирования;

3.1.4 Используемые технические средства

Программа работает на компьютерах IBM PC и совместимых с ними.

Минимальные системные требования для программы:

* процессор Pentium 300 Mhz;
* оперативная память от 128 Mb RAM;
* около 5 Mb свободного места на жестком диске для хранения исходного и исполняемого файлов, а также файла с результатами работы;
* операционная система Windows XP/Vista/Seven;
* монитор;
* манипулятор типа «мышь»;
* клавиатура.

Реализация программы выполнена в среде программирования Embarcadero C++ Builder XE.

3.1.5 Установка и удаление программы

Для установки программы необходимо скопировать исходную директорию в директорию установки.

Для удаления программы необходимо удалить исходную директорию из директории установки.

3.1.6 Вызов и загрузка

Для вызова программы «Синтез самоподобного трафика» необходимо запустить файл «STS.exe» из текущей папки программы. По запросу пользователя через стандартные диалоговые окна возможен выбор файлов данных.

3.1.7 Входные данные

Входными данными для программы служит файл с элементами временного ряда, разделенными символом переноса строки, параметры модели FARIMA. Разделителем дробной части является точка. Пример входного файла представлен на рисунке 8.

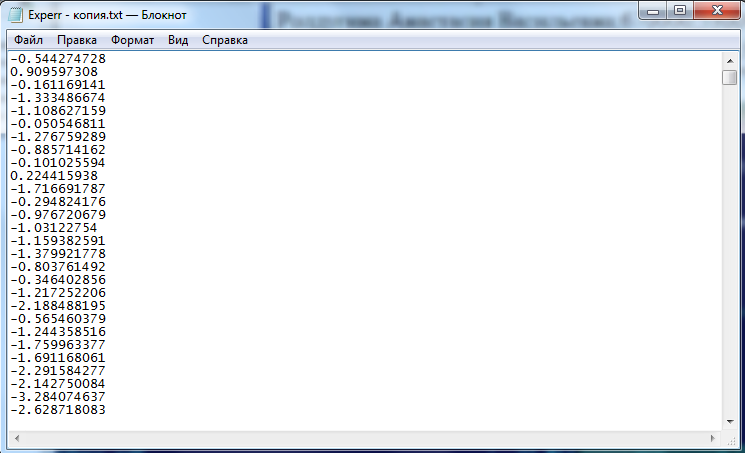


Рисунок 8. Пример файла входных данных

3.1.8 Выходные данные

Выходными данными являются график исходного ряда и ряда, полученного с помощью модели FARIMA, оценки параметров процесса, значения сгенерированного трафика, которые можно сохранить в файл (рисунок 9).

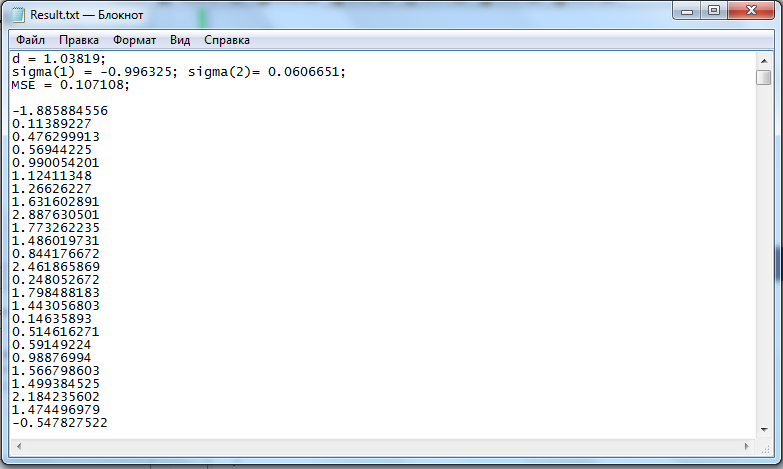


Рисунок 9. Пример выходного файла

3.2 Руководство оператора

3.2.1 Назначение программы

Программа «Синтез самоподобного трафика» предназначена для расчета параметров модели FARIMAпо введенным данным о наблюдениях самоподобного трафика, синтеза трафика по данной модели и предсказывания поведения самоподобного процесса.

Разработанный программный модуль может применяться для расчета характеристик СМО в системах с самоподобным трафиком и для исследования свойств самоподобных процессов.

3.2.2 Условия выполнения программы

Программа работает на компьютерах IBM PC и совместимых с ними.

Минимальные системные требования для программы:

* процессор Pentium 300 Mhz;
* оперативная память от 128 Mb RAM;
* около 5 Mb свободного места на жестком диске для хранения исходного и исполняемого файлов, а также файла с результатами работы программы;
* операционная система Windows XP/Vista/Seven;
* монитор;
* манипулятор типа «мышь»;
* клавиатура.

3.2.3 Выполнение программы

При запуске программы «Синтез самоподобного трафика» на экран выводится главное окно приложения (рисунок 10). Оно содержит строку меню, окно для вывода оценок параметров модели, поля для ввода параметров модели, кнопки «Оценить» и «Сгенерировать». Меню включает следующие пункты: «Файл», «Справка». Элемент меню «Файл» содержит позиции: «Открыть», которая позволяет загрузить входной файл данных; «Сохранить», которая позволяет сохранить все измененные данные в требуемом формате, и «Выход», служащую для выхода из программы.

Элемент меню «Справка» вызывает справку к программе (рисунок 11).

Кнопка «Оценить» становится активной после загрузки данных из файла. При клике на кнопку «Оценить» на экране появляется окно «График» (рисунок 12). Оно содержит график исходного ряда и настроенной на него модели. Также в окне формы 1 появляются значения оценок параметров настроенной модели.

При клике на кнопку «Сгенерировать» на экране появляется окно «Параметры модели» (рисунок 13), содержащее поля для задания параметров модели.

При клике на кнопку «Продолжить» на экране появляется окно «Синтез трафика» (рисунок 14), содержащее график синтезированного по модели ряда и значения ряда.

Для исследования временного ряда необходимо открыть файл, содержащий значения исследуемого временного ряда, используя кнопку меню «Файл->Открыть». При успешном открытии файла можно оценить параметры модели FARIMA для исследуемого ряда. Для этого необходимо задать количество параметров авторегрессии и скользящего среднего, а также параметр интегрирования d или установить флажок «Оценить» для оценки параметра *d*. Для оценки параметров необходимо нажать кнопку «Оценить».

Для синтеза самоподобного трафика необходимо задать количество параметров авторегрессии и скользящего среднего, а также параметр интегрирования *d* и нажать кнопку «Сгенерировать». Далее необходимо ввести значения параметров авторегрессии и скользящего среднего в окне «Параметры модели». Если в программе уже была произведена оценка для временного ряда, значения оценок параметров подставляются автоматически. По нажатию на кнопку «Продолжить» будет показан график сгенерированного ряда и значения членов ряда.

Для сохранения результатов необходимо выбрать пункт меню «Файл->Сохранить»

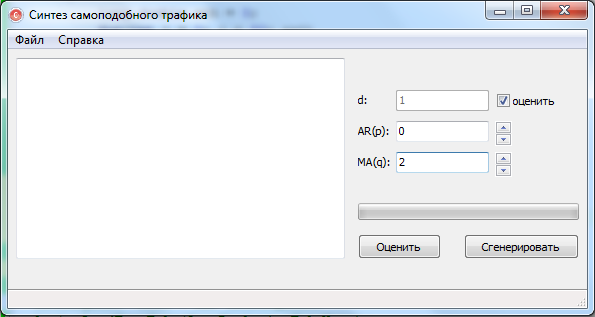


Рисунок 10. Главное окно программы

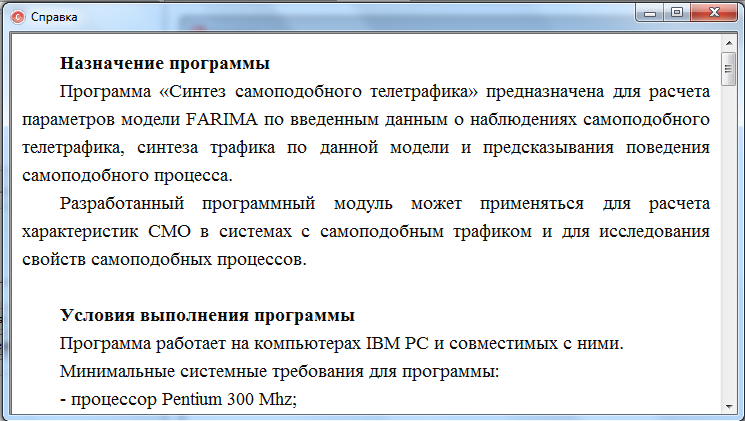


Рисунок 11. Окно «Справка»

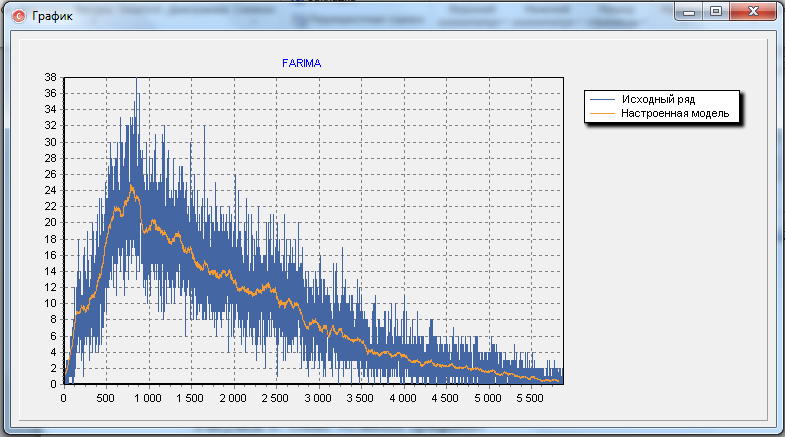


Рисунок 12. Окно «График»

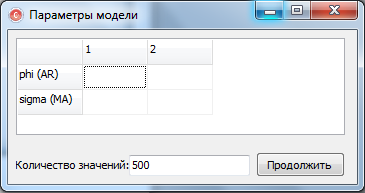


Рисунок 13. Окно «Параметры модели»

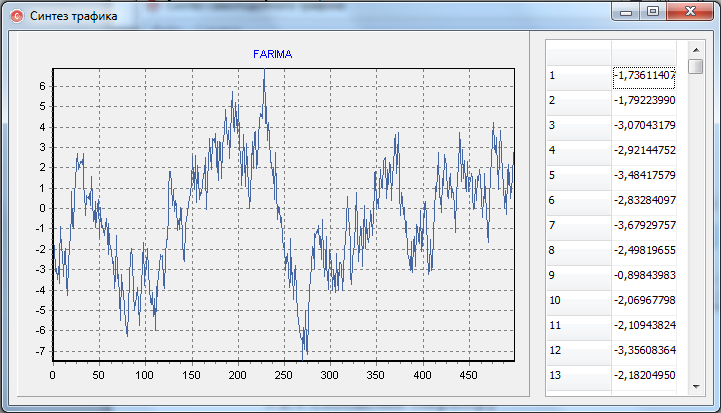


Рисунок 14. Окно «Синтез трафика»

3.2.4 Сообщения оператору

Перед всеми этапами вычислений происходит проверка корректности введенных данных. В случае если входная строка имеет неверный формат, то выдается сообщение об ошибке (рисунок 15).

В случае если формат загружаемого файла не соответствует требуемому или данный файл недоступен, выдается сообщение об ошибке.

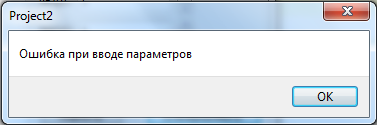


Рисунок 15. Формат сообщения об ошибке

**4 Моделирование трафика**

В качестве тестовых данных был взят ряд, каждый член которого соответствует количеству попаданий звонящих в очередь в десятисекундном интервале (ряд 1), а также ряд, отражающий количество пакетов в сетях компании Bellcore, передаваемых в единицу времени (ряд 2). Вид ряда 1 и 2 показан на рисунках 16, 17.

Автокорреляционная функция и спектральная плотность исходных рядов показывают, что их можно рассматривать как самоподобные случайные процессы (рисунки 18, 19, 20, 21).

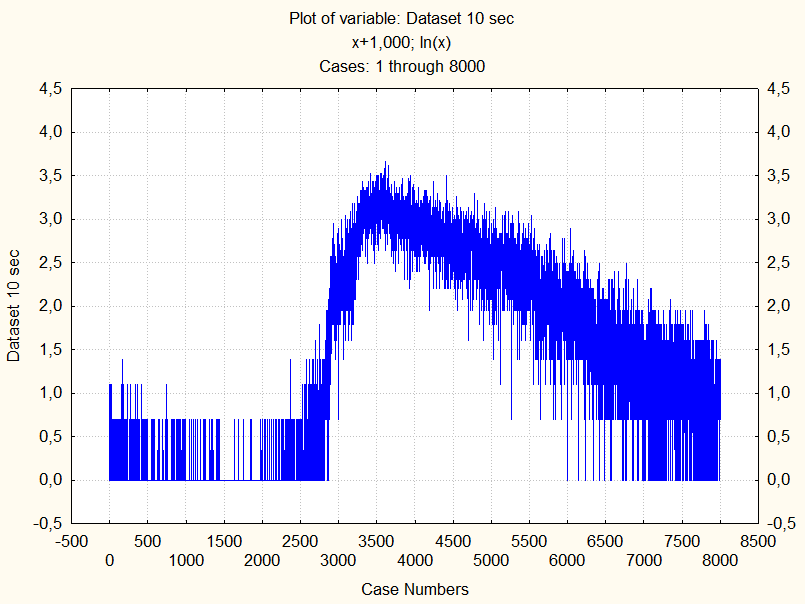


Рисунок 16. График временного ряда 1

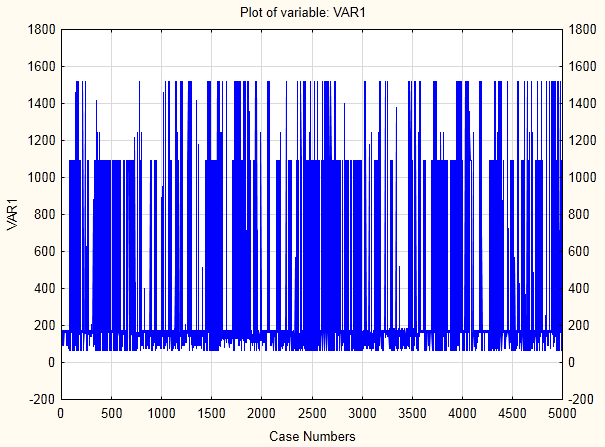


Рисунок 17. График логарифма временного ряда 2

Сгенерируем дробно-разностный шум (FARIMA (0, *d*, 0)) при *d* равном . Как видно из графиков (рисунки 18, 19), автокорреляционная функция медленно убывает и спектральная плотность сосредоточена в окрестности нуля.

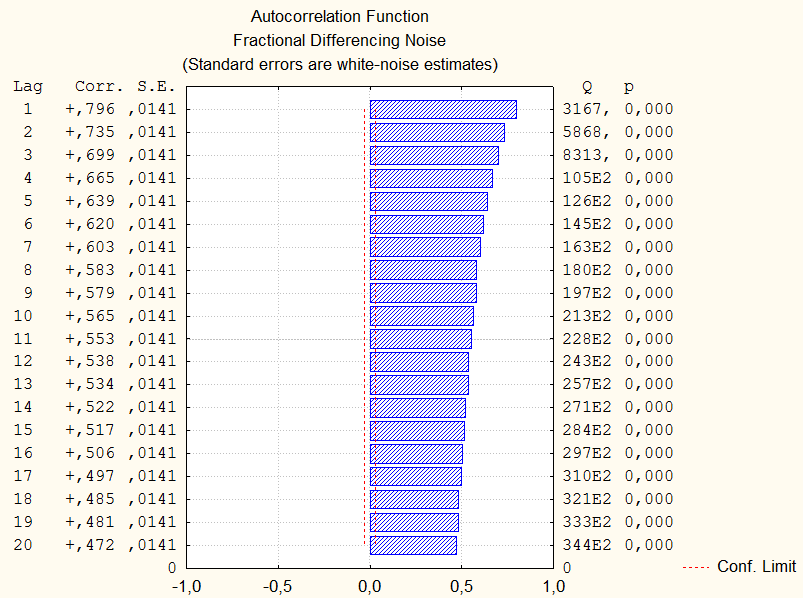


Рисунок 18. Автокорреляционная функция дробно-разностного шума

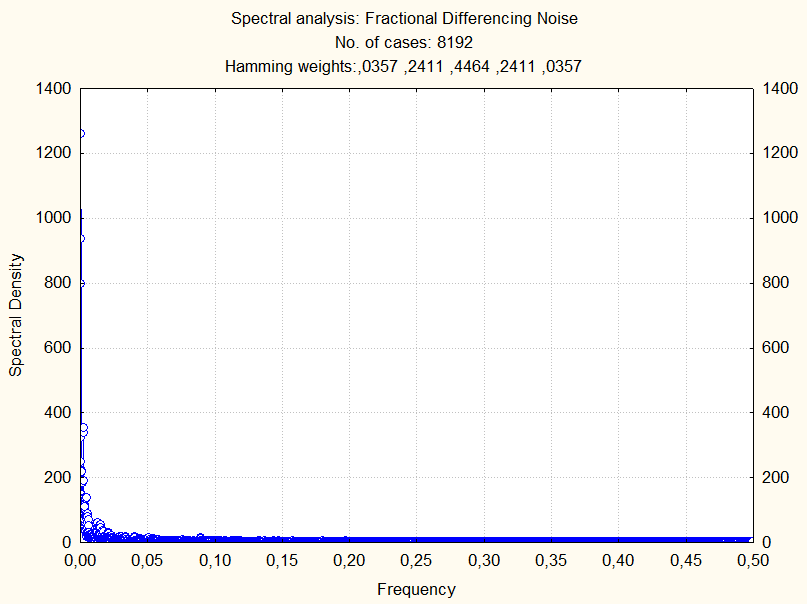


Рисунок 19. Спектральная плотность дробно-разностного шума

Смысл операции взятия дробной разности заключается в удалении из временного ряда долгосрочных зависимостей. Если взять дробную разность от дробно-разностного шума, то мы получим белый шум, т.к. процесс FARIMA (0, *d*, 0) не содер­жит краткосрочных зависимостей (рисунки 20, 21).

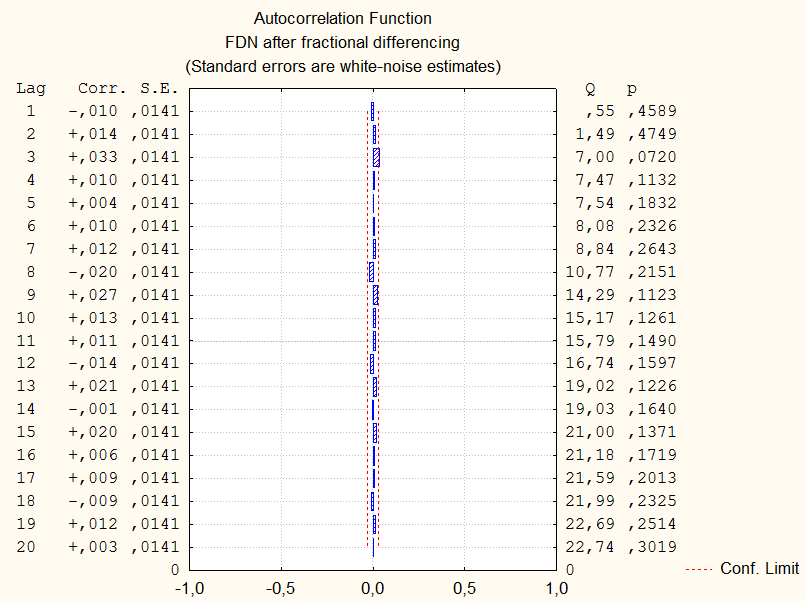


Рисунок 20. Автокорреляционная функция дробно-разностного шума после дифференцирования

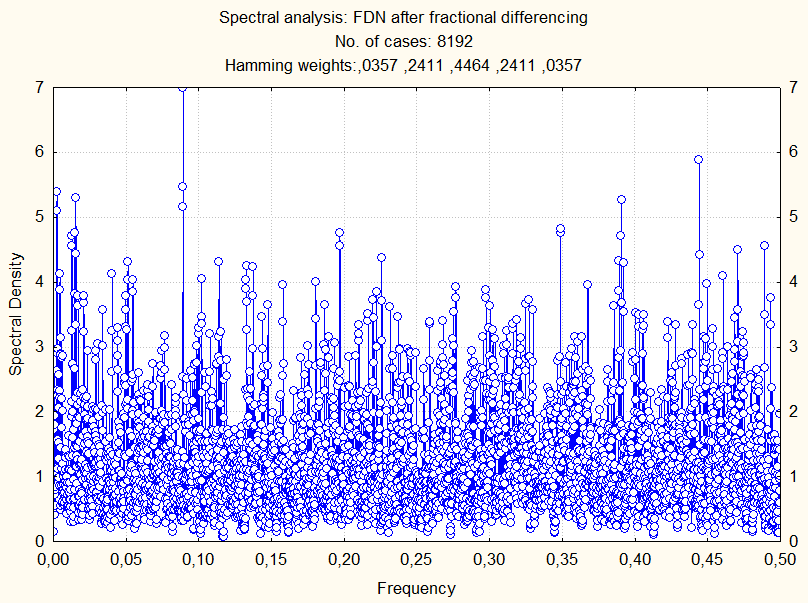


Рисунок 21. Спектральная плотность дробно-разностного шума после дифференцирования

Если сгенерировать и взять дробную разность от процесса FARIMA (*p*, *d*, *q*), мы получим процесс, имеющий свойства ARMA-процесса и содержащий краткосрочные зависимости (рисунки 22, 23).

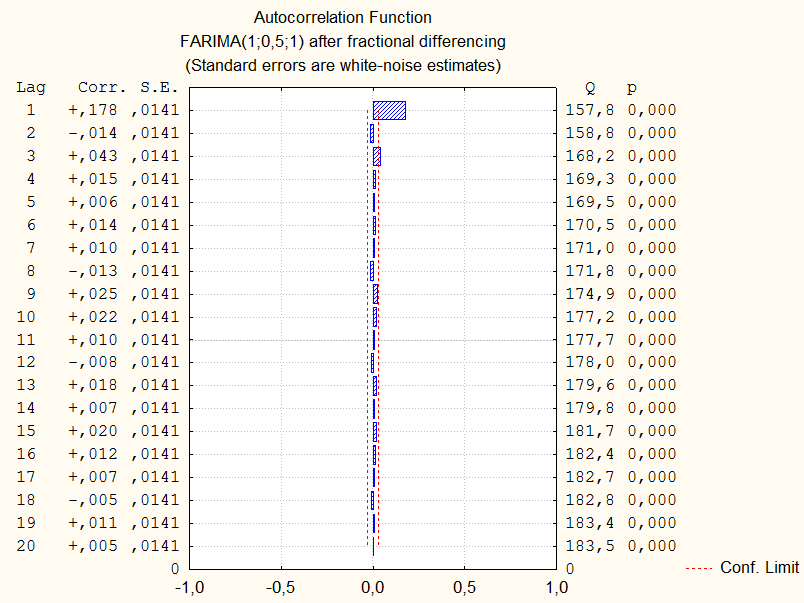


Рисунок 22. Автокорреляционная функция процесса FARIMA (*p*, *d*, *q*) после дифференцирования. *p* = 1, *q* = 1, *d* = 0,5

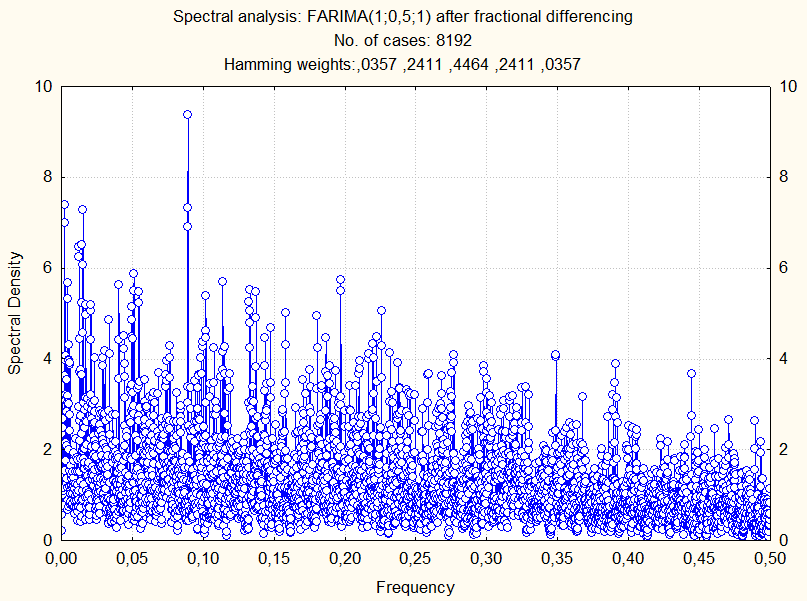


Рисунок 23. Спектральная плотность процесса FARIMA (*p*, *d*, *q*) после дифференцирования. *p* = 1, *q* = 1, *d* = 0,5

Построим графики АКФ и спектральной плотности исследуемых временных рядов (рисунки 24, 25, 26, 27). Как видно из рисунков, исследуемые ряды имеют характерные для самоподобных процессов свойства – медленное убывание АКФ и сосредоточенность спектральной плотности в окрестности нуля.

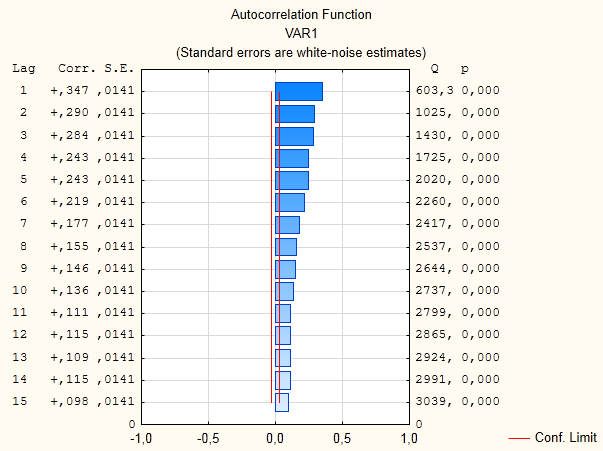


Рисунок 24. Автокорреляционная функция временного ряда 1

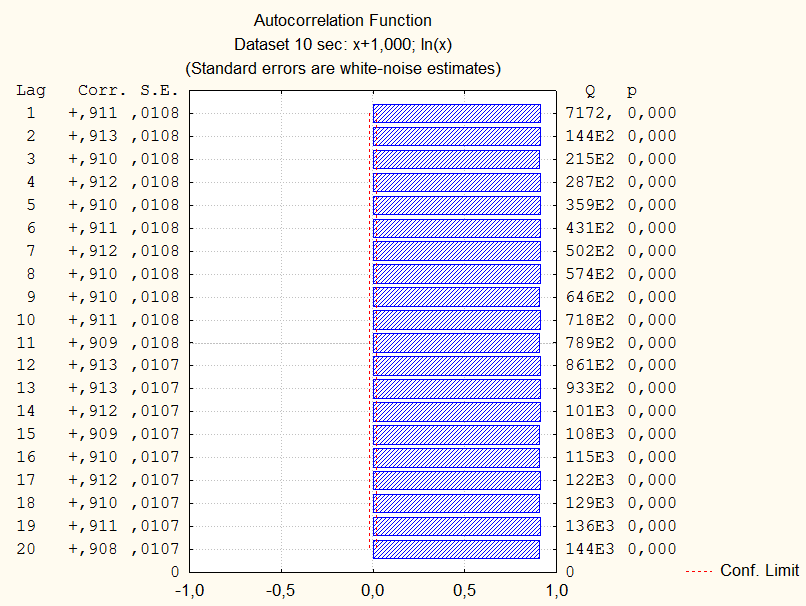


Рисунок 25. Автокорреляционная функция временного ряда 2

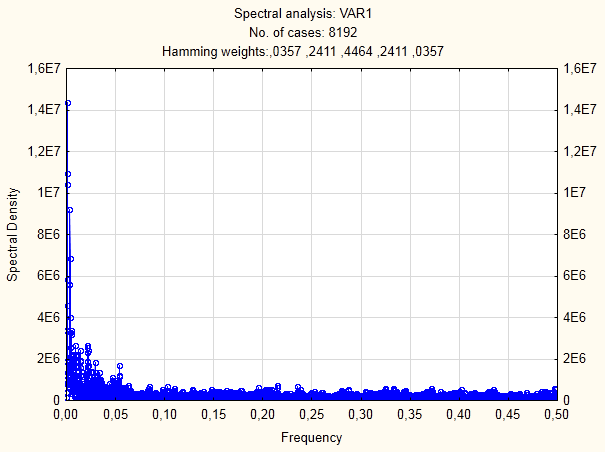


Рисунок 26. Спектральная плотность временного ряда 1

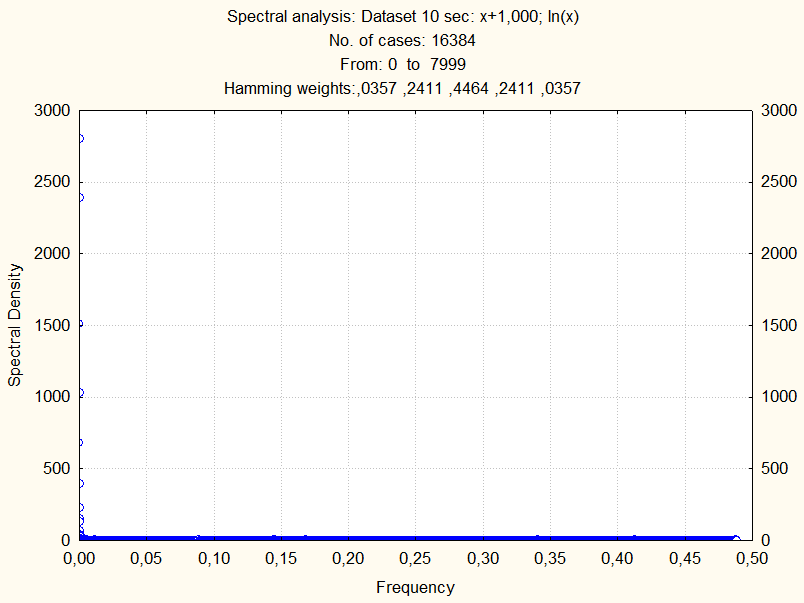


Рисунок 27. Спектральная плотность временного ряда 2

Оценим параметры FARIMA-модели для исследуемых рядов.

Для ряда 1 среди моделей с наилучшей оказалась FARIMA(0, 1,03819, 2) с остаточной дисперсией σ2 равной 0,107108 и параметрами MA(1) равным минус 0,996325, MA(2) равным 0,0606651 (рисунок 28).

Из-за близости значения *d* к единице, ряд так же хорошо моделируется процессом ARIMA(0, 1, 2) с остаточной дисперсией σ2 равной 0,107166.

Для ряда 2 среди моделей с  наилучшей оказалась FARIMA(1, 0,7542, 1) с остаточной дисперсией σ2 равной 32,65486 и параметрами MA(1) равным 0,54687, AR(1) равным 0,608684 (рисунок 29).

Наилучшая модель ARIMA(1, 1, 1) плохо описывала представленный ряд с остаточной дисперсией σ2 равной 114,84675.

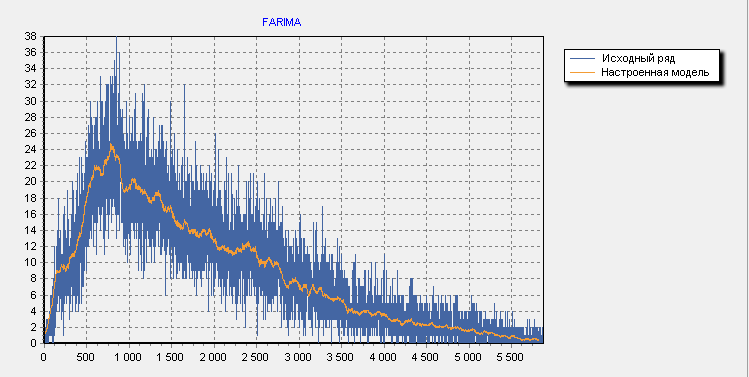


Рисунок 28. Ряд 1 и настроенная на него модель FARIMA(0, 1,03819, 2)

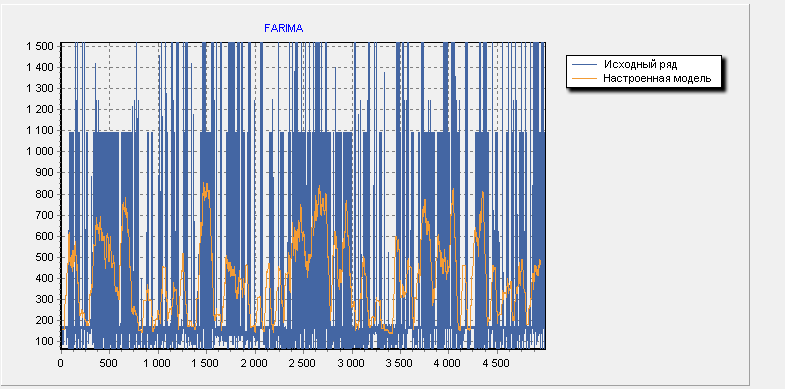


Рисунок 29. Ряд 2 и настроенная на него модель FARIMA(1, 0,7542, 1)

График ряда, сгенерированного по оценкам FARIMA-модели, показан на рисунках 30, 31.

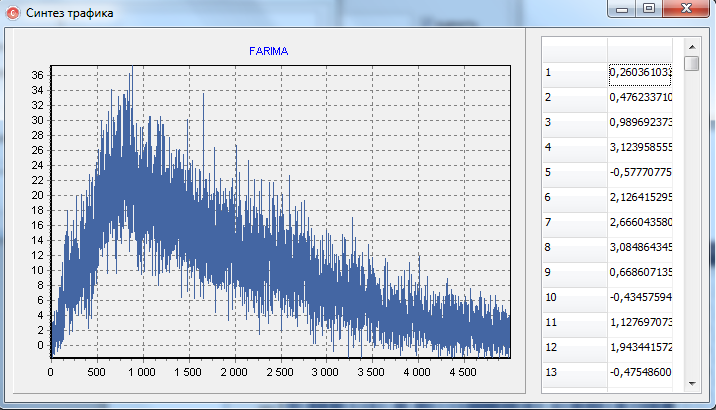


Рисунок 30. Синтез трафика по оценке параметров ряда 1

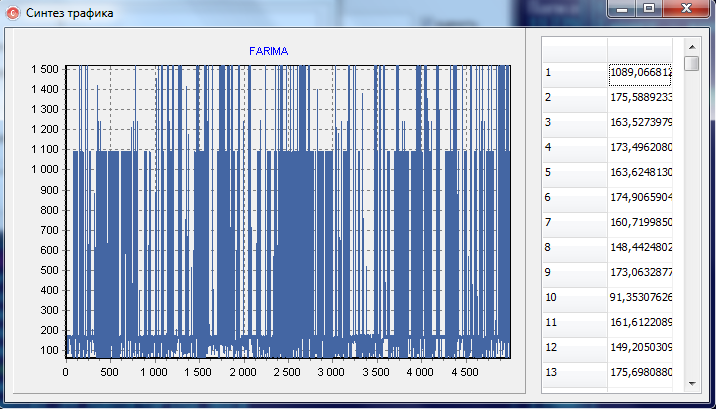


Рисунок 31. Синтез трафика по оценке параметров ряда 2

**Заключение**

В данной работе рассмотрены основные методы и подходы к моделированию самоподобного трафика, обсуждены вопросы применимости, достоинства и недостатки. Особое внимание уделено модели FARIMA, обсуждены алгоритмы моделирования.

В работе приведено описание программного продукта,разработанного для моделирования самоподобных процессов.

Было проведено исследование и апробация модели на реальных данных.

**Список источников**

1) Цыбаков Б.С. Модель телетрафика на основе самоподобного случайного процесса // Радиотехника. 1999. № 5. – С. 24-31.

2) Бокс Дж., Дженкинс Г.М. Анализ временных рядов. Прогноз и управление. Пер. сангл. // М.: Мир. 1974. – 640 с.

3) Taqqu M., Willinger W., Sherman R. Proof of a fundamental result in self–similar traffic modeling. // Computer Communication Review, 1997, vol.27, no. 2, pp. 5-23.

4) Шелухин О.И., Тенякишев А.М., Осин А.В. Фрактальные процессы в телекоммуникациях. Монография / Под ред. О.И. Шелухина. - М.: Радиотехника, 2003.- 480 с.

5) Добровольский Е.В., Нечипорук О.Л. Моделирование сетевого тра-фика с использованием контекстных методов // Науковіпраці ОНAЗ ім. О.С. Попова. – 2005. – № 1. – С. 24–32.

6) Медных А.В. Разработка методов моделирования самоподобного сетевого трафика // Зв’язок. – 2007. – № 6. – С. 20–22.

7) Leland W., Taqqu M., Willinger W. On the self-similar nature of IP-traffic // IEEE/ACM Transaction on Networking. – 1997. – № 3. – P. 423-431.

8) Петров В.В. То, что вы хотели знать о самоподобных процессах, но стеснялись спросить / В.В. Петров – М.: Радиотехника, 2003. – 112с.

9) Костромицкий А.И. Подходы к моделированию самоподобного трафика / А.И. Костромицкий, В.С. Волотка // Восточно-Европейский журнал передовых технологий. - 2010. – №46. - с. 46-49

10) Крылов В.В., Самохвалова С.С. Теория телетрафика и ее приложе-ния. – СПб.:БХВ-Петербург, 2005. – 288с.

11) Paxon V., Floyd S. Wide-area traffic: The failure of Poisson modeling. // IEEE/ACM Transactions on Networking #3, 1995. – pp. 226–244.

12) HoskingJ.R.M. Fractionaldifferencing // Biometrika 68. 1981. – С. 165-176.

13) JiakunLiu, YantaiShu, Lianfang Zhang, FeiXue, Yang Oliver W.W.Traffic modeling based on FARIMA models // Electrical and Computer Engineering. 1999. – С. 162-167.

14) McCarthy Joseph, DiSario Robert, SaraogluHakan A Recursive Algorithm For Fractionally Differencing Long Data Series // Journal of Modern Applied Statistical Methods. 2003.– С. 272-278.

15) Li Song, Pascal Bondon Piecewise FARIMA models for long-memory time series Proc. 1998 ACM SIGCOM, pp. 42-55.