

УДК 519.216

И.В. Дегтяренко (канд. техн. наук, доц.), А.М. Гарматенко, О.А. Ярошенко
ДВНЗ «Донецкий национальный технический университет», г. Донецк
кафедра автоматике и телекоммуникаций
E-mail: ilya_degtyarenko@ukr.net, alexgarmatenko@gmail.com, olyoaolyoa@gmail.com

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ ARFIMA МОДЕЛИ ФРАКТАЛЬНОГО ПРОЦЕССА

В статье рассмотрены вопросы построения прогностической ARFIMA модели фрактального процесса. Проведен анализ взаимосвязи параметров ARFIMA модели с показателем Херста. Описана методика определения параметров данной модели на основе использования метода «Detrended Fluctuation Analysis». Исследована эффективность применения данной методики при построении прогностических моделей.

Ключевые слова: фрактальный процесс, ARFIMA, показатель Херста, прогностическая модель, Detrended Fluctuation Analysis.

Общая постановка проблемы

Процессы, обладающие фрактальными свойствами достаточно часто встречаются в природных и технических системах [1-8]. Описание и анализ таких процессов достаточно проблематично проводить в рамках классических теорий стохастических и детерминированных процессов.

Теория фрактального анализа – одно из наиболее быстро развивающихся направлений современной науки. Если раньше она развивалась в основном только в рамках фундаментальных исследований, то в последние годы все большее распространение методы теории фрактального анализа получают при решении прикладных задач. К таким примерам относятся сжатие изображений [2], анализ трафика телекоммуникационных сетей [3-5], прогноз разрушений земной коры [6,7], создание антенн с изменяющимися диаграммами направленности [8], исследование динамики рынка [9] и т.д.

Понятие фрактал было введено Бенуа Мандельбротом в 1975 году. Согласно его определению «фракталом называется структура, состоящая из частей, которые в каком-то смысле подобны целому» [1]. Соответственно одним из главных свойств фрактальных процессов является самоподобие. Именно это свойство позволяет связать между собой поведение процесса в больших и малых масштабах приближения. Для описания таких процессов применяются распределения с «тяжелыми хвостами» [1,3]. Т.е. для фрактальных процессов характерно наличие долговременной памяти. Именно это свойство обосновывает возможность построения прогностических моделей описывающих поведение фрактальных процессов.

Одним из наиболее эффективных подходов к описанию процессов с долговременной памятью является метод построения авторегрессионной фрактально-интегрируемой модели скользящего среднего – ARFIMA [9]. Данная модель является развитием интегрированной авторегрессионной модели скользящего среднего (ARIMA) Бокса-Дженкинса [10]. Особенностью модели ARFIMA является то, что она допускает дробные полюсы, а также проявляет свойства как кратковременной так и долговременной зависимости [3].

© Дегтяренко И.В., Гарматенко А.М., Ярошенко О.А., 2013

Адекватная параметрическая идентификация параметров данной модели является ключевой задачей при построении эффективных методов прогнозирования поведения фрактальных процессов.

Постановка задач исследований

Данная работа посвящена разработке методики определения параметров ARFIMA модели, позволяющей строить прогностические модели фрактальных процессов, обладающих долговременной памятью. Основные задачи, которые решаются в работе:

- 1) анализ взаимосвязи параметров фрактального процесса с параметрами ARFIMA модели;
- 2) разработка рекомендаций по способу определения параметров ARFIMA модели на основе оценки степени самоподобности фрактального процесса;
- 3) исследование эффективности разработанной методики при формировании прогностической ARFIMA модели фрактального процесса.

Решение задач и результаты исследований

Модель ARFIMA(p, d, q) является очень удобным средством описания процессов, как с долговременной, так и с кратковременной памятью [3,9,10]. Параметры p и q соответствуют порядку полиномов соответственно авторегрессионной модели (AR) и модели скользящего среднего (MA) [10]. Параметр фрактальной интегрируемости d определяет степень долговременности памяти процесса [3]. Математическая модель ARFIMA(p, d, q) процессов может быть представлена в виде [8, 9, 11]:

$$\Psi(L)(1-L)^d x_t = \Theta(L)\varepsilon_t; \quad (1)$$

где $\Psi(L) = 1 - \psi_1 L - \dots - \psi_p L^p$; $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$;

$$(1-L)^d = 1 - dL - \frac{d(1-d)}{2!} L^2 - \frac{d(1-d)(2-d)}{3!} L^3 - \dots = \\ = \sum_{k=0}^{\infty} \frac{G(k-d)}{G(-d)G(k+1)} L^k;$$

L - оператор обратного сдвига, определяемый как $L^j x_t = x_{t-j}$; ψ_i - коэффициенты модели авторегрессии; d - параметр фрактальной интегрируемости; p - порядок авторегрессионной модели; q - порядок модели скользящего среднего; θ_i - коэффициенты модели скользящего среднего; ε_t - белый шум; $G(\cdot)$ - гамма-функция.

Гамма-функция может быть представлена как:

$$G(z) = \int_0^{\infty} t^{z-1} e^{-t} dt. \quad (2)$$

Методы определение параметров моделей авторегрессии и моделей скользящего среднего являются хорошо изученными [10].

ARFIMA-процесс является стационарным в широком смысле при значениях параметра $|d| < 0.5$. Если $0 < d < 0.5$, то считается, что процесс обладает долговременной памятью или другими словами является персистентным. При $-0.5 < d < 0$ процесс проявляет свойства антиперсистентности. Если параметр $d = 0$, то ARFIMA модель трансформируется в простую ARMA модель, служащую для описания процессов с кратковременной памятью.

При построении ARFIMA моделей идентификация параметра d является первичной. Традиционно параметр d оценивается по методу Виттла [13], который основан на поиске минимума функции $R(d)$:

$$d = \operatorname{arg\,min}_{d \in [\Delta 1; \Delta 2]} R(d), \tag{4}$$

где $R(d) = \log \frac{1}{m} \sum_{j=1}^m I_u(\lambda_j) - 2 \cdot d \cdot \frac{1}{m} \sum_{j=1}^m \log(\lambda_j); \quad \lambda_j = \frac{2 \cdot \pi \cdot j}{m};$

m - длина данных; $I_u(\lambda_j)$ - функция трансформации (аппроксимации) входных данных.

Известно [3], что параметр d тесно связан с показателем Херста (H), который описывает степень самоподобия фрактального процесса (см. рис. 1):

$$d = H - 0.5. \tag{5}$$

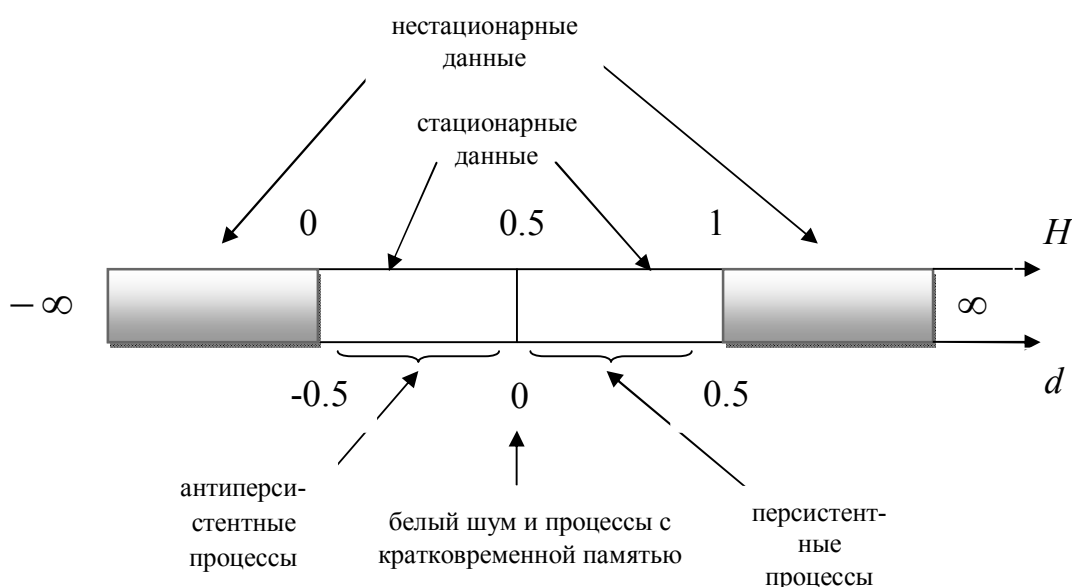


Рисунок 1 – Зависимость свойств ARFIMA-процессов от значений фрактально-интегрируемой составляющей d и параметра Херста

Показатель Херста как правило принимает значения в интервале от -0.5 до 1.5 [12]. Процессы с $H=0.5$ описываются моделью белого шума, т.е. это процессы без памяти либо процессы с очень короткой памятью. Наибольший интерес с точки зрения построения прогностических ARFIMA моделей имеют процессы с $0.5 < H \leq 1$, т.к. для них характерно наличие долговременной памяти. Процессы с $H > 1$ являются нестационарными и непредсказуемыми, но их, как правило, можно привести к стационарному виду, вычислив разности (производные) 1-го или 2-го порядка.

На сегодняшний день существует множество методов определения параметра Херста [3,4]. Как показывают результаты практических исследований [3,4] все они в разной мере чувствительны к параметрам исходных данных (размеру выборки, степени самоподобия, структуры данных и т.д.). Одним из методов дающих наиболее устойчивые и адекватные результаты оценки параметра Херста является метод анализа флуктуационных данных с исключенным трендом (Detrended Fluctuation Analysis – DFA) [14]. В данной работе предлагается использовать этот метод для определения параметра d ARFIMA модели.

Алгоритм вычисления параметра d описано ниже.

- 1) Вычисляется среднее из первоначального временного ряда по формуле:

$$y(k) = \sum_{i=1}^k (x(i) - \bar{x}), \quad (6)$$

где \bar{x} - среднее значение временного ряда; $x(i)$ - i -е значение временного ряда.

- 2) Вычисляется величина, характеризующая флуктуацию:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - y_n(k))^2}. \quad (7)$$

- 3) Строится зависимость $\log(F(n))$ от $\log(n)$ и определяется угол наклона аппроксимирующей прямой (φ). Тангенс данного угла определяет значение показателя Херста:

$$H = \text{tg}(\varphi). \quad (8)$$

- 4) Определяется параметр d по формуле (5).

После того как параметр d определен производится идентификация параметров авторегрессионной модели и модели скользящего среднего. Для этого производится синтез ARMA(p, q) процессов для серий порядков моделей p и q [10,15]. Главным критерием выбора оптимальных значений p и q является быстро спадающая нормированная АКФ процесса.

Для оценки эффективности предложенной методики был проведен анализ тестового фрактального процесса (см. рис. 2) с известным показателем Херста ($H=0.9$). На рисунке 3 представлен график нормированной АКФ ($\rho_s(\tau)$). Из графика можно увидеть, что изменения АКФ имеют циклический характер, и функция не достигает нуля даже в конце интервала рассмотрения. Это говорит о наличии в процессе долговременной памяти.

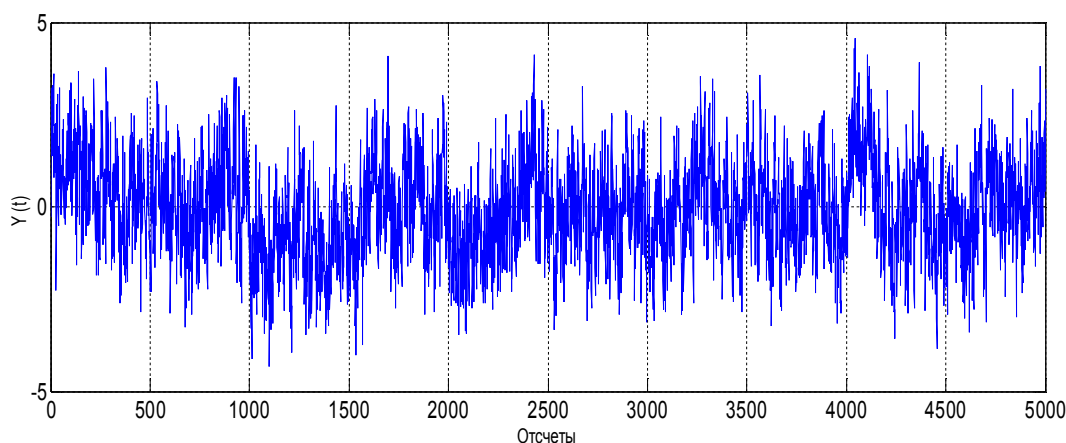


Рисунок 2 – Тестовая реализация фрактального процесса

Теоретический горизонт прогнозирования данных можно оценить по интервалу корреляции τ_k . Формула для определения τ_k как для монотонных, так и для знакопеременных АКФ имеет вид [16]:

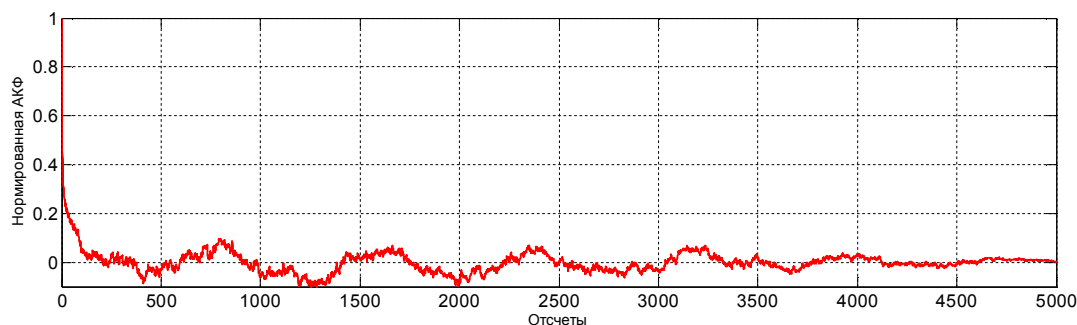


Рисунок 3 – График нормированной АКФ исследуемого процесса

$$\tau_k = \int_0^{\infty} |\rho_s(\tau)| d\tau. \quad (9)$$

Из-за того, что при решении практических задач использование больших объемов входных данных значительно затрудняет вычислительный процесс, рекомендуется определить минимальное количество отсчетов, в которых будет наблюдаться эффект долговременной памяти и по которому параметр Херста будет оценен с минимальной погрешностью. Для этого был проведен анализ зависимости среднего интервала корреляции (см. рис. 4) и точности определения параметра Херста (см. рис. 5) от размера блока анализируемых данных рассматриваемого процесса.

Зависимость на рисунке 4 имеет явно нелинейный характер. Видно, что длина входных данных в 1600 значений является достаточной, т.к. дальше интервал корреляции изменяется не значительно. Так, при длине скользящего окна в 1600 значений теоретический горизонт прогноза равен 128 отсчетам, а при длине окна в 5000 значений – 150 отсчетам. Проведенные исследования показали, что при окне в 1600 отсчетов относительная погрешность не превышает 5%. Поэтому в дальнейших исследованиях при идентификации параметров ARFIMA модели использовались входные данные длиной 1600 отсчетов.

В качестве входных данных для вычисления параметров модели ARFIMA рассмотрены пять интервалов исследуемого процесса. В таблице 1 приведены результаты идентификации параметров ARFIMA модели. Параметр d определялся методом Витгла и методом DFA. Видно, что метод DFA несколько ближе к теоретическому значению $d^*=0.4$.

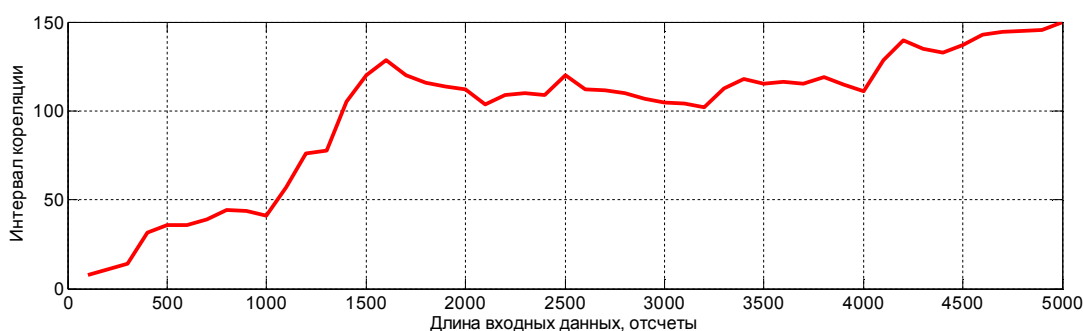


Рисунок 4 – Зависимость среднего интервала корреляции от размера блока анализируемых данных

Для оценки качества прогнозирования использован метод относительной ошибки аппроксимации [17]. Суть метода состоит в вычислении средней ошибки прогнозирования модели по формуле:

$$\bar{\delta} = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{x_t} 100\%, \tag{10}$$

где x_t - значения исходного ряда; e_t - точечная абсолютная ошибка прогноза.

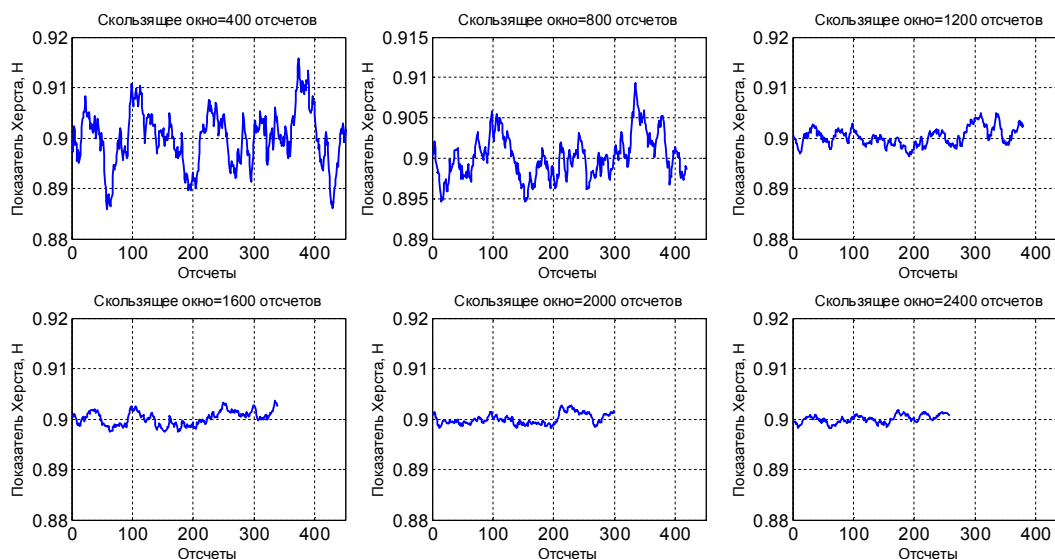


Рисунок 5 – Результаты оценки показателя Херста методом DFA при использовании различных объемов входных данных

Таблица 1

Результаты идентификации параметров ARFIMA модели при использовании различных участков исследуемого процесса

Участок входных данных	Значение параметра p	Значение параметра q	Значение параметра d (метод Витгла)	Значение параметра d (метод DFA)
[2400;4000]	8	9	0.4212	0.4009
[2600;4200]	6	9	0.4515	0.4013
[2800;4400]	7	7	0.441	0.4008
[3000;4600]	8	10	0.4242	0.4027
[3200;4800]	7	9	0.4569	0.3999

Точность прогноза при этом делится на четыре категории [17]:

- высокая - $\bar{\delta} < 10\%$;
- средняя - $\bar{\delta} = 10...20\%$;
- удовлетворительная - $\bar{\delta} = 20...50\%$;
- неудовлетворительная - $\bar{\delta} > 50\%$.

Это деление принято использовать при оценке качества прогноза социально-экономических данных. Для других задач границы интервалов качества прогноза могут отличаться.

На рисунках 6 и 7 представлены фрагменты реализаций исследуемого процесса и результаты работы синтезированных прогностических моделей с использованием метода

Виттла и DFA. График зависимости средней относительной ошибки от величины горизонта прогноза для обоих методов представлен на рисунке 8.

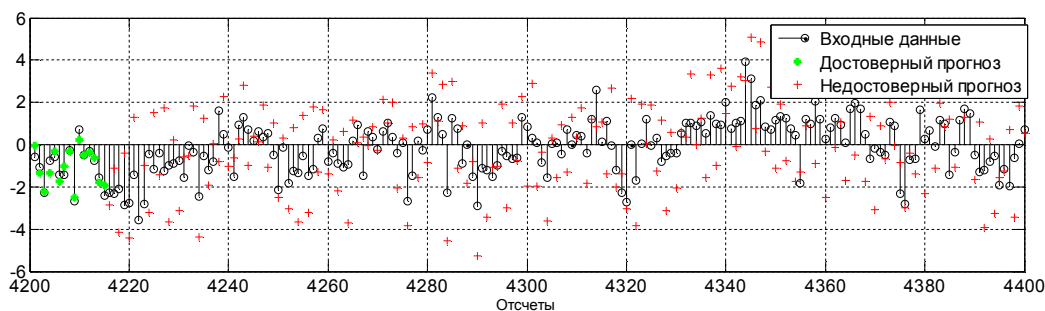


Рисунок 6 – Результаты численного моделирования работы прогностической модели построенной с использованием метода Виттла

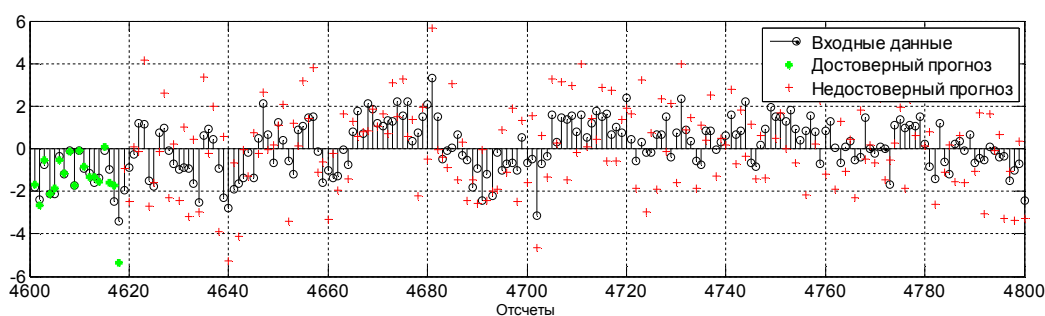


Рисунок 7 – Результаты численного моделирования работы прогностической модели построенной с использованием метода DFA

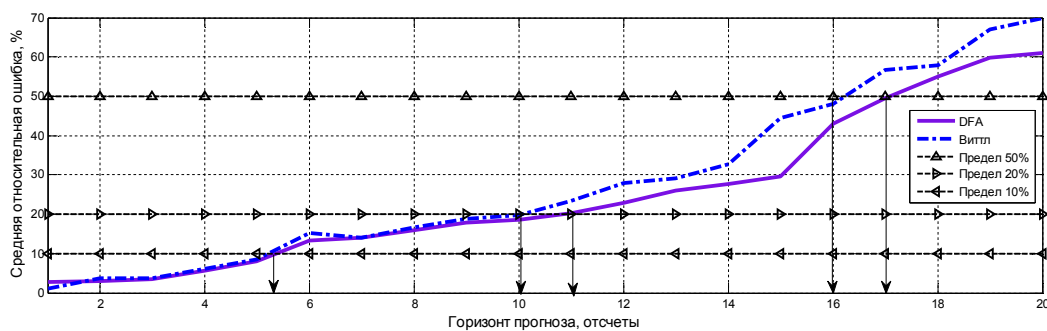


Рисунок 8 – График зависимости средней относительной ошибки от величины горизонта прогноза

Из данного графика видно, что при высокой точности прогнозирования длина достоверного прогноза примерно одинакова; при средней и удовлетворительной точности прогнозирования достоверный прогноз с использованием методики DFA выше. Критерием эффективности использования методики DFA является получение большего горизонта достоверного прогноза. Так, при использовании методики DFA достоверный прогноз составил 18 отсчетов, а при использовании метода Виттла – 16 отсчетов. Это подтверждает эффективность предложенной методики (DFA) расчета параметра d .

При применении данной методики следует особое внимание уделять точности расчета показателя Херста процесса. Как показали описанные выше исследования, одним из ключевых параметров, влияющих на точность определения показателя, является размер окна обрабатываемых данных (чем больше – тем точнее). Однако, для реальных процессов

показатель Херста может изменяться во времени и соответственно использование большого окна выборки будет вносить существенную инертность в его оценку. Таким образом, выбор размера окна должен проводиться в зависимости от решаемой задачи и априорных сведениях об анализируемом процессе.

Выводы

Рассмотрены вопросы построения прогностической модели фрактального процесса. Описана общая концепция построения ARFIMA модели. Проведен анализ взаимосвязи свойств фрактального процесса с параметрами ARFIMA модели.

Предложена новая методика определения параметра d ARFIMA модели, основанная на вычислении показателя Херста методом DFA. Даны рекомендации по практическому применению данной методики для построения прогностических моделей фрактальных процессов.

Проведена оценка эффективности предложенной методики путем численного моделирования. Показано, что данная методика позволяет увеличить горизонт удовлетворительного прогноза поведения фрактального процесса на 6% по сравнению с применением классического подхода, базирующегося на методе Виттла.

Список использованной литературы

1. Федер Е. Фракталы / Е.Федер. - пер. с англ. Ю.А. Данилова, А. Щукуров. – М.: Мир, 1991. – 254 с.
2. Короленко П.В. Новационные методы анализа стохастических процессов и структур в оптике / П.В. Короленко, М.С. Маганова, А.В. Меснянкин // Фрактальные и мультифрактальные методы вейвлет преобразования. – М.: МГУ им. Ломоносова, –2004.– 82 с.
3. Шелухин О.И. Самоподобие и фракталы. Телекоммуникационные приложения. / О.И. Шелухин, А.В. Осин, С.М. Смольский // - М.: ФИЗМАТЛИТ, 208.- 368 с.
4. Дегтяренко І.В. Аналіз методів визначення параметра самоподібності трафіку телекомунікаційних мереж / І.В. Дегтяренко, А.Г. Шепеленко // Наукові праці інституту проблем модулювання в енергетиці ім. Г.Є. Пухова. Серія: Моделювання та інформаційні технології. - Київ, 2007. – Випуск 42. – С. 68 – 73.
5. Воропаєва В.Я. Оцінка показників якості NGN-мереж з урахуванням фрактальності вхідного трафіку / В.Я. Воропаєва // Наукові праці ДонНТУ. Серія: Обчислювальна техніка та автоматизація. - Донецьк: ДонНТУ, 2008. – Випуск 15 (130). – С. 23-29
6. Булат О.Ф. Фракталы в геомеханике / О.Ф. Булат, В.І. Дирда. – К.: "Наукова думка", 2005. - 356 с.
7. Дегтяренко І.В. Прогностична модель послідовностей імпульсів акустичної емісії вугільних пластів / І.В. Дегтяренко, О.М. Гарматенко //Моделювання та інформаційні технології – Київ: Інститут проблем моделювання в енергетиці ім. Г.Є. Пухова, 2013. – С. 111-118.
8. Потапов А.А. Фракталы в радиофизике и радиолокации. Топология выборки / А.А. Потапов. - М.: Университетская книга, 2005. – 848 с.
9. Перцовский О.Е. Моделирование валютных рынков на основе процессов с длинной памятью: Препринт WP2/2004/03 / О.Е. Перцовский. — М.: ГУ ВШЭ, 2003. — 52 с.
10. Бокс Дж., Дженкинс Г. Анализ временных рядов, прогноз и управление: Пер. с англ. // Под ред. В.Ф. Писаренко. – М.: Мир, 1974, кн. 1. – 406 с.
11. Sowel F. Modelling long-run behaviour with fractional ARIMA-models / F. Sowel // Journal of Monetary Economics. – 1992. – P.277-302.
12. Ihlen Espen A.F. Introduction to multifractal detrended fluctuation analysis in Matlab / Espen A. F. Ihlen. – Norway: Department of Neuroscience. Norwegian University of Science and technology, 2012. – 18 p.

13. Shimotsu K. Exact local Whittle estimator of fractional integration / K. Shimotsu, C.B. Philips // Yale University. - 2005. – V. 33. - P. 2-6.
14. Peng CK Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series/ K Peng, S Havlin, HE Stanley, AL Goldberger Chaos. – 1995. – P. 82-87.
15. Beveridge S. Estimating time-series models / S. Beveridge. – 1993. – P. 137-142.
16. Куликов Е.И. Прикладной статистический анализ / Е.И. Куликов. – М.:Телеком, 2008. – 464 с.
17. Назаров М.Г. Курс социально-экономической статистики / М.Г. Назаров. – М.:ЮНИТИ-ДАНА, 2000. – 771с.

Надійшла до редакції:
12.04.2013 г.

Рецензент:
д-р техн. наук, проф. Воронцов О.Г.

І.В. Дегтяренко, О.М. Гарматенко, О.А. Ярошенко
ДВНЗ «Донецький національний технічний університет»

Ідентифікація параметрів ARFIMA моделі фрактального процесу. У статті розглянуті питання побудови прогностичної ARFIMA моделі фрактального процесу. Проведений аналіз взаємозв'язку параметрів ARFIMA моделі з показником Херста. Описана методика визначення параметрів даної моделі на основі використання методу «Detrended Fluctuation Analysis». Досліджена ефективність застосування даної методики для побудови прогностичних моделей.

Ключові слова: фрактальний процес, ARFIMA, показник Херста, прогностична модель, Detrended Fluctuation Analysis.

I.V. Degtyarenko, A.M. Garmatenko, O.A. Yaroshenko
Donetsk National Technical University

Identification of ARFIMA Model's Parameters for Fractal Process. The questions of developing a predictive model for fractal process are considered. Self-similarity and long-term memory are inherent for fractal processes. This is the basis for building predictive models of these processes. The general concept of developing a predictive ARFIMA model is described in the article. The interrelation of the parameters of ARFIMA model with fractal properties of the process is analyzed. The recommendations for the practical application to describe stationary processes within long-term memory are examined in the article. The procedure for calculating the parameter d of ARFIMA models using Whittle is described. A new method of determining this parameter is based on calculating the Hurst parameter using "Detrended Fluctuation Analysis" is described. This method provides the most stable and adequate Hurst parameter estimation for fractal processes. The dependences of Hurst calculation accuracy parameters and average interval of correlation with the length of data's window are analyzed. The recommendations about selection of the minimum length of data's window are given. The efficiency of the proposed method is evaluated by numerical simulation. Two prediction models of fractal process were constructed. The parameter d of first ARFIMA model was calculated by Whittle method. The parameters of the second model were calculated by the proposed method. The method of the relative error of approximation was used in assessing the quality of prediction. The simulation results showed that the developed method increases the horizon of forecasting fractal process by 6% in comparison with application of the classical approach, based on the method of Whittle.

Keywords: fractal process, ARFIMA, Hurst parameter, predictive model, Detrended Fluctuation Analysis.