

УДК 681.51

ВЫБОР ОПТИМАЛЬНЫХ БАЗОВЫХ ПАРАМЕТРОВ ИСТОЧНИКОВ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ ПРИ ИХ ИДЕНТИФИКАЦИИ

К.Г. Кирьянов

Нижегородский государственный университет им. Н.И. Лобачевского
Россия, 603009, Нижний Новгород, Проспект Гагарина, 23, Радиофизический факультет
E-mail: kkg@rf.unn.ru

Ключевые слова: идентификация систем, математическая модель, базовые параметры, свободные параметры, динамическая система, фазовая траектория, фазовое пространство, источник экспериментальных данных, источник текста, алгебраическая структура, генетическая карта данных

Key words: identification of systems, mathematical model, base parameters, free parameters, dynamic system, phase trajectory, phase space, source of the experimental data, source of the text, algebraic frame, genetic map of the data

Вводятся понятия набора базовых параметров и критериев выбора базовых параметров математических моделей Источников нестационарных экспериментальных данных, позволяющих с единых методических позиций объединить и «сшить» воедино традиционно разрозненные этапы идентификации, требующие для своей реализации, как правило, дополнительных априорных сведений и предположений о модели Источника данных. Приводятся примеры БП математических моделей Источников различных данных.

CHOICE OF OPTIMAL BASE PARAMETERS OF SOURCES OF THE EXPERIMENTAL DATA AT THEIR IDENTIFICATION / K.G.Kiryanov (The Nizhny Novgorod State University by N.I. Lobachevsky, Russia, Nizhny Novgorod 603009, 23, Gagarina av., Radiophysical Department, E-mail: kkg@rf.unn.ru). The concepts of a gang the base parameters and criteria of choice the base parameters of mathematical models of the Source of the non-stationary experimental data are introduced. The concept allows from uniform methodical positions to unite together traditionally separate stages of identifications requiring for the implementation, as a rule, padding prior informations and the suppositions about model of the Source. The examples BP of mathematical models of the Source of the various data are reduced.

1. Введение

Предлагается строгий подход (метод) получения численных оценок параметров ММ нестационарного источника выборочных значений экспериментальных данных, основанный на анализе, наблюдении и минимизации по этим параметрам определенных критериев (найденных нами функционалов). Излагаемые далее результаты носят научно-методический характер и могут эффективно использоваться не только в теоретических конструкциях, но и, что может быть более важно, в прикладных задачах. К решению таких задач относится цикл работ сотрудников ННГУ и членов проблемного Совета «Физико-математические основы наукоемких технологий» в МВВО АТН РФ и исследовательской лабора-

тории при этом Совете по определению так называемых Базовых параметров (БП) математической модели (ММ) Источников (И) экспериментальных данных фактически любой природы.

Общеизвестно, что стоимость любой исследовательской и научно-технической продукции определяется не только комплектующими изделиями, весом, габаритами и т.п. Стоимость такой продукции определяется, прежде всего, интеллектуальными усилиями, неизбежно затрачиваемыми при создании наукоемких изделий, определяющими ее пользовательские, метрологические и надежность характеристики (достоверность, погрешность, ресурс и т.д.).

Поэтому далее уместно рассмотреть один из результатов работы направленный на оптимальное решение наукоемких задач в чрезвычайно актуальной области Идентификации и Моделирования [1-6, 19-22], предложенный и совершенствованный на протяжении последних лет (см., например, [7-16, 24-27]) членами проблемного Совета «Физико-математические основы наукоемких технологий» в МВВО АТН РФ и исследовательской лаборатории при этом Совете.

Излагаемые результаты носят научно-методический характер и могут эффективно использоваться не только в теоретических конструкциях, но и, что может быть более важно, в прикладных задачах. К таким задачам относится цикл работ сотрудников и членов упомянутого Совета по определению т.н. Базовых параметров (БП) математической модели (ММ) Источников (И) экспериментальных данных фактически любой природы.

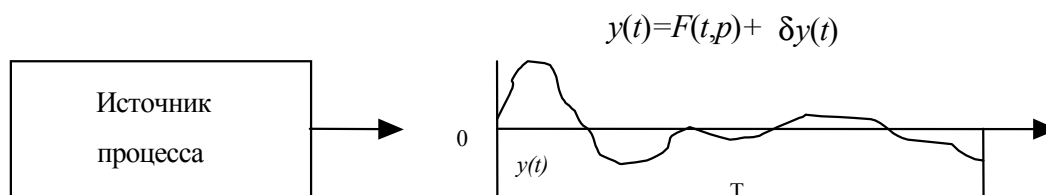


Рис. 1. Объект рассмотрения – необходимое число и значения компонент векторных параметров « p » математической модели $y(t)=F(t,p)$ источника аналогового или дискретного процесса (известных экспериментальных данных на интервале $t \in [0, T]$ с погрешностью $\delta y(t)$).

Далее предлагается строгий подход (метод) получения численных оценок параметров ММ нестационарного источника выборочных значений экспериментальных данных, основанный на анализе, наблюдении и минимизации по этим параметрам определенных критериев (найденных нами функционалов)

- либо минимума условной энтропии (Критерий1),
- либо минимума числа возможных генераторов, которые способны породить (вычислить с оптимальной точностью) эти данные (Критерий2),
- наглядное строгое графическое решение задачи выбора БП основанное на этом подходе (методе).

Графическое решение задачи бывает иногда важнее, чем аналитическое, так как интерес к образному представлению информации и «картинкам» у человека не пропадет никогда.

2. Определение базовых параметров математических моделей

К БП стационарного процесса нами отнесены три параметра – $(\Delta t, \Delta y, n)$ или иногда более удобные на практике и эквивалентные им

$$(1) \quad БП = (\Delta t, q, n).$$

К БП не стационарного процесса в (1) следует добавить «правильные» границы $t^{(i)}$ и/или длины

$$m^{(i)} = t^{(i)} - t^{(i-1)}$$

участков стационарности экспериментальных данных и индексы участков

$$(1^*). \quad БП = (\Delta t^i, q^i, n^i, t^i).$$

Далее для выбранного участка индексы указывать не будем. Смысл и польза БП заключается в том, что

- для нахождения их численных оценок требуются только сами данные с минимальной априорной информацией о ММ самого Источника данных (алфавиты символов данных и другие минимальные сведения об их алгебраической структуре (АС) [11-15, 23]);
- не зная определенных выше (по наборам (1), (1*)) или эквивалентным им БП стационарного участка нельзя перейти, хотим мы этого или не хотим, к дальнейшим этапам (раундам, шагам) процедуры идентификации;
- БП позволяют правильно выбрать «все остальные» сопутствующие им «Рабочие» параметры (РП) или, как их часто называют, «Свободные» параметры (СП) ММ;
- БП позволяют, располагая *дополнительной* априорной информацией о характеристиках Источника данных, правильно вычислить «все остальные» сопутствующие им РП (СП) любым из известных способов идентификации (методом наименьших квадратов (МНК), максимума правдоподобия (МП) и др.).

3. Пояснения к вычислению величин, входящих в набор БП

Любые исходные (аналоговые или дискретные) данные $y(t)$ при $t \in [0, T]$, не зависимо от нашего желания должны, быть проквантованы по уровню и времени и преобразованы в числовой вектор (рис. 2)

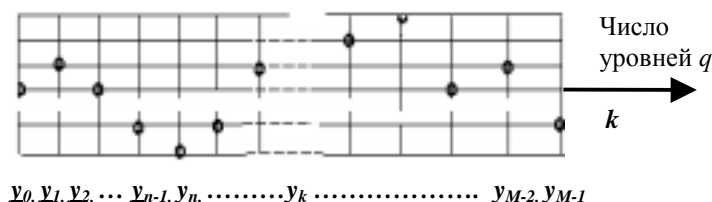


Рис. 2. Пример проквантованного по уровню и времени одномерного процесса $y(t)$.

$$(2) \quad y = (y_0, y_1, y_2, \dots, y_{n-1}, y_n, \dots, y_{M-2}, y_{M-1})$$

по формулам:

$$\begin{aligned}
 t &= \Delta t \times k + \Theta, \Theta = 0 \text{ (не нарушая общности);} \\
 y_k &= y(t \times k + \Theta); M = \lceil T / \Delta t \rceil; \\
 (3) \quad q(T) &= 1 + \lceil |y_{\max} - y_{\min}| / \Delta y \rceil,
 \end{aligned}$$

где $(\Delta t, \Delta y)$ – параметры дискретизации по времени и уровню квантования, [...] – округление содержимого скобок до максимального целого числа, а M – целое число выборок по времени, $q(T)$ – целое число уровней квантования, зависящее от длины текста T .

Проквантованный по времени и уровню набор данных далее будем называть *текстом*.

Достаточность размера $(\Delta t, \Delta y)$ – шагов для отображения процесса $y(t)$ на узлы сетки может определяться либо априорными сведениями о процессе, либо вычисляться на основе Критериев оптимальности.

Не снижая общности на рис. 2 и далее будем рассматривать, где не оговорено противное, только одномерные, скалярные данные. Распространение результатов подхода определения БП ММ настоящей работы на многомерный, векторный случай данных см., например, в [16] и Приложении 2.

4. Исходные Положения метода

Исходными Положениями (условиями справедливости) метода, дающими возможность использования его и предлагаемых далее критериев являются:

- 1) возможность получения квантованного текста с произвольными шагами $(\Delta t, \Delta y)$ по времени и уровню из исходного аналогового (непрерывного, разрывного) или дискретного (цифрового, символьного закодированного цифрами) одномерного или многомерного процесса $y(t)$, $t \in [0, T]$ и
- 2) существование ДДС (абстрактного детерминированного автомата) n -го порядка, являющегося ММ Источника текста, который может безошибочно породить (прогенерировать) все M символов текста по его n начальным символам (начальным условиям, начальной «затравке»).

Видим, что рассматриваемая ММ Источника текста носит *динамический характер и описывается* поведением фазовых траекторий (ФТ) в фазовом пространстве (ФП).

5. Определение стационарных участков экспериментальных данных восстановлением ФТ в ФП

Процедура определения стационарных участков экспериментальных данных основана на последовательном адаптивном восстановлении ФТ в ФП.

Полагая, что любые n последовательных выборок (« n -ка» данных) образуют одну точку в дискретном « $q - n$ » ФП ДС, являющейся источником экспериментальной последовательности $\{y_k\}$, восстанавливаем последовательными шагами точка за точкой траекторию в ФП при заданном числе уровней квантования q . Фазовая траектория будет находиться в гиперкубе со сторонами из q дискрет (рис. 3а). В силу конечности числа точек в дискретном ФП вновь построенная точка ФТ может совпасть с ранее восстановленной на каком-то из

предшествующих шагов. Т.е. траектория замкнется (выделится участок « $q - n$ стационарности»). Далее можем при фиксированном размере стороны гиперкуба q увеличивать размерность пространства вложения (окна данных) n на единицу, до тех пор пока вся фазовая траектория из M точек не перестанет зацикливаться (рис.3б). При этих n и q весь текст будет « $q - n$ стационарным».

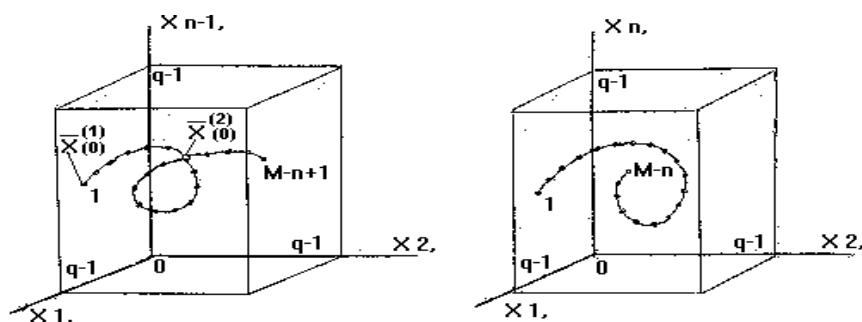


Рис. 3. Процедура восстановления фазовой траектории по экспериментальным данным. Случаи а) с самопересечением («зацикливанием») ФТ в ФП – выделением «стационарного» участка данных, б) без самопересечения.

Метод восстановления траектории текста (а не исходных данных!) позволяет не только единообразно, определять все БП, но и устанавливать расширенный набор БП сложных « $q - n$ » не стационарных сигналов - границы ($t^{(i)}$) и длины участков ($m^{(i)} = t^{(i)} - t^{(i-1)}$) стационарности экспериментальных данных, т.е. БП, а, следовательно, далее РП (СП) на соответствующих участках:

$$(4) \quad y(t)_{t \in [0, T]} \Rightarrow \{y_k\}_{k=0, \dots, M-1}$$

$$(5) \quad \{y_k\}_{k=0, \dots, M-1} \Leftrightarrow GK \equiv \{t^{(i)}, m^{(i)}, s^{(i)}, c^{(i)}, cc^{(i)}, f^{(i)}, x_0^{(i)}; I, J / n, q, a\},$$

$$i = 1, 2, \dots, I;$$

Здесь:

$$(6) \quad m(1) + \dots + m(I) + J = M,$$

$$(7) \quad x_{k+1} = f^{(i)}(x_k, p^{(i)}, x_0^{(i)}).$$

ММ генератора i -го участка данных, записанная в форме стационарного разностного уравнения Коши, не зависящего от дискретного времени k и позволяющая по n начальным последовательным выборкам данных $x_0^{(i)}$ (начальной « n -ке») последовательно вычислить все остальные выборки участка, $p^{(i)}$ - параметры $f^{(i)}$. Решением этого уравнения при начальных условиях $x_0^{(i)}$ являются функция $y^{(i)}(t) = F^{(i)}(t, p^{(i)})$, соответствующая i -му участку «стационарности, автономности», как и функция $F(t, p) = F^{(1)}(t, p^{(1)})$, при $I=1$, изображенная на рис.1;

- $f^{(i)}$ – дискретная по уровню и времени стационарная, автономная (не зависящая от дискретного времени k) функция последования или т.н. прогнозирующий оператор i -го участка данных;
- $k = 0, 1, 2, \dots, m^{(i)} - n + 1$ – «локальная» координата выборок i -го участка ФТ;

- $t^{(i)} = m^{(1)} + \dots + m^{(i)}$ – «абсолютная» координата конца i -го участка ФТ;
- $s^{(i)}$ – число точек до входа в цикл на i -м участке ФТ,
- $c^{(i)}$ – число точек в цикле i -го участка ФТ,
- $cc^{(i)}$ – полное число точек на цикле i -го участка ФТ,
 $s^{(i)} + c^{(i)} \leq s^{(i)} + cc^{(i)} = m^{(i)}$,
- $A = (AH, AI)$ – алгоритмы нарезки и идентификации.

Алгоритмы нарезки – AH : AH_0 – конец участка определяется первым *входом* на цикл, AH_1 – конец участка определяется первым *выходом* с цикла. Алгоритм идентификации AI : нахождения РП и СП на нарезанных, стационарных участках для любых подходящих ММ участков данных.

БП и родственные им функционалы, как и параметры ГК и ФТ $t^{(i)}$, $s^{(i)}$, $c^{(i)}$ и $cc^{(i)}$ участков ГК далее представлены на ряде примеров (подрисовочных подписях, графиках – см., например, Приложения) без дополнительных пояснений.

Видим, что наше определение стационарности данных и текстов, основано на общепринятой классификации разностных и дифференциальных уравнений.

Из сказанного следуют условия стационарности, автономности текста:

$$(8) \quad I = 0,$$

или

$$(9) \quad J = M.$$

После определения БП на каждом участке стационарности РП (СП) могут уже определяться любым из известных способов – МНК, МП и др. [1-6].

Отметим, что при экспресс-анализе текстов весьма полезны быстро оцениваемые неполные ГК (НГК) – $\{I\}, \{m^{(i)}; I\}, \dots$ и функционалы ГК (ФГК) – $\{m_{Av} \approx M/I\}$ и т.д., но НГК не позволяют восстановить однозначно текст. Полная ГК взаимно однозначна по отношению к « $(\Delta t, q, n)$ -округленным» на интервале $[0, T]$ исходным данным. По ней, в принципе, можно восстановить однозначно только текст, но не исходные данные $y(t)$ до их округления. При нахождении ГК наиболее трудоемким оказывается вычисление оператора $f^{(i)}$ [15].

6. Возможность сжатия (компрессии) текста

Из определения ГК (5) следует, что сжатие (компрессия) с возможностью последующего восстановления (декомпрессии) текста определенными техническими средствами возможно, если объем закодированного текста (левая часть соотношения (5)) больше объема закодированной ГК (правая часть соотношения (5)).

Из определения ГК также следует, что сжатие (компрессия) без возможности последующего восстановления (декомпрессии) текста более простыми техническими средствами возможно, если используются НГК (раздел 5).

7. Условия случайности и детерминированности текста

Когда конкретная реализация текста не известна и непосредственное получение ГК с помощью процедуры восстановления ФТ невозможно, остается теоретическая возможность оценки параметров ГК (средней длины гена, числа участков и т.п.) по *априорно известным* БП q и n текста, входящим в условие нарезки (5). При этом оказывается чрезвычайно полезна теоретическая вероятность заикливания ФТ точно на m -м шаге $p_T(m)$ в предположении независимости и равной вероятности «букв» y_m текста, так как ее при больших значениях I и M можно сравнить с экспериментальной гистограммой $p_\Sigma(m)$ длин $m^{(i)}$ участков ГК и проверить, например, по расхождению

$$(10) \quad d_{pb} = \sum_{m=1..I} |p_T(m) - p_\Sigma(m)|^b, b = 1, 2,$$

упомянутых вероятностей степень «не детерминированности» или «случайности» текста.

Более того, можно для заданного текста подобрать БП q и n , соответствующие «наиболее случайной ГК». Заметим, что в разделе 5 описан вариант-процедуры восстановления ФТ и условия стационарности (8), (9) для определения БП *полного* текста длины M .

Для независимых случайных равновероятных чисел y_k (последовательностей Бернулли) $p_T(m)$ дается точными формулами [10,15]:

$$(11) \quad \begin{aligned} p_T(m) &= (m / N^{m+1}) [N(N-1)(N-2)(N-3) \dots (N-m+1)] = \\ &= (m / N^{m+1}) [N! / (N-m)!] = m \cdot m! / N^{m+1} \cdot C_N^m = \\ &= m [N]_m / N^{m+1}, m = 1, 2, \dots, N, \quad N = q^n \end{aligned}$$

и вероятность $p_T(m)$ равна нулю при остальных значениях m .

Условие нормировки выполняется в силу тождества Абеля:

$$(12) \quad (1/N) \cdot \sum_{m=1..N} (m \cdot m! N^{(-m)} C_N^m) = 1.$$

Замена факториалов в (46) по формуле Стирлинга дает хорошее приближение:

$$(13) \quad p_T(m) = (m / N^m) (1 - 1/N)^{(m-N-1/2)} \Psi(m, N),$$

где множитель

$$(14) \quad \Psi(m, N) = (e^{-m} \cdot (1 - m/N)^{-N}) \rightarrow 1,$$

при фиксированном m и $N \rightarrow \infty$.

Примеры распределений $p_T(m)$ и $p_{\Sigma}(m)$ представлены на рис.4.

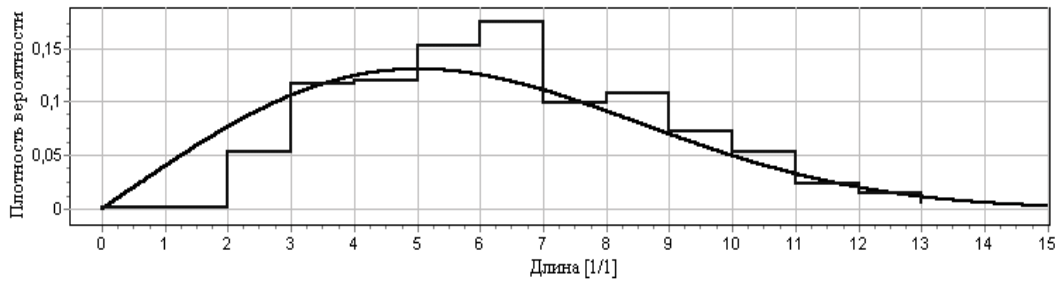


Рис. 4. Распределение длин генов в файле вируса гепатита В – «Гепатит HBV320.txt» с $M=3183$ и числовой кодировкой нуклеотидов в генетическом тексте: $T \leftrightarrow 0$, $C \leftrightarrow 1$, $A \leftrightarrow 2$, $G \leftrightarrow 3$ [27]. Расхождения (10) $dp1=0,0128$ и $dp2=0,0003$ получаются для конкретного текста при значениях $q=5$ и $n=2$. При этом $I=451$, $m_{\max}=5,0$, $m_{\text{перегиба}}=8,7$, $m_{\text{ср}(\Sigma)}=7,1$, $m_{\text{ср}(T)}=6,3$, $m_{\text{ско}(\Sigma)}=2,5$, $m_{\text{ско}(T)}=3,3$.

Точка максимума теоретического одномодального распределения $p_T(m)$ «высоты» $1/(eN)^{1/2}$ расположена примерно при $m_{\max} \cong (N)^{1/2}$, а точка перегиба – при значении

$$m_{\text{перегиба}} \cong (3 * N)^{1/2},$$

что при $q=5$ и $n=2$ соответствует распределению на рис.4.

Параметр N распределения $p_T(m)$ – единственный и важнейший. Характеризует объем гиперкуба рис.3 и, следовательно, сложность Источника текста.

Для экспресс-идентификации по гистограмме длин $m^{(i)}$ параметра N сложности текста или сложности Источника, породившего этот текст, полезны приближенные соотношения для среднего и дисперсии длины участка:

$$(15) \quad m_{Av} = M \{m^{(i)}\} = \sqrt{\pi/2} N^{1/2} \cong 1.253 N^{1/2} \cong N^{1/2}$$

$$(16) \quad \sigma_m = \left\{ M \left[(m^{(i)})^2 - (m_{Av})^2 \right] \right\}^{1/2} \cong 0.656 \cdot N^{1/2}.$$

Для текста Бернулли все средние длины предпериода m_{nn} и периода m_n , как при нарезке типа H_0 , так и H_1 примерно совпадают со средней длиной гена

$$(17) \quad m_{nn} \cong m_n \cong N^{1/2},$$

что соответствует закону пройденного пути броуновской частицы в теории Эйнштейна-Смолуховского или длине случайно порождаемого при самосборке полимерного клубка [15]. При этом, согласно уравнению (11), при малых длинах текста m вероятность $p_T(m)$ заикливания или, другими словами, встречи n -конца с ранее встречавшимся похожим n -участком, растет линейно.

Важным следствием предложенной ММ ГК является факт, следующий из формул (11)-(16), что подавляющее большинство «коротких» текстов длины

$$M \leq N^{1/2} = q^{n/2},$$

при восстановлении ФТ не приводит к ее заикливанию.

8.Оптимальные БП

8.1. Критерий 1

Критерий 1 выбора оптимальных базовых параметров ММ полного текста или текста на найденном участке стационарности.

При восстановлении ФТ вычисляя функцию условной энтропии (18) в выбранном из априорных сведений интервале при каждом $q \in [q_{\min}, q_{\max}]$

(18) $H(q/n, M, A) \equiv H(q) \equiv E(q) \equiv \log_2 N(q) \equiv \log_2(q^{n(q)}) \equiv n(q) * \log_2 q$
получаем график $H(q) \equiv E(q)$, который при достаточном размере области поиска $q_{\max} - q_{\min} + 1$ часто имеет характерный вид представленный на Рис.5.

Функция $E(q)$, как это следует из процедуры восстановления участка ФТ, характеризует сложность $N(q)$ ФП ММ ДДС, или, что в случае описанного алгоритма тоже самое, эффективность «плотной укладки» стационарного участка траектории ФП вложенной в гиперкуб рис.3 за счет касания его границ.

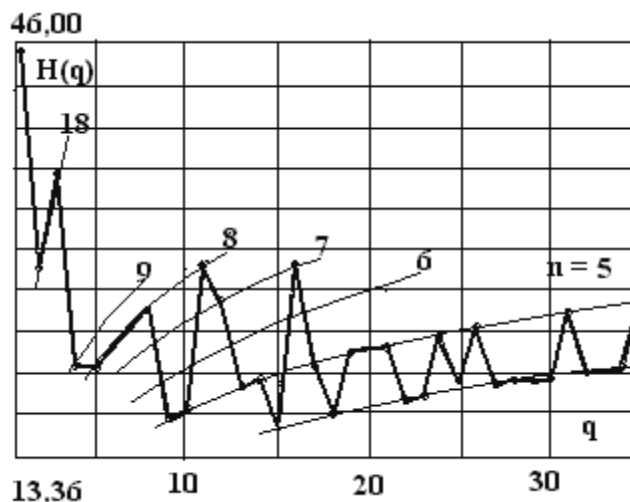


Рис. 5. Изменения энтропии $E(q)$ оптической прозрачности атмосферы в г. Москве за январь-март 1988 г.

По точке E_{\min} последовательно (см. рис.5) находим сначала $q_{opt} = 15$ ($\approx 7\%$ – так можно оценить точность исходных данных) из уравнения $E_{\min} = E(q_{opt})$, а затем по «логарифмическим огибающим с постоянными значениями n » (на рис.5 отмечены для удобства тонкими линиями) находим $q_{opt} = n(q_{opt}) = 4$. Такое элементарное пояснение процедуры определения БП можно было дать позднее в разделе 8.3, посвященному графической форме критерия выбора оптимальных базовых параметров ММ полного текста или текста на найденных участках стационарности. Здесь же укажем, что многие полезные зависимости БП от q , как и функционалы этих зависимостей, оптимальные при $q = q_{opt}$ БП по разным критериям (см. Приложение 1) реализуются программно по простым формулам, связанным с алгоритмами восстановления ФТ текстов в дискретном ФП.

Для одновременного определения не только q , n , но и Δt_{opt} изначально заданного аналогового процесса $y(t)$, из которого получен q – значный текст по формулам (2)-(3), указанная процедура повторяется с разными Δt из ожидаемого интервала их значений.

8.2. Критерий 2

Существуют и другие имеющие информационный смысл критерии оптимальности БП текста, эквивалентные критерию (9) на найденном участке стационарности. Например, Критерий 2 мыслимого числа $N_A(q)$ всех Источников текста – абстрактных автоматов, порождающих тексты тождественные экспериментальному (см., например, [14,15])

(19) $E_A(q) \equiv \log_2 \log_2 N_A(q) = n(q) * \log_2 q + \log_2 \log_2 q = E(q) + \log_2 \log_2 q$
с отличным от $E(q)$ значением энтропии, но, как ясно, с точно такими же или весьма близкими значениями БП – q_{opt} , n_{opt} , Δt_{opt} .

8.3. Критерий 3

Это строгий, наглядный, удобный «Графический Критерий 3» – способ определения оптимальных БП следующий непосредственно из пояснений к критериям 1 и 2. При этом следует отметить, что для разных входных последовательностей (см., например, рис.6 с последовательностью y_k ежегодного прироста колец деревьев в средней полосе России [17], рисунки в приложениях) графики энтропии $E(q)$ могут иметь самое различное поведение.

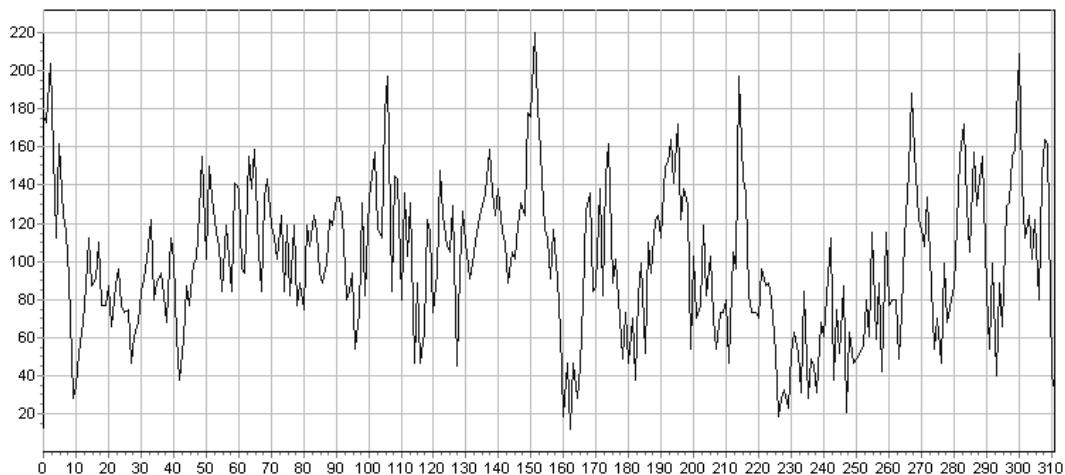


Рис. 6. Входная последовательность ежегодного прироста колец деревьев за $M=312$ лет.

Например, когда график $E(q)$, в отличие от рис.5, содержит сравнительно малое количество точек (см. например, рис.7) с трудно отмечаемыми глазом, но обязательно существующими согласно виду формулы (18) «логарифмическими огибающими постоянных n », БП n_{opt} определяется по формуле (20).

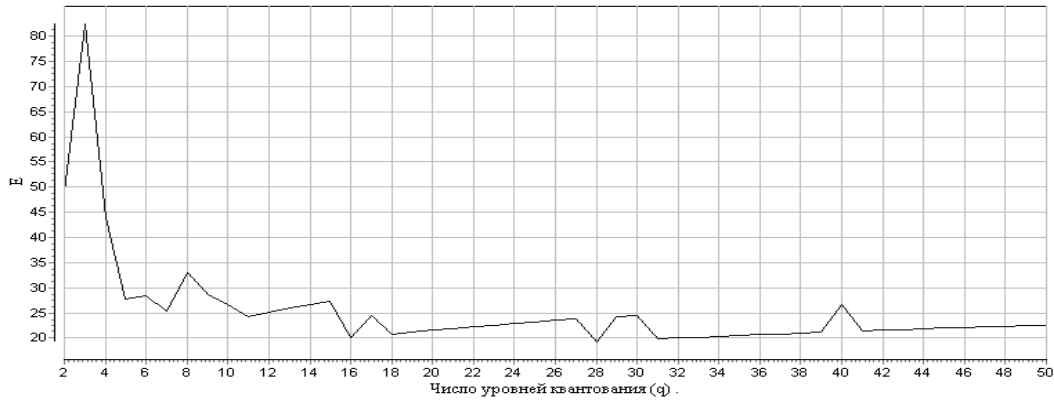


Рис. 7. $E(q) = n(q) \cdot \log_2 q$. $2 \leq q \leq 50$ – Энтропия для последовательности прироста колец деревьев.

$$(20) \quad n_{opt} = E_{\min} / \log_2 q_{opt} = E(q_{opt}) / \log_2 q_{opt}.$$

Действительно (см. значения $E(q_{opt})$ и q_{opt} на рис.7),

$$n_{opt} \approx 19,23 / 4,8 \approx 4.$$

Сделаем еще ряд общих замечаний по поведению $E(q)$, в зависимости от характера выборок y_k самого текста и выбора области поиска q при восстановлении ФТ дискретной ММ Источника конкретной реализации текста.

- Диапазон области поиска $q \in [q_{\min}, q_{\max}]$ следует выбирать с учетом априорной информации либо об уровне зашумленности исходных данных/текста, либо о изначально возможной его квантовой природе.
- При анализе аналоговых процессов типа полигармонических сигналов следует учитывать как возможную точность порождения этих сигналов, так и, дополнительно, соотношение «фаз» выборок текста и гармонических компонент исходного аналогового процесса. Например, при не синхронных, не кратных периоду выборках из точно аналогового синусоидального сигнала можно получить (в полном согласии с теорией разностных уравнений) БП для n отличный от $n = 2$, так как в этом случае текст из выборок y_k может слабо напоминать исходный аналоговый синусоидальный процесс.
- Функция $E(q)$ при «малом» q_{opt} и «большом» n_{opt} имеет, как правило, значительные изменения при $q < q_{opt}$, а при $q > q_{opt}$ до q_{\max} протяженную логарифмически медленно возрастающую часть, что говорит о малоуровневом потоке событий в исходном тексте.
- Функция $E(q)$ при «большом» q_{opt} и «малом» n_{opt} имеет, как правило, значительные изменения (быстрый ступенчатый спад) при $q < q_{opt}$, а при $q > q_{opt}$ до q_{\max} – протяженную часть, изменяющуюся возрастающими логарифмическими ступенями, что говорит о многоуровневом тексте, насыщенном изменениями выборок (например, со следующими друг за другом импульсными, полигармоническими или случайными выборками).

9. Замечания по известным методам определения некоторых БП

Анализ известных ММ Источников данных (разделы 10.1-10.3) показывает, что для них, по-видимому, не существует алгоритмов, позволяющих «одновременно, за одну процедуру» идентифицировать по выборкам данных все перечисленные БП и их функционалы, не прерывая ее дополнительными трудно формализуемыми решениями.

Кроме того, каждый из приводимых далее примеров ММ источников для сравнения с предложенным подходом определения БП, справедлив без дополнительных предположений и ограничений лишь в своей алгебраической структуре, т.е. работает «хорошо» с каким либо одним видом данных:

- аналоговыми непрерывными процессами,
- аналоговыми разрывными процессами,
- аналоговыми дискретными по времени данными,
- квантованными по уровню параметрами и переменными состояния,
- символьными текстами,
- векторными процессами, у которых компонентами являются процессы перечисленного характера,
- последовательностями, «кадров», «сцен», и т.д.), и не фиксирует границы участков стационарности экспериментальных данных.

Далее (раздел 10) рассматриваются примеры работы предложенного подхода, кратко отмечаются характерные особенности и трудности определения некоторых БП для ряда хорошо известных ММ источников данных. Из примеров ясно, что предложенный подход и метод определения БП справедлив и годится при сделанных выше не стеснительных и не обремененных требованиях для всех без исключения перечисленных типов (видов) данных.

После определения БП на участках стационарности РП (СП) на этих участках могут определяться, как уже отмечалось, любым из известных способов [1-5,18-21]. Для этого требуется дополнительный и не связанный с определением БП этап обоснования уже критериев выбора РП (СП) на основе дополнительных априорных сведениях о функционировании Источника экспериментальных данных и целей использования результатов идентификации.

10. Возможности определения Базовых параметров в известных ММ Источников данных

Ввиду важности проблемы выбора ММ источника данных существует много подходов к задаче выбора и определяющей ее актуальность [1-4]. Однако, например, как отмечается в работе [4], характеризующей различные подходы к выбору не всех, а только одного БП – порядка ММ, «выбор порядка модели для данных, получаемых из реальных записей неизвестных процессов, пока еще носит субъективный, а не точный научный характер».

Нами не обнаружено общих методов и алгоритмов, позволяющих «одновременно» идентифицировать по выборкам данных все три перечисленные БП (см., например, разделы 10.1-10.3). Без дополнительных критериев нестационарности они не определяют и четвертый – вектор–параметр границ участков стационарности экспериментальных данных, незнание которых может приводить к неправильным оценкам БП и РП.

10.1. ММ В.А. Котельникова

Ясно, что Δt (в нашей терминологии – это один из БП) большинство исследователей определяют по чрезвычайно широко используемой в теории сигналов и практике проектирования информационно измерительных систем «теореме В.А.Котельникова» [1-6]. Предполагая, что спектр $Y(f)$ процесса $y(t)$ ограничен интервалом

$$(21) \quad Y(f) = \begin{cases} |Y(f) * \exp(2\pi j \varphi(f))|, & \text{при } -f_c \leq f \leq f_c \\ 0, & \text{вне интервала } -f_c \leq f \leq f_c \end{cases}$$

оказывается, что справедливо точное представление процесса $y(t)$

$$(22) \quad y(t) = \sum_{k=-\infty}^{\infty} y_k * u_k(t, f_c),$$

где ортогональные базисные функции и точки t_k расположения выборочных значений даются соотношением (23):

$$(22) \quad u_k(t, f_c) = \frac{\sin(2\pi f_c(t - t_k))}{2\pi f_c(t - t_k)}, \quad t_k = \frac{k\pi}{\omega} = \frac{k}{2f_c}$$

Трудности определения даже одного БП $\Delta t = 1/2f_c$ с помощью теоремы В.А.Котельникова обусловлены сложностью практической проверки условия ее справедливости (21) для конкретных данных $y(t)$ конечной длины T (см. рис. 1): конечность интервала $[0, T]$ и разрывы $y(t)$ на его границах могут приводить к значению $f_c \rightarrow \infty$ и, следовательно, неоправданному измельчению Δt ; Всегда имеющийся в данных аддитивный или мультипликативный «шум» разной природы (погрешности выборок, очень малые по величине, но резкие скачки, неопределенность, вызванная случайной квантовкой и т.д.) также приводит к значению $f_c \rightarrow \infty$ и, также, неоправданному измельчению Δt ; Сглаживание разрывов «окнами» (Хемминга, Ханна, Бартлета, Чебышева и др.[4,5]) может приводить к значительному искажению исходных данных Δy ; Ортонормированность базисных функций (23) не позволяет связать последовательные значения y_k непосредственно в ряде Котельникова (22) и получить динамическую ММ, как это удастся сделать в полигармонической ММ.

10.2. Полигармоническая ММ

Полигармоническая ММ процесса (отрезок ряда Фурье с некрратными частотами) [1-4, 17] часто используется для прогнозирования природного процесса $y(t)$ с погрешностью $|\delta y(t)| \leq \Delta y$:

$$(24) \quad y(t) = \sum_{i=1}^d a_i \cdot \sin(\omega_i t + \psi_i) + \delta y(t), \quad t \in [0, T]$$

зависит от $3d$ РП (СП): по d – частот, амплитуд и фаз. Когда имеются уравнения связи между параметрами, например, когда Источник представляет собой связанные генераторы (и при наличии синхронизации его частей) независимых рабочих параметров может быть меньше чем $3d$.

При значениях $\omega_i = i \cdot \omega_1$ уравнение (14) становится уже отрезком ($2d$) – параметрического ряда Фурье, который при большом позволяет упростить формулу (14), и сокращает число параметров ММ, ускоряя их идентификацию с помощью МНК или других методов оптимизации.

Разделение $3d$ параметров на БП и РП, вообще говоря, зависит от «искусства» исследователя и может быть связана с поиском других ММ эквивалентных форме (14). Переход к эквивалентной ММ может существенно упростить процедуру идентификации, уменьшения числа и нахождения рабочих параметров ММ. Действительно, если взглянуть на уравнение (24) как на решение дифференциального уравнения (напомним, что ищем БП динамических форм ММ), то, как нетрудно показать непосредственными выкладками, эквивалентная статическому уравнению (14) ММ будет иметь вид (25)

$$(25) \quad y^{2d}(t) + (\omega_1^2 + \omega_2^2 + \omega_3^2 + \dots + \omega_d^2) y^{2d-2}(t) + \\ (\omega_1^2 \omega_2^2 + \omega_1^2 \omega_3^2 + \omega_2^2 \omega_3^2 + \dots + \omega_{d-1}^2 \cdot \omega_d^2) y^{(2d-4)}(t) + \\ \dots + (\omega_1^2 \omega_2^2 \omega_3^2 \dots \omega_d^2) y(t) = \varepsilon(t),$$

и содержать уже p параметров (коэффициентов дифференциального уравнения), которые зависят только от частот ω_i , а другие $2p$ параметров (a_i и ψ_i) определяются из (14) как начальные условия (!). Из (25) при p гармониках имеем сравнительно простую оценку одного БП n (теперь уже другого БП, а не Δt , как это было в разделе 10.1) для динамической полигармонической модели:

$$(26) \quad n_{opt} \approx 2 * d + 1.$$

Однако, не смотря на соотношение (26) проблема выбора порядка полигармонической ММ источника данных остается фактически из-за субъективной оценки величины d .

10.3. Другие известные динамические ММ

Известны методы Х.Акаике, Е.Парзена также пригодны для оценки только одного из параметров ММ – размерности, но они ограничены только ММ корреляционно-спектральных процессов [1-4].

Следует отметить, что аналоговый вариант процедуры восстановления ФП непрерывной ДС с выходом в виде скалярного непрерывного временного ряда был предложен Н.Х. Паккардом и уточнен В. Такенсом [19-22] только применительно к гладким, дифференцируемым (!) отображениям, лежащим в основе ММ ДС, имеющим при определенных параметрах режим странного аттрактора. Не останавливаясь здесь на деталях теоремы Такенса заметим, что в ней, как справедливо отмечается, например, в [22], «доказывается только возможность

восстановления ФТ аттрактора размерности d непрерывной ДС по $n \geq (2d+1)$ временным выборкам с шагом Δt одной из координат и «не предлагается эффективной практически реализации».

Теорема Такенса не накладывает никаких ограничений на величину шага Δt и на величину допустимой погрешности выборок Δy из непрерывного (!) процесса, что должно приводить к ошибкам определения размерности d непрерывной, даже не очень сложной ДС [22]. Эти обстоятельства, по-видимому, стимулировали создание метода оценки размерности аттрактора, использующего процесс усреднения реализаций данных путем вычисления так называемого «корреляционного интеграла» П.Грасбергера и И.Прокаччо, который эффективно используется преимущественно для вычисления размерности источника только для весьма представительного массива данных [19-22].

Несмотря на некоторое внешнее сходство рассматриваемой нами процедуры восстановления ФТ в ФП ММ дискретных по уровню и времени ДС с процедурой изложенной в работе Пакарда-Такенса справедливой для «гладких, дифференцируемых» аналоговых ДС, отметим их существенные отличия, которые благодаря принятым в разделе 4 «Положениям» позволили реализовать принципиально новые ее свойства и возможности:

- использования процедуры как для исходных непрерывных ($q \gg 2$), дискретных ($q \geq 2$) данных, так и для символьных текстов и т.п. [10,15,26];
- введения и нахождения оптимальных БП произвольного отрезка текста по энтропийным и другим, родственным им, критериям [11-15,24];
- введения понятия ГК данных, являющейся по сути дела ММ нестационарного текста (процесса), и интерпретации участков генов, как интервалов стационарной, бесперебойной работы Источников между их «переключениями», «разладками», «восстановлениями» и т.п. [10,15,25];
- проведения предварительной идентификации текста (процесса) путем определения БП ММ ГЕНераторов участков, без какой бы то не было конкретной априорной информации о генераторах и возможностью последующего произвольного выбора вида ММ Источника (авторегрессионной, полигармонической и т.д.) в зависимости от целей дальнейших этапов идентификации остальных параметров ММ, зависящих от базовых [11-15];
- оценки случайности или детерминированности текста (процесса) сравнением гистограммы длин участков генов с теоретическим распределением длин участков у абсолютно случайного текста [11-15];
- удобной оценки влияния не только изменений текста, но и других его мутаций (удалений, вставок) на ГК [10, 15, 27];

и т.д.

Отмеченные особенности БП, основанные на ММ ГК, позволили использовать их в самых различных областях знаний и приложениях: идентификации, теории надежности, генетике и т.д. и т.п. (Приложения 1, 2).

11. Выводы

Возможность расчленения задачи на части в науке и практике, как правило, приводит либо к принципиальной возможности ее решения, либо к существенному упрощению решения такой задачи (например, при разделении уравнений в

частных производных на части, зависящие только от координат и только от времени).

Предложенный метод расчленения задачи идентификации на предварительную стадию (стадию нахождения основных, базовых параметров), справедливую при любой ММ Источника экспериментальных данных, и возможность отделения этой стадии от последующих стадий, связанных с принципиальной необходимостью использования дополнительной априорной информации о ММ Источника данных с возможностью нахождения по базовым параметрам всех остальных («свободных», «рабочих») параметров ММ Источника данных имеет большое значение в любой конкретной теоретической и прикладной областях исследований.

Работа выполнена при поддержке РФФИ (грант 02-02-17573).

Автор от имени своих коллег и себя лично благодарит организаторов и участников семинаров радиофизического факультета ННГУ, факультета ИСТ НГТУ и ВВО МАТН РФ и других организаций Баранова В.Г., Грязнова М.И., Гурбатова С.Н., Ивенина А.Я., Кислякова А.Г., Малахова А.Н., Маркова А.А., Мироновского Л.А., Моругина С.Л., Орлова И.Я., Роткова Л.Ю., Рябинина Ю.А., Ульянова А.А., Шалфеева В.Д., Хранилова В.П., Якимова А.В. за проведенные в разное время полезные обсуждения и поддержку работ в области идентификации базовых параметров математических моделей Источников экспериментальных данных по их генетическим картам.

Автор также благодарит своих коллег Багрова С.А., Вершинина П.А., Лебедева О.Л., Семечукова И.В., Таланова В.А., Исева Р.Р., Петрунина Д.А., Рычкова В.А., Шабельникова А.В. и др., с которыми ему посчастливилось в разное время плодотворно сотрудничать.

Приложения

Краткое изложение материала работы не позволяет остановиться на многих дополнительных весьма интересных возможностях метода БП, а также его теоретических и практических приложениях. Поэтому далее кратко рассматриваются:

- полезные функционалы БП скалярных и векторных исходных данных, реализованные в программных системах (например, в [26]) с удобными форматами встраиваемых в тексты статей рисунков и графиков (Приложение 1),
- некоторые способы исследования БП аналоговых динамических систем в прикладных задачах идентификации и моделировании (Приложение 2) с помощью упомянутого в разделе 3 обобщения процедур определения БП на векторные процессы, так как исследование дискретных динамических систем с помощью упомянутых в Приложении 1 функционалов БП дискретных ДС не встречает особых затруднений.

Приложение 1. Полезные функционалы БП скалярных и векторных исходных данных [11,15, 26]

№	Функции $q \in [q_{min}, q_m]$	Название функции	Формула при $q=q_{орб}$ соответствующая минимуму $E(q)$ и $Ea(q)$, Сокращения $\log_2 q = \log q$, $\log_e q = \ln q$
1.	N	Число точек в ФП	$= q^{n(q)}$
2.	E	Энтропия– Критерий 1	$= \log N(q) \equiv \log q^{n(q)} = n(q) \log q$,
3.	Ea	Энтропия– Критерий 2	$= \log \log Na(q) \equiv E(q) + \log(\log q)$,
4.	n	Размерность вложения данных	$= E(q) / \log_2 q = \ln f$
5.	f_c	Верхняя частота спектра данных	$= 1/2 \cdot \Delta t = n/2 \cdot \tau_n$
6.	τ_n	Характеристическое время изменения данных / текста	$= n \cdot \Delta t$
7.	$\ln f$	Информационная емкость данных	$= 2 \cdot \tau_n \cdot f_c = n$
8.	Δt	Шаг дискретизации по времени	$= \tau_n / n$
9.	E_M	Энтропия на символ текста	$= E(q) / M = n(q) \cdot \log q / M$
10.	$E_{\log M}$	Энтропия на логарифм длины всего текста	$= E(q) / \log M = n(q) \cdot \log q / \log M = n(q) / \log_q M =$ $n(q) \cdot \log_M q = n(q) \cdot H(q)$
11.	H	параметра Херста (дискретный аналог)	$= \log_2 q / \log_2 M$ $= 1 / \log_q M = \log_M q$
12.	R	Избыточность на символ текста. $\ln 2 \cdot R(q) = 0,693 \cdot R(q) \approx E_k(q) = \lambda_l = \ln q$ – дискретный аналог энтропии Колмогорова, λ_l - тах показатель Ляпунова	$= 1/2 \cdot \log q = \log(q)^{1/2}$
13.	Q	Избыточность алфавита уровней текста	$= q_{max} / q$
14.	N_2	Число точек в ФП (зависимые выборки)	$= q_m \cdot q^{n(q) \cdot -1} =$ $(q_m/q) \cdot N(q) = Q(q) \cdot N(q)$
15.	m_2	Средняя длина гена (зависимые выборки)	$= [N_2(q)]^{1/2}$
16.	m_{Av}	Средняя длина гена с учетом $n(q)$	$= m_2(q) + n(q)$
17.	I	Число участков генов с учетом $n(q)$	$= M / m_{Av}(q)$
18.	D	Потенциальная предсказательная способность на участке гена	$= m_{Av}(q) / n(q)$
19.	dn/dq	Коэффициент пересчета БП q в n	$= - (n-1) / q \ln q$ $= (n-1) \cdot dn/dq_{max}$
20.	$\ln m_2$	«порог и наклон» – характеризуют: 1) избыточность $Q(q)$ уровней, 2) размерность вложения $n(q)/2$	$= \ln Q^{1/2}(q) + n(q)/2 \cdot \ln q + \ln [1 + n(q)/N_2]$
21.	n_f	Дробная размерность ФП	$= [1/2(n(q)-1) + 1/m_2(q)] \bullet [1 + n(q)/m_2(q)] =$ $= [1/2(n(q)-1) + 1/m_2(q)] \bullet [1 + (D)^{-1}]$
22.	E_K	Энтропия Колмогорова	$= [\ln 2 \cdot R(q) + 1/m_2(q)] \bullet [1 + n(q)/m_2(q)]$, $E_k 2(q) \approx [\ln 2 \cdot R(q) + \dots] \bullet [1 + \dots]$, $\ln 2 = 0,693$
23.	n_{ph}	Фоновая размерность ФП	$= 1/2 [1 + n(q)/m_2(q)] \approx 1/2$, при $n(q) \ll m_2(q)$
24.	$(n_f), (n), (2 \cdot n_f + 1)$	Соотношение размерностей ФП - дискретный аналог неравенства Манэ	$n_f(q) \leq n(q) \leq [2 \cdot n_f(q) + 1]$

Приложение 2. Векторные процессы [16]

Изучение поведения аналоговых динамических систем с помощью использования обобщенных процедур определения БП на векторные процессы (наборы файлов данных, интерпретируемых как компоненты совокупной ДС) можно изучать с определения БП каждой из предполагаемых компонент. Например, рассмотрим в качестве Источника известную ММ «аттрактора Э. Лоренца» [19-22] в стандартной форме

$$(1-П2) \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \cdot \begin{pmatrix} x \cdot y \\ x \cdot z \\ y \cdot z \end{pmatrix},$$

где СП ММ – это коэффициенты $b_{11} = -b_{12} = -a$, $b_{21} = -b$, $b_{22} = -1$, $b_{33} = -c$, $c_{22} = -1$, $c_{31} = 1$, а остальные – нули. Отметим, что в этом случае мы можем иметь очень точные «экспериментальные данные» $x(t)$, $y(t)$, $z(t)$ с достаточно малым шагом Δt и большим количеством знаков заранее вычислив их на ЭВМ при стандартных параметрах и начальных условиях [19], например, как это показано на рис.1-П2.

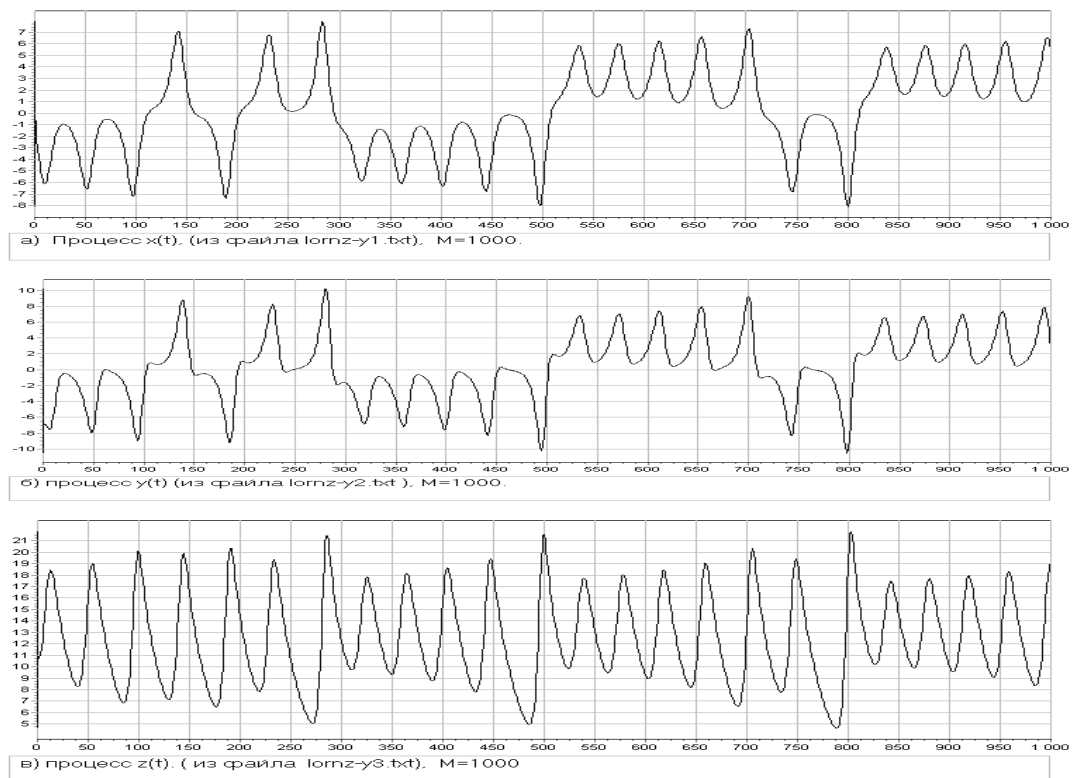


Рис.1-П2. Процессы $x(t)$, $y(t)$, $z(t)$ (фрагменты а), б) и в) рисунка) как компоненты вектора совокупного состояния $C(t) = (x(t), y(t), z(t))$ аттрактора Э. Лоренца.

Результаты определения БП реализаций каждой из компонент, взятой по отдельности, а также вектора совокупного состояния аттрактора $C(t) = (x(t), y(t), z(t))$, представлены на рис.2-П2 с областью поиска $q \in [2-1600]$.

Из фрагментов а)-в) этого рисунка следует, что точностей использованных данных и алгоритмов вычисления БП (Приложение 1) оказалось достаточно, чтобы получить следующие оценки БП каждой из компонент: $q_{opt} \approx 1200-1500$, $n_{opt} = 3$. Значение оценки q_{opt} соответствует точности лучше 0,1%, а значение n_{opt} указывает на возможность определения правильного порядка ДС по одной реализации. Из фрагмента г) рис.2-П2 следует, что набор $(x(t), y(t), z(t))$ из трех компонент образует состояние исследуемой ДС, так как $n_{opt} = 1$. Поэтому исследуемая ДС при моделировании может по начальной тройке выдавать в каждый следующий момент времени вектора из троек компонент совокупного состояния с оцененной по значению БП точностью $\approx 0,3\%$ ($q_{opt} \approx 311$).

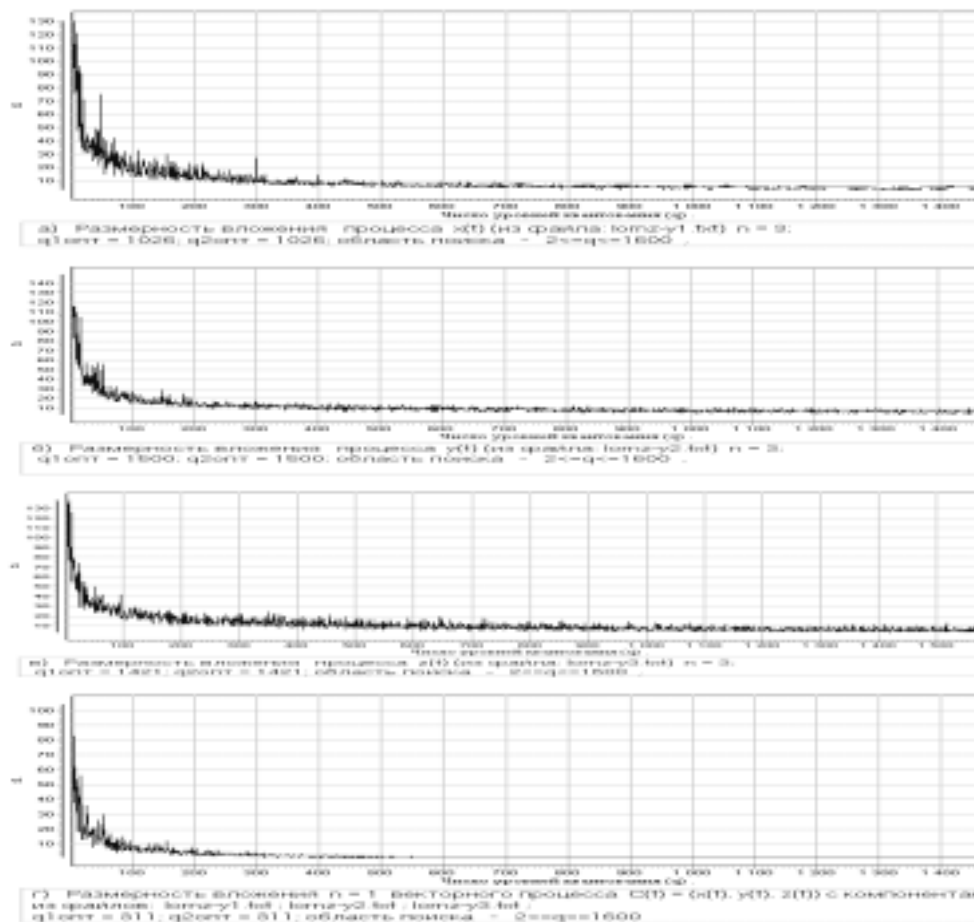


Рис. 2-П2. БП компонент состояния $x(t), y(t), z(t)$, как отдельных независимых процессов (фрагменты а), б) и в) рисунка), а также БП совокупного состояния $C(t) = (x(t), y(t), z(t))$ (фрагмент г).

На рис.3-П2 и рис.4-П2 представлены, соответственно, ГК скалярного процесса $x(t)$ (рис.3-П2) и векторного процесса $C(t) = (x(t), y(t), z(t))$ (рис.4-П2) для странного аттрактора Лоренца при ряде значений пар (q, n) , задающих «разрешающую способность» ГК установкой средней длины гена $m_{Av} \cong N^{1/2} = q^{n/2}$

(см. свойства распределения $p_T(m)$, раздел 7) или, что тоже, параметров $I > 0$ и J в соотношении (6). Для выбора n по q из (6) и (15) имеем связь

$$(2\text{-П}2) \quad n^* \equiv n(I, J, q) \equiv 2 \cdot \lceil \log_q (M/2I)(1 - J/M) \rceil, I > 0.$$

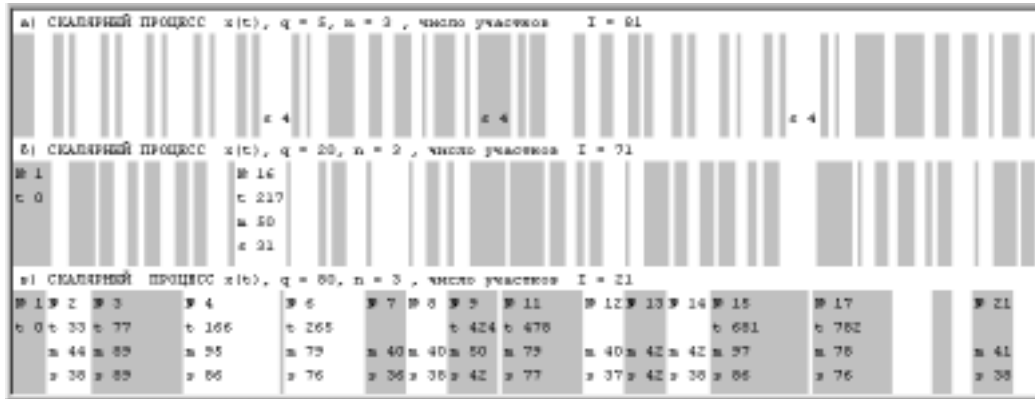


Рис. 3-П2. Участки ГК (чередующиеся светлые и темные полосы) скалярного процесса $x(t)$ для странного аттрактора Лоренца при значениях пар (q, n) – (5,3), (20,3), (80,3) (фрагменты а), б), в) соответственно).

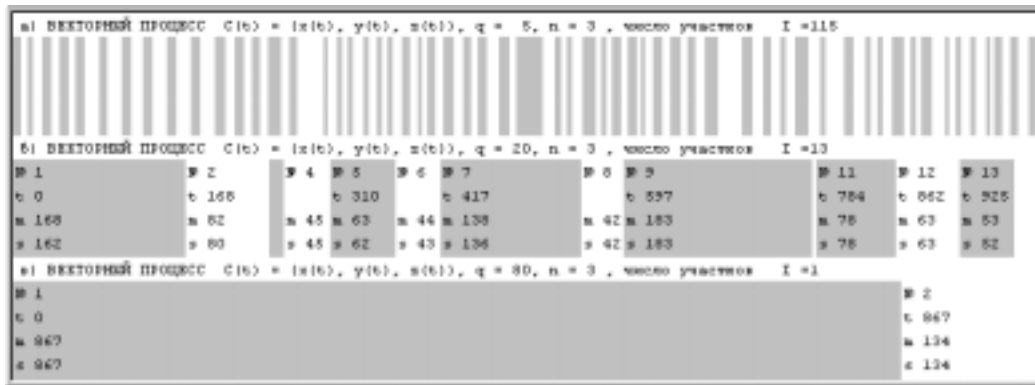


Рис.4-П2. ГК векторного процесса $C(t) = (x(t), y(t), z(t))$ для странного аттрактора Лоренца при значениях пар (q, n) – (5,3), (20,3) и (80,3) (фрагменты а), б), в) соответственно).

Подбор параметров $m_{Av} \cong N^{1/2} = q^{n/2}$ позволяет выделять однородные участки текста. Выделение эффективнее, когда m_{Av} меньше длины однородности. Для ДС « σ -го порядка» параметр q переходит в $q \cdot \sigma$, а среднее значение m_{Av} возрастает примерно в $\sigma^{n/2}$ раз.

После оценки БП для ДУ третьего порядка типа (1-П2) (уравнения Лоренца, Росслера [19-21], динамики «среднего класса» и т.п.) в качестве СП иногда удобно взять коэффициентами уравнения нелинейной регрессии одной из компонент состояния [28] и оценить их методом МНК или МП.

Список литературы

1. Эйкхофф П. Основы идентификации систем управления. М.: Мир, 1975. Райбман Н.С. Методы нелинейной и минимаксной идентификации // Современные методы идентификации систем / Под ред. П. Эйкхоффа. М.: Мир, 1983. С. 177-277.
2. Льюинг. Идентификация систем. Теория для пользователя. М.: Наука, 1991.
3. Гроп Д. Методы идентификации систем. М.: Мир, 1979.
4. Марпл. С.Л.-мл. Цифровой спектральный анализ и его приложения. М.: Мир, 1990. 584 с.; Key S.M., Marple S.L. Modern methods of the spectral analysis. Review // Proceedings of the IEEE. 1981. Vol. 69, No 11. P. 5.
5. Корн Г., Корн Т.. Справочник по математике (для научных работников и инженеров). М.: Наука, 1973. 831 с.
6. Баскаков С.И. Радиотехнические цепи и сигналы. М.: Высшая школа., 1983. 536 с.
7. Кирьянов К.Г. «Смысловая» синхронизация (привязка) потоков данных // Тезисы докладов V Всесоюзного симпозиума «Нано- и пикосекундная импульсная техника и ее применение в радиоизмерениях». Горький, 1983. С. 82-83.
8. Кирьянов К.Г. «Смысловая» синхронизация (привязка) потоков данных // Техника средств связи. Сер. РИТ, М., ЦОНИ «Экос», 1985. Вып. 2. С. 1-20.
9. Кирьянов К.Г. Идентификация и диагностирование цифровых и символьных последовательностей данных. Результаты экспериментов с последовательностями с технических и биологических объектов // В сб. «Тезисы докладов IV Всесоюзного симпозиума «Методы теории идентификации в задачах измерительной техники и метрологии». Новосибирск, 1985, С. 247.
10. Кирьянов К.Г. Диагностирование последовательностей данных по их генетическим картам. // В сб. «Комбинаторно-алгебраические методы и их применение». Межвузовский сборник. Горький: ГГУ, 1987. С. 40-46.
11. Кирьянов К.Г. Исследование сложных объектов и процессов по их генетическим картам. Синергетические измерения (ч. 1) // Техника средств связи. Серия РТ. М.: ЭКОС, 1991. Вып. 4. С. 45-78.
12. Кирьянов К.Г., Китаева Е.К., Разумкова Е.Н. Выбор основных параметров математической модели источника экспериментальных данных методом минимальной энтропии // Вестник ВВО АТН РФ. Серия «Высокие технологии в радиоэлектронике». Н. Новгород, 1995. Вып. 1. С. 101-105.
13. Vershinin P.A., Rasumkova E.N., Kiryanov K.G. About the identification and analysis of analogue non-stationary processes // Theses of reports of the Second International Scientific School-Seminar "Dynamic and stochastic wave phenomena". N. Novgorod, 1994. P. 116.
14. Kiryanov K.G. To a choice of the basic parameters of mathematical model of experimental data source // In: Proceedings of the 5th International Specialist Workshop in area Nonlinear Dynamics of Electronic Systems (NDES '97). Moscow, Russia, 26-27 June, 1997. P. 400-403.
15. Кирьянов К.Г. Генетический код и тексты: динамические и информационные модели сложных систем. Нижний Новгород: ТАЛАМ, 2002. 100 с.
16. К.Г. Кирьянов. Идентификация базовых параметров динамики «сцен» // Четвертая научная конференция по радиофизике. 7 мая 2000 г. Н. Новгород: Талам, 2000. С. 176-177.
17. Кирьянов К.Г., Шабельников А.В. Новый метод прогноза временных изменений природных процессов // Радиотехника и электроника. 1995. Вып. 5. С. 752-762.
18. Колмогоров А.Н. Теория информации и теория алгоритмов. М.: Наука, 1987. 304 с.
19. Shuster G. Determined chaos. Physik-Verlag GmbH, Weinheim (F.R.G.), 1984 / Шустер Г. Детерминированный хаос. М.: Мир, 1998. 240 с.
20. Рабинович М.И., Трубецков Д.И. Введение в теорию колебаний и волн. НИЦ «Регулярная и хаотическая динамика». 2000. 560 с.
21. Неймарк Ю.И., Ланда П.С.. Стохастические и хаотические колебания. М.: Наука, 1987.
22. Анищенко В.С. Знакомство с нелинейной динамикой. Саратов: Изд. ГУНЦ «Колледж», 2000. 180 с.
23. Яблонский С.В. Введение в дискретную математику. М.: Наука, 1979.
24. Кирьянов К.Г. Измерение динамической сложности процессов по их генетическим картам // Сб. «Развитие и внедрение новой техники радиоприемных устройств и обработки сигналов». М.: Радио и связь. 1989. С. 100.

25. Кирьянов К.Г. Вычисление характеристик надежности сложных систем на основе новых математических моделей отказов // Системы обработки информации и управления. Межвузовский сборник научных трудов. Вып. 5. Н. Новгород: Изд-во НГТУ, 1999. С. 23-31.
26. Кирьянов К.Г., Исаев Р.Р. Программная система анализа основных характеристик математических моделей источников экспериментальных данных // Труды научной конференции по радиофизике. Н. Новгород, Изд-во ННГУ, 1998. С. 53.
27. Кирьянов К.Г., Лебедев О.Л., Грэн Э.Я. «Применение алгоритмов для аппаратуры диагностирования и картирования кольцевых последовательностей данных» // Техника средств связи. Сер. РИТ. М.: ЦОНТИ «Экос», 1986. Вып. 2. С. 34-45.
28. Манцеров С.А. Нелинейные АР-модели источников экспериментальных данных для прогнозирования // В кн.: Тр. 5-й научн. конф. по радиофизике. 7 мая 2001 г. / Ред. А.В. Якимов. Н. Новгород: ННГУ, 2001. С. 327.