

АНАЛИЗ РЕКУРРЕНТНОСТИ ЗАШУМЛЕННЫХ ВРЕМЕННЫХ РЯДОВ

© 2018 г. А.В. Русаков, А.Б. Медвинский, Н.И. Нуриева

*Институт теоретической и экспериментальной биофизики РАН,
142290, Пушкино Московской области, ул. Институтская, 3*

E-mail: rusakov_a@rambler.ru

Поступила в редакцию 19.04.18 г.

Предложен метод численной оценки характера динамики зашумленных временных рядов с использованием рекуррентных диаграмм в условиях, когда внешний шум трансформирует рекуррентные диаграммы, соответствующие качественно разным динамическим режимам, таким образом, что рекуррентные диаграммы становятся практически неразличимыми. На примере разностной логистической математической модели показано, что этот метод позволяет отличать случайные флуктуации от детерминированных колебаний, как регулярных, так и хаотических, даже при сравнительно высоких значениях отношения шум/сигнал. Кроме того, показано, что увеличение уровня шума может приводить к сближению рекуррентных свойств регулярных и хаотических колебаний.

Ключевые слова: временные ряды, рекуррентные диаграммы, случайный шум, предсказуемость.

DOI: 10.1134/S0006302918040129

Исследование характера динамики временного ряда, полученного в результате измерений или наблюдения за параметрами исследуемой системы, необходимо для оценки предсказуемости состояний системы и дальнейшего изучения механизмов ее функционирования. В последние десятилетия был разработан богатый набор как линейных, так и нелинейных методов исследования временных рядов. Поскольку большинство природных биологических и экологических процессов существенно нелинейны [1–3], нелинейные методы оценки предсказуемости и выявления механизмов, лежащих в основе таких явлений, оказываются более эффективными по сравнению с линейными [2,4]. При этом, однако, нет такого универсального метода, который подходил бы для всех случаев. Одним из существенных недостатков многих методов нелинейного анализа при оценке временных рядов, полученных в ходе наблюдений за природными процессами, является зависимость этих методов от длины исследуемого временного ряда. Например, нелинейные методы оценки горизонта предсказуемости [1] и вычисления ляпуновской экспоненты [2] дают приемлемые результаты при длине временных рядов не менее чем в тысячу значений, полученных в результате компьютерного моделирования [5–7]. В случае более коротких временных рядов (в несколько сотен значений или меньше) эти методы дают недостоверные результаты. Ис-

следование природных процессов, особенно динамики биологических популяций, часто проводится на основании временных рядов, полученных в ходе полевых наблюдений, длительность которых во многих случаях составляет всего лишь порядка ста точек (см., например, работы [8,9]).

Одним из немногих нелинейных методов, который дает возможность проводить численную оценку характеристик процессов, представленных короткими временными рядами, является численный рекуррентный анализ. В основе этого метода лежит фундаментальное свойство детерминированных колебательных процессов – повторяемость состояний, или рекуррентность. Рекуррентность предполагает тесное сближение достаточно продолжительного фрагмента фазовой траектории с более ранним фрагментом той же траектории. В случае периодических колебаний тесное сближение означает полное совпадение и происходит периодически. В ходе нерегулярных, аperiodических детерминированных колебаний сближение фрагментов фазовой траектории происходит нерегулярно. Наличие рекуррентности является существенным отличием детерминированных колебательных процессов по сравнению со случайными колебаниями.

Несмотря на то что данный метод активно развивается с конца восьмидесятых годов XX века [10,11], четкие правила выбора пара-

метров при использовании метода анализа рекуррентности до сих пор не выработаны, исследователю приходится пользоваться некоторыми рекомендациями, сформулированными на основании решения ряда конкретных задач [12].

В данной работе нами предлагается метод численной оценки характера динамики временных рядов с использованием рекуррентных диаграмм и обосновывается выбор численных значений некоторых параметров, используемых при построении рекуррентных диаграмм. Основное внимание при этом уделяется анализу рекуррентных диаграмм, соответствующих логистическим отображениям при добавлении шума, как модельного объекта, воспроизводящего влияние случайных внешних воздействий на результаты реальных измерений, проводимых в лаборатории или в ходе полевых наблюдений.

МОДЕЛЬ

С целью получения модельных временных рядов с уже исследованной ранее динамикой использовали разностную модель, логистическое отображение, широко используемое при решении задач математической экологии (см., например, работу [13]):

$$N_{t+1} = rN_t(1 - N_t). \quad (1)$$

Изменение модельного параметра r уравнения (1) в диапазоне 3,1–3,9 приводит к изменению характера динамики временного ряда N_t , а именно к переходу от стационарного режима к регулярным колебаниям и далее, через удвоение периода регулярных колебаний, – к хаотическим колебаниям [13,14].

В данной работе параметр r принимали равным 3,2 и 3,55 (при этом имели место регулярные колебания), а также равным 3,9 (при этом динамика колебаний N_t приобретает хаотический характер). При $r = 3,2$ период колебаний равен 2, при $r = 3,55$ – равняется 8. Для регулярных колебаний доминантный показатель Ляпунова равен нулю; для хаотических колебаний, которые возникают при $r = 3,9$, доминантный показатель Ляпунова положителен и примерно равен 0,4.

С целью получения зашумленных рядов к временному ряду, полученному по формуле (1), добавляли случайные значения, полученные с помощью генератора случайных чисел:

$$N_t \rightarrow N_t + aR_{\text{rand}}, \quad (2)$$

где R_{rand} – случайное число в диапазоне от 0 до 1, a – масштабирующий коэффициент. Коэффициент a выбирали в диапазоне от 0 до 0,4, что соответствует уровню шума по амплитуде от 0 до 40%, так как численные значения N_t , полученные с помощью уравнения (1), могут

лежать в диапазоне от 0 до 1 (при начальных значениях, лежащих в этом же диапазоне).

РЕКУРРЕНТНЫЕ ДИАГРАММЫ, ИХ ПАРАМЕТРЫ И ХАРАКТЕРИСТИКИ

Для оценки динамики исследуемых временных рядов нами строились рекуррентные диаграммы по следующей формуле:

$$R_{ij}(\epsilon) = H(\epsilon - \|\vec{x}_i - \vec{x}_j\|), \quad i, j = 1, \dots, L, \quad (3)$$

где векторы $\vec{x}_k = (x_k, x_{k-h}, \dots, x_{k-(d-1)h})$, $k = i, j$, h – временной лаг, d – размерность пространства вложения [1,10], ϵ – размер окрестности точки в фазовом пространстве в момент i , $H(x)$ – функция Хэвисайда: $H\theta = 0$, если $x < 0$, и $H(x) = 1$ в остальных случаях, $\|\dots\|$ – норма, L – длина временного ряда.

Метод рекуррентных диаграмм реализует отображение d -мерной фазовой траектории состояний системы $x(t)$ на двумерную квадратную двоичную матрицу размером $L \times L$. В такой матрице значение 1 (на рекуррентной диаграмме обычно отображается черной точкой) соответствует совпадению (с точностью до ϵ) состояний системы в моменты времени i и j . На рекуррентной диаграмме как по оси абсцисс, так и по оси ординат откладывается время.

На рис. 1 показаны рекуррентные диаграммы для случайного процесса (рис. 1а), полученного с помощью генератора случайных чисел, а также регулярного (рис. 1б) и хаотического (рис. 1в) процессов, полученных по формуле (1) при $r = 3,55$ и $r = 3,9$ соответственно. Длина L временных рядов, служивших базисом для построения этих рекуррентных диаграмм, равна 150. Поскольку разные типы поведения системы имеют разные распределения точек на рекуррентной диаграмме, то по характеру распределения этих точек на рекуррентной диаграмме в принципе можно сделать вывод о характере исследуемой динамики. Из рис. 1а видно, что случайному процессу соответствует беспорядочное распределение точек рекуррентной диаграммы; диагональные, горизонтальные и вертикальные линии отсутствуют либо редки и имеют малую длину. Для строго регулярных процессов (рис. 1б) характерны диагональные линии, параллельные основной диагонали рекуррентной диаграммы – линии идентичности, для которой $i = j$ (3). Хаотические процессы (рис. 1в) характеризуются наличием диагональных линий, расположенных нерегулярно. Несмотря на, казалось бы, явные различия в характере распределения точек на рекуррентных диаграммах для временных рядов, характеризующихся разной динамикой, оценка динамики реальных временных рядов, полученных в ре-

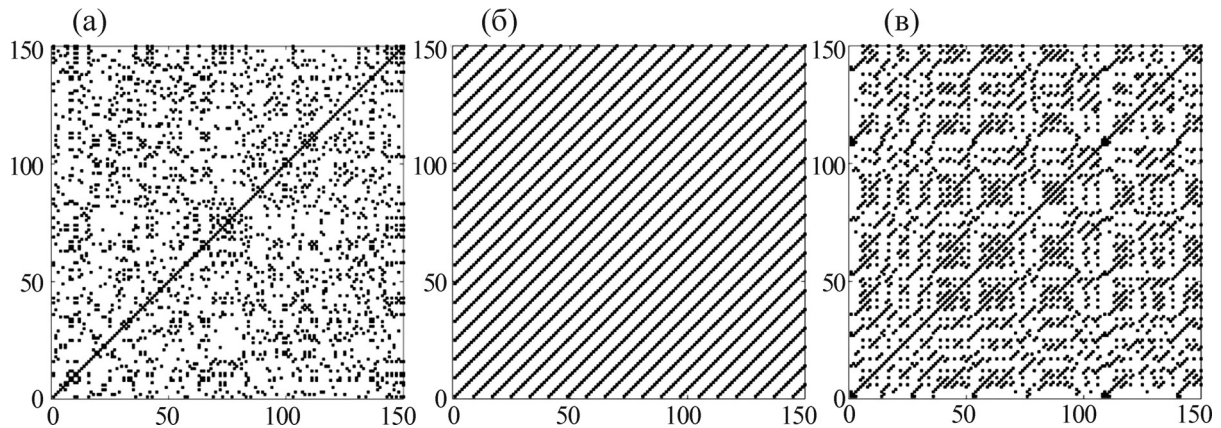


Рис. 1. Рекуррентные диаграммы для случайного процесса, полученного с помощью генератора случайных чисел (а), регулярного процесса (б) и хаотического процесса (в), полученных по формуле (1) при $r = 3,55$ и $r = 3,9$ соответственно. На всех рекуррентных диаграммах $\epsilon = 0,05$.

зультате наблюдений, часто затруднена. С одной стороны, это связано с тем, что правила выбора параметров при построении рекуррентных диаграмм (размерность пространства вложения, размер ϵ -окрестности) имеют рекомендательный характер и варьируют в зависимости от задачи. С другой стороны, ряды, полученные в ходе экспериментов и полевых наблюдений, часто сильно зашумлены. В этом случае рекуррентные диаграммы демонстрируют смешанный характер, их интерпретация в значительной мере субъективна. Немаловажна и длина исследуемого временного ряда. Чем короче временной ряд, тем меньше рекуррентных свойств успевает продемонстрировать система на исследуемом временном отрезке, и поэтому тем меньше точек на рекуррентной диаграмме, которые могут быть проанализированы с помощью численного рекуррентного анализа.

В данной работе при выборе численного значения ϵ используется плотность заполнения рекуррентной диаграммы:

$$P_f = n/L^2, \tag{4}$$

где n – количество черных точек на рекуррентной диаграмме (количество единиц в рекуррентной матрице), L – длина исследуемого ряда (длина стороны рекуррентной матрицы).

Суть предлагаемого нами метода, целью которого является численная оценка характера динамических режимов с использованием рекуррентных диаграмм, состоит в вычислении отношения количества черных точек, формирующих на рекуррентной диаграмме диагональные и параллельные основной диагонали отрезки, к общему количеству черных точек. Для этого вычисляется гистограмма распределения диагональных отрезков в зависимости от длины этих отрезков:

$$P(l) = \sum_{i,j=1}^L (1 - R_{i-1,j-1})(1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k}; i \neq j, \tag{5}$$

где l – длина диагонального отрезка, $R_{i-1,j-1} = 0$, если $R_{1j} = 1$ или $R_{i1} = 1$, и $R_{i+k,j+k} = 0$, если $R_{Lj} = 1$ или $R_{iL} = 1$ [15]. Затем вычисляется отношение количества всех точек рекуррентной диаграммы, входящих в диагональные отрезки с длиной, превышающей пороговое значение l_{min} , к общему количеству точек на диаграмме:

$$P_d = \frac{\sum_{l=l_{min}}^L lP(l)}{n}, \tag{6}$$

где n – общее количество черных точек на рекуррентной диаграмме, т.е. точек, для которых $R_{ij} = 1$.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

При исследовании динамики временных рядов методом рекуррентных диаграмм необходимо определить ряд параметров построения таких диаграмм. Как отмечено выше, строгих правил выбора параметров при построении рекуррентных диаграмм нет, а между тем важно, чтобы анализ временных рядов (сравнение их рекуррентных свойств) проходил при одинаковых или близких параметрах. Часть параметров в нашем исследовании определяется характером исследуемых временных рядов.

Длина модельных временных рядов выбрана равной 150, что соответствует длине рядов, которые были получены в ходе полевых наблюдений; их длина в среднем составляет примерно сто точек [8,9].

При выборе численного значения l_{\min} имеется возможность следовать рекомендациям, данным в работе [16]. Отметим, что при слишком малых численных значениях l_{\min} есть опасность некорректного учета ложных соседей, связанного с тангенциальным движением вдоль фазовой траектории, когда значения автокорреляционной функции все еще далеки от нуля [12]. При этом соседние значения временного ряда могут быть ложно интерпретированы как точки на пространственно близких, но удаленных друг от друга во времени участках траектории в фазовом пространстве. Полевые наблюдения за численностью, например, планктонных популяций обычно производятся не чаще одного раза в неделю [9], а за численностью грызунов и хищников – обычно раз в год [8,17]. В таких случаях корреляция между изменениями измеряемых значений от точки к точке фазовой траектории, очевидно, отсутствует и проблема ложных соседей снимается. Все же нельзя исключить появления коротких (например, $l = 2$) диагональных отрезков на рекуррентной диаграмме в результате воздействия внешнего шума, и это обстоятельство накладывает ограничение на величину l_{\min} . С другой стороны, если численное значение l_{\min} достаточно велико, то гистограмма $P(l)$, определяемая уравнением (5), может оказаться разреженной настолько, что оценка величины P_d с помощью уравнения (6) будет ненадежной. В нашем случае $l_{\min} = 3$. При такой величине порогового значения l влияние ложных соседей даже при довольно значительном уровне шума невелико. Кроме того, распределение диагональных отрезков в зависимости от их длины, которое описывается гистограммой (5), предоставляет достаточно информации для того, чтобы можно было эффективно отличить случайные флуктуации от периодических и хаотических колебаний.

Размерность d пространства вложения принята равной единице. Возможность выбора такого значения обосновывается в работе [12]. При выборе больших значений построение рекуррентных диаграмм, соответствующих сравнительно коротким временным рядам, затруднительно.

Значение временного лага h выбрано равным единице. Значение временного лага, большее единицы, обычно выбирают в тех случаях, когда состояние системы меняется медленно на длительном промежутке времени и в результате соседние значения временного ряда могут быть ложно определены как точки на удаленных друг от друга во времени участках фазовой траектории. Поскольку, как отмечено выше, частота проведения полевых наблюдений обычно невелика, состояние системы между наблюдениями успевает существенно измениться и указанная выше проблема не возникает.

Выбор параметра ϵ в уравнении (3) может значительно влиять на распределение точек на рекуррентной диаграмме. С увеличением параметра ϵ все больше точек исследуемого временного ряда попадает внутрь ϵ -окрестности, в результате увеличивается количество точек на рекуррентной диаграмме, а значит, растет значение параметра P_f (4). При некорректно заданной величине ϵ эти точки могут как отражать реальные рекуррентные свойства системы, так и являться ложными соседями. На рис. 2 представлены зависимости параметров P_f (рис. 2а) и P_d (рис. 2б) от численного значения ϵ для случайного процесса (сплошная линия), хаотического процесса в отсутствие случайной компоненты (кружок) и зашумленных хаотических процессов с уровнем шума 10 (звездочка), 20 (квадрат) и 30% (крестик). Согласно рис. 2а, плотность заполнения рекуррентной диаграммы P_f возрастает одинаково для всех процессов (как в присутствии, так и в отсутствие шума) на интервале значений ϵ до 0,1 и достигает 0,2 при $\epsilon = 0,1$. При $\epsilon > 0,1$ скорость роста P_f с увеличением ϵ оказывается различной для разных уровней зашумленности хаотического сигнала, что говорит о существенном воздействии шума на вид рекуррентной диаграммы, а значит, и на возможное влияние ложных соседей.

Согласно рис. 2б, значения параметра P_d зашумленных хаотических временных рядов для $\epsilon < 0,05$ практически неотличимы от параметра P_d случайного временного ряда. Для значений ϵ на интервале 0,05–0,1 исследуемые хаотические ряды отличаются по параметру P_d от случайного временного ряда. При этом для $\epsilon = 0,1$ это различие наиболее выражено, но значения P_d для хаотического ряда в отсутствие шума и хаотического ряда с уровнем шума 10% различить не представляется возможным. В то же время значения P_d для всех хаотических рядов различаются при $\epsilon = 0,05$, однако значения P_d для временных рядов с уровнем шума 20 и 30% близки к значению P_d для случайного процесса. Таким образом, выбор значения параметра ϵ лежит в диапазоне 0,05–0,1.

Рис. 1 демонстрирует существенные различия в распределении точек на рекуррентной диаграмме между случайным, регулярным и хаотическим процессами в отсутствие внешнего шума. Видно, что для случайного процесса точки на рекуррентной диаграмме не образуют длинные диагональные отрезки (рис. 1а); численное значение P_d при этом мало (рис. 2б). Для регулярного процесса все точки рекуррентной диаграммы находятся на диагональных отрезках (рис. 1б), при этом значение P_d близко к единице. На рекуррентной диаграмме хаотического процесса присутствуют как одиночные

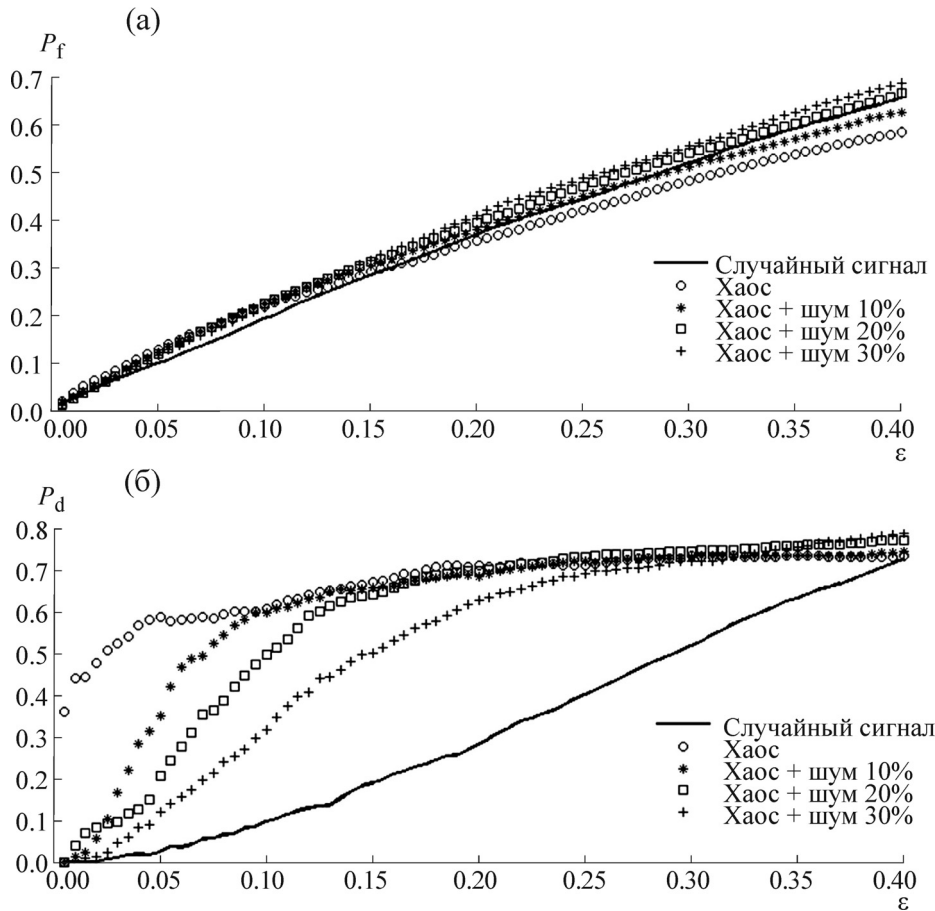


Рис. 2. Зависимость параметров P_f (а) и P_d (б) от значения ϵ для случайного процесса (сплошная линия), хаотического процесса (кружки), зашумленных хаотических процессов ($r = 3,9$) с уровнем шума 10 (звездочки), 20 (квадраты) и 30% (крестики).

точки, так и диагональные отрезки, расположенные нерегулярным образом (рис. 1в); значение параметра P_d при этом близко к 0,6 (рис. 2б). Для сравнения на рис. 3 показаны рекуррентные диаграммы для зашумленных регулярных временных рядов при $r = 3,2$ (первая колонка) и при $r = 3,55$ (вторая колонка), а также для зашумленного хаотического ряда при $r = 3,9$ (третья колонка); $\epsilon = 0,05$. Уровень шума здесь составляет 10 (а–в), 20 (г–е) и 30% (ж–и). Видно, что увеличение уровня шума может приводить к сближению рекуррентных свойств регулярных и хаотических процессов, что выражается в очевидном сходстве распределения точек на соответствующих рекуррентных диаграммах.

Для сравнения рекуррентных свойств зашумленных регулярных и хаотических процессов было получено по 1000 временных рядов длиной $L = 150$ по формуле (2) для каждого из исследуемых случаев: регулярные процессы ($r = 3,2$ и $r = 3,55$) и хаотический процесс ($r = 3,9$) для уровней шума 10, 20, 30 и 40%. Для

всех временных рядов вычисляли параметр P_d по формуле (6). На рис. 4 представлены средние значения P_d , для описанных выше случаев, а также для 1000 случайных временных рядов ($\epsilon = 0,05$ (рис. 4а) и $\epsilon = 0,1$ (рис. 4б)). Также показаны максимальное и минимальное значения для каждого случая. Видно, что для $\epsilon = 0,05$ (рис. 4а) при небольшом уровне шума, равном 10% ($a = 0,1$), значения P_d для регулярных и хаотических процессов существенно различаются. С увеличением уровня шума различия между регулярными и хаотическими процессами нивелируются. При максимальном уровне шума ($a = 0,4$) численные значения P_d сближаются настолько, что разные динамические режимы: случайные флуктуации, регулярные колебания и хаотические колебания, становятся практически неразличимы. При $\epsilon = 0,1$ значения P_d для зашумленных регулярных и хаотических процессов слабо различимы, однако регулярные и хаотические процессы при всех уровнях шума хорошо отличаются от случайных флуктуаций (рис. 4б).

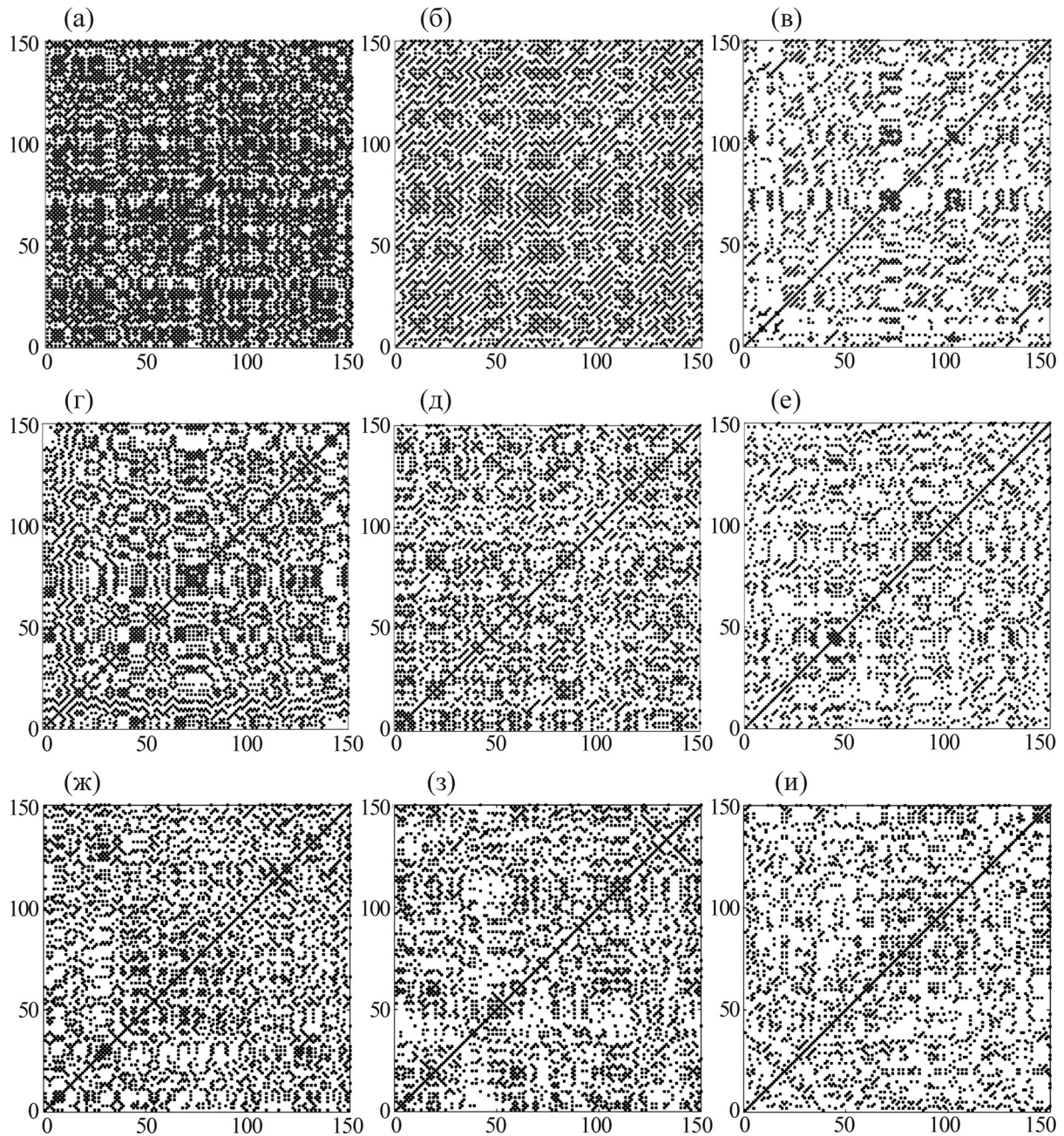


Рис. 3. Рекуррентные диаграммы для зашумленных регулярных временных рядов при $r = 3,2$ (слева), $r = 3,55$ (посередине) и зашумленного хаотического ряда при $r = 3,9$ (справа) при $\epsilon = 0,05$. Уровень шума составляет 10 (а–в), 20 (г–е) и 30% (ж–и). Длина временных рядов L равна 150.

Таким образом, как показано на рис. 4, предложенный нами подход позволяет отличать случайные процессы от детерминированных, как регулярных, так и хаотических, при анализе сравнительно коротких временных рядов даже в тех случаях, когда уровень внешнего шума делает рекуррентные диаграммы, соответствующие случайным, регулярным и хаотическим процессам качественно сходными между собой (ср. рис. 1а и рис. 3). Несомненный интерес

представляет использование этого подхода в рамках численного рекуррентного анализа результатов мониторинга динамики гидробионтов и других природных популяций с целью выявления факторов, определяющих характер колебаний популяционного обилия.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 17-04-00048).

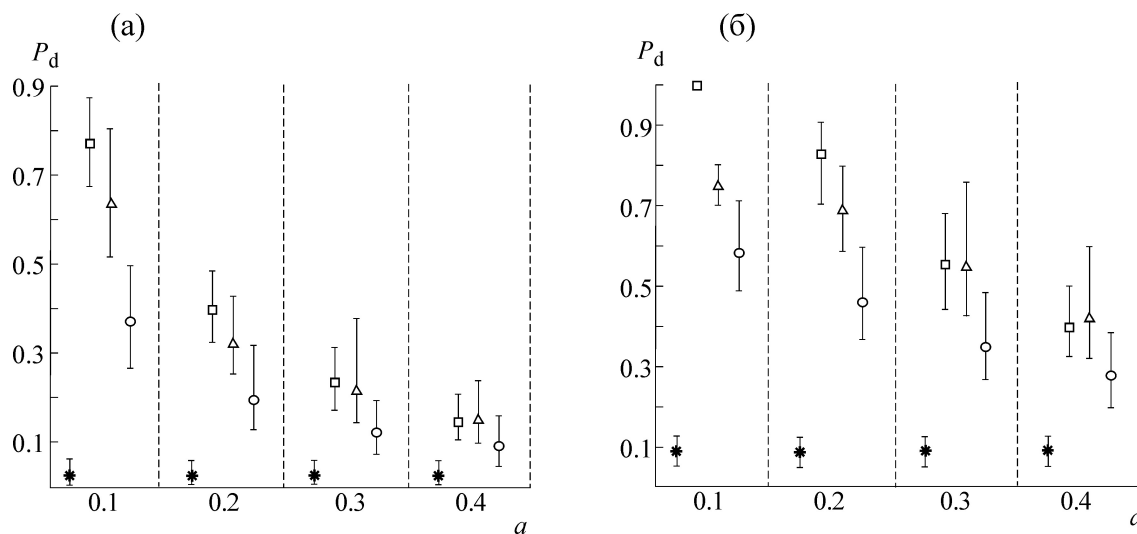


Рис. 4. Зависимости параметра P_d от уровня зашумленности временных рядов a для случайных временных рядов (звездочки), регулярных временных рядов при $r = 3,2$ (квадраты) и $r = 3,55$ (треугольники) и хаотического ряда при $r = 3,9$ (кружки) для $\epsilon = 0,05$ (а) и $\epsilon = 0,1$ (б). Для каждого случая получили 1000 значений P_d для временных рядов длиной $L = 150$. Маркерами на диаграммах показаны средние значения, также обозначены максимальные и минимальные значения.

СПИСОК ЛИТЕРАТУРЫ

1. D. Kaplan and L. Glass, *Understanding Nonlinear Dynamic* (Springer, New York, 1995).
2. Н. Kantz and Т. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University, Cambridge, 1997).
3. S. P. Otto and Т. Day, *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution* (Princeton University, Princeton, 2007).
4. Т. В. Жукова, Н. П. Радчикова, Т. М. Михеева и др., Биол. внутр. вод, № 3, 7 (2017).
5. А. В. Medvinsky and А. V. Rusakov, *Chaos, Solitons & Fractals* **44**, 390 (2011).
6. А. В. Medvinsky, А. V. Rusakov, and N. I. Nurieva, *Ecol. Complexity* **14**, 108 (2013).
7. А. Б. Медвинский, Н. И. Нуриева, А. В. Русаков и Б. В. Адамович, *Биофизика* **62**, 107 (2017).
8. S. Ellner and P. Turchin, *The American Naturalist* **145**, 343 (1995).
9. А. В. Medvinsky, B. V. Adamovich, R. R. Aliev, et al., *Ecol. Complexity* **32**, 90 (2017).
10. J.-P. Eckmann, S. O. Kamphorst, and D. Ruelle, *Europhys. Lett.* **4**, 973 (1987).
11. N. Marwan, *Eur. Phys. J. Special Topics* **164**, 3 (2008).
12. N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, *Phys. Reports* **438**, 237 (2007).
13. M. Kot, *Elements of Mathematical Ecology* (Cambridge University, Cambridge, 2001).
14. R. M. May, *Nature* **261**, 459 (1976).
15. А. В. Medvinsky, B. V. Adamovich, A. Chakraborty, et al., *Ecol. Complexity* **23**, 61 (2015).
16. J. Theiler, *Phys. Rev. A* **34**, 2427 (1986).
17. C. Elton, *Voles, mice and lemmings: problems in population dynamics* (Oxford University, Oxford, 1942).

The Analysis of Recurrence of the Noisy Time Series

A.V. Rusakov, A.B. Medvinsky, and N.I. Nurieva

*Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences,
ul. Institutskaya 3, Pushchino, Moscow Region, 142290 Russia*

We propose a method for numerical evaluation of the character of the dynamics of noisy time series with the use of recurrence plots. Under influence of external noise the recurrence plots, which correspond to qualitatively different dynamic regimes, are transformed in such a way that they become virtually indistinguishable. By the example of the logistic map we show that this method allows one to distinguish random fluctuations from deterministic oscillations, both regular and chaotic oscillations, even at relatively high values of the signal to noise ratio. Besides, we show that an increase in the noise level can lead to convergence of the recurrent properties of regular and chaotic oscillations.

Keywords: time series, recurrence plots, random noise, predictability