



## ВЛИЯНИЕ ПАРАМЕТРОВ МЕТОДА ЛОЖНЫХ СОСЕДЕЙ НА ОЦЕНКУ МИНИМАЛЬНОЙ РАЗМЕРНОСТИ ВЛОЖЕНИЯ

*А.В. Красков, С.А. Ракитин*

На простых тестовых примерах иллюстрируется идея метода ложных соседей, используемого при обработке временных рядов. Рассматривается влияние параметров процедуры (выбор метрики, длина реализации, разрядность представления данных, время задержки) на оценку размерности множеств, восстановленных в фазовом пространстве.

В настоящее время исследование поведения нелинейных систем в фазовом пространстве, восстановленном по временной реализации одной динамической переменной, становится общепринятым [1,2]. Это связано с тем, что в эксперименте для наблюдения часто доступна лишь одна переменная, а такой подход позволяет дополнить информацию об исследуемой системе. Восстановление аттрактора является, в частности, составной частью процедуры построения динамических моделей по временным рядам. Наиболее популярным приемом восстановления фазовых траекторий является метод задержек. Из точек временной реализации доступной наблюдению переменной, представленной в виде дискретного ряда  $x(n)=x(n\Delta t)$ , где  $\Delta t$  – время дискретизации, составляются вектора состояния системы в восстанавливаемом фазовом пространстве в моменты  $n$  дискретного времени

$$y(n) = [x(n), x(n+\tau), \dots, x(n+\tau(d-1))], \quad n = 0, 1, 2, \dots, N-\tau(d-1), \quad (1)$$

где  $\tau$  – время задержки,  $N$  – количество точек временной реализации,  $d$  – размерность восстанавливаемого фазового пространства.

Узловым моментом в решении задачи восстановления является определение минимальной размерности  $d_E$  фазового пространства, обеспечивающей однозначность задания состояния (точке восстановленной фазовой траектории должно соответствовать единственное состояние системы). При моделировании значение  $d_E$  позволяет судить о необходимом числе уравнений конструируемой математической модели. Выбор слишком высокой размерности ведет к неоправданному усложнению модели и усилению влияния шумов [3]. Занижение размерности может привести к неправильному описанию поведения системы.

По теореме Такенса [1] достаточной для однозначного восстановления считается размерность  $d \geq 2d_A+1$ . Размерность аттрактора  $d_A$  обычно определяют как корреляционную размерность. В [2] было показано, что для восстановления достаточна размерность равная дополненной до ближайшего целого значения

корреляционной размерности. Принципиально другой подход реализуется в методе ложных соседей [3], который опирается непосредственно на запрет самопересечений фазовых траекторий динамической системы. Если у восстановленной траектории наблюдаются самопересечения, то это говорит о недостаточной размерности фазового пространства. Последовательно увеличивая размерность вложения можно добиться удаления всех пересечений. Минимальная размерность пространства, в котором отсутствуют пересечения, является размерностью вложения  $d_E$ , определяемой методом ложных соседей.

Проиллюстрируем суть проблемы реализации метода ложных соседей на примере восстановленной фазовой траектории гармонической временной реализации  $x(t)=\sin t$  (периода  $T=2\pi$ ), оцифрованной с некоторым временным шагом  $\Delta t$  (рис. 1, а). При  $d=1$  восстановленное по  $x(t)$  множество расположено на отрезке прямой (рис. 1, б). На нем точка  $k$  имеет соседями точки  $l$  и  $s$ , причем  $k$  и  $s$  соответствуют двум качественно разным состояниям системы (в них разные по знаку производные). При восстановлении в двумерном пространстве все точки «разойдутся» (рис. 1, в), но  $l$  и  $k$  – слабо, а  $k$  и  $s$  сильно ( $\Delta_{lk} < \Delta_{ks}$ ). По этому признаку точки  $k$  и  $s$  называют «ложными» соседями, а точки  $l$  и  $k$  «истинными». В [3] для каждой точки находится один, ближайший, сосед. Итак, последовательно увеличивая  $d$ , добиваются удаления всех ложных соседей, а следовательно, и пересечений. Относительное число пересечений фиксируется по графику зависимости отношения числа ложных соседей к общему числу точек в восстановленном пространстве от размерности вложения  $d$  (рис 1, г). В рассматриваемом примере, как можно показать, все ложные соседи исчезают при  $d=2$ , то есть минимальная размерность вложения  $d_E$  равна двум. В общем случае оценка ложности требует применения специальной процедуры.

В качестве основной количественной меры «истинности–ложности соседства» в [3] предложено условие

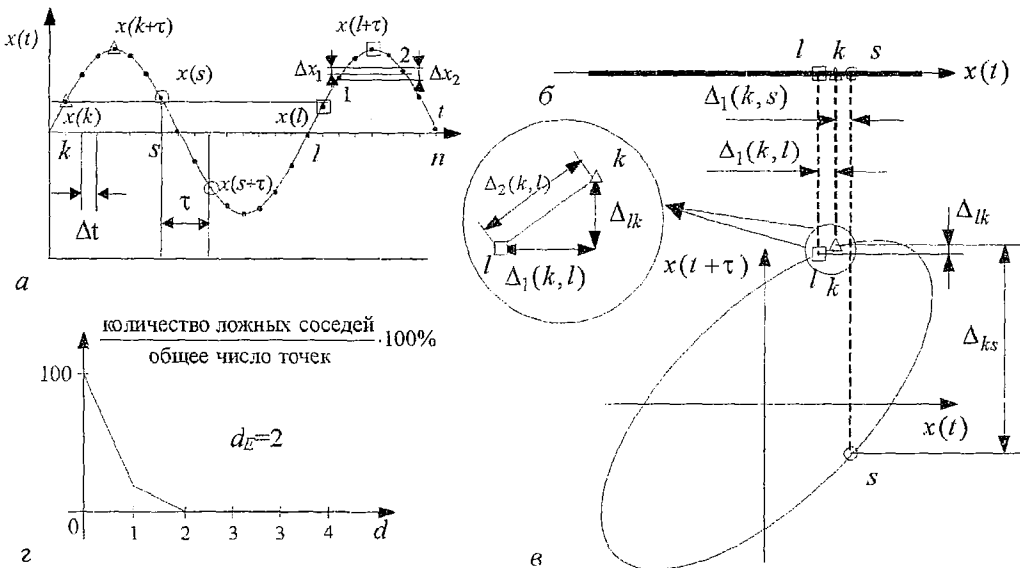


Рис. 1. Иллюстрация метода ложных соседей: а – временная реализация  $x(t)$  с дискретными отсчетами  $x(n\Delta t)$ . Выделены точки с  $n=k, s, l$ , имеющие близкие значения  $x(t)$ , и точки с  $n=k+3, s+3, l+3$ , задержанные на  $\tau=3\Delta t$  ( $\Delta t \approx T/16$ ), а также точки 1 и 2, совпадающие при разрядности  $p-1$  ( $\Delta x_2=1/2^{p-1}$ ) и отличающиеся при разрядности  $p$  ( $\Delta x_1=1/2^p$ ); б – фазовый портрет, восстановленный в одномерном пространстве; в – фазовый портрет, восстановленный в двумерном пространстве; г – график зависимости отношения числа ложных соседей к общему числу точек в восстановленном пространстве от размерности вложения  $d$

$$\Delta_{ks}/\Delta_d(k,s) > R_{кр} \quad (\text{либо } \Delta_{ks} > R_{кр} \Delta_d(k,s)), \quad (2)$$

где  $\Delta_d(k,s)$  – расстояние между точками  $k$  и  $s$  в  $d$ -мерном пространстве, которое определяется на основе той или иной метрики, например, евклидовой;  $\Delta_{ks}=|x(k+d\tau)-x(s+d\tau)|$  – характеризует приращение расстояния между соседними точками при переходе в  $(d+1)$ -мерное пространство;  $R_{кр}$  – эмпирический параметр, характеризующий величину расхождения точек в восстановленном пространстве при увеличении его размерности на 1 (например, в [3]  $R_{кр}=15$ , в [4]  $15 < R_{кр} < 20$ ). Если при восстановлении в фазовом пространстве возникают соседи, которые идеально совпадают, то  $\Delta_d(k,s)$  обращается в ноль и для любого, даже незначительного, приращения  $\Delta_{ks}$ , которое может наблюдаться и у истинных соседей, точки  $k$  и  $s$  будут идентифицироваться как ложные. При проведении натурального эксперимента из-за ограниченной разрядности аналого-цифрового преобразователя (АЦП) вероятность совпадения соседей повышается. При разрядности АЦП равной  $p$  различаются  $2^p$  значений наблюдаемой переменной ( $\Delta x=1/2^p$  – наименьшая различимая величина). Так на рис. 1, а при разрядности  $p$  ( $\Delta x_1=1/2^p$ ) значения оцифрованной реализации  $x(t)$  в точках 1 и 2 различаются, а при разрядности  $p-1$  ( $\Delta x_2=1/2^{p-1}=2/2^p=2\Delta x_1$ ) – совпадают. Из сказанного следует, что условие (2) в случае идеального совпадения точек необходимо модифицировать. В данной работе это делалось заменой  $\Delta_d(k,s)$  на половину наименьшей различимой величины

$$\Delta_{ks}/(0.5\Delta x) > R_{кр}. \quad (2a)$$

*Целью работы* является иллюстрация и эмпирическая оценка влияния выбора разрядности  $p$  представления данных, метрики  $L^1$ , количества точек  $N$  реализации и времени задержки  $\tau$ , используемого при восстановлении, на результаты определения размерности вложения  $d_E$ . Сначала для выявления особенностей метода исследуется простой тестовый пример с заранее известным значением размерности  $d_E=2$  (гармоническая временная реализация  $x(t)=\sin t$ ). Затем выделенные особенности проверяются и уточняются на более сложных примерах систем с хаотической динамикой:

трехмерный поток Ресслера

$$\begin{cases} \dot{x} = -(y + z), \\ y = ay + x, \\ \dot{z} = b + z(x - \mu), \end{cases} \text{ при } a=0.2, b=0.2, \mu=4.6$$

и аттрактор Лоренца

$$\begin{cases} \dot{x} = \sigma(y - z), \\ y = Rx - y - xz, \\ \dot{z} = -bz + xy, \end{cases} \text{ при } R=42.92, b=4, \sigma=16.$$

Результаты представляются графически на плоскости параметров  $(\tau, R_{кр})$ . Линии на плоскости разделяют области, приводящие к значению  $d_E$ , указанному цифрами. Для синусоиды областью параметров, обеспечивающих удовлетворительные результаты, являются значения, для которых  $d_E=2$ .

Рассмотрим результаты численного исследования влияния параметров метода на определяемое значение  $d_E$  для синусоиды.

**1. Время задержки.** На рис. 2, а представлены результаты расчета

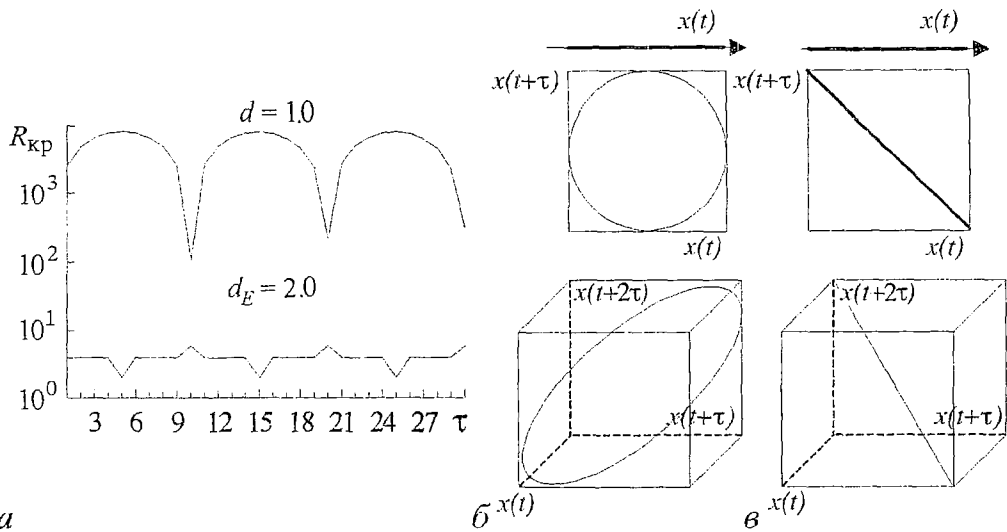


Рис. 2. Влияние времени задержки  $\tau$  на значение размерности  $d_E$  в первом тестовом примере  $x(t) = \sin t$ : а – области различных значений  $d$  на плоскости  $(\tau, R_{кр})$ ,  $N=20000$ ,  $\Delta t = 2\pi/(11.5 \cdot 3^{1/2})$ ,  $p=12$ ; эволюция фазовых портретов для времени задержки  $\tau \cong T/4$  (б) и  $\tau \cong T/2$  (в)

размерности вложения при разных значениях времени задержки  $\tau$ . Метод дает оптимальную размерность  $d_E=2$  в некотором диапазоне  $R_{кр}$ , наиболее широком (примерно от 10 до 10000) при  $\tau = T/4, 3T/4, \dots$ , и занижает ее до  $d=1$  при  $\tau$  кратном половине периода  $T$  ( $\tau \cong mT/2$ ). Это можно объяснить тем, что в последнем случае точки фазовой траектории, восстановленной в пространстве любой размерности  $d$ , располагаются на отрезке (диагональ  $d$ -мерного куба) (рис. 2, б) и при переходе из  $d$  в  $d+1$  точки не расходятся сильно. Таким образом, для удовлетворительной работы метода желательно избегать  $\tau \cong mT/2$ , а выбирать время задержки около  $T/4$  (ср. рис. 2, б, в), что эквивалентно выбору  $\tau$  в первом нуле автокорреляционной функции (АКФ). При таком выборе  $\tau$  компоненты вектора состояния не коррелированы между собой. Если же у АКФ отсутствуют нули, то  $\tau$  можно выбирать так, чтобы компоненты были наименее коррелированы, то есть в ее первом минимуме.

**2. Количество точек на реализации.** Если, не меняя шага  $\Delta t$ , увеличить число точек  $N$  (увеличивая длину временной выборки), то на фазовом портрете в  $d$ -мерном пространстве рядом с ближайшими соседями  $l$  и  $k$ , которые при переходе в  $(d+1)$ -мерное пространство оказываются ложными, может появиться какая-нибудь точка  $s$ , причем так, что расстояние между  $s$  и  $l$  меньше, чем между  $l$  и  $k$ . В этом случае ближайшим соседом точки  $l$  будет не  $k$ , а  $s$ . Тогда количество ложных соседей может:

- *остаться* тем же самым, если при переходе в  $(d+1)$ -мерное пространство  $l$  и  $s$  оказываются ложными, а  $k$  и ее ближайший сосед – истинными;
- *увеличиться*, если при переходе в  $(d+1)$ -мерное пространство оказываются ложными точки  $l$  и  $s$ ,  $k$  и ее ближайший сосед;
- *уменьшиться*, если при переходе в  $(d+1)$ -мерное пространство оказываются истинными точки  $l$  и  $s$ ,  $k$  и ее ближайший.

При равновесии процессов уменьшения и увеличения количества ложных соседей результаты перестают зависеть от количества точек реализации. На рис. 3, а представлены кривые для синусоиды. Видно, что уже начиная с 4000 точек результаты практически не меняются, область размерности  $d_E=2$  остается постоянной примерно в диапазоне  $10 < R_{кр} < 1000$ . Для системы Ресслера при

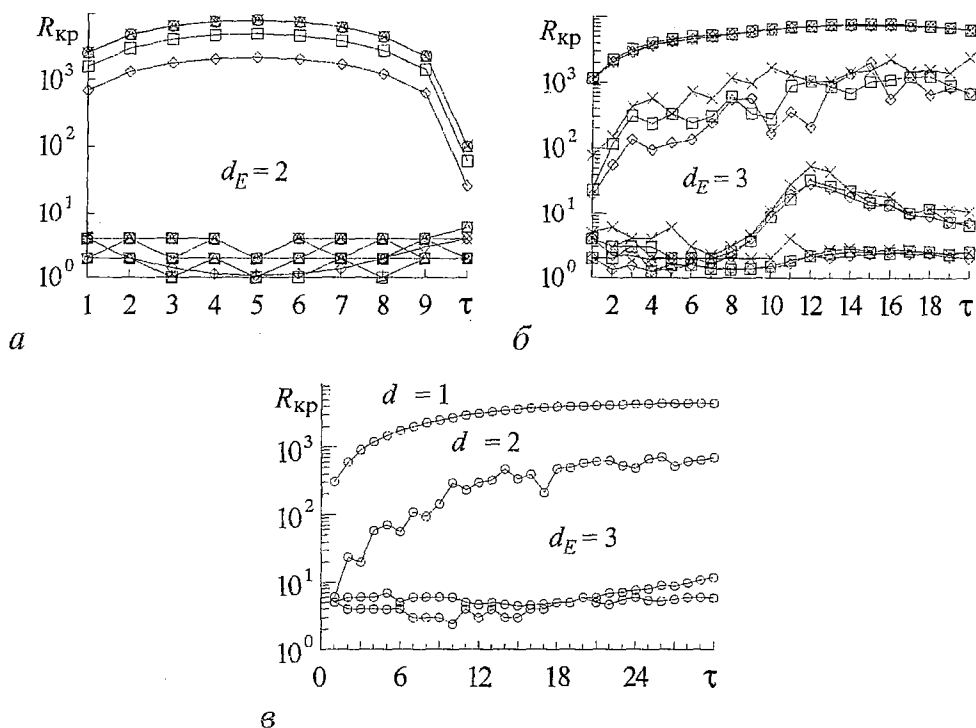


Рис. 3. Границы области оптимального значения размерности вложения  $d_E$  в зависимости от длины временной реализации  $N$ : *а* –  $x(t)=\sin t$ ,  $p=12$ ,  $\Delta t=2\pi/(11.5 \cdot 3^{1/2})$ ,  $\square$  – 2000,  $\circ$  – 3000,  $\Delta$  – 4000,  $\times$  – 25000; *б* – система Ресслера шаг интегрирования 0.1,  $p=12$ ,  $\diamond$  – 5000,  $\square$  – 10000,  $\times$  – 25000; *в* – система Лоренца шаг интегрирования 0.004,  $p=12$ ,  $N=10000$

увеличении числа точек кривые, отделяющие область размерности  $d_E=3$ , поднимаются, но со стороны низких размерностей рост больше, тем самым область оптимального значения  $d_E$  увеличивается и уже при  $N=10000$  она становится достаточно большой, примерно от 10 до 100 (рис. 3, б). На рис. 3, а, б выделены области только оптимальных размерностей вложения  $d_E=2$  и  $d_E=3$ , соответственно. На рис. 3, в для системы Лоренца показаны все области размерности вложения, включая оптимальное значение  $d_E=3$ .

**3. Выбор метрики.** В работах [3,4] расстояние между точками в критерии ложности рассчитывалось в евклидовой метрике  $L^2$ :  
в [3] по формуле

$$\Delta_d(k,s) = \{\sum_{i=1}^d [x(k+(i-1)\tau) - x(s+(i-1)\tau)]^2\}^{1/2},$$

а в [4] – с использованием нормировки

$$\Delta_d(k,s) = (1/d)\{\sum_{i=1}^d [x(k+(i-1)\tau) - x(s+(i-1)\tau)]^2\}^{1/2}.$$

Рассмотрим влияние на результат определения  $d_E$  изменения метрики на  $L^1$ . При этом по аналогии с [3,4] будем использовать нормированное и ненормированное межточечное расстояние

$$\Delta_d(k,s) = \sum_{i=1}^d |x(k+(i-1)\tau) - x(s+(i-1)\tau)|.$$

Такой выбор обеспечивает сокращение времени счета. Качественное различие этих метрик состоит в скорости роста расстояния между точками с увеличением размерности вложения. Медленнее всего растет расстояние в метрике, предложенной в [4], быстрее всего в  $L^1$ . При применении евклидовой метрики  $L^2$  и

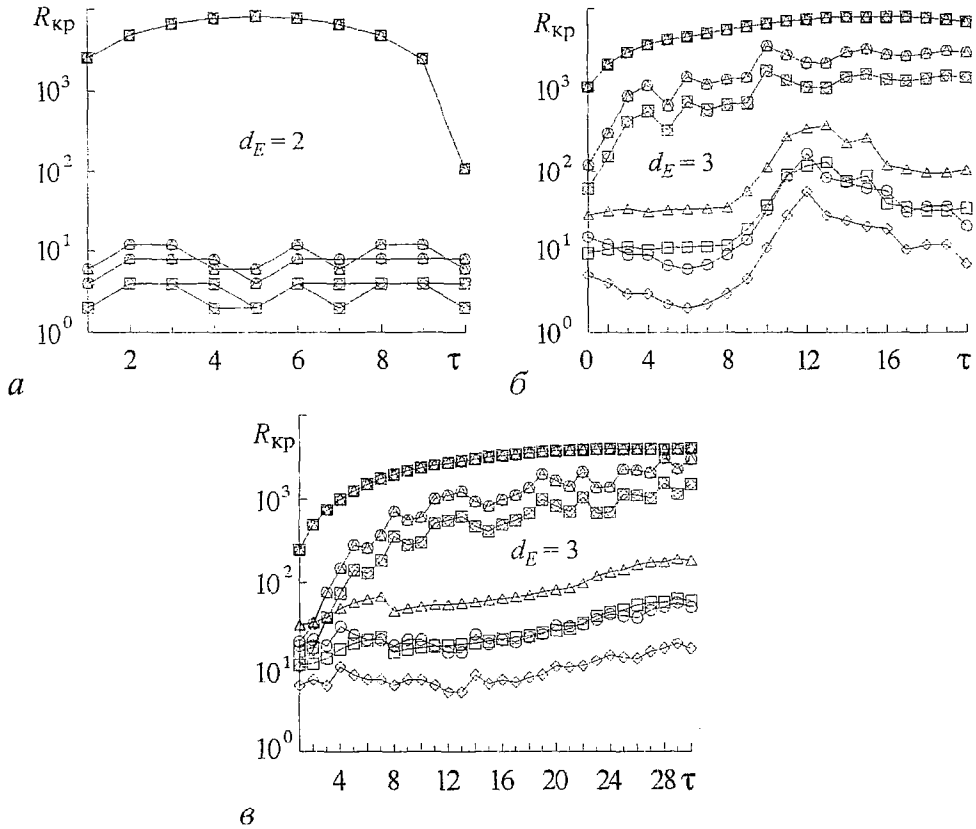


Рис. 4. Границы области оптимального значения размерности в зависимости от выбора метрики: *a* –  $x(t)=\sin t$ ,  $p=12$ ,  $\Delta t=2\pi/(11.5 \cdot 3^{1/2})$ ,  $N=15000$ ; *б* – система Ресслера шаг интегрирования 0.1,  $p=12$ ,  $N=20000$ ; *в* – система Лоренца шаг интегрирования 0.004,  $p=12$ ,  $N=20000$ . Здесь  $\diamond$  – метрика  $L^1$ ;  $\square$  – метрика  $L^2$  (евклидова);  $\circ$  – метрика  $L^1/d$ ;  $\times$  – метрика  $L^2/d$

метрики  $L^1$  была получена одинаковая область параметров, в которой  $d_E=2$  (рис. 4, *a*). Метрики  $L^1/d$  и  $L^2/d$  несколько сузили эту область по параметру  $R_{кр}$ . Это произошло из-за того, что расстояние в нормированных метриках растет медленнее, чем в  $L^1$  и  $L^2$ , так как знаменатель условий (2) и (2а) уменьшился, а значения  $R_{кр}$ , при которых ближайшие соседи идентифицируются как ложные, возросли. Таким образом нет никакого основания отдавать предпочтение более «долгой» евклидовой метрике, а даже наоборот, на примере системы Ресслера (рис. 4, *б*) видно, что при расчетах в метрике  $L^1$  область параметров с оптимальной размерностью  $d_E=3$  увеличилась в сторону высоких размерностей.

**4. Разрядность цифрового преобразователя.** При исследовании зависимости размерности вложения от разрядности  $p$  АЦП было обнаружено, что при увеличении  $p$  область параметров с  $d_E=2$  расширяется по параметру  $R_{кр}$  (рис. 5, *a*). Это можно объяснить тем, что при малой разрядности практически все ближайшие соседи совпадают, то есть работает условие (2а). При увеличении разрядности идеально совпадающих точек становится меньше и начинает работать условие (2). Одновременно растет количество различных значений  $x(t)$ , а так как при оцифровке наименьшая различимая величина  $\Delta x$  принимается равной единице, то растут и оцифрованные значения  $x(t)$ , а вместе с ними и  $\Delta_{kss}$ , поэтому ближайшие соседи будут идентифицироваться как ложные при больших  $R_{кр}$ . Отметим, что кривая, ограничивающая снизу область оптимальных значений  $d_E$  на плоскости параметров (со стороны высоких размерностей), опускается с ростом  $p$ , так как

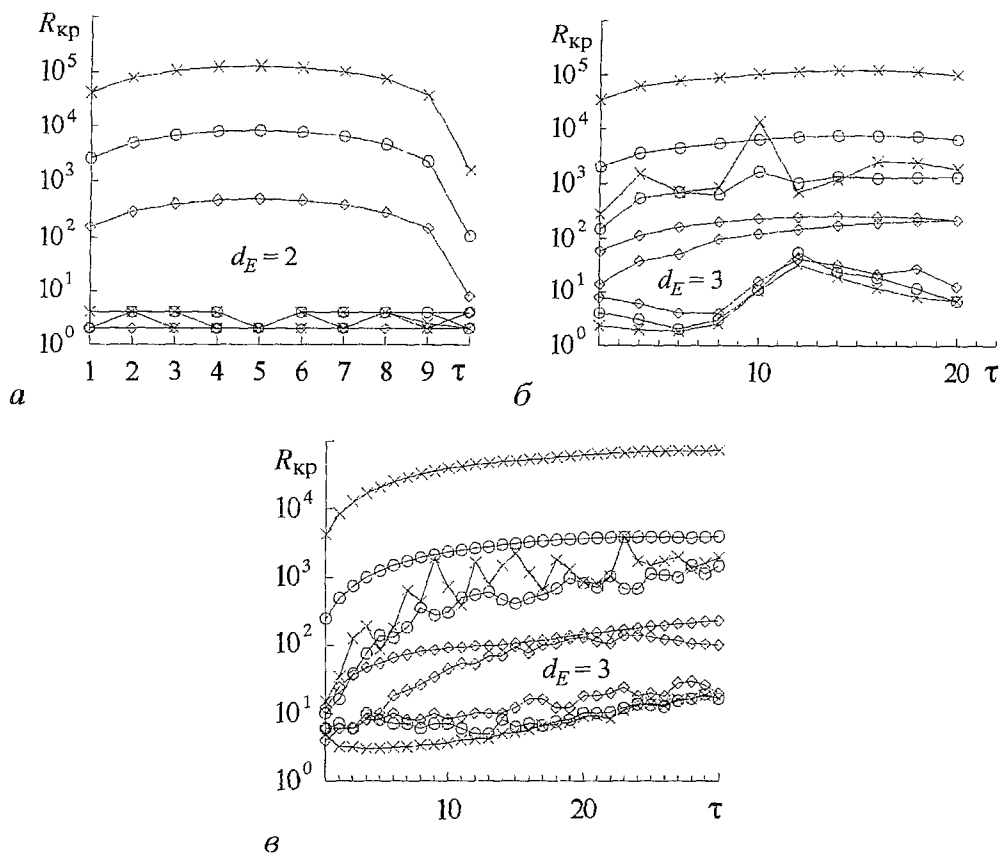


Рис. 5. Влияние разрядности  $p$  АЦП на размер области оптимального значения размерности вложения  $d_E$ : а -  $x(t)=\sin t$ ,  $\Delta t=2\pi/(11.5 \cdot 3^{1/2})$ ,  $N=15000$ ; б - система Ресслера шаг интегрирования 0.1,  $N=20000$ ; в - система Лоренца шаг интегрирования 0.004,  $N=20000$ . Здесь  $\diamond$  -  $p=8$ ,  $\circ$  -  $p=12$ ,  $\times$  -  $p=16$

уменьшается количество ложных соседей, появившихся из-за недостаточной разрядности. При увеличении разрядности область оптимальной размерности вложения будет увеличиваться, но при определенном значении  $p$  (которое зависит и от сложности сигнала) наступает момент, когда дальнейшее наращивание разрядности не имеет смысла, так как ширина этой области по параметру  $R_{кр}$  выходит за пределы диапазона  $10 \div 100$ . Для синусоиды (рис.5, а) это происходит при  $p=8$ , а для систем Ресслера и Лоренца при  $p=12$  (рис. 5, б, в).

Изложенные результаты для простого примера (синусоида) и более сложных (хаотического режима систем Ресслера и Лоренца) позволяют сказать, что применительно к малоразмерным системам при выборе параметров метода ложных соседей целесообразно придерживаться следующих соображений.

1.  $R_{кр}$  желательно выбирать в диапазоне  $10 \div 100$ .

2. С увеличением разрядности оцифровки область оптимальной размерности вложения увеличивается. В рассмотренных малоразмерных примерах приемлемым является значение  $p=12$ , что соответствует реальным АЦП.

3. Увеличение числа точек (длины реализации) при постоянном шаге  $\Delta t$  расширяет область допустимых значений  $R_{кр}$ , одновременно с этим увеличивается нижний предел  $R_{кр}$ .

4. При соблюдении рекомендаций п. 1-3 выбор метрики ( $L^1$  или  $L^2$ ) практически не существен, но для некоторых систем (например, Ресслера) метрика  $L^1$  дает лучшие результаты.

Вопрос о времени задержки наиболее сложный. Выбор  $\tau$  для гармонического

сигнала не вызывает трудностей: около  $T/4$ , что эквивалентно выбору  $\tau$  в первом нуле АКФ. Для рассматриваемого хаотического режима в системе Ресслера выбор  $\tau$  в первом нуле АКФ ( $\tau_{\text{АКФ}}=14$ ) поднимает (от 10 до 40) нижнюю границу области оптимальной размерности, тем самым уменьшая ее ширину. Аналогичная процедура для системы Лоренца ( $\tau_{\text{АКФ}}=103$  при  $10 < R_{\text{кр}} < 100$ ) дает более высокую размерность. Выбор  $\tau$  в первом минимуме функции взаимной информации (mutual information function) [5] для системы Лоренца ( $\tau_{\text{МИФ}}=4$ ) несколько улучшает ситуацию, но не на много (диапазон  $R_{\text{кр}}$  уменьшается со стороны низких размерностей). Для системы Ресслера значение  $\tau_{\text{АКФ}} \approx \tau_{\text{МИФ}}=15$  также не дает желаемого результата. Визуальный выбор  $\tau$ , при котором фазовый портрет имеет наибольший поперечный размер, но еще без «перехлестов», дает достаточно удовлетворительные результаты. В этом случае для системы Лоренца  $10 < \tau < 25$ , а для системы Ресслера  $3 < \tau < 11$ .

*Работа выполнена при поддержке РФФИ (грант № 96-02-16755).*

### **Библиографический список**

1. *Takens F.* // Dynamical Systems and Turbulence / Eds. by D. Rand and L.S. Young. Warwick, 1980.
2. *Casdagli M., Sauer T., Yorke J.A.* Embedology // J. Stat. Phys. 1991. Vol. 65. P. 579.
3. *Kennel M.B., Brown R., Abarbanel H.D.I.* Determining Minimum Embedding Dimension using a Geometrical Construction // Phys. Rev. A. 1992. Vol. 45. P. 3403.
4. *Abarbanel H.D.I. and Kennel M.B.* Local False Nearest Neighbors and Dynamical Dimensions from Observed Chaotic Data // Phys. Rev. E. 1993. Vol. 47. P. 3057.
5. *Fraser A.M., Swinney H.L.* Independent coordinates for strange attractors from mutual information. // Phys.Rev. A. 1986. Vol. 33, №2. P. 1134.

*Саратовский государственный  
университет*

*Поступила в редакцию 15.11.97  
после переработки 26.12.97*

## **FALSE NEAREST NEIGHBORS: INFLUENCE OF METHOD PARAMETERS ON MINIMUM EMBEDDING DIMENSION ESTIMATION**

*A.V. Kraskov, S.A. Rakitin*

In the paper, simple test systems are used to illustrate the false nearest neighbors method which is among the tools for analyzing time series. We study the influence of some parameters of the above method (data set length, number of bits in binary data representation, metrics chosen, and delay time) on minimum embedding dimension estimation.

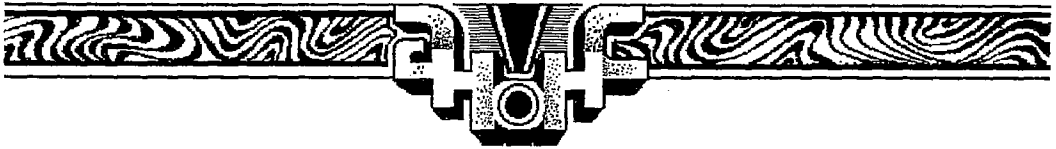


*Красков Александр Владимирович – студент третьего курса Высшего колледжа прикладных наук СГУ. Соросовский студент. Область научных интересов – методы обработки временных рядов.*





*Ракитин Сергей Александрович* – родился в 1973 году в Саратове, окончил физический факультет Саратовского государственного университета (1995). В настоящее время является аспирантом кафедры электроники и волновых процессов СГУ. Область научных интересов – анализ временных рядов.



*Вышел в свет журнал*

ISSBN: 1026-0226 DDSFA

**DISCRETE DYNAMICS IN NATURE AND SOCIETY**

**An International Multidisciplinary Research and Review Journal**

**Editors-in-Chief**

Vladimir Gontar and Michael Sonis

**Contents**

Vol. 1, N 3 (1997) 179–242

Synchronization of Spatiotemporal Chaos Using Nonlinear Feedback Functions. *M.K. Ali and J.Q. Fang* . . . . . 179

Catastrophe Models and the Expansion Method: a Review of Issues and an Application to the Econometric Modeling of Economic Growth. *E. Casetti* . . . . 185

Endogenous Oscillations in a Discrete Dynamic Model with Inventory. *A. Matsu-moto* . . . . . 203

Well-Posedness of Difference Elliptic Equation. *P.E. Sobolevskii* . . . . . 219

About Chaotization Mechanisms of the Distributed Dynamical Systems Which are Close to Discrete. *Yu.P. Bliokh, M.G. Lyubarsky and V.O. Podobinsky* . . . . 233

