

11.4;01.1

Влияние шума наблюдений на эффект ложной связи между осцилляторами при оценивании по временным рядам

© С.Н. Крылов^{1,2}, Д.А. Смирнов¹, Б.П. Безручко^{1,2}¹ Саратовский филиал Института радиотехники и электроники им. В.А. Котельникова РАН, Саратов, Россия² Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, Саратов, Россия
E-mail: krylov.sn@nxt.ru

Поступило в Редакцию 8 апреля 2019 г.

В окончательной редакции 8 апреля 2019 г.

Принято к публикации 30 апреля 2019 г.

Практика выявления структуры связей между элементами сложных систем по экспериментальным записям их колебаний (временным рядам) методом причинности по Винеру–Грейнджеру обнаружила ряд проблем, мешающих получению достоверных результатов. В частности, наличие шума наблюдений может привести к „эффекту ложной связи“ — получению вывода о взаимной связи двух элементов, реально связанных однонаправленно. Проведен количественный анализ этого явления и даны рекомендации для снижения его вероятности. Показано, что эффект проявляется, как правило, лишь при сильных шумах, сопоставимых с уровнем наблюдаемых колебаний, но отмечены и более редкие ситуации, в которых он имеет место при значительно меньших шумах.

Ключевые слова: связанные осцилляторы, временные ряды, оценки направленных связей.

DOI: 10.21883/PJTF.2019.16.48146.17829

Необходимость выявления и количественной оценки связи между двумя колебательными системами по их наблюдаемым временным рядам $\{x_1(t)\}$ и $\{x_2(t)\}$, где $t = 1, 2, \dots$ — дискретное время, возникает в различных областях физических и междисциплинарных исследований (см., например, [1–8]). При этом во многих ситуациях важен вопрос о том, является ли связь одно- или двунаправленной [1–3, 9–11]. Для ответа широко используется метод причинности по Винеру–Грейнджеру, развитый в прикладной математике и эконометрике [1, 2] и становящийся все более популярным среди физиков [3, 5, 6, 8–15]. В нем оценка силы воздействия системы j на систему i основана на сравнении прогноза состояния (характеризующей величины) системы i в данный момент времени с помощью двух авторегрессионных моделей: совместной и индивидуальной. В совместной модели для прогноза используются предыдущие состояния обеих систем:

$$x_i(t) = a_{i,0} + \sum_{n=1}^{p_i} a_{i,n} x_i(t-n) + \sum_{n=1}^{q_i} b_{i,n} x_j(t-n) + \eta_i(t),$$

$$i, j = 1, 2, i \neq j, \quad (1)$$

где $[\eta_1(t), \eta_2(t)]$ — двумерный белый шум. В индивидуальной модели используются прошлые данные только от самой прогнозируемой системы:

$$x_i(t) = A_{i,0} + \sum_{n=1}^{p_i} A_{i,n} x_i(t-n) + \xi_i(t), \quad i = 1, 2, \quad (2)$$

где ξ_i — нормальные белые шумы. В качестве оценки величины направленной связи обычно принимается

нормированное улучшение прогноза:

$$G_{j \rightarrow i} = (\sigma_i^2 - \sigma_{i|j}^2) / \sigma_i^2, \quad (3)$$

где $\sigma_i^2, \sigma_{i|j}^2$ — дисперсии ошибок прогноза индивидуальной и совместной моделей, которые оцениваются вместе с коэффициентами авторегрессионных моделей по временному ряду методом наименьших квадратов. Величина $G_{j \rightarrow i}$ может принимать значения от 0 до 1. Если обе величины $G_{1 \rightarrow 2}$ и $G_{2 \rightarrow 1}$ статистически значимо больше нуля, то делается вывод о наличии двунаправленной (взаимной) связи.

Описанная методика часто и продуктивно используется разными авторами, однако практика продемонстрировала и возможность получения ложных выводов [9–15] из-за влияния различных факторов, в частности шума наблюдений — измерительного шума [10, 12–15]. С учетом этого были предложены специальные тесты на однонаправленность связи [9, 13, 14], которые, однако, требуют детальных априорных данных об исследуемых системах и трудоемки с вычислительной точки зрения. Поэтому на практике важно уметь оценить заранее, вероятен ли эффект ложной связи, т. е. можно ли обойтись описанным традиционным методом или нужны специальные тесты. Для выработки эмпирических правил такой предварительной оценки далее исследуются условия проявления эффекта ложных связей для системы из двух осцилляторов в зависимости от уровня шума наблюдений и различных параметров. Рассматривались линейные диссипативные осцилляторы, заданные стоха-

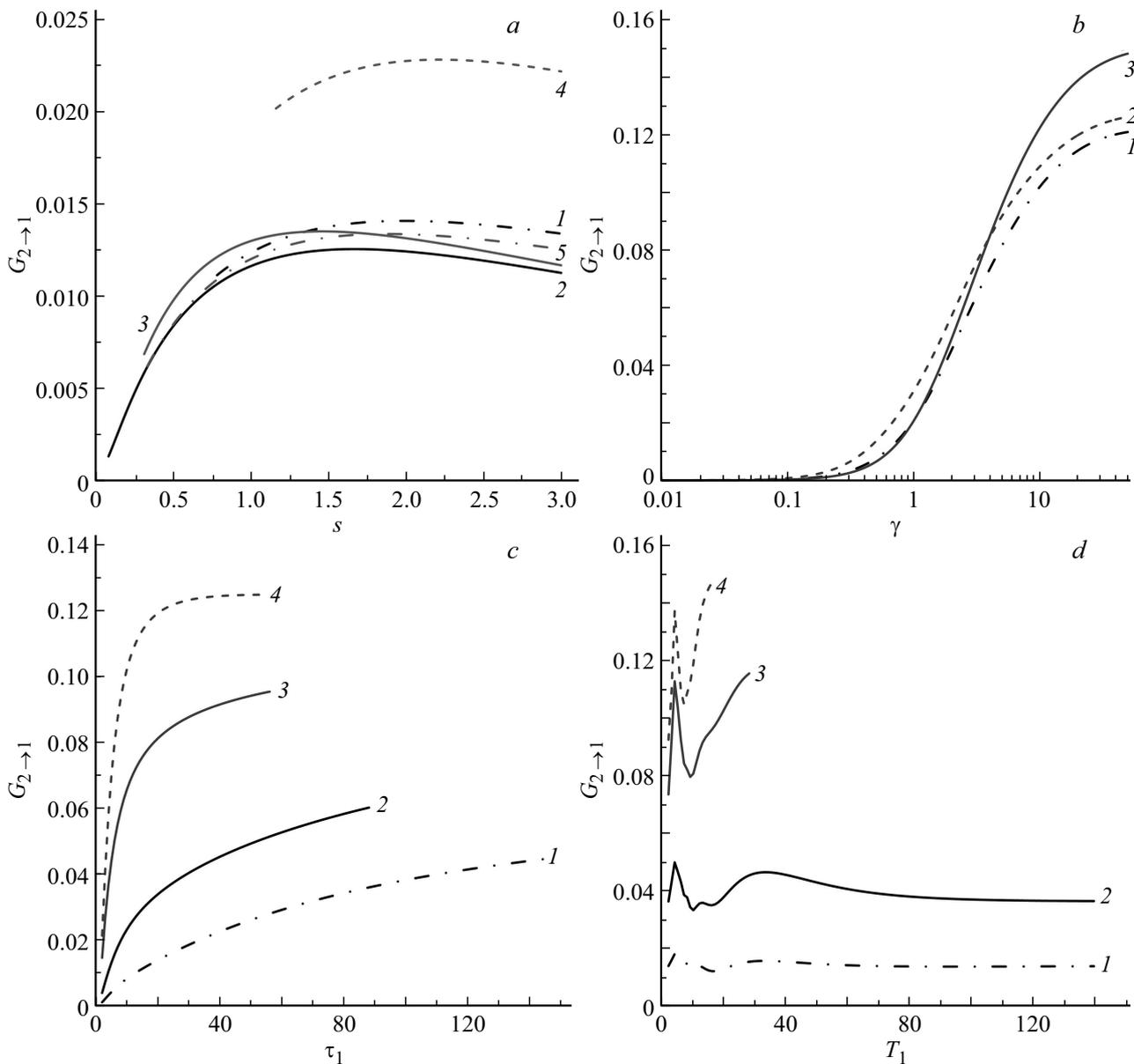


Рис. 1. Величина улучшения прогноза в ложную сторону $G_{2 \rightarrow 1}$ при умеренной корреляции $\rho^2 < 0.5$ и базовом наборе параметров $T_1 = T_2 = 10$, $\tau_1 = \tau_2 = 20$ (если не указаны специально другие значения) в зависимости от уровня шума наблюдений s при $\gamma = 0.75$ (a) (1 — для базового набора, 2 — $T_2 = 15$, 3 — для $T_1 = 15$, 4 — для $\tau_1 = 40$, 5 — для $\tau_2 = 40$), от параметра связи γ при $s = 2$ (b) (1 — для базового набора, 2 — для $\tau_1 = 40$, 3 — для $T_1 = 15$), от времени релаксации ведущего осциллятора τ_1 при $s = 2$ (c) (1 — для $\gamma = 0.75$, 2 — для $\gamma = 1.5$, 3 — для $\gamma = 5$, 4 — для $\gamma = 35$) и от периода ведущего осциллятора T_1 при $s = 2$, $T_2 = 10$, $\tau_1 = 20$, $\tau_2 = 20$ (d) (1–4 — для тех же значений γ , что и на части c).

стическими разностными уравнениями

$$\begin{aligned}
 X_1(t) &= a_1 X_1(t-1) + b_1 X_1(t-2) + \xi_1(t), \\
 X_2(t) &= a_2 X_2(t-1) + b_2 X_2(t-2) + \xi_2(t) + c X_1(t-1),
 \end{aligned}
 \tag{4}$$

где c — коэффициент связи, $a_i = 2 \cos(2\pi/T_i) \exp(-1/\tau_i)$, $b_i = -\exp(-2/\tau_i)$, T_i и τ_i — собственные периоды и времена релаксации колебаний, выраженные в единицах числа шагов дискретного времени, ξ_i — гауссовы взаимно некоррелированные белые шумы с

дисперсиями $\sigma_{\xi_i}^2$ ($i = 1, 2$). Наблюдаемые величины $x_i(t) = X_i(t) + \xi_i(t)$ — это суммы переменных состояния системы (4) и независимых друг от друга гауссовых белых шумов. Такой выбор модельной системы обусловлен ее общностью и простотой. Она допускает точное аналитическое и численное исследование величин $G_{j \rightarrow i}$ без статистического оценивания и соответствующих погрешностей, а именно значения $G_{j \rightarrow i}$ рассчитывались через матрицу ковариационных функций процесса (x_1, x_2) [10,11].

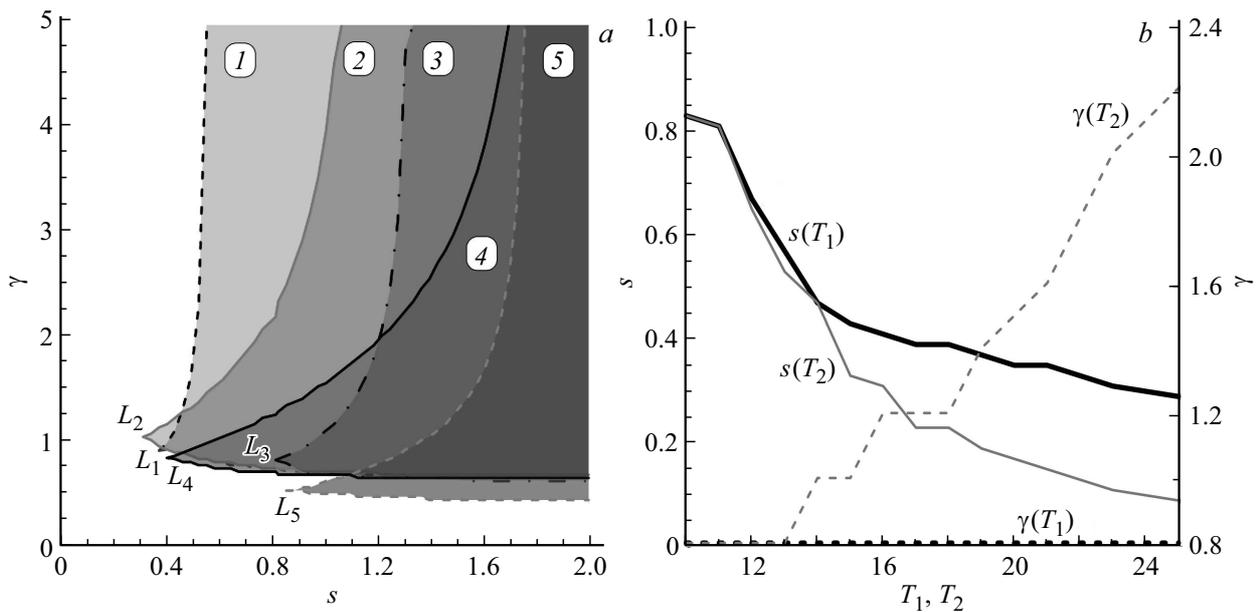


Рис. 2. Области эффекта ложной связи $G_{2 \rightarrow 1} \geq 0.01$ при $\rho^2 \leq 0.5$ и базовом наборе параметров $T_1 = T_2 = 10$, $\tau_1 = \tau_2 = 20$ (если не указаны специально другие значения). *a* — на плоскости (γ, s) : область 1 — для $\tau_2 = 40$, область 2 — для $T_2 = 15$, область 3 — для базового набора, область 4 (ограничена жирной линией) для $T_1 = 15$, область 5 — для $\tau_1 = 40$; *b* — координаты точки „сборки“ L при вариации T_1 или T_2 относительно базового набора: зависимости $\gamma(T_1)$, $s(T_1)$, $\gamma(T_2)$ и $s(T_2)$.

Для оценки типичности и величины эффекта ложных связей значения $G_{2 \rightarrow 1}$ и $G_{1 \rightarrow 2}$ рассчитывались в широкой области значений параметров осцилляторов и уровней шума: собственные периоды T_i менялись от 2 до 40 шагов дискретного времени, времена релаксации τ_i — от 2 до 60 шагов, величина $\gamma = c \sqrt{\langle x_1^2 \rangle} \sigma_{\xi_2}$ — от 0.01 до 50 (это относительный вклад слагаемого связи в уравнение (4), т.е. некий параметр связи [11], его значения задавались путем задания нужного значения s при прочих заданных величинах, угловые скобки означают математическое ожидание). Далее представлены результаты для случая зашумленного ряда x_1 и „чистого“ ряда $x_2 = X_2$, так как именно такой вариант шума наблюдений приводит к наибольшим значениям „ложной связи“ $G_{2 \rightarrow 1}$. Отношение s дисперсии шума наблюдений ξ_1 к дисперсии „чистого“ сигнала X_1 менялось от 0.01 до 30. Проявлением эффекта ложных связей считалось значение $G_{2 \rightarrow 1} > 0.01$ [11]. Далее представлены лишь ситуации, где квадрат максимума взаимной корреляционной функции между x_1 и x_2 составляет $\rho^2 < 0.5$ (умеренные корреляции), так как превышение этого порога на практике предупреждает о трудностях и делает эту ситуацию не столь опасной с точки зрения ложных выводов о связи.

Согласно полученным результатам, проявление эффекта ложной связи наиболее сильно зависит от τ_1 и в меньшей степени от других параметров (рис. 1). Так, на рис. 1, *a* ($\gamma = 0.75$) кривая 1 представляет зависимость „ложного“ $G_{2 \rightarrow 1}$ от уровня шума наблюдений s при базовом наборе параметров $T_1 = T_2 = 10$, $\tau_1 = \tau_2 = 20$.

Остальные кривые показывают результат вариации параметров. Большой эффект вызывает лишь изменение времени релаксации ведущего осциллятора $\tau_1 = 40$: кривая 4 далека от кривой 1, а изменения собственных периодов (кривые 2 и 3) и времени релаксации ведомого осциллятора τ_2 (кривая 5) не дают таких отклонений. Зависимость $G_{2 \rightarrow 1}(T_1)$ существенна лишь при малых периодах (менее десяти точек на периоде) и больших γ (рис. 1, *d*), но и тогда относительные изменения $G_{2 \rightarrow 1}$ при вариации T_1 (рис. 1, *d*) гораздо слабее, чем при вариации τ_1 (рис. 1, *c*). Зависимость $G_{2 \rightarrow 1}$ от других параметров еще слабее: относительное изменение $G_{2 \rightarrow 1}$ при вариации τ_2 составляет не более 1%, а при вариации T_2 — не более 1.5% (при $\gamma \leq 1.5$). На рис. 1, *a* эффект ложной связи проявляется (т.е. $G_{2 \rightarrow 1} > 0.01$) при большом уровне шума наблюдений ($s > 0.5$), а максимален при весьма сильном (примерно при $s = 2$). Зависимости $G_{2 \rightarrow 1}(\gamma)$ при $s = 2$ показывают монотонный рост $G_{2 \rightarrow 1}$ с ростом силы связи и насыщение при больших связях $\gamma \approx 35$ (рис. 1, *b*), максимальное значение $G_{2 \rightarrow 1}$ составляет 0.15 при $\gamma = 50$. Однако достаточно большие $G_{2 \rightarrow 1}$ (до 0.04) имеют место и при гораздо меньших связях, например при $\gamma = 0.75$ в случае $s = 2$ и большого τ_1 (рис. 1, *c*, кривая 1).

На рис. 2, *a* оттенками серого и ограничивающими кривыми показаны области в пространстве параметров s и γ , где проявляется эффект ложной связи при $\rho^2 \leq 0.5$. Точки „сборки“ L соответствуют минимальному значению уровня шума наблюдений, при котором проявляется эффект. Положение точки L сложным образом зависит от параметров осцилляторов. Несмотря на указанную

выше доминирующую зависимость $G_{2 \rightarrow 1}$ от τ_1 , зависимость положения L от τ_1 не так сильна. Наиболее сильно она зависит от T_1 , T_2 , графики координат точки L — $\gamma(T_1)$, $s(T_1)$ и $\gamma(T_2)$, $s(T_2)$ — на рис. 2, b показывают, что эффект ложной связи возможен и при умеренных уровнях шума наблюдений, например при $s = 0.1$, т.е. при 10% шуме наблюдений (рис. 2, b , линия для $s(T_2)$).

Отметим, что полученное в работе [15] для систем с двухленточным хаотическим аттрактором или с циклом удвоенного периода повышение чувствительности метода (рост улучшения прогноза в правильную сторону) с ростом шума наблюдений при достаточно больших шумах представляется более близким к эффекту ложных связей, чем к улучшению работы метода. Действительно, в этом случае должен иметь место и рост улучшения прогноза в ложную сторону (что не анализировалось в [15]) из-за отмеченного в [15] ухудшения индивидуальных моделей и значительной взаимной корреляции сигналов, обусловленной синхронным чередованием состояний систем между лентами аттрактора. Более того, из-за этой синхронности даже для несвязанных систем должно проявляться ненулевое улучшение прогноза в обе стороны при наличии шума наблюдений. Так что более полно данную ситуацию при росте шума наблюдений можно описать как „эффект ложных выводов о направленной связи из-за синхронной компоненты в сигналах“, а повышение чувствительности метода здесь, скорее, случайное обстоятельство.

Итак, в работе показано, что эффект ложной связи из-за шума наблюдений проявляется даже при умеренной взаимной корреляции наблюдаемых временных рядов, если уровень шума больше некоторого значения, зависящего от индивидуальных параметров осцилляторов. Эффект типичен и силен при очень больших уровнях шума, когда дисперсия шума лежит в диапазоне от половины до удвоенной дисперсии чистого сигнала. Это не обычная на практике ситуация, так что мы делаем вывод, что чаще всего шум наблюдений не вынуждает отказываться от обычного метода оценки связей. Однако при определенных параметрах осцилляторов, о которых можно судить по данным рис. 2, эффект может проявляться и при гораздо меньших уровнях шума — порядка 10% от дисперсии чистого сигнала. По этим результатам и имеющимся сведениям о параметрах исследуемых систем на практике можно заранее оценить, ожидается ли заметный эффект ложных связей, и если все же ожидается, то нужно применять специализированные методы оценки связи и дополнительные тесты, основанные на имеющейся априорной информации об исследуемых системах [9,13,14].

Финансирование работы

Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант № 17-02-00307).

Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

Список литературы

- [1] Wiener N. // Modern mathematics for the engineer / Ed. E.F. Beckenbach. N.Y.: McGraw–Hill, 1956. P. 165–190.
- [2] Granger C.W.J. // Information and Control. 1963. V. 6. P. 28–48.
- [3] Hung Y.-C., Hu C.-K. // Phys. Rev. Lett. 2008. V. 101. P. 244102.
- [4] Sugihara G., May R., Ye H., Hsieh C., Deyle E., Fogarty M., Munch S. // Science. 2012. V. 338. P. 496–500.
- [5] Runge J., Petoukhov V., Donges J.F., Hlinka J., Jajcay N., Vejmelka M., Hartman D., Marwan N., Paluš M., Kurths J. // Nature Commun. 2015. V. 6. P. 8502.
- [6] San Liang X. // Phys. Rev. E. 2016. V. 94. P. 052201.
- [7] Stokes P.A., Purdon P.L. // Proc. Natl. Acad. Sci. USA. 2017. V. 114. P. 7063–7072.
- [8] Faes L., Nollo G., Stramaglia S., Marinazzo D. // Phys. Rev. E. 2017. V. 96. P. 042150.
- [9] Смирнов Д.А., Безручко Б.П. // Изв. вузов. Радиофизика. 2012. Т. 55. № 10–11. С. 736–749.
- [10] Smirnov D.A. // Phys. Rev. E. 2013. V. 87. P. 042917.
- [11] Крылов С.Н., Смирнов Д.А., Осипов Г.В., Безручко Б.П. // Письма в ЖТФ. 2015. Т. 41. В. 11. С. 94–102.
- [12] Smirnov D.A., Marwan N., Breitenbach S.F.M., Lechleitner F., Kurths J. // Europhys. Lett. 2017. V. 117. P. 10004.
- [13] Nalatore H., Ding M., Rangarajan G. // Phys. Rev. E. 2007. V. 75. P. 031123.
- [14] Nalatore H., Rangarajan G. // Phys. Rev. E. 2014. V. 90. P. 062127.
- [15] Медведева Т.М., Парамонов Ф.Б., Сысоев И.В. // Нелинейная динамика. 2015. Т. 11. № 4. С. 657–670.