

УДК 550.3

ПРОБЛЕМА ФЛИККЕР-ШУМА ПРИ ИЗУЧЕНИИ ПРИЧИННО-СЛЕДСТВЕННЫХ СВЯЗЕЙ МЕЖДУ ПРИРОДНЫМИ ПРОЦЕССАМИ

© 2003 г. А. В. Дещеревский, А. Я. Сидорин

Представлено академиком В.Н. Страховым 21.02.2003 г.

Поступило 05.03.2003 г.

При изучении причинно-следственных связей между природными процессами широко используется взаимный корреляционный анализ временных рядов данных. Однако при этом очень часто упускается из вида, что экспериментальные серии, как правило, не соответствуют тем моделям данных, на которых основаны используемые методы анализа. В работе представлены результаты многолетнего мониторинга электротеллурического поля (ЭТП) и электрической активности (ЭА) слабозлектрической рыбы. Обнаружено сходство спектров и высокая (до -0.44 при длине рядов 1394 отсчета) корреляция между указанными процессами, сохраняющаяся после фильтрации сезонного хода и вычитания тренда. Установлено, что оба ряда имеют фликкер-шумовую структуру. Показано, что вследствие этого невозможно отвергнуть альтернативную гипотезу о независимости анализируемых процессов.

ЭКСПЕРИМЕНТАЛЬНЫЕ ДАННЫЕ

Мониторинг (ЭТП и ЭА) слабозлектрической рыбы проводился на Гармском полигоне в рамках эксперимента по прогнозу землетрясений [1]. Нильский слоник *Gnathonemus leopoldianus* – это небольшая рыбка, обитающая в мутной воде и использующая для ориентации в пространстве и коммуникации электрические импульсы, причем чувствительность электрорецепторной системы нильского слоника достигает 10^{-6} В/м. Рыба содержалась в специальном помещении при постоянной температуре 26°C в бескаркасном аквариуме из оргстекла размерами $0.5 \times 0.5 \times 0.5$ м. Наблюдения продолжались с января 1983 г. по октябрь 1986 г. ЭА рыбы определяли частотой генерируемых ей электрических импульсов [2]. Анализировали ряд количества импульсов, гене-

рируемых рыбой в течение часа, осредненных по суточным интервалам (рис. 1, 1).

Измерения ЭТП проводили с 1979 по 1992 г. в 16 км от пункта наблюдений за рыбами [1]. В качестве датчиков использовали свинцовые электроды, закопанные на глубину 2 м и образующие измерительную линию субширотной ориентации длиной 500 м (рис. 1, кривая 2).

ВИЗУАЛЬНЫЙ, СПЕКТРАЛЬНЫЙ И КОРРЕЛЯЦИОННЫЙ АНАЛИЗ ИСХОДНЫХ ДАННЫХ

Как показывает визуальный анализ исходных данных, вариации ЭА рыбы и ЭТП практически точно противофазны. Совпадают как отдельные пики и бухты, так и общие тренды активности. Весьма схожи и спектры: на фоне общего преобладания длиннопериодных вариаций отчетливо выделяются совпадающие пики на периодах 1 год, 3.0 мес., 3.7 мес., 1.0 мес. и близкие пики на периодах 0.7, 5.5–6 и 8 мес. (рис. 2).

Формально рассчитанный коэффициент корреляции между исходными реализациями равен -0.39 (табл. 1). По техническим причинам ряд бионаблюдений содержал пропуски общим числом 77. В случае линейной интерполяции пропущенных данных корреляция возрастает до -0.44 .

Ключевой элемент интерпретации – оценка значимости результатов. Мог ли такой коэффициент корреляции получиться случайно или вероятность этого слишком мала и гипотезу о независимости сопоставляемых процессов необходимо отвергнуть? Стандартные оценки 95%-го доверительного уровня [3] приводят к значению 0.05, что указывает на ничтожно малую вероятность случайно получить значение 0.39 или 0.44. Однако такое заключение, как известно, обусловлено требованиями стационарности и гауссовости сопоставляемых процессов. Если обратиться к экспериментальным данным, легко установить, что ни первое, ни второе условие для рассмотренных реализаций не выполняются. Так, коэффициенты

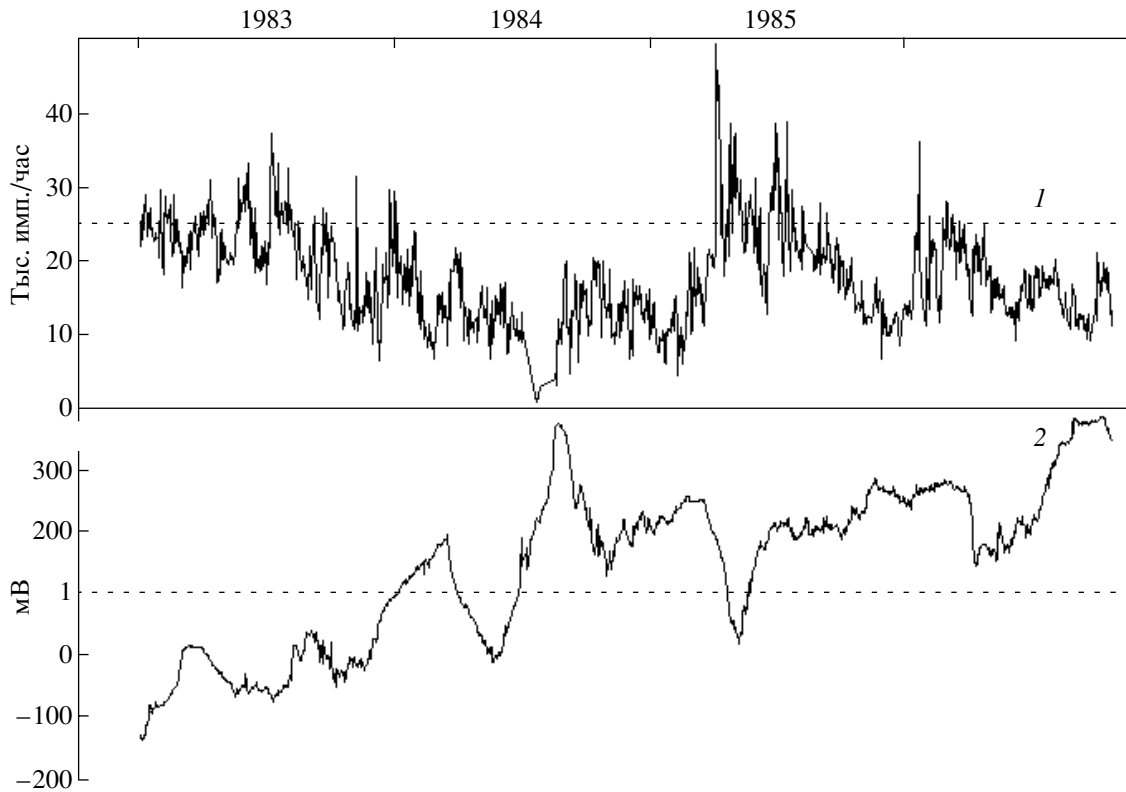


Рис. 1. Ряды среднесуточных значений количества импульсов, генерируемых рыбой в течение часа (1), и ЭТП (2).

асимметрии и эксцесса [3] для нормально распределенной случайной величины должны быть близки к 0, в действительности же для ЭА рыбы они равны 0.63 и 0.92, для ряда ЭТП –0.25 и –0.93. Критерий согласия χ^2 для рыбы утверждает отличие функции распределения от гауссовой с вероятностью не менее 0.997, для ряда ЭТП – с вероятностью не менее 0.999999.

Насколько это критично с точки зрения анализа значимости? В столбцах 3 и 4 табл. 1 приведены значения ранговых коэффициентов корреляции [3]. Как видно, разница соответственно с 1-м и 2-м столбцом небольшая, следовательно отклонением функции распределения от гауссовой в данном случае можно пренебречь.

Хорошо известно, что в наибольшей степени повлиять на результаты корреляционного анализа могут отклонения от стационарности. Строго говоря, в силу ограниченности длины реализации условие стационарности никогда не выполняется на практике, но можно проверить, насколько существенно проявления нестационарности, наблюдаемые в пределах рассматриваемого интервала времени.

Анализ рис. 1 показывает, что среднее значение ЭТП и ЭА рыбы в ходе наблюдений постоянно менялось. Простейший и достаточно широко распространенный способ борьбы с такой “неста-

ционарностью” – вычитание линейного тренда. Как видно из табл. 1, коэффициент корреляции между рядами в этом случае возрастает, что вновь приводит нас к выводу о неслучайности обнаруженной корреляции.

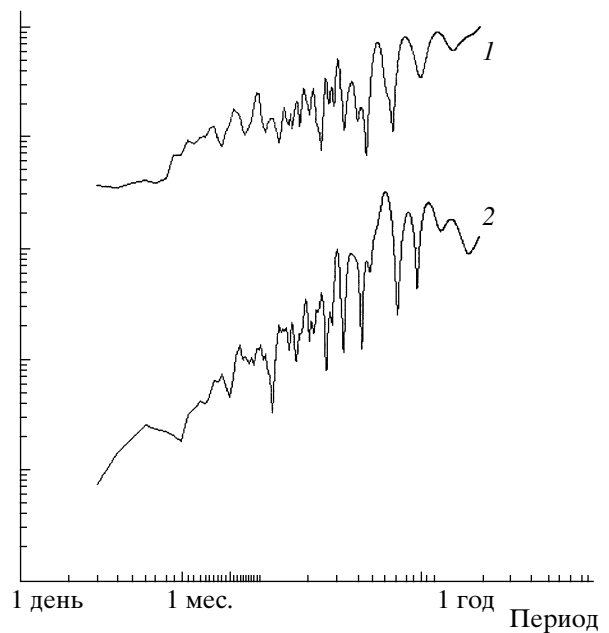


Рис. 2. Спектры реализаций, показанных на рис. 1.

Таблица 1. Корреляция между среднесуточными значениями ЭА рыбы и ЭТП, разные варианты расчета

Вариант расчета	1	2	3	4
Исходные реализации, линейная интерполяция пропусков	-0.39	-0.44	-0.38	-0.42
То же, отфильтрован сезонный ход	-0.38	-0.46	-0.37	-0.47
Исходные реализации, без интерполяции пропусков	-0.38	-0.39	-0.39	-0.38
То же, отфильтрован сезонный ход	-0.38	-0.38	-0.38	-0.38

Примечание. 1 – коэффициент корреляции между реализациями, описанными в графе “вариант расчета”; 2 – коэффициент корреляции между этими же реализациями после вычитания линейного тренда; 3, 4 – то же, что 1 и 2, ранговый коэффициент корреляции.

ВЛИЯНИЕ СЕЗОННЫХ ЭФФЕКТОВ И ДЕЛЕНИЕ РЯДА НА УЧАСТКИ

Присутствие в рядах сезонной периодичности также может нарушать условие стационарности, поэтому мы выполнили анализ реализаций с отфильтрованным сезонным ходом (см. табл. 1). Как видно, после вычитания сезонной компоненты корреляция между остатками не уменьшается. В зависимости от порядка выполнения операций: вычитания сезонного хода, линейного тренда и интерполяции пропусков, – а также особенностей методики расчета сезонной составляющей [4] коэффициент корреляции между отфильтрованными реализациями заметно варьируется (в табл. 1 показаны не все варианты расчета), но во всех случаях он многократно превосходит формальный доверительный уровень 0.05.

Отметим, что заметная антикорреляция между рассматриваемыми процессами сохраняется и в случае анализа данных по двухлетним интервалам времени. В зависимости от предварительной обработки коэффициент корреляции по данным за 1983–1984 гг. меняется в пределах от -0.3 до -0.6, по данным за 1985–1986 гг. от -0.3 до -0.4. И только в случае разбиения рядов на более мелкие (годовые и менее) подынтервалы коэффициент корреляции начинает сильно варьироваться (меняется в разы или даже меняет знак).

Итак, представленные результаты, казалось бы, не оставляют сомнений в том, что ЭА рыбы и ЭТП меняются согласованно и что с вероятностью, близкой к 1, такая синхронизация не может быть результатом случайного совпадения.

ПРОБЛЕМА ФЛИККЕР-ШУМА ПРИ КОРРЕЛЯЦИОННОМ АНАЛИЗЕ

В геофизике, как и в большинстве других областей знаний, большое место отводится поиску причинно-следственных связей между различными явлениями и процессами. Такие задачи решаются, например, при поиске предвестников землетрясений, изучении солнечно-земных связей и взаимодействия геосферных оболочек, анализе

влияния геофизических и космических процессов на биосферу.

Ключевой метод большинства подобных исследований – взаимный корреляционный анализ временных рядов данных. Однако при этом очень часто упускается из вида, что анализируемые данные, как правило, не соответствуют базовым статистическим моделям. Например, для чрезвычайно широко распространенных в природе фликкер-шумовых процессов (см., например, [6–14]), в том числе для рассматриваемых в настоящей работе реализаций (рис. 2), наблюдается рост спектральной мощности в низкочастотном пределе спектра, что прямо указывает на отсутствие стационарности. А ведь классические оптимальные процедуры статистических оценок становятся совершенно непригодными уже при малейших нарушениях строгих модельных допущений [5]. Насколько это существенно в нашем случае? Обратимся к численному эксперименту.

На компьютере был сгенерирован массив реализаций белого шума длиной 1394 отсчета каждая. Как показали тесты, все они неотличимы от гауссова процесса; коэффициент парной корреляции во всех случаях не превосходил по модулю 0.06, что вполне согласуется с теоретическими оценками для независимых случайных рядов.

Далее ряды были “проинтегрированы”: каждое значение ряда заменено суммой всех предшествующих значений исходного ряда. Как известно [15], спектральные свойства построенных таким образом модельных рядов приближаются к свойствам рассматриваемых в настоящей работе экспериментальных реализаций, что дает основание использовать их для численного эксперимента.

Выполненные расчеты показали, что коэффициент парной корреляции для всех пар “проинтегрированных” модельных рядов (всего было рассмотрено 152 пары) почти с одинаковой вероятностью принимает любые значения от -0.8 до + 0.8 (см. табл. 2). Средний модуль коэффициента корреляции с хорошей точностью равен 0.40, что на порядок превосходит стандартный 95%-ный доверительный уровень.

Вне всяких сомнений полученный результат не связан с неидеальностью генератора случайных чисел или какими-либо аналогичными техническими причинами. Все гораздо проще. Хотя связь между рядами отсутствует, внутри каждого ряда последовательные значения сильно коррелированы, причем эта взаимозависимость распространяется на неограниченные интервалы времени. Этим модельные ряды в корне отличаются от белого шума, для которого последовательные отсчеты полностью независимы. Если разложить модельные реализации в ряд Фурье, то окажется, что при любой длине ряда наибольшую амплитуду будут иметь вариации с периодом, сравнимым с продолжительностью наблюдений (это прямо следует из структуры спектров). Именно вклад подобных гармоник определяет значение коэффициента парной корреляции для подобных реализаций. Если у двух независимо полученных модельных или экспериментальных реализаций указаны колебания-доминанты случайно окажутся синфазны (или противофазны), то коэффициент корреляции для этой пары может быть сколь угодно близок к единице. И только присутствие менее длиннопериодных составляющих не позволяет достичь точного равенства.

Ясно, что полученный результат не зависит от длины реализации. Рассмотренный эксперимент наглядно показывает, что для рядов со свойствами фликкер-шума обычные методы оценки доверительных уровней абсолютно неприменимы, и их нельзя использовать даже для весьма приближенной, грубой проверки значимости рассчитанных коэффициентов корреляции.

ИЗМЕНЕНИЕ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ ПРИ ЧАСТОТНОЙ ФИЛЬТРАЦИИ

В табл. 3 приведены данные о коэффициенте корреляции реализаций, отфильтрованных от низкочастотных трендов. Тренд оценивался методом скользящего среднего при различной ширине сглаживающего окна. Для улучшения амплитудно-частотной характеристики фильтра использовалась гауссова весовая функция окна.

Как видно, по мере перехода к более высокочастотным компонентам рядов корреляция быстро падает, а вот тренды демонстрируют хорошее согласие. В принципе, это могло бы означать, что рыбы реагируют на медленные, продолжительностью несколько лет, изменения ЭТП и не реагируют на более быстрые вариации. Однако такая гипотеза представляется маловероятной в силу как физических, так и чисто статистических обстоятельств. Напротив, уменьшение коэффициента корреляции по мере перехода к более высоким частотам, наблюдающееся для экспериментальных рядов (табл. 3), практически точно

Таблица 2. Результаты модельного эксперимента. Гистограмма распределения значений модуля коэффициента корреляции для независимо сгенерированных случайных серий со свойствами фликкер-шума

K	N
0.00–0.10	26
0.10–0.20	20
0.20–0.30	18
0.30–0.40	12
0.40–0.50	14
0.50–0.60	25
0.60–0.70	18
0.70–0.80	13
0.80–0.90	6
0.90–1.00	0

Примечание. K – диапазон значений модуля коэффициента корреляции, N – число пар.

Таблица 3. Корреляция между среднесуточными значениями ЭА рыбы и ЭТП после фильтрации низкочастотных трендов

Величина сглаживающего окна	1	2
Исходные реализации	–0.38	–
Линейный тренд	–0.39	–1.00
730 сут	–0.23	–0.56
365 сут	–0.14	–0.54
180 сут	–0.10	–0.51
90 сут	–0.04	–0.47
45 сут	–0.03	–0.45
Межсуточные изменения	0.04	–
Ранги межсуточных изменений	0.03	–

Примечание. 1 – высокочастотные компоненты, 2 – трендовые компоненты.

соответствует аналогичному уменьшению коэффициента корреляции для модельных реализаций (т.е. той ситуации, когда высокая корреляция между фликкер-шумовыми процессами обусловлена случайным совпадением трендов).

Как известно, ряды приращений ЭТП и ЭА рыбы (или, другими словами, их межсуточных изменений) в первом приближении можно рассматривать как реализации стационарного случайного процесса [9, 15]. Как показывает анализ (табл. 3), для этих рядов величина коэффициента корреляции не выходит за границу 95%-го доверительного интервала, что вполне согласуется с гипотезой о независимости процессов. (Отметим, что для еще более высокочастотных вариаций с периода-

ми менее суток, которые в данной работе не рассматривались, картина может быть иной.)

Итак, несмотря на идентичность спектров сопоставляемых рядов ЭА рыбы и ЭТП и высокую и стабильную корреляцию между ними, полученные результаты не могут считаться доказательством существования физически обусловленных причинных связей между этими процессами, поскольку допускают объяснение и в рамках альтернативной гипотезы о независимости анализируемых процессов.

Полученные результаты наглядно демонстрируют хорошо известный, но часто не принимаемый во внимание при анализе геофизических данных факт, что формальное использование стандартных методов корреляционного анализа при исследовании фликкер-шума может привести к ошибочным выводам. Это особенно актуально в связи тем, что подобные процессы исключительно широко распространены в природе. Необходимо в каждом случае тщательно проверять соответствие свойств анализируемых данных той модели, на которой основан метод, используемый для их анализа.

Работа выполнена при поддержке Российского фонда фундаментальных исследований по гранту 01-05-65503.

СПИСОК ЛИТЕРАТУРЫ

1. Гармский геофизический полигон / Под ред. А.Я. Сидорина. М.: ИФЗ АН СССР. 1990. 240 с.
2. Боровиков А.Г., Сидорин А.Я. Сейсмические приборы. М.: Наука, 1986. В. 18. С. 30–34.
3. Корн Г., Корн Т. Справочник по математике. М.: Наука, 1984. 831 с.
4. Дещереvский А.В., Сидорин А.Я. Некоторые вопросы методики оценки среднесезонных функций для геофизических данных. М.: ОИФЗ РАН, 1999. 40 с.
5. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике: подход на основе функций влияния. М.: Мир, 1989. 512 с.
6. Дещереvский А.В., Журавлев В.И., Сидорин А.Я. // ДАН. 1996. Т. 346. № 6. С. 815–818.
7. Дещереvский А.В., Лукк А.А., Сидорин А.Я. // Физика Земли. 1997. №7. С. 3–19.
8. Дещереvский А.В., Сидорин А.Я. // ДАН. 2001. Т. 376. № 1. С. 100–105.
9. Лукк А.А., Дещереvский А.В., Сидорин А.Я., Сидорин И.А. Вариации геофизических полей как проявление детерминированного хаоса во фрактальной среде. М.: ОИФЗ РАН, 1996. 210 с.
10. Тимашев С.Ф. // ЖФХ. 1993. Т. 67. № 4. С. 798–799.
11. Lapenna V., Macchiato M., Telesca L. // Phys. Earth and Planet. Inter. 1998. V. 106. P. 115–127.
12. Бердыев А.А., Мухамедов В.А. // ДАН. 1987. Т. 297. № 5. С. 1077–1081.
13. Шустер Г. Детерминированный хаос. М.: Мир, 1988. 240 с.
14. Bak P., Tang C., Wiesenfeld K. // Phys. Rev. Lett. 1987. V. 59. № 4. P. 381–384.
15. Дещереvский А.В., Журавлев В.И. Тестирование методики оценки параметров фликкер-шума. М.: ОИФЗ РАН, 1996. 12 с.