



УДК 519.711.3

© 2015 г. **Ф.Ф. Пащенко**, д-р техн. наук
(Институт проблем управления им. В.А. Трапезникова РАН, Москва),
О.С. Амосов, д-р техн. наук, профессор,
Н.В. Муллер
(Комсомольский-на-Амуре государственный технический университет)

СТРУКТУРНО-ПАРАМЕТРИЧЕСКАЯ ИДЕНТИФИКАЦИЯ ВРЕМЕННОГО РЯДА С ПРИМЕНЕНИЕМ ФРАКТАЛЬНОГО И ВЕЙВЛЕТ-АНАЛИЗА

Предложены метод структурно-параметрической идентификации временного ряда с использованием фрактального и вейвлет-анализа и математическая модель временного ряда с учетом возможных диапазонов изменения показателей хаотичности.

Ключевые слова: временной ряд, математическая модель, идентификация, фрактальный и вейвлет-анализ.

Введение

Для анализа временных рядов (ВР) наиболее широко используются традиционные методы статистического анализа случайных величин и функций. Наряду с ними в последние годы получают распространение способы обработки сигналов, основанные на фрактальном и вейвлет-преобразовании [1 – 4]. Отличительная особенность последних состоит в том, что они позволяют вскрыть особенности локальной структуры и выявить различные свойства сложного сигнала, невидимые при обычном представлении в режиме реального времени.

Поэтому появляется необходимость в разработке алгоритма идентификации модели временного ряда с использованием для выявления его структуры и локальных особенностей фрактального и вейвлет-анализа.

Целью работы является структурно-параметрическая идентификация модели временного ряда с применением фрактального и вейвлет-анализа.

Постановка задачи

Структурная схема идентификации временного ряда представлена на рис. 1, она является частной задачей идентификации без наблюдаемого входного сигнала [5], где $y(t)$ – выходные данные ВР; $y_M(t)$ – выходные данные модели ВР; $\varepsilon(t)$ –

разность (невязка) между выходными данными ВР и модели; $g(t)$ – неконтролируемое возмущающее воздействие.

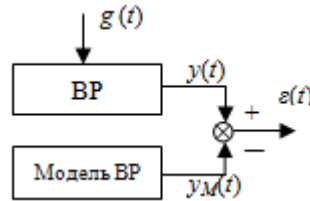


Рис. 1. Структурная схема идентификации ВР.

Сформулируем задачу структурно-параметрической идентификации.

Пусть наблюдаются выходные значения $y(t)$ ВР. Необходимо выбрать тип математической модели $y_M(t)$ временного ряда (структурная идентификация), оптимизировать ее параметры (параметрическая идентификация) путем вычисления ошибки идентификации $\varepsilon(t) = y(t) - y_M(t)$ и определения соответствия выходных и моделируемых значений ВР с использованием некоторого критерия.

В большинстве практических случаев задачу решают из условия минимума средней квадратической ошибки [5]:

$$M \{ \rho[y(t), y_M(t)] \} = M \{ \varepsilon^2(t) \} \rightarrow \min ,$$

где $\varepsilon(t) = y(t) - y_M(t)$ – ошибка идентификации; $\rho[y(t), y_M(t)] = \varepsilon^2(t)$ – функция потерь; M – знак математического ожидания.

В работе предлагается метод идентификации модели временного ряда на основе фрактального, вейвлет и корреляционного анализов.

Фрактальный анализ

Суть фрактального анализа состоит в определении степени хаотичности временного ряда и оценивается с помощью *показателей хаотичности*: показателя Херста H и фрактальной размерности D , которые отражают природу анализируемого процесса. Показатель H связан с фрактальной размерностью следующим выражением [6]: $D = 2 - H$.

Для вычисления показателей H и D имеются разработанные алгоритмы и компьютерные программы, – например, программа FRACTAN [7].

Анализ показателей позволяет правильно выбрать тип математической модели ВР.

1. Если $H \approx 0.5$, $D \approx 1.5$, временной ряд совпадает с классическим броуновским движением. Для его моделирования могут с успехом использоваться марковские модели процессов, а для анализа и идентификации достаточно традиционных статистических методов.

2. Если $0.5 < H < 1$, $1 < D < 1.5$, то имеем дело с персистентным временным рядом. Например, если наблюдается восходящая тенденция, то в будущем будет отмечаться динамика роста. С ростом H от 0.5 до 1 устойчивость становится все заметнее.

3. Для $0 < H < 0.5$; $1.5 < D < 2$ временной ряд является антиперсистентным, когда восходящая тенденция сменяется нисходящей или наоборот.

В случаях 2 и 3 для моделирования временного ряда может быть использована модель фрактального броуновского движения (ФБД) [6].

Случайный процесс $X(t)$ называется ФБД с параметром H (показатель Херста), $0 < H < 1$, если приращения случайного процесса $\Delta X(\tau) = X(t + \tau) - X(t)$ имеют гауссовское распределение вида [6]:

$$P(\Delta X < x) = \frac{1}{\sqrt{2\pi} \sigma \tau^H} \cdot \int_{-\infty}^x \exp \left[-\frac{z^2}{2\sigma^2 \tau^{2H}} \right] dz,$$

где σ – стандартное (среднеквадратичное) отклонение случайной величины X .

ФБД с параметром $H = 0.5$ совпадает с классическим броуновским движением. Приращения ФБД называются фрактальным гауссовским шумом, дисперсия которого подчиняется соотношению [6]

$$D[X(t + \tau) - X(t)] = \sigma^2 \tau^{2H}.$$

Математическая модель временного ряда

С учетом возможных диапазонов изменения показателей хаотичности предлагается математическая модель временного ряда вида $y_t = \eta_t + \chi_t + \varepsilon_t$, где η_t – трендовая составляющая, для построения которой применяются следующие функции [8]: для линейного тренда $\eta = ax + b$; для логарифмического тренда $\eta = (a \lg(x)) + b$; для степенного тренда $\eta = ax^b$; для экспоненциального тренда $\eta = ae^{bx}$; полином степени n $\eta = (a_n x^n) + \dots + (a_2 x^2) + (a_1 x^1) + b$; $t = t_0, t_1, \dots, t_n$ – моменты времени; χ_t – компонента, выражающая меру хаотичности ряда, описывающая характер процесса и зависящая от H – показателя Херста или D – показателя фрактальной размерности; ε_t – случайный шум.

Для определения количественного значения H Херст вывел эмпирический закон в виде [6]:

$$H = \text{Ln} \left(\frac{R}{S} \right) / \text{Ln} \left(\frac{n}{2} \right),$$

где R – максимальный размах исследуемого ряда; S – среднеквадратическое отклонение наблюдений; n – количество наблюдений.

Предложенная модель является универсальной и может быть применена для идентификации и прогнозирования временных рядов из разных предметных областей. Входными данными модели являются временные ряды, которые могут содержать любую информацию, описывающую исследуемый процесс, – например: оцифрованные сигналы, полученные с различного вида датчиков (виброакустика, различные виды колебаний); временные последовательности биржевых котировок (курсы валют, акций и т.д.); различные медицинские показатели и сигналы (кардиограммы, энцефалограммы, различные статистические данные); данные сейсмической активности; синоптические данные; статистические данные социальных и производственных процессов, бытового и производственного травматизма.

Вейвлет-анализ

Вейвлет-анализ практически незаменим для идентификации временного ряда при наличии у него локальных особенностей и неоднородностей. Кратко рассмотрим суть вейвлет-преобразований.

Вейвлет-преобразования.

Непрерывные прямое и обратное вейвлет-преобразование сигнала $s(t)$ имеют вид [3]:

$$W_s(a, b) = (s(t), \psi_{ab}(t)) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt, \quad s(t) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_s(a, b) \psi_{ab}(t) \frac{dadb}{a^2},$$

где $\psi_{ab}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$ – вейвлет, порождаемый материнским вейвлетом $\psi(t)$ за счет операций сдвига во времени (b) и изменения временного масштаба (a); C_ψ – нормирующий коэффициент; (\cdot, \cdot) – скалярное произведение соответствующих сомножителей.

Вейвлет-спектр $W_s(a, b)$ (wavelet spectrum, или time-scale-spectrum – масштабно-временной спектр) в отличие от фурье-спектра является функцией двух аргументов: первый аргумент a (временной масштаб) аналогичен периоду осцилляций, т.е. обратен частоте, а второй b – аналогичен смещению сигнала по оси времени. Следует отметить, что $W_s(b, a_0)$ характеризует временную зависимость (при $a = a_0$), тогда как зависимости $W_s(a, b_0)$ можно поставить в соответствие частотную зависимость (при $b = b_0$).

Дискретное вейвлет-преобразование – вейвлет-разложение аппроксимации j -го уровня разрешения $\tilde{s}(t)$ для глубины разложения m имеет вид [9]:

$$\begin{aligned} \tilde{s}_j(t_i) &= \tilde{s}_{j-m}(t_i) + \tilde{s}_{j-m}^d(t_i) + \dots + \tilde{s}_{j-1}^d(t_i), \\ \tilde{s}_j(t_i) &= \sum_{k \in \mathbf{Z}} a_{j-m,k} \varphi_{j-m,k}(t_i) + \sum_{k \in \mathbf{Z}} d_{j-m,k} \psi_{j-m,k}(t_i) + \dots + \sum_{k \in \mathbf{Z}} d_{j-1,k} \psi_{j-1,k}(t_i), \end{aligned}$$

где j характеризует уровень разрешения; $\varphi_{j-m,k}(t_i), \psi_{j-m,k}(t_i)$ – соответственно масштабирующая (аппроксимирующая) и вейвлет-функция (детализирующая функция); $\mathbf{a}_{j-m,k} = \{a_{j-m,k}\}$, $\mathbf{d}_{j-\gamma,k} = \{d_{j-m,k}\}, \gamma = \overline{1, m}$ – наборы аппроксимирующих и детализирующих коэффициентов разложения $(j-\gamma)$ уровня разрешения; $\mathbf{Z} = \{\dots, -1, 0, 1, \dots\}$ – множество целых чисел.

Дополнительные показатели вейвлет-преобразования.

Как отмечается в [1], коэффициенты вейвлет-преобразования содержат комбинированную информацию об анализирующем вейвлете и анализируемом сигнале. Несмотря на это, вейвлет-анализ позволяет получить и объективную информацию об анализируемом сигнале, так как некоторые свойства вейвлет-преобразования не зависят от выбора анализирующего вейвлета. Независимость от анализатора делает эти простые свойства преобразования очень важными.

Не останавливаясь на элементарных свойствах вейвлет-преобразования, рассмотрим два таких важных свойства.

Показатель частотно-временной локализации. Частотно-временная локализация, наличие частотно-временного окна и угла влияния позволяют проигнорировать, например, крупномасштабные полиномиальные составляющие и проанализировать особенности высокого порядка или мелкомасштабные вариации функции. Безразлично, дифференцировать ли нужное число раз анализирующий вейвлет или саму функцию. Это очень полезное свойство, если учесть, что часто функция представляет собой ряд чисел, а анализирующий вейвлет задан формулой [1].

Для вейвлет-преобразования $W[f] = W(a, b)$ функции $f(t)$

$$W(a, b) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt$$

например, первая производная по переменной a равна:

$$\frac{\partial W}{\partial a} = -\frac{1}{2} |a|^{-\frac{3}{2}} \text{sign } a \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt + |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi'\left(\frac{t-b}{a}\right) \left(-\frac{t-b}{a^2}\right) dt =$$

$$-\frac{1}{2} |a|^{-\frac{5}{2}} \left(|a|^2 \text{sign } a \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \right) + 2 \int_{-\infty}^{\infty} f(t) \psi'\left(\frac{t-b}{a}\right) (t-b) dt.$$

$$\frac{\partial W}{\partial \alpha} = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) \psi'\left(\frac{t-b}{a}\right) \left(-\frac{1}{a}\right) dt = -|a|^{-\frac{3}{2}} \int_{-\infty}^{\infty} f(t) \psi'\left(\frac{t-b}{a}\right) dt.$$

$$\text{sign } a = \begin{cases} 1, & \text{если } a > 0; \\ 0, & \text{если } a = 0; \\ -1, & \text{если } a < 0, \end{cases}$$

где t – ось времени функции $f(t)$; b – момент времени; a – временной масштаб; $\Psi(t)$ – вейвлет-функция.

Данная величина дает наглядную картину наиболее резко изменяющихся компонент сигнала. Этот показатель позволяет оценивать внезапность возникновения сигнала или длительное нарастание в течение времени.

Корреляционный анализ вейвлет-спектров. Для его проведения по аналогии можно воспользоваться известной зависимостью между корреляционными функциями и спектрами сигналов подобно тому, как это делается для преобразования Фурье [10].

Так как результаты вейвлет-анализа имеют вид многомерных массивов вейвлет-коэффициентов, несущих большой объем информации, то для корректного сравнения таких данных следует применять корреляционные методы анализа. Они позволят достоверно выявить закономерности и степени соответствия при сравнении вейвлет-скалограмм путем получения функций корреляционных зависимостей для более точных прогнозных оценок.

Алгоритм идентификации

Для того, чтобы идентифицировать модель временного ряда, нами предлагается использовать алгоритм, состоящий из следующих этапов:

1. Предобработка временного ряда, которая заключается в выборке статистических данных для формирования интересующего нас временного ряда.
2. Проведение для синтеза временного ряда фрактального анализа.
3. Вейвлет-анализ.
4. Применение дополнительных показателей частотно-временной локализации и корреляционный анализ вейвлет-спектров. В качестве дополнительного показателя используем первую производную от результатов частотно-временного распределения (первая производная вейвлет- спектра), которая позволяет оценить скорость изменения возникновения этих компонентов.
5. Синтез структуры и параметров математической модели ВР.
6. Проверка критерия.

Результатом решения будет полноценная объемная информация о временном ряде, с выявленными характеристиками, по которым можно идентифицировать математическую модель ВР. Блок-схема алгоритма, отражающая последовательность всех этапов метода идентификации, представлена на рис. 2.

Иллюстрирующий пример (анализ производственного травматизма в Комсомольск-на-Амуре)

Пример иллюстрирует первые четыре этапа идентификации.

Первый этап. Для формирования интересующего нас временного ряда выполняем выборку статистических данных – из данных отдела социального страхования по производственному травматизму г. Комсомольска-на-Амуре за период с 2001 г. по 2003 г.

Второй этап. Проводим обработку временного ряда методом фрактального анализа.

Определение показателей хаотичности H и D :

С 2001 г. по 2002 г. $H = 0.58$ и $D = 1.42$.

С 2002 г. по 2003 г. $H = 0.6$ и $D = 1.4$.

Показатели хаотичности находятся в диапазоне $0,5 < H < 1$; $1 < D < 1,5$, поэтому исследуемый временной ряд является персистентным с восходящей тенденцией, так как наблюдается динамика роста H .

Третий этап. Вейвлет-преобразование дает возможность получить более полноценную информацию об исследуемом процессе. На рис. 3, 4 по каждому году отдельно представлен вейвлет-спектр, отражающий появление несчастного случая у работающих.

Из рис. 3 и 4 видно, что мощность спектров находится в прямо пропорциональной зависимости со значениями фрактальной размерности.

На рис. 3 на вейвлет-спектре наблюдается резкий скачок возникновения нестационарности сигнала в январе, марте, апреле, октябре и ноябре исследуемого периода, а на рис. 4 – в феврале, апреле, мае и сентябре 2002 г.

В итоге появляется возможность выделить наиболее нестационарные области и в последующем по выделенным областям более детально провести анализ причин возникновения выявленных нестационарностей в сигнале.

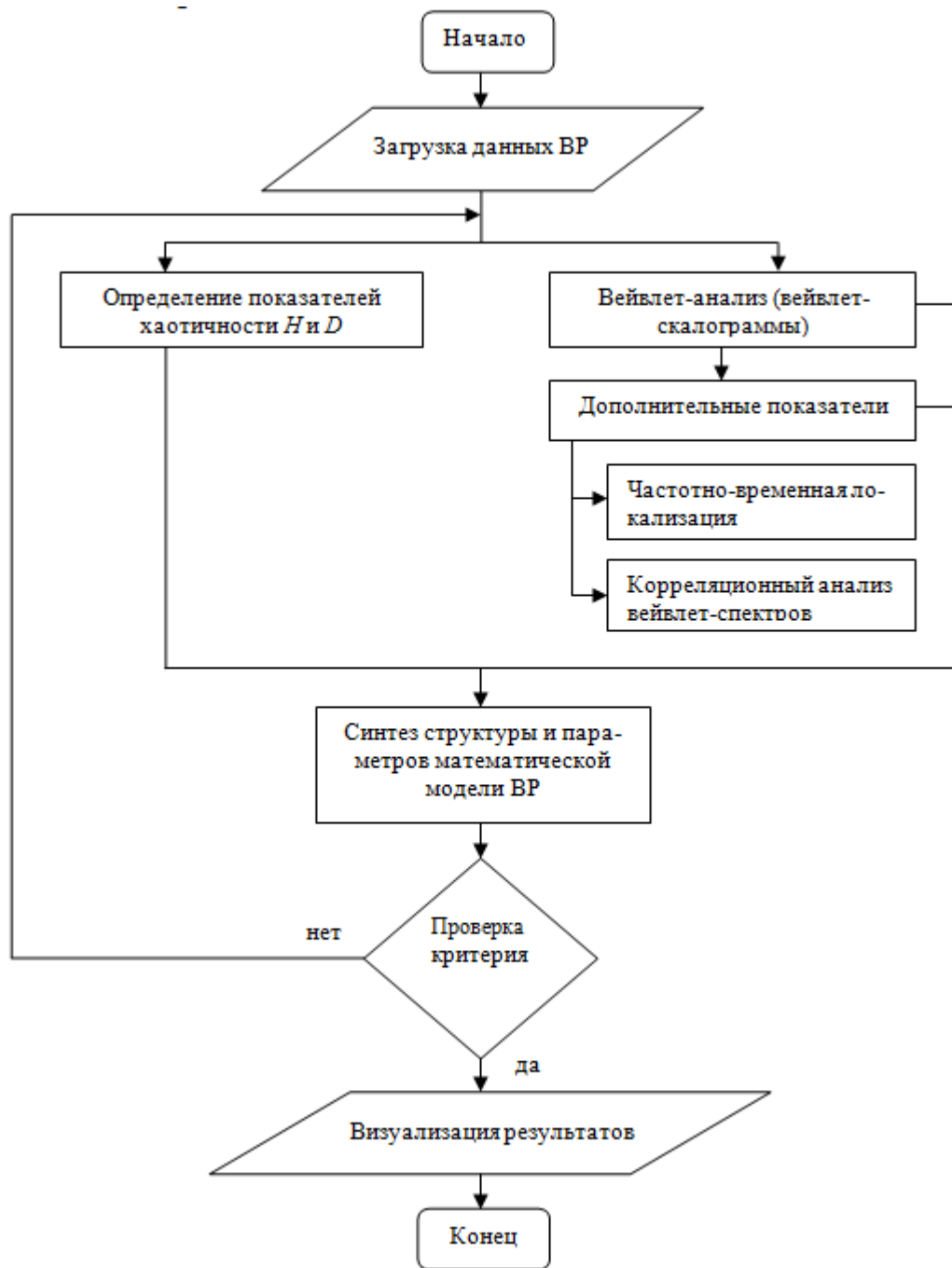


Рис. 2. Блок-схема алгоритма идентификации.

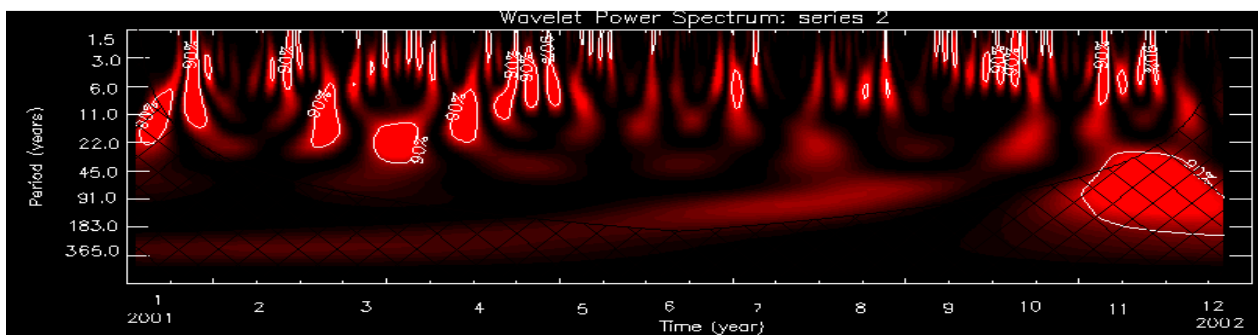


Рис. 3. Вейвлет-спектр временного ряда травматизма за 2001 г.

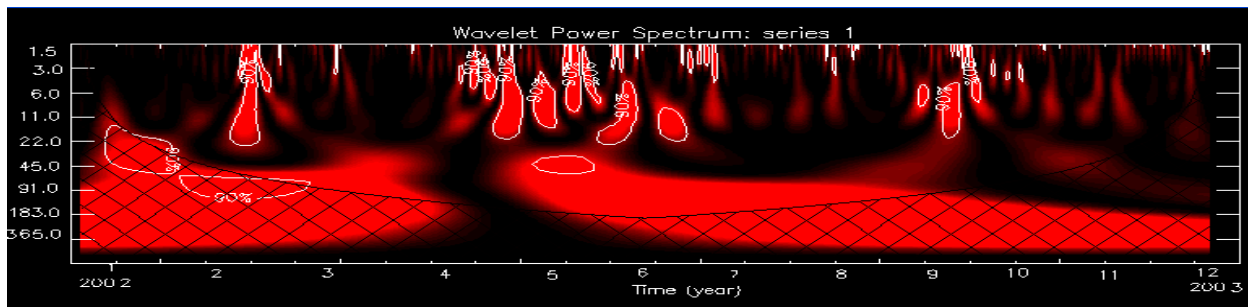


Рис. 4. Вейвлет-спектр временного ряда травматизма за 2002 г.

Четвертый этап. Определим частотно-временную локализацию за период с 2001 г. по 2002 г. (рис. 5).

Показатель частотно-временной локализации также демонстрирует, что в январе, апреле, октябре и ноябре 2001 г. наблюдается рост травматизма.

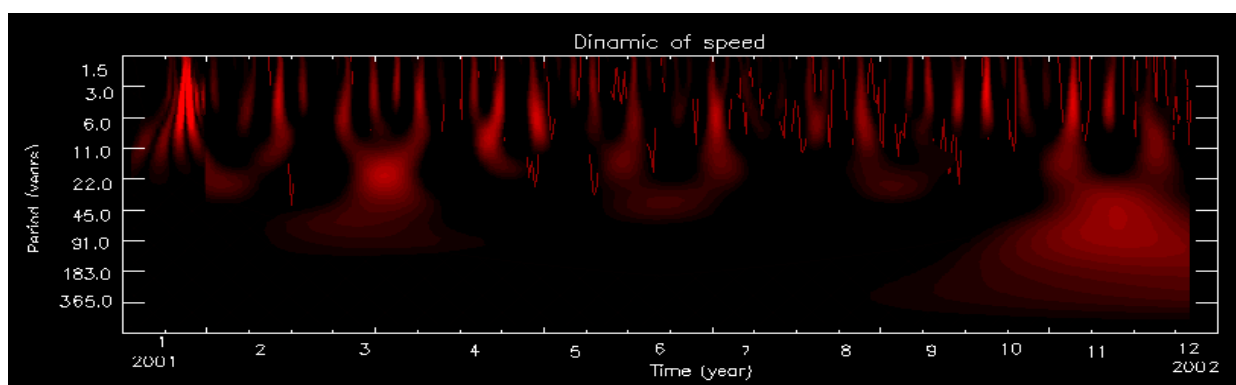


Рис. 5. Дополнительный показатель частотно-временной локализации с 2001 г. по 2002 г.

Далее проводим корреляционный анализ вейвлет-спектров путем получения численных значений корреляции двух сравниваемых процессов для получения качественного (в виде графической визуализации) и количественного сравнения (в виде числового значения корреляции) с целью выявления дополнительной информации и закономерностей (рис. 6).

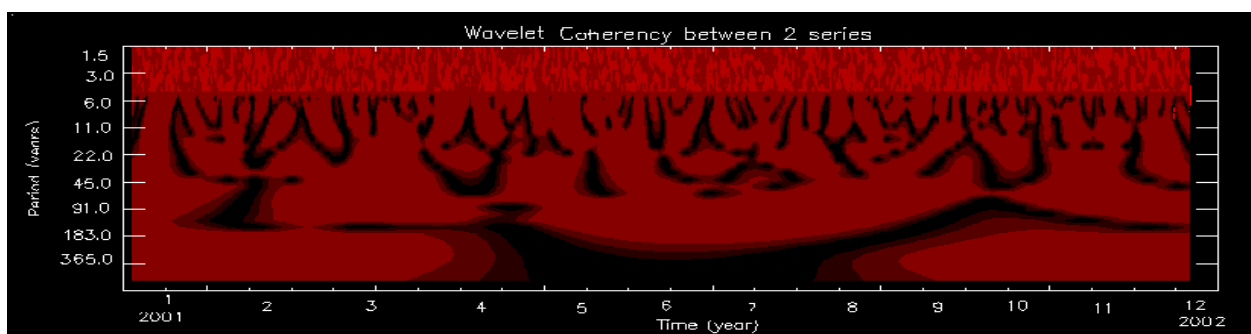


Рис. 6. Корреляция вейвлет-спектров временного ряда травматизма 2001 г. и 2002 г.

Рис. 6 отражает степень корреляции двух исследуемых процессов (рис. 3 и 4) в виде скалограммы, где наложение и совпадение рисунков дает ярко-красный цвет, что говорит о максимальном совпадении и корреляции исследуемых процессов.

Заключение

Дана постановка задачи структурно-параметрической идентификации временного ряда и предложен метод ее решения с использованием фрактального и вейвлет анализа.

Разработана математическая модель временного ряда с учетом возможных диапазонов изменения показателей хаотичности.

Предложен дополнительный показатель частотно-временной локализации временных рядов, позволяющий оценить угол влияния и скорость возникновения нестационарности.

Расширены возможности предложенных видов анализа введением корреляционного анализа (вторичных) параметров аппроксимации – результатов вейвлет-скалограмм.

Рассмотрен практический пример временного ряда с показателем Херста, отличным от 0.5, для его обработки используются вейвлет-преобразование и дополнительные показатели.

В дальнейшем планируется проиллюстрировать синтез структуры и параметров математической модели ВР.

ЛИТЕРАТУРА

1. Астафьева Н.М. Вейвлет-анализ: Основы теории и примеры применения // Успехи физических наук. – 1996. – Т.166, № 11. – С.1145-1170.
2. Смоленцев Н.К. Основы теории вейвлетов. Вейвлет в MATLAB. – М.: ДМК Пресс, 2005.
3. Яковлев А.Н. Введение в вейвлет-преобразование. Учебное пособие. – Новосибирск: Изд-во НГТУ, 2003.
4. Амосов О.С., Муллер Н.В. Исследование временных рядов с применением фрактального и вейвлет-анализа // Интернет-журнал «Науковедение». – 2014. – № 3 (22). – С. 1-14.
5. Пащенко Ф.Ф., Пикина Г.А. Основы моделирования энергетических объектов. – М.: Физматлит, 2011.
6. Короленко, П.В., Маганова, М.С., Меснянкин А.В. Новационные методы анализа стохастических процессов и структур в оптике. Фрактальные и мультифрактальные методы, вейвлет-преобразования. Учебное пособие. – М.: МГУ им. Ломоносова, Научно-исследовательский институт ядерной физики им. Д.В. Скобельцына, 2004.
7. Сычев В.В. Программа «FRACTAN 3.2» [Электронный ресурс] URL: <http://www.iki.rssi.ru/magbase/RESULT/APPENDIX/fractan.boom.ru/soft.htm> (дата обращения 27.04.2015).
8. Эконометрика. Учебник / под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2003.
9. Амосов О.С., Амосова Л.Н., Магола Д.С. Оценивание случайных последовательностей с использованием регрессии и вейвлетов // Информатика и системы управления. – 2009. – №3 (21) – С.101-109.
10. Сергиенко А.Б. Цифровая обработка сигналов. – СПб.: Питер, 2006.

E-mail:

Пащенко Федор Федорович – feodor@ipu.ru;

Амосов Олег Семенович – osa18@yandex.ru;

Муллер Нина Васильевна – only_nina@mail.ru.