

УДК 621.391

П.Ю. Костенко, К.С. Васюта, А.Н. Барсуков, С.Н. Симоненко

Харьковский университет Воздушных Сил им. И. Кожедуба, Харьков

## НОВЫЙ ПОДХОД К НЕПАРАМЕТРИЧЕСКОМУ ОБНАРУЖЕНИЮ ХАОТИЧЕСКИХ СИГНАЛОВ НА ФОНЕ БЕЛОГО ШУМА С ИСПОЛЬЗОВАНИЕМ "НЕЛИНЕЙНОЙ ДИНАМИЧЕСКОЙ СТАТИСТИКИ"

В статье выясняются причины недостаточной эффективности энергетического подхода к обнаружению сигналов неизвестной "формы", основанного на использовании  $\chi^2$ -статистики. Формализуется понятие "форма" сигнала с помощью корреляционной размерности и связанной с ней BDS-статистики. Анализируется подход к непараметрическому обнаружению хаотического сигнала на фоне белого шума с неизвестной плотностью распределения его значений с использованием BDS-статистики. Предлагаются алгоритм обнаружения хаотического сигнала, структурная схема его реализации, а также ATS-алгоритм формирования суррогатных сигналов, которые используются для эмпирической оценки отношения правдоподобия.

**Ключевые слова:** непараметрический обнаружитель, BDS-статистика, ATS-алгоритм формирования суррогатных сигналов, "форма" сигнал.

### Введение

Последние два десятилетия активно ведутся исследования методов передачи информации с использованием достижений хаотической динамики [1 – 3] в надежде с их помощью решить вопросы защищенности систем передачи информации (СПИ) от несанкционированного доступа и их работы "под шум" для обеспечения скрытности функционирования. При этом предполагалось, что повышение скрытности таких систем может быть достигнуто благодаря визуальному сходству хаотических и случайных процессов, а также их практической неразличимости в рамках корреляционного и спектрального анализа. Не в последнюю очередь оптимизм в отношении разведзащищенности таких СПИ был связан с отождествлением свойств хаотических и шумоподобных сигналов для которых установлена их высокая энергетическая скрытность. Тем не менее, оказалось, что многие системы связи, использующие хаотические сигналы, характеризуются в действительности ограниченной скрытностью и конфиденциальностью. Это обстоятельство объясняется тем, что динамические системы, операторы эволюции которых задаются дифференциальными уравнениями или точечными отображениями (поточковые и каскадные), используемые для формирования хаотических сигналов, генерируют "низкоразмерный" хаос с просто структурированным аттрактором, нелинейное строение которого нетрудно обнаруживается, так как он легко отличим от аттракторов соответствующих случайным процессам с независимыми значениями. Учет этого свойства хаотических процессов может скомпрометировать скрытность функционирования СПИ, рассчитанную с использованием энергетического критерия, который приводит к завышенной ее оценке [4]. Дело в том, что энергетический критерий обнаружения не использует информацию о "форме" процесса или зависимости его значений. В результате можно говорить о снижении статуса скрытности хаотических процессов.

Для учета "формы" сигнала при его обнаружении в работе [5] было предложено использовать BDS-статистику наблюдаемого процесса, которая связана с его корреляционной размерностью. В результате сравнительного анализа эффективности обнаружения сигнала с использованием BDS-статистики и традиционной  $\chi^2$ -статистики, учитывающей только энергетические признаки сигнала, было показано, что во многих случаях  $\chi^2$ -статистика дает заниженную вероятность обнаружения не только хаотических, но и шумоподобных сигналов и, как следствие, неоправданно оптимистичную его скрытность.

**Целью данной работы** является выяснение причины недостаточной эффективности энергетического подхода к обнаружению сигналов неизвестной "формы". Формализация понятия "форма" сигнала с помощью корреляционной размерности и связанной с ней BDS-статистикой, а также представление одного из вариантов структурной схемы непараметрического обнаружителя хаотических сигналов, реализация которого предполагает эмпирическую оценку отношения правдоподобия по суррогатным сигналам, сохраняющим свойства аттрактора наблюдаемого хаотического сигнала. На наш взгляд, эти вопросы рассматриваются впервые и могут представлять интерес для специалистов по обработке сигналов.

### Основной раздел

В дальнейшем под скрытностью будем понимать [4] способность противостоять мерам радиотехнической разведки: обнаружению сигнала и определению его структуры на основе оценки ряда его параметров без учета возможности раскрытия смысла информации. В качестве критерия скрытности можно принять величину  $p_{\text{скр}} = 1 - p_p$ , которая определяется вероятностью разведки  $p_p = p_{\text{обн}} p_{\text{стр}}$ , характеризующей вероятностью правильного обнаружения сигнала  $p_{\text{обн}}$  и вероятностью раскрытия его структуры  $p_{\text{стр}}$ .

Традиционно для обнаружения детерминированных сигналов с неизвестной “формой” и случайных процессов используют энергетические обнаружители (радиометры), базирующиеся на  $\chi^2$ -статистике. В качестве признака процесса используется его энергия. Фактически оптимальный обнаружитель представляет собой измеритель мощности процесса, позволяющий выявлять энергетические приращения над мощностью шумов при наличии сигнала в анализируемом диапазоне частот. В тоже время понятие “форма” процесса можно рассматривать как лингвистическую характеристику, которую можно формализовать, пользуясь, например следующей цепочкой: “форма” процесса  $\rightarrow$  структурированность аттрактора процесса  $\rightarrow$  зависимость значений процесса  $\rightarrow$  критерий зависимости (динамический или статистический)  $\rightarrow$  мера зависимости (например, динамические инварианты: показатели Ляпунова, корреляционная размерность или энтропия). Корреляционной размерностью можно характеризовать структурированность аттрактора, связанного с анализируемым процессом. Например, случайный I.I.D (independent and identically distributed) процесс неструктурирован, его аттрактор полностью “заполняет” пространство вложения, а корреляционная размерность совпадает с его размерностью. Особенностью аттрактора хаотического процесса является его структурированность, недостаточное “заполнение” пространства вложения, а также насыщение корреляционной размерности по мере увеличения размерности пространства вложения. В результате значение корреляционной размерности оказывается меньше размерности пространства вложения. Различия в “наполняемости” фазового пространства аттракторами случайного и хаотического процессов и, как следствие, в зависимостях корреляционной размерности от размерности пространства вложения подсказывает один из способов классификации случайных и хаотических процессов, а также решения задачи обнаружения хаотического сигнала на фоне шума.

Проявление структурированности (образование кластеров) аттрактора указывает на наличие взаимосвязей элементов наблюдаемого процесса. В такой интерпретации классы процессов (случайные, хаотические, регулярные) можно “метризовать”. При необходимости возможно дальнейшее расслоение упомянутых классов на подклассы (например, линейные и нелинейные, стационарные и нестационарные) и с помощью числовых показателей упорядочивание их имен, т.е. шкалирование. Следовательно, лингвистическую переменную “форма” процесса можно формализовать и рассматривать её как более информативный признак, чем энергия сигнала. Упомянутые классы и их расслоения будут иметь различную “форму”, которую удобно характеризовать с помощью числовых показателей их корреляционных размерностей. Таким образом “форма” процесса отражает ограничения на возможные взаимосвязи его значений, которые в частном случае детерминированных сигналов могут быть определены для каждого значе-

ния параметра известными функциональными зависимостями. В теории сигналов “форму” некоторых из них отождествляют с геометрическими образами, представленными графиками функций: гармонических, импульсных и других.

Метризуем упомянутые классы процессов, отличающиеся своей “формой”, с использованием шкалы корреляционной размерности и связанной с ней BDS-статистикой.

BDS-статистика и построенная на ее основе относительно новая процедура – BDS-тест были предложены в результате анализа финансовых рынков экономистами Броком, Дечертом и Шейнкманом (B. Brock, W. Dechert и J. Scheinkman) в 1987 [6] и представляют один из мощных методов выявления зависимостей во временных рядах, интенсивно разрабатываемых в последнее десятилетие в рамках их нелинейного анализа.

Его цель состоит в том, чтобы различить данные I.I.D. и любой вид зависимости – проверить нулевую гипотезу  $H_0$  о независимости и тождественном распределении значений временного ряда  $\xi = (\xi_1, \xi_2, \dots, \xi_N)$ , используя для этого критерий значимости. Согласно этому критерию для принятия гипотезы  $H_0$  необходимо выбрать критическую область  $G_\alpha$ , удовлетворяющую условию  $P(g \in G) = \alpha$ , где  $g(\xi_1, \xi_2, \dots, \xi_N)$  – статистика наблюдения, а  $\alpha$  – устанавливаемый уровень значимости. Отклонения от нулевой гипотезы могут быть обнаружены и с помощью вычислений других нелинейных статистик, для которых критическую область  $G_\alpha$  можно выбирать разными способами. Поэтому следует выбирать  $g$  и  $G$  в некотором смысле наилучшими из возможных, а именно такими, чтобы полученный критерий был наиболее мощным. Заметим, что BDS-тест, может выявлять и нелинейность (нелинейный детерминизм), при условии, что любая линейная зависимость была удалена из данных (например, используя традиционные модели типа ARIMA или взятие первой разности натуральных логарифмов). Однако не может тестировать хаос непосредственно [6]. Поэтому, с помощью непараметрической BDS-теста можно обнаружить один из признаков хаоса – нелинейность. В тоже время для анализа свойств нелинейности временного ряда были предложены и другие статистические тесты, обладающие большей мощностью различения [7,8]. Однако эти тесты (измерители) нелинейности в одних случаях дают хорошую характеристику различения (дискриминации), но сбоят в других. Если мы фактически ищем детерминированную структуру в сигнале, то простые статистики, которые основаны на кумулянтах высших порядков, не очень привлекательны, потому что они также весьма чувствительны к тем отклонениям от нулевой гипотезы, которые мы не ищем. В таком случае мы выберем дискриминационную статистику  $g$  соответственно, из арсенала методов динамических систем, т.е. BDS-статистику.

При такой постановке задачи различения конкурирующая гипотеза  $H_1$  оказывается сложной, а ошибка второго рода  $\beta$  неизвестна, поэтому она в расчет не принимается, а гипотеза  $H_0$  отвергается, если выборка такова, что когда справедлива эта гипотеза, получение подобной выборки маловероятно. Иными словами, гипотеза  $H_0$  отвергается, если статистика выборки попадает в критическую область вероятность попадания в которую (уровень значимости  $\alpha$ ) мала, когда справедлива проверяемая гипотеза. Следует отметить, что обе гипотезы относятся к разным семействам гипотез.

BDS-тест основан на статистической величине  $w(\bar{\xi})$  (BDS-статистике)

$$w_{m,N}(\varepsilon) = \sqrt{N-m+1} \frac{C_{m,N}(\varepsilon) - C_{1,N-m}(\varepsilon)^m}{\sigma_{m,N}(\varepsilon)}. \quad (1)$$

В работе [9] были предложены очень быстрые алгоритмы для её оценки. Числитель BDS-статистики определяется корреляционными интегралами  $C_{m,N}(\varepsilon)$ ,  $C_{1,N}(\varepsilon)$ , а знаменатель среднеквадратическим отклонением  $\sigma_{m,N}(\varepsilon)$  числителя. Для вычисления  $C_{m,N}(\varepsilon)$  ( $m > 1$ ) необходимо выполнить «вложение» временного ряда в  $m$ -мерное псевдофазовое пространство, элементами которого, на основании теоремы Такенса (Takens) [8], являются точки  $\xi_i^m = (\xi_i, \xi_{i+1}, \dots, \xi_{i+m})$  с координатами  $\{\xi_{i+k}\}_{k=1}^m$  заданными  $m$  последовательными значениями исходного временного ряда. Корреляционный интеграл определяет частоту попадания произвольной пары точек фазового пространства в гиперсферы радиуса  $\varepsilon$ :

$$C_{m,N}(\varepsilon) = \frac{2}{(N-m+1)(N-m)} \sum_{s=m}^N \sum_{t=s+1}^N \prod_{j=0}^{m-1} I_\varepsilon(\xi_{s-j}^m, \xi_{t-j}^m);$$

$$I_\varepsilon(\xi_i^m, \xi_j^m) = \begin{cases} 1, & \|\xi_i^m - \xi_j^m\| \leq \varepsilon \\ 0, & \|\xi_i^m - \xi_j^m\| > \varepsilon \end{cases}, \quad (2)$$

в котором  $I_\varepsilon(\xi_i^m, \xi_j^m)$  – функция Хевисайда для всех пар значений  $i$  и  $j$ , где  $0 \leq i \leq N$  и  $0 \leq j \leq N$ ;  $N$  – число элементов временного ряда  $\{\xi_i\}_{i=1}^N$ . Его значение стремится к определенному пределу по мере уменьшения  $\varepsilon$ . Рекомендуются выбирать  $\varepsilon$  таким, что  $\varepsilon = 0.5\sigma \div 2\sigma$ , где  $\sigma$  – среднеквадратическое отклонение процесса  $\{\xi_i\}_{i=1}^N$ . В соответствии с теорией, зависимость корреляционного интеграла от  $\varepsilon$  имеет степенной вид  $C_{m,N}(\varepsilon) \sim \varepsilon^{D_c}$ , где  $D_c$  – корреляционная размерность временного ряда. Для  $m = 1$  имеем:

$$C_{1,N}(\varepsilon) = \frac{2}{N(N-1)} \sum_{s=1}^N \sum_{t=s+1}^N I_\varepsilon(\xi_s, \xi_t). \quad (3)$$

Брок и др. показали, что  $C_{m,N}(\varepsilon) \Rightarrow C_{1,N}(\varepsilon)^m$  с 100% вероятностью при  $N \rightarrow \infty$ , а  $(C_{m,N}(\varepsilon) - (C_{1,N}(\varepsilon))^m) \cdot \sqrt{N-m+1}$  является случайной асимптотически нормально распределенной величиной с нулевым средним и среднеквадратическим отклонением  $\sigma_{m,N}(\varepsilon)$ , которое определяется как:

$$\sigma_{m,N}(\varepsilon) = 2 \sqrt{k^m + 2 \sum_{j=1}^{m-1} k^{m-j} \cdot (C_{1,N}(\varepsilon))^{2j} + (m-1)^2 \times} \\ \times \sqrt{(C_{1,N}(\varepsilon))^{2m} - m^2 k (C_{1,N}(\varepsilon))^{2m-2}}, \quad (4)$$

где  $k = \frac{1}{(N-1)(N-2)N} \left[ \sum_{t=1}^N \left[ \sum_{s=1}^N I_\varepsilon(\xi_t, \xi_s) \right]^2 - \right. \\ \left. - 3 \sum_{s=1}^N \sum_{t=s+1}^N I_\varepsilon(\xi_t, \xi_s) + 2N \right].$

BDS-статистика  $w(\bar{\xi})$  является нормально распределенной случайной величиной при условии, что оценка  $\hat{\sigma}_{m,N}(\varepsilon)$  близка к ее теоретическому значению  $\sigma_{m,N}(\varepsilon)$  [6].

Задача обнаружения хаотического сигнала рассматривается как непараметрическая проверка одной из двух гипотез:  $H_0$  – наблюдаемые данные  $\bar{\xi} = (\xi_1, \xi_2, \dots, \xi_N)$  независимы и одинаково распределены (белый шум) т.е. плотность (функция) распределения факторизуется  $F_N(\xi_1, \xi_2, \dots, \xi_N) = \prod_{i=1}^N F(\xi_i)$  и  $H_1$  – данные не I.I.D., что возможно в случае, когда они являются аддитивной смесью шума и сигнала, значения которого зависимы.

Согласно гипотезе  $H_0$  статистика  $w(\bar{\xi})$  асимптотически распределена как  $N(0,1)$ , если число наблюдений асимптотически стремится к бесконечности. В. Brock, Hsieh, и LeBaron [10] сообщают об обширных моделированиях и показывают, что асимптотическое распределение хорошо приближается, когда наблюдений больше чем 500. В качестве теста на достоверность гипотезы  $H_0$  об отсутствии в наблюдении хаотического процесса принимается выполнение неравенства  $|w_{m,N}(\varepsilon)| \leq 1,96$  для значения статистики  $w_{m,N}(\varepsilon)$ , что соответствует уровню значимости  $\alpha = 0,05$  (вероятности ошибки первого рода), тогда с 95% уверенностью можно принять гипотезу  $H_0$  (I.I.D.). Критическая область уровня  $\alpha = 0,05$ , состоит из двух бесконечных полуинтервалов  $(-\infty, -1.96]$  и  $[1.96, \infty)$ , вероятность попадания в каждый из которых статистики  $w$  равна  $\alpha/2 = 0,025$ . Когда  $w_{m,N}(\varepsilon)$  меньше 3,0 эта уверенность более 99%. В отсутствии шумов наблюдения применение критерия значимости к статисти-

ке  $w_{m,N}(\varepsilon)$  позволяет эффективно решать задачу классификации наблюдения.

До сих пор мы говорили о свойствах критерия в предположении, что верна гипотеза  $H_0$ . А что происходит, когда верна альтернативная гипотеза  $H_1$ ? В этом случае распределение статистики критерия  $w(\bar{\xi})$  изменится. Поэтому при проверке статистических гипотез недостаточно ориентироваться на значение  $\alpha$ . Следует определить мощность критерия  $1-\beta$  или вероятность ошибки второго рода  $\beta$  при принятии альтернативной гипотезы  $H_1$ : данные не I.I.D., что подразумевает, зависимость (возможно нелинейную) временного ряда, если были взяты первые разности натуральных логарифмов. Мощность критерия – это вероятность принятия при применении данного критерия альтернативной гипотезы  $H_1$  при условии, что она верна, т.е. его способность обнаружить имеющееся отклонение от нулевой гипотезы (выявлять различия, если они есть). Очевидно, что при фиксированной ошибке 1-го рода (ее мы задаем сами, и она не зависит от свойств критерия) критерий будет тем лучше, чем больше его мощность (т.е. чем меньше ошибка 2-го рода). Для расчета мощности критерия  $1-\beta$  ( $\beta = p(w(\bar{\xi}) \in G_\alpha | H_1)$ ),  $G_\alpha$  – критическая область при заданном уровне значимости  $\alpha$ ) необходимо знать условную плотность распределения  $p(w(\bar{\xi}) | H_1)$ . Мощность критерия (теста) определяется эмпирическим путем. Одни и те же задачи могут быть решены с помощью разных критериев (тестов), при этом обнаруживается, что некоторые критерии позволяют выявить различия там, где другие оказываются неспособными это сделать, или выявляют более высокий уровень значимости различий. BDS-тест выбран потому, что позволяет обнаружить много типов отклонений от независимости и идентичного распределения, и может служить общим образцовым тестом классификации процессов (временных рядов)  $\bar{\xi}$ , особенно в присутствии нелинейной динамики. Hsieh показал [11], что у BDS-теста хорошая мощность обнаружить четыре типа non-I.I.D. поведения: линейная и нелинейная зависимость, нестационарность и хаос.

BDS-тест можно считать непараметрическим тестом (критерием), характеризующимся тем, что в качестве его статистик используются нелинейные функции  $w(\bar{\xi})$  от наблюдений, распределение которых не зависит от вида распределения наблюдаемых величин  $\bar{\xi}$ . В этом случае мы получаем возможность получить некоторую информацию о многомерной функции (плотности) распределения  $F_N(\xi_1, \xi_2, \dots, \xi_N)$  анализируя одномерную эмпирическую функцию распределения  $p(w)$  статистики  $w$ .

Задача статистической теории решений с использованием BDS-статистики, если ее излагать в

терминах радиотехнических систем (РТС), состоит в следующем. На вход системы поступает наблюдение  $\{\xi(t)=as(t)+n(t), t \in (0, T)\}$  сигнала и помехи  $n(t)$ , аппроксимируемой белым шумом с нулевым средним и дисперсией  $\sigma_n^2$ . Наблюдение обычно представляется вектором  $\bar{\xi} = \xi_1, \dots, \xi_N$  своих отсчетов  $\xi(t_i)$  в моменты  $t_i = ih$  с шагом  $h$ , заданных на интервале времени  $T_\xi = Nh$ . В модель наблюдения входит двоичный параметр  $a \in (0, 1)$  наличия хаотического сигнала ( $a=1$ ) с дисперсией  $\sigma_s^2$ , тогда  $\xi = s+n$  и его отсутствия ( $a=0$ ):  $\xi = n$ . Требуется по принятой реализации наблюдения  $\{\xi(t), t \in (0, T)\}$  принять решение, имеется ли на входе сигнал ( $\hat{a} = 1$  или  $\hat{a} = 0$ ). В такой постановке задача проверки гипотезы  $H_1$  может интерпретироваться как задача обнаружения хаотического процесса на фоне шума с I.I.D. Качество обнаружения будем характеризовать двумя условными вероятностями: вероятностью ложной тревоги  $\alpha$  и вероятностью правильного обнаружения  $D$ .

Для нахождения оптимального алгоритма обработки BDS статистики принимаемого колебания  $\{\xi(t), t \in (0, T)\}$  при обнаружении используем так называемый критерий Неймана-Пирсона. Согласно этому критерию оптимальной считается система, которая при заданном значении вероятности ложной тревоги  $\alpha = \alpha_{\text{зад}}$  максимизирует вероятность правильного обнаружения  $D$ . Это задача на условный экстремум. Ее решение находится из уравнения

$$D - l_0 \alpha = \max, \quad (5)$$

где  $l_0$  – множитель Лагранжа, определяемый из условия  $\alpha = \alpha_{\text{зад}}$ .

Введенные вероятности можно выразить следующими формулами:

$$\alpha = \text{Вер} \{w > w_\Pi | a = 0\} = \int_{w_\Pi}^{\infty} p_n(w) dw;$$

$$D = \text{Вер} \{w > w_\Pi | a = 1\} = \int_{w_\Pi}^{\infty} p_{sn}(w) dw, \quad (6)$$

где  $p_n(w) = p(w | a = 0)$  и  $p_{sn}(w) = p(w | a = 1)$  – условная плотность вероятности BDS-статистики  $w(\bar{\xi})$  принимаемого колебания при условии, что  $a = 0$  (сигнал отсутствует) и  $a = 1$  (сигнал имеется). Эту задачу на условный экстремум можно записать в виде

$$J(U_1) = \int_{U_1} [p_{sn}(w) - l_0 p_n(w)] dw = \max_{U_1} \{ \alpha = \alpha_{\text{зад}} \}. \quad (7)$$

Выбором области ненулевого решения  $U_1$  максимизируем этот интеграл. Решение можно интерпретировать так: по принятому колебанию  $\{\xi(t), t \in (0, T)\}$  нужно сформировать BDS-статистику и два числа

$p_{sn}(w)$  и  $p_n(w)$ , затем сравнить их между собой  $l = p_{sn}(w)/p_n(w)$  с учетом весового множителя  $l_0$ :

$$l \begin{matrix} \hat{a}=1 \\ > \\ < \\ \hat{a}=0 \end{matrix} l_0; l_0: \alpha = \alpha_{зад}. \quad (8)$$

Вероятность ложной тревоги  $\alpha$  желательно уменьшать. Для этого необходимо увеличивать пороговый уровень  $w_n$ . Однако последнее приводит к уменьшению вероятности правильного обнаружения  $D$ , что нежелательно. Поэтому допустимые значения вероятностей  $\alpha$  и  $D$  выбираются из компромиссных соображений и задаются техническими условиями (ТУ):  $\alpha_{зад}$  и  $D_{зад}$ . После того, как выбран порог  $w_0$ , обеспечивающий требуемое ТУ  $\alpha = \alpha_{зад}$ , нужно еще обеспечить требуемое значение  $D \geq D_{зад}$ , которое будет выполняться для одного значения BDS-статистики при условии превышения отношением сигнал/помеха порогового значения  $q(\alpha_{зад}, D_{зад})$ . Расчет обнаружителя Неймана Пирсона при этом состоит в поиске комбинации трех переменных: решающего порога  $w_n$ , приходящегося на один отсчет BDS-статистики, отношения сигнал/помеха  $q$  и объема выборки  $N$ , при которых обеспечиваются заданные вероятности ошибок первого и второго рода  $\alpha$  и  $\beta$ . Взаимосвязь между вероятностями ошибок, значением решающего порога и пороговым сигналом носит сложный характер, поэтому расчет возможен только численными методами, например, методом последовательных приближений.

Придерживаясь концепции построения обнаружителя хаотического сигнала основанного на использовании отношения правдоподобия необходимо иметь выражения для плотностей  $p_n(w)$  и  $p_{sn}(w)$ , связанных с нулевой  $H_0$  и с альтернативной гипотезой  $H_1$ . Как уже отмечалось гипотеза  $H_0$  – простая, она однозначно фиксирует стандартное нормальное распределение BDS-статистики. Появление на фоне белого шума хаотического сигнала с неизвестными параметрами изменяет вероятностные свойства не только наблюдения  $\xi$ , но и BDS-статистики. По мере увеличения уровня хаотического сигнала начинает проявляться зависимость между значениями наблюдения. Аналитически описать трансформацию наблюдения  $\xi$  и статистики  $w$  в терминах функций их распределения или плотностей вероятности в зависимости от соотношения между компонентами сигнала и шума не представляется возможным. Здесь мы сталкиваемся со случаем, когда альтернативная гипотеза  $H_1$  оказывается сложной, и мы не можем определить величину ошибки второго рода. Поэтому в ситуации, когда статистика критерия не попадает в критическую область, не утверждают категорично, что “нулевая гипотеза принимается”, а формулируют вывод более осторожно: “нулевая гипотеза не отвергается”. Тем

самым подчеркивается, что хотя мы и не обнаружили отклонения от нулевой гипотезы, мы могли его при верной  $H_1$  ошибочно не обнаружить с вероятностью  $\beta$ , которую мы не знаем и которая, возможно, довольно значительна. Если же ошибка  $\beta$  действительно велика, то утверждение “нулевая гипотеза принимается” не представляет большой ценности. Учитывая, что сложные альтернативы могут быть двусторонними и односторонними при наличии достоверной информации о направлении отклонения от нулевой гипотезы предпочтительнее использование односторонней альтернативы, поскольку это повышает мощность критерия.

Когда альтернативная гипотеза оказывается сложной, требуются специальные походы к принятию решения. Например, при достаточно длительном наблюдении временной ряд можно сегментировать и построить эмпирическую функцию распределения  $p_{sn}(w)$  по ансамблю сегментов. Число сегментов ориентировочно можно выбрать равным  $L = 50$ . Если учесть, что для применения BDS-статистики рекомендуется чтобы  $N > 500$ , а  $N/m > 200$ , то объем выборки  $N > L \cdot 500$ . Однако, часто имеющиеся выборочные данные слишком малочисленны чтобы по выборке ограниченного объема  $N$  получить хорошую эмпирическую оценку плотности распределения  $p_{sn}(w)$ .

Рассмотрим альтернативную возможность оценки плотности  $p_{sn}(w)$ , которая, как показывают исследования, может заметно отличаться от нормальной и существенно зависит от вида отображения используемого для генерирования хаотического процесса. В современной математической статистике получили интенсивное развитие подходы к повышению достоверности расчета вероятностных характеристик за счет использования информации о фактическом законе распределения, содержащейся в самой ограниченной по объему выборке данных опытным путем (методом Монте-Карло). В основе используемого нами подхода к оценке плотности лежит идея формированию псевдо ансамблей суррогатных сигналов по выборке ограниченного объема и повторения BDS-теста много раз над различными реализациями суррогатных данных. Основное условие при формировании суррогатных данных это сохранение тех свойств сигналов, которые не соответствуют гипотезе  $H_0$ , в нашем случае это их зависимость. Ключевыми моментами при этом являются выбор процедуры построения множества суррогатных данных (псевдо выборок)  $\left\{ \xi_n^k \right\}$ ,  $k = 1, \dots, B$ , функции (статистики)  $w_k = g\left(\left\{ \xi_n^k \right\}\right)$  от наблюдения  $\xi = \left\{ \xi_n^0 \right\}$ , а также метода оценки выборочной функции распределения вероятностей выбранной статистики. Отметим, что в вычислительном отноше-

нии это дорогой подход, однако, ему нет никакой альтернативы. В тоже время, чтобы ограничить вычислительные усилия (ресурсы), тесты можно выполнить на довольно низком уровне значимости, для которого необходимым является достаточно небольшой набор суррогатных данных. Этот подход начал практически реализовываться в последние два десятилетия благодаря развитию современных вычислительных средств.

Понятие “суррогатных данных”, введенное Theiler и др. [12] в 1992 году, экстенсивно использовалось в контексте статистического тестирования временного ряда на нелинейность, которую принимали за нулевую гипотезу. При этом определялось статистическое различие между статистикой, вычисленной для реального временного ряда и для ансамбля испытательной статистики (псевдо выборки), вычисленной на линеаризованных версиях данных, так называемых суррогатных данных, или “заемистелях”. Рассматривались два основных вида нулевых гипотез: простая и сложная. Простая нулевая гипотеза утверждает, что данные произведены определенным и известным (линейным) процессом. Сложная нулевая гипотеза утверждает, что неизвестный основной процесс – член определенного семейства процессов.

Среди известных способов построения множества псевдо выборок можно выделить следующие. Бутстреп- процедуры [8], основанные на многократном тиражировании исходной эмпирической выборки, хаотическом перемешивании полученной огромной совокупности данных и извлечении из нее псевдо выборок (бутстреп-копий). В этом случае суррогатный временной ряд  $\{\xi_n^k\}$ ,  $k = 1, \dots, B$  получен из скалярного временного ряда  $\bar{\xi} = \{\xi_n^0\}$ , случайным перемешиванием его элементов. Процесс разрушает временные корреляции, которые не ожидаются для процесса I.I.D., но сохраняет то же самое распределение вероятности, которое определено для реального процесса с I.I.D.. Вторая процедура позволяет сформировать ряд  $\{\xi_n^k\}$ , в результате преобразования Фурье данных  $\bar{\xi} = \{\xi_n^0\}$ , рандомизации реальных значений фаз компонент полученного спектра (перемешивая или заменяя их на случайные, выбранные из интервала  $[0, \pi]$ ) и взятия обратного преобразования Фурье. В этом случае суррогатные данные сохраняют линейные корреляции реального процесса (спектры мощности), но любая нелинейная структура разрушается. Третья процедура преобразует данные таким образом чтобы достичь двух целей: во-первых, так же, как и при использовании второй процедуры, сохранить в суррогатах спектры мощности (линейную корреляцию) данных и, во-вторых, идентичность вероятности распределения реальных и суррогатных данных.

Эти три процедуры соотносятся с тремя гипотезами шума: I.I.D., линейно фильтрованный шум и монотонное нелинейное преобразование линейно фильтрованного шума. Следовательно, эти методы линейные суррогатные методы. Поэтому каждый из этих трех алгоритмов добавляет другой слой к иерархии гипотезы.

Для рассматриваемой нами задачи обнаружения, когда нулевая гипотеза простая, а альтернативная – сложная, упомянутые технологии получения суррогатных данных не представляют интереса, так как не сохраняют свойства аттрактора, связанного с обнаруживаемым сигналом, а, следовательно, изменяют свойства BDS-статистики. Поэтому необходимо расширить арсенал алгоритмов формирования суррогатных данных на “более интересные” гипотезы. Обратимся к алгоритмам формирования суррогатных траектории аттрактора (attractor trajectory surrogates – ATS). Эти суррогаты имитируют поведение динамической системы, то есть, сохраняя не только линейные, но также и нелинейные её свойства. Их ансамбль может использоваться, для оценки распределения значений любой статистики, полученной из временных рядов.

Один из возможных алгоритмов формирования ATS может быть описан следующим образом [9]:

1. Определим векторный временной ряд  $\{z_t\}_{t=1}^{N-d_w}$  с элементами  $z_t = (x_t, x_{t+\tau}, x_{t+2\tau}, \dots, x_{t+d_e\tau})$  вложением скалярного временного ряда  $\{x_t\}_{t=1}^N$  в псевдофазовое пространство с временной задержкой  $\tau$ . Для простоты выбираем окно вложения  $d_w = d_e\tau - 1$ , где  $d_e$  размерность вложения.
2. Выберем начальное состояние  $s_1 \in \{z_t \mid t = 1, \dots, N - d_w\}$ .
3. Положим  $i = 1$ .
4. Выберем наугад для  $s_i$  одного соседа из множества  $\{z_t \mid t = 1, \dots, N - d_w\}$ , например,  $z_j$ .
5. Установим  $s_{i+1} = z_{j+1}$  и увеличим  $i$ .
6. Повторим эту процедуру от шага 4 пока  $i = N$ .
7. Возьмем  $\{(s_t)_1 : t = 1, 2, \dots, N\}$ , как суррогатный временной ряд, в котором  $(\cdot)_1$  обозначает скалярную первую координату вектора.

Применяя этот алгоритм, получаем векторный временной ряд  $(s_t)_t$ , стохастическая траектория которого на аттракторе приближает  $(z_t)_t$ . Таким образом, суррогатная реализация представляет случайное блуждание на аттракторе. У векторного суррогатного временного ряда  $(s_t)_t$  основная динамика такая же, как и у первоначальной системы  $(z_t)_t$ , но она искажена шумом. Также отметим, что, как  $\{s_t \mid t = 1, 2, \dots, N\}$  так и  $\{z_t \mid t = 1, \dots, N - d_w\}$  приближают один и тот же аттрактор, но не реконструируют его.

Рассмотрим более простой, с вычислительной точки зрения, алгоритм формирования ATS не требующий вложения скалярного временного ряда  $\{\xi_k\}$  в псевдо фазовое пространство. Предложенный нами алгоритм может быть описан следующим образом:

1. Подвергнем множество значений временного ряда (отсчетов наблюдаемого сигнала)  $\xi_k \in (\xi_{\min}, \xi_{\max})$  равномерному квантованию дискретным множеством уровней  $\{\xi^{(1)}, \dots, \xi^{(m)}\}$  с интервалом между соседними уровнями квантования  $\Delta_i = \xi^{(i+1)} - \xi^{(i)} = \text{const}$ .

2. Из множества  $\{\xi_k\}$  выделим подмножества  $\{\bar{\xi}_{\Delta_i}\}$ , принадлежащие интервалам  $\{\Delta_i\}_{i=1}^I$ .

3. Хаотически (или случайным образом) перемешаем (переставим) элементы подмножеств  $\bar{\xi}_{\Delta_i}$ . Максимально возможное число перестановок  $I_i!$ , где  $I_i$  – число элементов в интервале  $\Delta_i$  (размерность вектора  $\bar{\xi}_{\Delta_i}$ ). Операторы перестановки выполняют взаимно однозначное отображение подмножеств, принадлежащих интервалам  $\{\Delta_i\}_{i=1}^I$ , на себя. Подмножества до и после перемешивания состоят из одних и тех же элементов, но отличаются порядком их следования, т.е. выполняется пере присваивание номеров элементов временного ряда, составляющих все возможные пары каждого из подмножеств  $\Delta_i$ .

4. Объединяем перемешанные таким образом элементы подмножеств и получаем первую реализацию  $\{\xi_n^1\}$  суррогатной последовательности, сохраняющей (имитирующей) свойства аттрактора сигнальной составляющей наблюдения.

5. Повторяем эту процедуру начиная с шага 3 пока не получим ансамбль суррогатных последовательностей  $\{\xi_n^k\}_{k=1}^B$ .

Максимальная мощность  $I_i!$  множества суррогатных последовательностей зависит от длины исходного временного ряда  $N$ , от числа уровней квантования  $I$  и может значительно превышать минимально необходимую мощность  $B$ , т.е.  $I! \gg B$ .

На рис. 1, а, б показаны фазовые портреты соответственно хаотической последовательности, сформированной итерированием палаточного ото-

бражения  $\Delta(x) = r(1 - 2|x/2 - 1|)$ ,  $x \in [0, 1]$  с управляющим параметром  $r = 0,9999$ , и полученной из неё суррогатной последовательности при двадцати уровнях квантования.

Размытие фазового портрета, построенного для суррогатной реализации, можно интерпретировать как результат добавления к хаотической последовательности суррогатного шума. Такого рода шум оказывает на процесс передачи информации с использованием хаотической несущей по существу такое же влияние, как и помехи в канале связи.

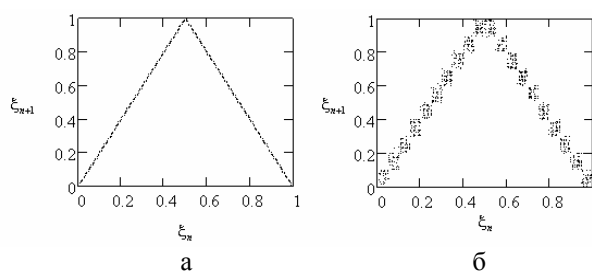


Рис. 1. Фазовые портреты: а – исходного хаотического отображения, б – суррогата

Для наглядности в первой строке табл. 1 показаны семь первых значений исходной реализации хаотической последовательности, а в остальных строках по семь первых значений каждой из шести суррогатных реализаций. Структурная схема одного из вариантов BDS-обнаружителя изображена на рис. 2.

В его состав входят формирователь суррогатных сигналов (ФСС) и формирователь непараметрических статистик (ФНС), который в свою очередь содержит блоки формирования корреляционных интегралов  $C_1$  и  $C_m$ , невязки  $\Delta = (C_1)^m - C_m$ , СКО невязки –  $\hat{S}_{m,N}(\varepsilon)$  и делитель ( $\div$ ). Кроме того, обнаружитель содержит формирователь оценки плотности вероятности  $\hat{p}(w | H_1)$  BDS-статистик  $w$  в присутствии сигнала, блок формирования отношения правдоподобия  $l(w) = \hat{p}(w | H_1) / N(w; 0, 1)$ , генератор случайной последовательности (ГСП) с независимыми и нормально распределенными значениями  $N(0, 1)$ , формирователь оценки плотности вероятности  $\hat{p}(1)$ , (если необходимо оценить оптимальный порог обнаружения), блок оценки второго момента плотности  $\hat{p}(1)$  и пороговое устройство (ПУ)  $l_0(w(\bar{\xi})) \underset{<}{>} \hat{\sigma}_1^2$ .

Таблица 1

Семь первых значений исходной реализации хаотической последовательности и семь первых значений каждой из шести суррогатных реализаций

	1	2	3	4	5	6	7
1	0,526	0,948	0,104	0,208	0,417	0,833	0,334
2	0,532	0,905	0,133	0,201	0,443	0,824	0,31
3	0,502	0,936	0,12	0,21	0,419	0,825	0,324
4	0,55	0,9	0,129	0,226	0,406	0,81	0,317
5	0,524	0,939	0,133	0,206	0,411	0,849	0,301
6	0,532	0,928	0,125	0,23	0,432	0,804	0,314
7	0,546	0,905	0,105	0,234	0,442	0,839	0,323

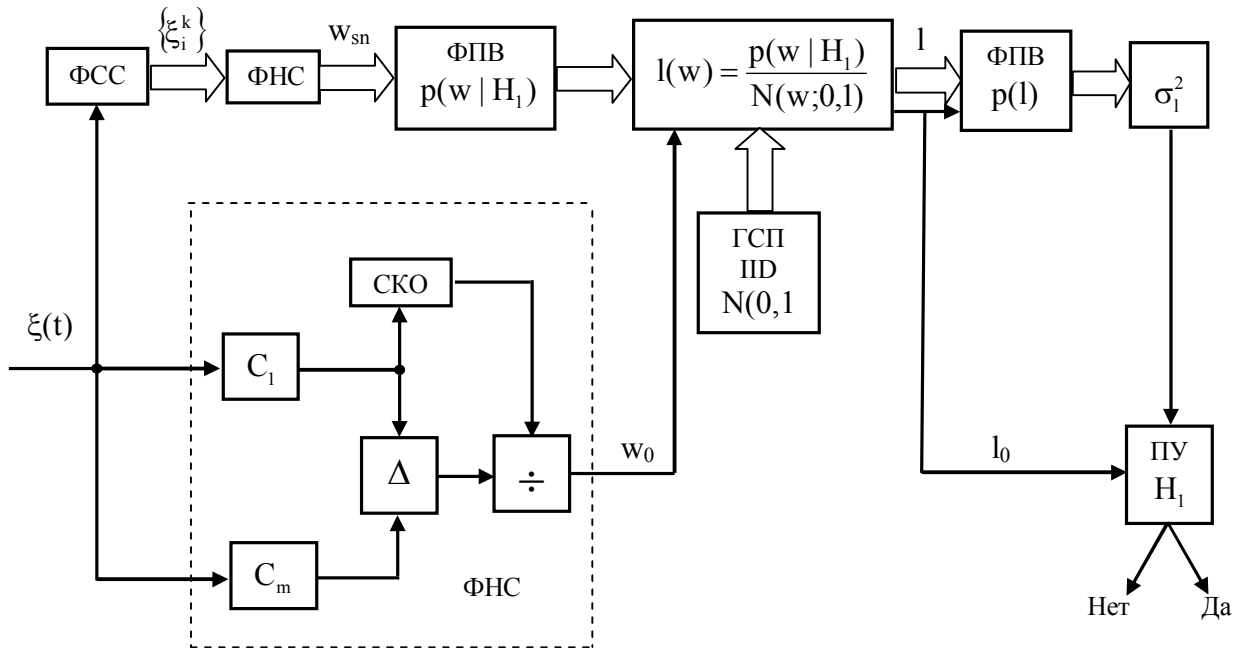


Рис. 2. Непараметрический BDS обнаружитель

На практике порог устанавливают по дисперсии шума на выходе приемника, реализующего обнаружитель Неймана-Пирсона. В нашем случае подобный порог, равный величине  $\hat{\sigma}_1^2$ , формируется в блоке оценки второго момента плотности  $\hat{p}(l)$ , однако он может быть оценен и непосредственно по вектору  $l$ . В этом случае блок оценки  $\hat{p}(l)$  может отсутствовать.

На рис. 3, а, б показаны гистограммы распределений BDS-статистик  $\{w_k\}$ , полученные по ансамблю хаотических последовательностей с  $\sigma_\xi = 0,28$ , сформированных палаточным отображением с управляющим параметром 0,9999 и разными начальными значениями  $\{\xi_0^k\}_{k=1}^B$  ( $B = 200$ ) в отсутствии и присутствии шума  $\sigma_n = 0,175\sigma_\xi$  в каждой последовательности.

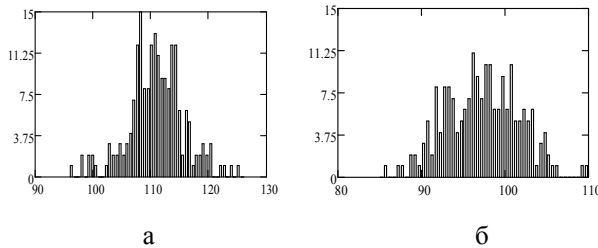


Рис. 3. Гистограммы распределений BDS-статистик

Легко заметить, что добавление шума смещает гистограмму в область меньших её значений. По мере увеличения шума следует ожидать стремление гистограммы распределения BDS-статистик к стандартному нормальному распределению.

На рис. 4 показана гистограмма распределения BDS-статистик  $\{w_k\}$ , полученная по ансамблю  $\{\xi_n^k\}_{k=1}^B$  ( $B = 200$ ) суррогатных последовательностей,

сформированных предложенным нами алгоритмом формирования ATS по реализации  $\{\xi_n^0\}$  с  $\sigma_\xi = 0,28$ , порожденной палаточным отображением с управляющим параметром 0,9999, и начальным значением  $\xi_0 = 0,7$ , к которой добавлен шум с  $\sigma_n = 0,175\sigma_\xi$ .

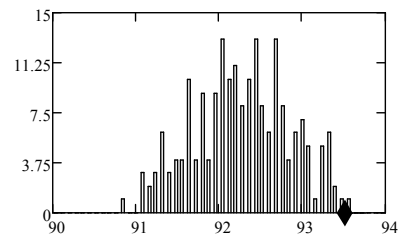


Рис. 4. Гистограммы распределений BDS-статистик

Без шума значение BDS-статистики равнялось  $w_0 = 110,053$ . Добавление шума уменьшает её значение  $w_0 = 93,5$ , выделенное ромбиком на горизонтальной оси.

Заметим, что анализ существующих в статистике методов оценки выборочной функции (плотности) распределения вероятностей показывает, что определенными преимуществами перед традиционными параметрическими и гистограммными обладают адаптивные ядерные оценки, способные извлекать апостериорную информацию о реальном законе распределения данных даже при обработке сравнительно небольших по объему выборок, насчитывающих всего несколько десятков реализаций. Поэтому блок ФПВ может быть реализован с использованием алгоритма ядерной оценки плотности  $p(w | H_1)$ .



## Выводы

Таким образом, предложенная в данной работе формализация понятия “форма” сигнала с помощью корреляционной размерности и связанной с ней BDS-статистикой позволяют предложить новый подход к непараметрическому обнаружению хаотического сигнала на фоне белого шума с неизвестной плотностью распределения его значений. Недостающая информация, необходимая для реализации критерия Неймана-Пирсона, извлекается из наблюдения с использованием ATS-алгоритма формирования суррогатных сигналов, которые используются для эмпирической оценки отношения правдоподобия. Структура обнаружителя принципиально отличается от обнаружителей использующих традиционной  $\chi^2$ -статистики, учитывающие только энергетические признаки сигнала, и дающие неоправданно оптимистичные (завышенные) значения скрытности не только хаотических, но и шумоподобных сигналов. В ситуации, когда получение необходимого ансамбля наблюдений не представляется возможным, мы предлагаем проверить гипотезу о наличии хаотического сигнала на фоне белого шума на основе заместителей (суррогатов), произведенных математическим алгоритмом. Предложенный алгоритм формирования суррогатных сигналов в отличие от других подходов, таких как линейно-подобные суррогаты, формирует суррогатные сигналы, сохраняющие нелинейные свойства хаотических сигналов, т.е. их траектории в псевдофазовом пространстве.

## Список литературы

1. Сверхширокополосная беспроводная связь на основе динамического хаоса / А.С. Дмитриев, А.В. Клецов, А.М. Лактюшкин, А.И. Панас, С.О. Старков, А.Д. Хилинский // *Радиотехника и электроника*. – 2006. – Т. 51. – № 10. – С. 1193–1209.

2. Кальянов Э.В. Скрытая передача информации при использовании запаздывания для выделения полезного сигнала из маскирующих колебаний / Э.В. Кальянов // *Письма в ЖТФ*. – 2009. – Т.35, №6. – С. 56–62.
3. Петерс Э. Фрактальный анализ финансовых рынков: применение теории Хаоса в инвестициях и экономике / Э. Петерс. – М.: Интернет-трейдинг. – 2004. – 304 с.
4. Помехозащищенность радиосистем со сложными сигналами / Г.И. Тузов, В.А. Сивов, В.И. Прытков и др. // М.: Радио и связь. – 1985. – 264 с.
5. Костенко П.Ю. Обнаружение хаотического процесса искаженного белым шумом с использованием BDS-статистик / П.Ю. Костенко, А.Н. Барсуков, К.С. Васюта, С.Н. Симоненко // *Радиоэлектроника*. – 2009. – Т. 52, № 11. – С. 41–50.
6. Brock W.A. A test for independence based on the correlation dimension / W.A. Brock, W.D. Dechert and J.A. Scheinkman // *Working Paper #8702. Department of Economics, University of Wisconsin* – 1987.
7. Small Michael Applied Nonlinear Time Series Analysis Applications in Physics, Physiology and Finance / Michael Small // *World Scientific Publishing Co. Pte. Ltd.* – 2005. – 245 P.
8. Holger Kantz Nonlinear time series analysis/ Holger Kantz and Thomas Schreiber // *Second edition Printed in the United Kingdom at the University Press, Cambridge*. – 2004. – 369 p.
9. Kanzler Ludwig Very Fast and Correctly Sized Estimation of the BDS Statistic / Ludwig Kanzler // *Christ Church and Department of Economics University of Oxford*. – 1999. – 95 P.
10. Brock, W. A. A Test of Nonlinear Dynamics, Chaos, and Instability / Brock, W. A., D. Hsieh, and B. LeBaron // *Cambridge: MIT Press*. – 1991.
11. Hsieh, D. A. / *Chaos and nonlinear dynamics: Applications to financial markets* / Hsieh, D. A. // *Journal of Finance*. – 1991, 46: 1839-1877.
12. Theiler, J. Testing for nonlinearity in time series: The method of surrogate data / Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. // *Physica D*, 1992, 58: 77-94.

Поступила в редколлегию 15.06.2010

Рецензент: д-р техн. наук, проф. О.В. Лемешко, Харьковский национальный университет радиоэлектроники, Харьков.

## НОВИЙ ПІДХІД ДО НЕПАРАМЕТРИЧНОГО ВИЯВЛЕННЯ ХАОТИЧНИХ СИГНАЛІВ НА ФОНІ БІЛОГО ШУМУ З ВИКОРИСТАННЯМ "НЕЛІНІЙНОЇ ДИНАМІЧНОЇ СТАТИСТИКИ"

П.Ю. Костенко, К.С. Васюта, А.Н. Барсуков, С.Н. Симоненко

У статті з'ясовано причини недостатньої ефективності енергетичного підходу до виявлення сигналів невідомої “форми”, заснованого на використанні  $\chi^2$ -статистики. Формалізується поняття “форма” сигналу з використанням кореляційної розмірності і пов'язаною з нею BDS-статистикою. Проаналізовано підхід до непараметричного виявлення хаотичного сигналу на тлі білого шуму з невідомою щільністю розподілу його значень з використанням BDS-статистики. Запропоновано алгоритм виявлення хаотичного сигналу, структурна схема його реалізації, а також ATS-алгоритм формування суррогатних сигналів, які використовуються для емпіричної оцінки відношення правдоподібності.

**Ключові слова:** непараметричне виявлення, BDS - статистика, ATS - алгоритм формування суррогатних сигналів, “форма” сигналу.

## THE NEW APPROACH TO NONPARAMETRIC DETECTION OF CHAOTIC SIGNALS AGAINST A WHITE NOISE WITH USE OF "NONLINEAR DYNAMIC STATISTICS"

P. Y. Kostenko, K. S. Vasuta, O. M. Barsukov, S. N. Symonenko

In paper reasons of insufficient efficiency of the energy approach to detection of signals of the unknown "shape", based on  $\chi^2$ -statistics use are become clear. The concept "shape" of a signal about the help of correlative dimensions of a quantity and the BDS-statistics related to it is formalized. The approach to nonparametric detection of a random signal against a white noise with an unknown density function of its significances with use BDS-statistics is analyses. Algorithms of detection of a random signal, the skeleton diagrammed of its implementation, and also ATS-algorithm of shaping of substitute signals which are used for an empirical estimate of the ration of likelihood are offered.

**Keywords:** nonparametric detector, BDS - statistics, ATS – attractor trajectory surrogate, signal "form", correlative dimensions.