

УДК 519.21

UDC 519.21

АНАЛИЗ МЕТОДОВ ГЕНЕРАЦИИ ВРЕМЕННЫХ РЯДОВ С ДОЛГОВРЕМЕННОЙ КОРРЕЛЯЦИОННОЙ СТРУКТУРОЙ

ANALYSIS OF METHODS OF GENERATION OF TIME SERIES WITH LONG-TERM CORRELATION STRUCTURE

Кесиян Грант Арутович
аспирант

Kesiyani Grant Arutovich
postgraduate student

Уртеннов Махамед Хусеевич
д.ф.-м.н., профессор

Urtenov Makhamet Khuseevich
Dr.Sci.Phys.-Math., professor

Шахмеликян Тимур Аркадьевич
аспирант
*Кубанский государственный университет,
Краснодар, Россия*

Shahmelikyan Timur Arkadyevich
postgraduate student
Kuban State University, Krasnodar, Russia

В данной статье производится сравнительный анализ шести методов моделирования фрактально дифференцированного шума, среди которых: прямой метод определения, метод Хоскинга, метод Дэвиса и Харта, метод циркулянтных вложений, метод Паксона и метод спектрального моделирования. Кроме того, приведены алгоритмы проведения анализа и подобран метод оценки параметра Херста, являющийся основным критерием качества при сравнении вышеуказанных методов моделирования фрактально дифференцированного шума

This article gives a comparative analysis of six modeling techniques of fractal differential noise, including: *direct definition method*, the *method of Hosking*, the *method of Davis and Hart's*, the *circulant embedding method*, *Paxson method* and *method of spectral simulation*. Furthermore, the algorithms are given for the analysis and the method of estimation of Hurst's parameter is selected, this method is the main criterion of quality when comparing the above mentioned methods of fractal differential noise

Ключевые слова: ВРЕМЕННЫЕ РЯДЫ, ДОЛГОВРЕМЕННАЯ КОРРЕЛЯЦИОННАЯ СТРУКТУРА, ЭКСПОНЕНТА ХЕРСТА, FDN-ПРОЦЕСС, ПОРЯДОК СЛОЖНОСТИ АЛГОРИТМА, СИНТЕТИЧЕСКИЕ ДАННЫЕ

Keywords: TIME SERIES, LONG-TERM CORRELATION STRUCTURE, EXPONENT HURST, FDN-PROCESS, ORDER COMPLEXITY OF ALGORITHM, SYNTHETIC DATA

Введение

Моделирование временных рядов является перспективным инструментом для изучения большого количества явлений.

Многие природные процессы, такие как данные о погоде, электрических или физиологических измерений, а так же техногенные явления, такие как сетевой трафик, проявляют спектры, которые соответствуют классу дробных шумовых процессов $1/f$ [8]. Дробные шумы представляют собой обобщение процессов броуновского движения [9]. Такие процессы обладают долговременной корреляционной структурой. Примерами моделей обладающих долговременной памятью являются

модель фрактального гауссовского шума (FGN) и модель фрактально дифференцированного шума (FDN).

В данной статье производится сравнительный анализ шести методов моделирования фрактально дифференцированного шума, среди которых: прямой метод определения, метод Хоскинга, метод Дэвиса и Харта, метод циркулянтных вложений, метод Пакссона и метод спектрального моделирования. Следует отметить, что для выполнения анализа предварительно производится выбор метода для оценки экспоненты Херста, который будет основным критерием сравнения.

1 Постановка задачи

В целях работы было выбрано шесть наиболее популярных методов моделирования фрактально дифференцированного шума: прямой метод определения, метод Хоскинга, метод Дэвиса и Харта, метод циркулянтных вложений, метод Пакссона и метод спектрального моделирования.

Необходимо провести сравнительный анализ вышеуказанных методов на качество моделирования фрактально дифференцированного шума с заданным параметром – экспонентой Херста.

Для проведения непосредственного анализ сначала необходимо подобрать метод оценки параметра Херста, который будет основным критерием качества методов моделирования фрактально дифференцированного шума.

В работе [6] был проведен анализ шести наиболее популярных методов оценки экспоненты Херста: метода нормированного размаха (R/S-анализ), метода совокупной дисперсии, метода абсолютного значения, метода дисперсии остатков, метода основанного на периодограмме и локального метода Уиттла.

Из анализа был сделан вывод, что из шести методов, метод Уиттла лучше всех производит оценку параметра Херста временных рядов с долговременной памятью, к которым добавлены различные типы шумов.

Но для «чистых» временных рядов с долговременной корреляционной структурой лучшим оказался метод дисперсии остатков, так как он производит несмещенные оценки почти для всех значений Херста ($0 < H < 1$).

Работа [10] показала, что лучшим методом оценки параметра Херста является метод Уиттла, но для временных рядов длиной 4000 и выше, метод дисперсии остатков так же является хорошим выбором.

Таким образом, учитывая, что рассмотренные выше методы генерируют временные ряды без наложения сторонних шумов, то для их анализа лучше всего подходит метод дисперсии остатков. Но при этом нужно учесть, что длины моделируемых рядов должны быть не меньше 4000.

2 Алгоритм анализа

Проведение анализа можно разбить на две части: анализ отклонения среднего и анализ дисперсии.

Алгоритм для проведения *анализа отклонения среднего* следующий:

1. Выбор анализируемого метода.
2. Генерация 100 реплик фрактально дифференцированного шума выбранным методом для заданного значения Херста H .
3. Получение оценок экспоненты Херста по каждому из 100 сгенерированных рядов фрактально дифференцированного шума с помощью метода дисперсии остатков.
4. Вычисление среднего значения по 100 оценкам параметра Херста.
5. Анализ отклонения среднего значения, полученного на шаге 4 от значения, заданного на шаге 2.
6. Повторяем все с шага 2 для каждого значения Херста в интервале от 0.01 до 0.99 с шагом 0.01.

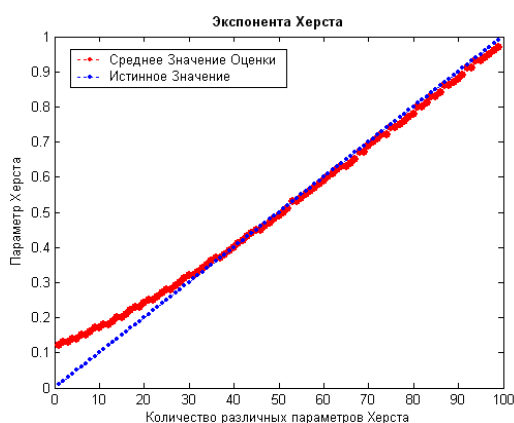
Отметим, что длина каждой выборки была выбрана 7560, так как это число имеет больше всего делителей из всех чисел, не превышающих 10000, что связано с качеством метода дисперсии остатков.

Алгоритм для проведения *анализа дисперсии* состоит из пяти шагов:

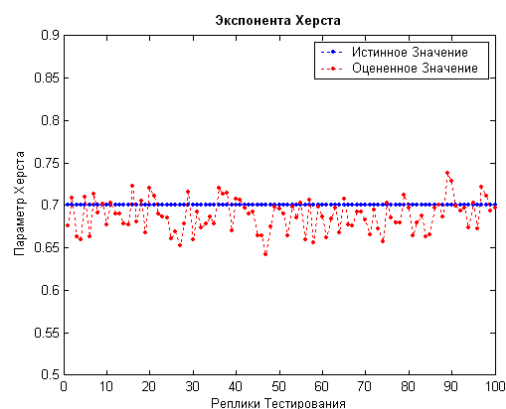
1. Выбор анализируемого метода.
2. Генерация 100 реплик фрактально дифференцированного шума выбранным методом для экспоненты Херста $H=0.7$.
3. Получение оценок экспоненты Херста по каждому из 100 сгенерированных рядов фрактально дифференцированного шума с помощью метода дисперсии остатков.
4. Вычисление среднего квадрата отклонений по 100 оценкам параметра Херста.
5. Повторяем все с шага 1 для каждого метода.

3 Обсуждение результатов анализа

Результаты *прямого метода определения* [12] представлены на рисунке 1 (а) и (б).



(а)



(б)

Рисунок 1. Оценки параметров Херста и дисперсия оценок для прямого метода определения: а) - Прямой метод определения $H=0.01\sim 0.99$, б) - Прямой метод определения при $H=0.7$

Из анализа мы можем увидеть, что прямой метод определения не всегда производит несмещенные оценки. Этот метод генерирует временные ряды с переоцененной экспонентой Херста для $0.01 \leq H < 0.26$. Для остальных значений параметра Херста он производит лишь незначительные отклонения (0.01~0.02). На рисунке 1(б) показана дисперсия 100 реплик моделирования *FDN*-процесса прямым методом определения для $H=0.7$. Средний квадрат отклонений для $H=0.7$ составил 0.00051.

На рисунке 2 (а) и (б) представлены результаты анализа *метода Хоскинга* [11, 12].

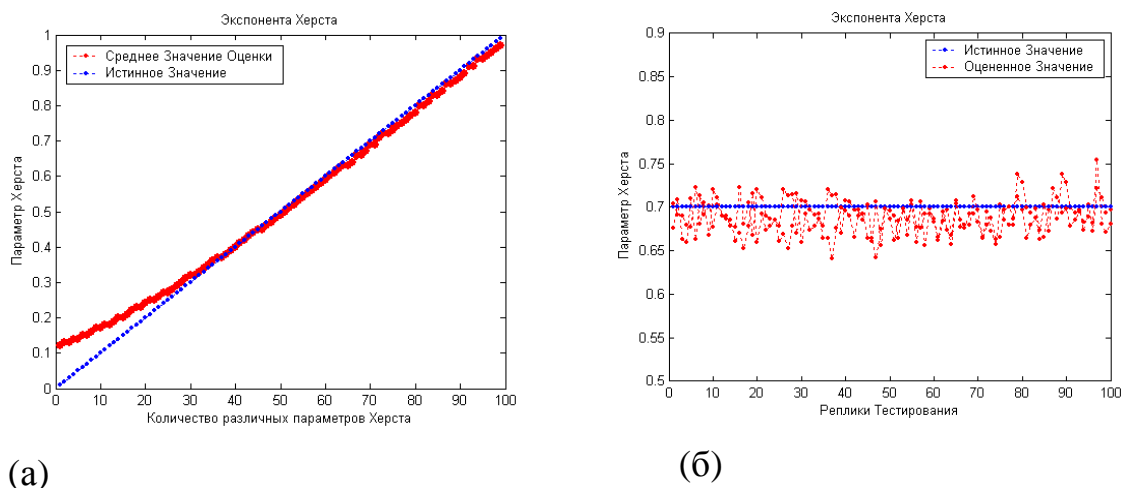
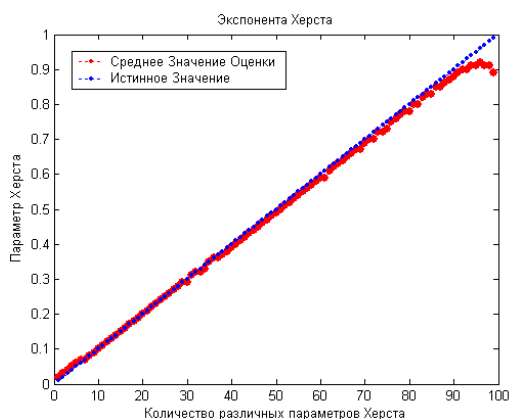


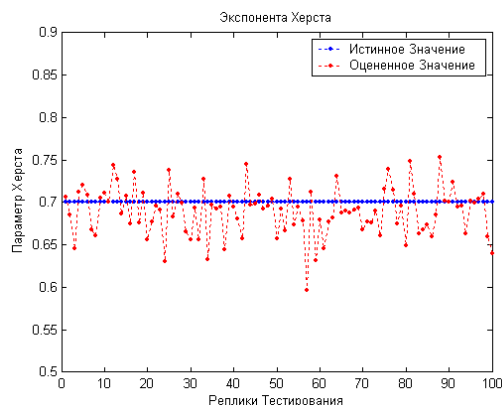
Рисунок 2. Оценки параметров Херста и дисперсия оценок для метода Хоскинга: а) - Метод Хоскинга $H=0.01 \sim 0.99$, б) - Метод Хоскинга при $H=0.7$

Из анализа видно, что метод Хоскинга незначительно лучше, чем прямой метод определения и, генерирует временные ряды с переоцененной экспонентой Херста для $0.01 \leq H < 0.25$. Для остальных значений параметра Херста он производит лишь малые отклонения (0.01~0.02). На рисунке 2(б) показана дисперсия 100 реплик моделирования *FDN*-процесса методом Хоскинга для $H=0.7$. Средний квадрат отклонений для $H=0.7$ составил 0.00051, что совпадает со средним квадратом отклонений для прямого метода определения.

Результаты *метода Дэвиса и Харта* [2, 4] представлены на рисунке 3 (а) и (б).



(а)

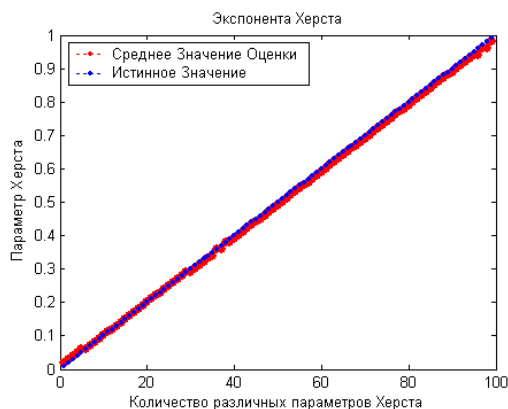


(б)

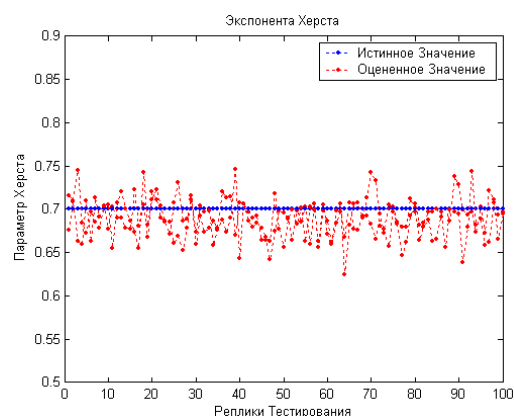
Рисунок 3. Оценки параметров и дисперсия оценок для метода Дэвиса и Харта: а) - Метод Дэвиса и Харта $H=0.01\sim 0.99$, б) - Метод Дэвиса и Харта при $H=0.7$

По рисунку 3 (а) видно, что метод Дэвиса и Харта лучше, чем прямой метод определения и метод Хоскинга. Он генерирует временные ряды с недооцененной экспонентой Херста лишь для $0.93 \leq H \leq 0.99$. Для остальных значений параметра Херста он показывает хорошие результаты, отклоняясь лишь на $0.01\sim 0.02$. На рисунке 3(б) показана дисперсия 100 реплик моделирования *FDN*-процесса методом Дэвиса и Харта для $H=0.7$. Средний квадрат отклонений для $H=0.7$ составил 0.00092, что больше чем средний квадрат отклонений для прямого метода определения и для метода Хоскинга.

На рисунке 4 (а) и (б) представлены результаты анализа *метода циркулянтных вложений* [1, 5]. Из анализа видно, что метод циркулянтных вложений производит практически всегда несмещенные оценки. Этот метод точно воспроизводит временные ряды с экспонентой Херста для $0.01 \leq H \leq 0.99$. Отклонения от истинных значений являются незначительными ($0.01\sim 0.02$).



(а)

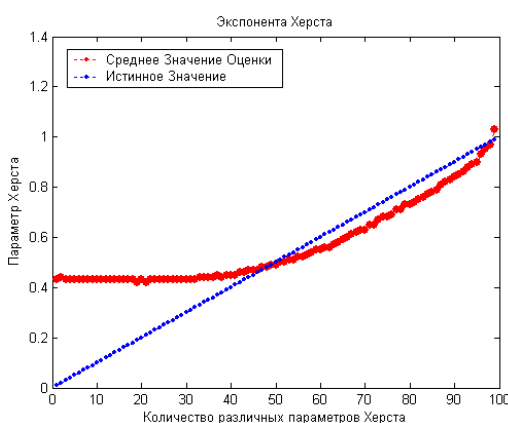


(б)

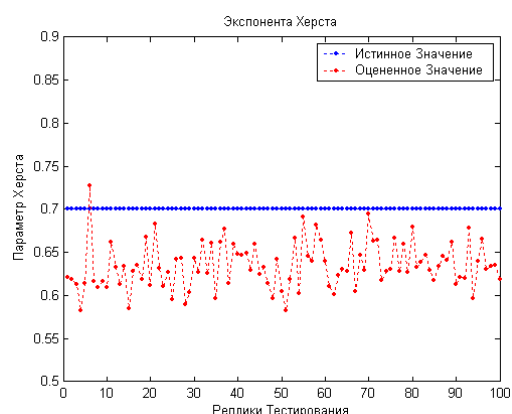
Рисунок 4. Оценки параметров Херста и дисперсия оценок для метода циркулянтных вложений: а) - Метод циркулянтных вложений $H=0.01\sim 0.99$, б) - Метод циркулянтных вложений при $H=0.7$

На рисунке 4(б) показана дисперсия 100 реплик моделирования *FDN*-процесса методом циркулянтных вложений для $H=0.7$. Средний квадрат отклонений для $H=0.7$ составил 0.00065, что немного больше, чем для прямого метода определения и для метода Хоскинга, но меньше чем для метода Дэвиса и Харта.

Результаты анализа *метода Пакссона* [7] продемонстрированы на рисунке 5 (а) и (б).



(а)

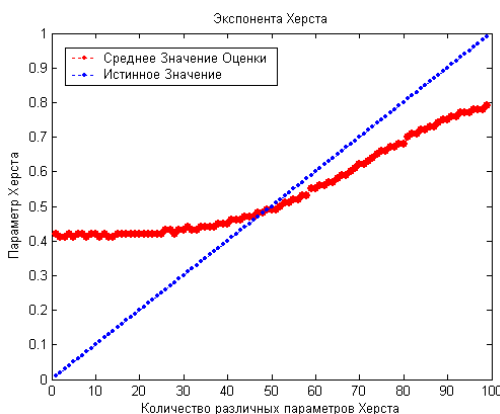


(б)

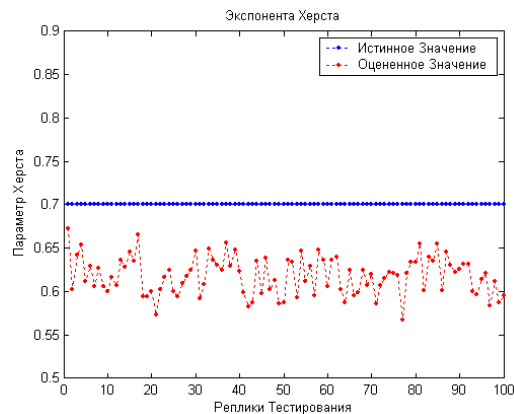
Рисунок 5. Оценки параметров Херста и дисперсия оценок для метода Пакссона: а) - Метод Пакссона $H=0.01\sim 0.99$, б) - Метод Пакссона при $H=0.7$

Анализ показал, что метод Пакссона недостаточно точно генерирует FGN-процесс. Этот метод генерирует временные ряды с переоцененной экспонентой Херста для $0.01 \leq H \leq 0.44$ и $H=0.99$ и недооцененной экспонентой Херста для $0.53 \leq H \leq 0.96$. Дисперсия 100 реплик моделирования FGN-процесса методом Пакссона для $H=0.7$ показана на рисунке 5(б). Средний квадрат отклонений для $H=0.7$ составил 0.00499, что больше, чем у выше рассмотренных методов.

Из рисунка 6 (а) видно, что *спектральный метод* [3] генерирует FGN-процесс хуже, чем все выше рассмотренные методы.



(а)



(б)

Рисунок 6. Оценки параметров Херста и дисперсия оценок для спектрального метода: а) - Спектральный метод $H=0.01 \sim 0.99$, б) - Спектральный метод при $H=0.7$

Этот метод генерирует временные ряды с переоцененной экспонентой Херста для $0.01 \leq H \leq 0.45$ и недооцененной экспонентой Херста для $0.54 \leq H \leq 0.99$. На рисунке 6 (б) показана дисперсия 100 реплик моделирования FGN-процесса спектральным методом для $H=0.7$. Средний квадрат отклонений для $H=0.7$ данным методом составил 0.00721, что больше, чем у остальных выше рассмотренных методов.

Полученные в результате экспериментов свойства, рассмотренных методов генерирования временных рядов с долговременной корреляционной структурой, представлены в таблице 1:

Таблица 1 – СВОДНАЯ ТАБЛИЦА СВОЙСТВ АНАЛИЗИРУЕМЫХ МЕТОДОВ

Свойство Метод	Количество точно воспроизведенных временных рядов с заданным H (0.01~0.99)	Средний квадрат отклонений для $H=0.7$	Порядок сложности алгоритма
Прямой метод определения	74	0.00051	$O(N^2)$
Метод Хоскинга	75	0.00051	$O(N^2)$
Метод Дэвиса и Харта	92	0.00092	$O(N \log(N))$
Метод циркулянтных вложений	99	0.00065	$O(N \log(N))$
Метод Пакссона	10	0.00499	$O(N \log(N))$
Спектральный метод	8	0.00721	$O(N \log(N))$

Заключение

Многие авторы при использовании тех или иных методов для оценки экспоненты Херста не тестируют эти методы на синтетических данных. Между тем, апробация различных алгоритмов сразу на реальных данных, может привести к неправильной оценке корреляционной структуры временного ряда ввиду несостоятельности используемых методов и, в конечном счете, к неправильным выводам. Кроме того, используемые методы должны быть протестированы на устойчивость относительно шумов и эффектов кратковременной памяти, которые могут привести к смещению оценок.

С другой стороны, для анализа методов синтетической генерации процессов с долговременной корреляционной структурой, необходима уверенность в алгоритмах, которые оценивают экспоненту Херста.

В данной статье был проведен сравнительный анализ шести методов генерации временных рядов с долговременной корреляционной структурой. При этом на основании работ [6] и [10] в качестве основного оценщика экспоненты Херста был выбран метод дисперсии остатков. В результате проведенных анализов можно сделать следующие выводы:

1. Лучшим методом оказался *метод циркулянтных вложений*, так как он точно воспроизводит временные ряды на всем интервале значений экспоненты Херста, но средний квадрат отклонений для $H=0.7$ у этого метода немного больше, чем для прямого метода определения и для метода Хоскинга. Кроме того, этот метод может быть реализован с помощью алгоритма с наименьшим порядком сложности $O(N \log(N))$;

2. На втором месте по количеству точно воспроизведенных временных рядов с заданной экспонентой Херста метод Дэвиса и Харта. Этот метод генерирует временные ряды с недооцененной экспонентой Херста лишь для 7 параметров Херста ($0.93 \leq H \leq 0.99$) из 99;

3. Хорошими методами моделирования временных рядов с экспонентой Херста больше 0.25 и 0.24 являются метод прямого определения и метод Хоскинга соответственно. Так же эти методы обладают самым меньшим средним квадратом отклонений для $H=0.7$ (0.00051). Но стоит учесть, что их реализация потребует $O(N^2)$ операций.

4. Недостаточно точными оказались метод Пакссона и спектральный метод. В результате проведенных экспериментов, эти методы точно воспроизвели временные ряды только для 10 (метод Пакссона) и 8 (спектральный метод) показателей Херста из 99. При этом метод Пакссона и спектральный метод имеют самые большие средние квадраты отклонений для $H=0.7$ (0.00499 и 0.00721 соответственно).

Методы генерации временных рядов с долговременной корреляционной структурой позволяют нам проводить эксперименты и изучать такого рода процессы. У нас появляется возможность оценивать их

количественные характеристики, выявлять закономерности изменений этих характеристик при добавлении различных шумов, анализировать корреляционную структуру при наложении фильтров кратковременной памяти и т.д.

С учетом результатов настоящей работы, для целей генерации временных рядов с долговременной корреляционной структурой, мы рекомендуем использовать метод циркулянтных вложений.

Библиографический список

1. Chan, G. (1999). An Effective Method for Simulating Gaussian Random Fields. American Statistical Association, *Proceedings of The Statistical Computing Section*, 133-138.
2. Davies, R. B. and Harte, D. S. (1987), Tests for Hurst Effect, *Biometrika*, 74, pp.96-101.
3. Dieker, A.B. (2002). Simulation of fractional Brownian motion. Master's thesis, Vrije Universiteit, Amsterdam, April 2002. See www.cwi.nl/~ton.
4. Dieker, A.B. and Mandjes, M. (2003) On spectral simulation of fractional Brownian motion. *Probability in the Engineering and Informational Sciences*, 17 (3). pp. 417-434. ISSN 0269-9648.
5. K. Kubilius, D. Melichov, On comparison of the estimators of the Hurst index of the solution of SDEs driven by a fBm, *Informatica* (in press).
6. Hu Sheng, YangQuan Chen Robustness analysis of the estimators for noisy long-range dependent time series IDETC/CE 2009. San Diego, USA (September 1, 2009).
7. V. Paxson, Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic, *Computer Communication Review*, 27 (1997), pp. 5-18.
8. P. Shan and M. Li, An EMD based simulation of fractional Gaussian noise, *International Journal of Mathematics and Computers in Simulation*, vol. 1, no. 4, pp. 312–316, 2007.
9. E. Peters, *Fractal analysis of financial markets: Application of chaos theory to investment and the economy*. M: Internet-trading, 2004 - 304 pp.
10. Rea, W., Oxley, L., Reale, M., & Brown, J. (2009). Estimators for Long Range Dependence: An Empirical Study. *Statistics*, (1980), 1-14. Retrieved from <http://arxiv.org/abs/0901.0762>.
11. Y. Shu, F. Xue, Z. Jin, and O.W.W. Yang, The impact of self-similar traffic on network delay, presented at J. Comput. Sci. Technol., 1999, pp.585-589.
12. Fei Xue, Jiakun Liu, Yantai Shu, Lianfang Zhang, Oliver W.W. Yang: Traffic Modeling Based on FARIMA Models, *Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering*, 1999, vol.1, pp. 162-167.