

УДК 551.58;519.23

ОЦЕНКА ВЗАИМОДЕЙСТВИЯ КЛИМАТИЧЕСКИХ ПРОЦЕССОВ: ЭФФЕКТ РЕДКОЙ ВЫБОРКИ РЯДОВ АНАЛИЗИРУЕМЫХ ДАННЫХ

© 2013 г. Д. А. Смирнов*, И. И. Мохов**

*Саратовский филиал Института радиотехники и электроники им. В.А. Котельникова РАН
410019 Саратов, ул. Зеленая, д. 38
E-mail: smirnovda@yandex.ru

**Институт физики атмосферы им. А.М. Обухова РАН
119017 Москва, Пыжевский пер., д. 3
E-mail: mokhov@ifaran.ru

Поступила в редакцию 29.10.2012 г., после доработки 28.02.2013 г.

Показано, что основанный на понятии “причинности по Грейнджеру” подход при однонаправленной связи и достаточно редкой выборке рядов анализируемых данных может приводить к ошибочным выводам о двунаправленной связи. Эффект выявлен при анализе связи вариаций потока солнечного излучения и глобальной приповерхностной температуры. Представлен статистический тест для подтверждения или опровержения предположений о характере связи (однонаправленном или двунаправленном). Соответствующий анализ связи явлений Эль-Ниньо и индийского муссона подтвердил ранее сделанные выводы об их взаимном воздействии.

Ключевые слова: временные ряды, оценки связей, однонаправленная связь, двунаправленная связь, причинность по Грейнджеру, температура, солнечная активность, Эль-Ниньо, Южное колебание, индийский муссон.

DOI: 10.7868/S0002351513050118

1. ВВЕДЕНИЕ

Все более актуальными становятся исследования не только связи различных глобальных и региональных климатических процессов, но и направления воздействия, которое может быть не односторонним и меняться во времени, в том числе и вследствие внешних естественных и антропогенных воздействий [1–28]. Для исследования причинно-следственных связей в земной климатической системе полезен подход, основанный на понятии “причинности по Грейнджеру” [29], позволяющий оценивать степень взаимного влияния климатических процессов с учетом внешних воздействий (см., например [13, 14, 18, 19, 21, 23, 25, 26, 28]). Связь двух процессов может быть как однонаправленной (ОС, с воздействием одного из двух исследуемых процессов на другой), так и двунаправленной (ДС, со взаимным воздействием), как, например, было выявлено при анализе причинности по Грейнджеру для явлений Эль-Ниньо и индийского муссона [25, 26].

Система Y влияет на систему X “по Грейнджеру”, если прогноз будущего поведения X улучшается при учете данных об Y по сравнению с прогнозом, основанным только на данных об X . Не-

нулевое улучшение прогноза (УП) ассоциируется с наличием воздействия Y на X , а ненулевые УП “в обе стороны” обычно интерпретируются как признак ДС. При этом прогноз осуществляется на один временной шаг Δt – интервал выборки. Однако возможна сложная зависимость УП от Δt , и в [30, 31] было отмечено, что при достаточно редкой выборке даже в случае однонаправленно связанных систем могут наблюдаться ненулевые УП в обе стороны, т.е. “ложные связи”. Это принципиальное обстоятельство необходимо учитывать, в частности, при анализе климатических данных.

В данной работе проведен анализ эффекта редкой выборки при определении взаимосвязи процессов по временным рядам данных на основе причинности по Грейнджеру. Наличие этого эффекта, приводящего к недостоверным выводам о ДС, выявлено при исследовании вариаций глобальной приповерхностной температуры и потока солнечного излучения. Возможность ложных выводов продемонстрирована на примере эталонных стохастических систем с ОС. Представлен статистический тест для различения ОС и ДС, основанный на учете эффекта редкой выборки. С

помощью предложенного теста проверен вывод о ДС между ЭНЮК и индийским муссоном [25, 26].

2. ПРИЧИННОСТЬ ПО ГРЕЙНДЖЕРУ

Пусть $(X(t), Y(t))$ — двумерный случайный процесс, реализации которого регистрируются в дискретные моменты времени с интервалом выборки Δt : $x_n = X(n\Delta t)$, $y_n = Y(n\Delta t)$, где n — целое. Для наборов значений x и y до момента времени n можно ввести обозначения $x_n^- = \{x_{n-k}\}_{k=1}^\infty$ и $y_n^- = \{y_{n-k}\}_{k=1}^\infty$. Из всех возможных способов индивидуального (без учета Y) прогноза величины x_n наименьший средний квадрат ошибки достигается при $x_n^{ind} = M[x_n | x_n^-]$, где $M[x_n | x_n^-]$ означает условное математическое ожидание x_n при условии x_n^- . Обозначим дисперсию этой ошибки $\sigma_{x,ind}^2 = M[(x_n - x_n^{ind})^2]$. Наилучший совместный (с учетом Y) прогноз дается формулой $x_n^{joi} = M[x_n | x_n^-, y_n^-]$ и имеет ошибку с дисперсией $\sigma_{x,joi}^2 = M[(x_n - x_n^{joi})^2]$. Нормированная величина УП $G_{y \rightarrow x} = (\sigma_{x,ind}^2 - \sigma_{x,joi}^2) / \sigma_{x,ind}^2$ характеризует причинность по Грейнджеру (влияние) в направлении $Y \rightarrow X$. Аналогично определяется влияние $X \rightarrow Y$.

Подобный подход был впервые реализован [29] для стационарных гауссовских процессов (x_n, y_n) . Использовалось то обстоятельство, что такой процесс единственным образом описывается двумерным линейным уравнением авторегрессии (АР) вида

$$\begin{aligned} x_n &= \sum_{k=1}^{\infty} a_{x,k} x_{n-k} + \sum_{k=1}^{\infty} b_{x,k} y_{n-k} + \xi_n, \\ y_n &= \sum_{k=1}^{\infty} a_{y,k} y_{n-k} + \sum_{k=1}^{\infty} b_{y,k} x_{n-k} + \psi_n, \end{aligned} \quad (1)$$

где (ξ_n, ψ_n) — двумерный гауссовский белый шум с нулевым средним, соответствующими дисперсиями компонентов σ_ξ^2 и σ_ψ^2 , ковариацией $M[\xi_n \psi_n] = \gamma$. Условие, что шум “белый”, эквивалентно минимуму ошибки прогноза [32], при этом $\sigma_\xi^2 = \sigma_{x,joi}^2$ и $\sigma_\psi^2 = \sigma_{y,joi}^2$. Далее, процесс x_n подчиняется и одномерному уравнению АР, т.е. первому уравнению (1) с нулевыми $b_{x,k}$ и белым шумом ξ_n' , дисперсия которого $\sigma_{\xi'}^2 = \sigma_{x,ind}^2$. Теперь по дисперсиям шумов $\sigma_{\xi'}^2$, σ_{ξ}^2 определяется величина УП $G_{y \rightarrow x}$. Аналогично определяется величина $G_{x \rightarrow y}$.

Чтобы оценить теоретические величины $G_{y \rightarrow x}$ и $G_{x \rightarrow y}$ по конечному временному ряду $\{x_n, y_n\}_{n=1}^N$, все суммы в уравнениях (1) ограничиваются членом $k = p$ (вместо $k = \infty$), и оцениваются коэффициенты и дисперсии шумов в АР моделях порядка p с помощью стандартного метода наименьших квадратов. В анализируемых далее численных примерах длина ряда N велика, так что величина p выбирается просто настолько большой, чтобы результаты оценивания практически перестали меняться с ее дальнейшим увеличением (а именно, $p = 10$ оказалось достаточным во всех примерах). При анализе климатических временных рядов для выбора порядка используется критерий Шварца [33]. Статистическая значимость отличия оценок $G_{y \rightarrow x}$ и $G_{x \rightarrow y}$ от нуля проверяется с помощью F -теста Фишера [34].

Следует отметить, что уравнения (1) при различных Δt являются различными справедливыми представлениями исходной системы (X, Y) . Но как величины $G_{y \rightarrow x}$ и $G_{x \rightarrow y}$ меняются при изменении Δt ? Если нет реального воздействия $Y \rightarrow X$, то логично было бы ожидать $G_{y \rightarrow x} = 0$ при любом Δt или, по меньшей мере, $G_{y \rightarrow x} \ll 1$ и $G_{y \rightarrow x} < G_{x \rightarrow y}$. Однако эти ожидания не обязательно оправдываются [35], более того, относительная мера «ложной причинности» $r = G_{y \rightarrow x} / G_{x \rightarrow y}$ может существенно превышать единицу.

3. ЭФФЕКТ РЕДКОЙ ВЫБОРКИ

Для оценки влияния интервала выборки на результаты оценки связей проведен анализ стохастических линейных диссипативных осцилляторов с дискретным временем:

$$\begin{aligned} X(t) &= A_{X,1} X(t-1) + A_{X,2} X(t-2) + B_X Y(t-1) + \Xi(t), \\ Y(t) &= A_{Y,1} Y(t-1) + A_{Y,2} Y(t-2) + B_Y X(t-1) + \Psi(t), \end{aligned} \quad (2)$$

где Ξ , Ψ — независимые гауссовские белые шумы с нулевым средним и дисперсиями σ_Ξ^2 и σ_Ψ^2 , B_X и B_Y — коэффициенты связи. Собственный период колебаний X (положение пика в спектре мощности) и его время релаксации (определяет ширину пика) даются выражениями $A_{X,1} = 2 \cos(2\pi/T_X) \exp(-1/\tau_X)$ и $A_{X,2} = -\exp(-2/\tau_X)$ [36]. Соответствующие выражения для Y аналогичны. В качестве стартовых значений для анализа использовались значения периодов колебаний $T_X = T_Y = 4.4$ и времен релаксации $\tau_X = \tau_Y = 4$, дисперсий шумов $\sigma_\Xi^2 = \sigma_\Psi^2 = 1$, коэффициентов связи $B_X = 0$ и $B_Y = 0.3$ (ОС $X \rightarrow Y$). Величины УП оценивались по временным рядам достаточно большой длины $N = 10^5$, так что статистические флуктуации были пренебрежимо малы. Согласно рис. 1 (кружки) ОС адекватно характери-

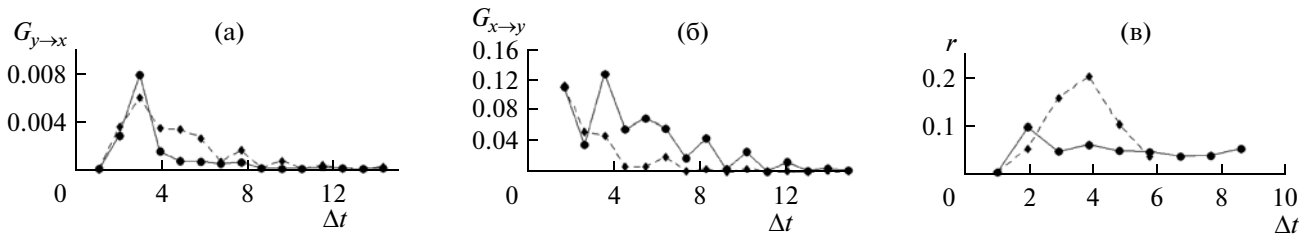


Рис. 1. Характеристики причинности по Грейнджеру для однонаправленно связанных осцилляторов (2) при $T_X = T_Y = 4.4$ с прореживанием (кружки, сплошные линии) и с прореживанием и осреднением (ромбики, штриховые линии).

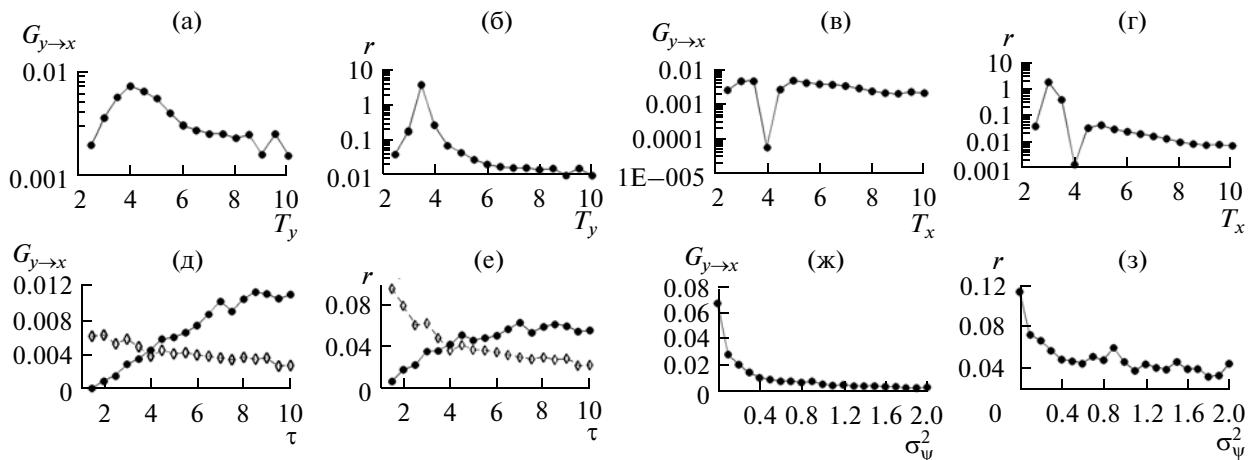


Рис. 2. Характеристики причинности по Грейнджеру для однонаправленно связанных осцилляторов (2) в зависимости от различных параметров системы при остальных фиксированных из базового набора $T_X = T_Y = 5$, $\tau_X = \tau_Y = 4$, $\sigma_\Xi^2 = \sigma_\Psi^2 = 1$, $B_X = 0$, $B_Y = 0.3$. д – $\tau = \tau_X$ (черные кружки), е – $\tau = \tau_Y$ (ромбики).

зуются величинами УП при $\Delta t = 1$: $G_{y \rightarrow x} = 0$ и $G_{x \rightarrow y} > 0$. Однако с ростом Δt положительными становятся и значения $G_{y \rightarrow x}$, эта “ложная связь” и есть проявление эффекта редкой выборки. Значение $G_{y \rightarrow x}$ максимально при $\Delta t = 3$, где $r = 0.1$ означает, что УП в направлении $Y \rightarrow X$, не соответствующем реальному воздействию, достаточно велико – составляет 10% от УП в направлении реального воздействия $X \rightarrow Y$. При $\Delta t \rightarrow \infty$ зависимость между значениями x , y в сильно удаленные последовательные моменты пропадает и величины $G_{y \rightarrow x}$, $G_{x \rightarrow y}$ близки к 0 (см. рис. 1а, 1б, $\Delta t = 15$) – процесс (x_n, y_n) становится белым шумом.

Следует отметить для интерпретации ненулевых $G_{y \rightarrow x}$ при $\Delta t > 1$, что X – марковский процесс второго порядка и вектор $(X(t), X(t-1))$ содержит полную информацию о распределении будущих значений $X(t+l)$ при любом $l > 0$. В этом случае $(X(t), X(t-1))$ полностью определяет состояние процесса X в момент времени t без УП процесса X

при учете Y . При $\Delta t > 1$ прогноз $X(t+l)$ на основе $\{X(t), X(t-\Delta t), X(t-2\Delta t), \dots\}$ не является наилучшим, так как значение $X(t-1)$ не может быть в точности восстановлено по прореженным данным для X . При этом дополнительная прогностическая информация может быть получена по данным об Y за счет корреляции между $X(t-1)$ и наблюдаемыми значениями Y , тем самым приводя к положительному УП $G_{y \rightarrow x}$, не соответствующему реальному воздействию. Таким образом, источник “ложных связей” – неполнота информации о состоянии ведущей системы в наблюдаемых данных.

Положительные $G_{y \rightarrow x}$ и r отмечаются в широких диапазонах значений параметров, как показано на рис. 2, где $\Delta t = 2$ и на каждом графике меняется один из параметров, а все остальные имеют фиксированные значения из базового набора $T_X = T_Y = 5$, $\tau_X = \tau_Y = 4$, $\sigma_\Xi^2 = \sigma_\Psi^2 = 1$, $B_X = 0$, $B_Y = 0.3$. При малых шумах в ведомой системе σ_Ψ^2 лучше восстанавливаются состояния ведущей системы по данным об Y , а поэтому велико оказыва-

ется и $G_{y \rightarrow x} > 0.06$, и $r > 0.1$ (рис. 2ж, 2з). Отношение r превышает единицу при $T_X = 3$ (рис. 2г) или $T_Y = 3.5$ (рис. 2б). Следует отметить, что r столь велико для неидентичных осцилляторов (в указанных случаях — при $T_X \neq T_Y$), тогда как для идентичных (т.е. при $T_X = T_Y$, $\tau_X = \tau_Y$, $\sigma_\Xi^2 = \sigma_\Psi^2$) оно остается в пределах $r < 1$. Согласно рис. 2д, 2е величина r по-разному зависит от двух времен релаксации: она растет с ростом τ_X и падает с ростом τ_Y , последнее связано с тем, что информация о состоянии X хуже восстанавливается по данным от Y . Поэтому еще большие значения $G_{y \rightarrow x}$ и r наблюдаются, когда время релаксации ведомой системы сравнительно велико, а ведомой мало. Например, при $\tau_X = 10$, $\tau_Y = 1$, $\sigma_\Xi^2 = 0.01$ и остальных параметрах из базового набора следует $G_{y \rightarrow x} = 0.14$ с превышением всех значений на рис. 2.

Представленные результаты свидетельствуют, что эффект ложных связей из-за редкой выборки типичен (проявляется не только при каких-то избранных значениях параметров) и может быть весьма сильным ($r > 1$). Оба эти вывода нетривиальны, и их важно учитывать при детектировании и анализе реальных связей. В целом, следует ожидать, что для систем с непрерывным временем эффект редкой выборки может проявляться при любом интервале выборки, так как это уже прореживание. Это подтверждают численные эксперименты с аналогичными стохастическими осцилляторами с непрерывным временем. Однако при уменьшении интервала выборки величина $G_{y \rightarrow x}$ уменьшается, так что при оценке по конечному временному ряду она не признается значимой. В связи с этим ложные связи фактически проявляются только при не очень малом интервале выборки, зависящем от длины временного ряда и теоретического значения $G_{y \rightarrow x}$.

Вопрос об общих условиях максимального проявления эффекта ложных связей и его зависимости от параметров систем требует особого рассмотрения. Отмеченный эффект не обязательно связан с соотношением между частотой выборки анализируемых прореженных данных и частотой Найквиста для более детальных данных. Его проявление должно зависеть от соотношения между интервалом выборки и ключевыми параметрами исследуемых систем, включая характерные периоды колебаний осцилляторов. Согласно полученным в данной работе результатам, в частности для осцилляторов с близкими периодами, максимальный эффект проявляется при интервале выборки в пределах от $1/3$ до $2/3$ характерного периода колебаний. Для случая, представленного на рис. 1а, с $T_X = 4.4$ это соотношение при максимуме для $\Delta t = 3$ близко к $2/3$.

Следует отметить, что в климатических исследованиях рассматриваются не только прореженные, а прореженные и осредненные за интервал выборки (временной шаг) данные. Для исходной системы с дискретным временем это сводится к следующему:

$$\begin{aligned} x_n &= \frac{1}{\Delta t} \sum_{i=1}^{\Delta t} X((n-1)\Delta t + i), \\ y_n &= \frac{1}{\Delta t} \sum_{i=1}^{\Delta t} Y((n-1)\Delta t + i). \end{aligned} \quad (3)$$

При этом имеет место аналогичный эффект редкой выборки. Величины “ложной связи” $G_{y \rightarrow x}$ и r оказываются близкими к случаю простого прореживания и даже могут превосходить последние (рис. 1, ромбики).

4. СВЯЗЬ ВАРИАЦИЙ СОЛНЕЧНОГО ИЗЛУЧЕНИЯ И ГПТ

Эффект редкой выборки для климатических процессов можно оценить при анализе связи вариаций потока солнечного излучения $x(t)$ и глобальной приповерхностной температуры $y(t)$, однонаправленный характер которой известен априори. Воздействие $x \rightarrow y$ было очевидно [18, 37, 38], а $y \rightarrow x$ нереалистично. Представленные на рис. 3а, 3б ежемесячные ($\Delta t = 1$ мес) данные для $x(t)$ и $y(t)$ (см., например, <http://climexp.knmi.nl>) охватывают период 1882–2008 гг., т.е. $N = 1524$ мес. Медленный рост обеих величин может быть связан со стохастическим трендом стационарных процессов [3], так что оценка причинности по Грейнджеру с помощью АР моделей применима. Далее рассматриваются оценки УП по исходным данным, а также по прореженным и осредненным данным согласно (3) для интервалов $\Delta t = 3, 6, 9, 12$ мес с длиной соответствующих рядов $N = 508, 254, 169, 127$.

При оценке причинности по Грейнджеру по рядам вариаций потока солнечного излучения и глобальной приповерхностной температуры использовались индивидуальные АР модели различного порядка авторегрессии p и с многочленами различных степеней K в правой части. Оптимальная модель процесса x выбиралась по критерию Шварца [33] из условия минимизации $S_x = \frac{N}{2} \ln \sigma_{x,ind}^2 + \frac{k}{2} \ln N$, где k — количество оцениваемых коэффициентов. Аналогично подбирались и индивидуальные порядки и степени для процесса y . Для анализированных данных оптимальными всегда оказывались линейные модели x и y , см. рис. 3в, 3г для $\Delta t = 1$, где оптимальный АР порядок для x равен $p_x = 6$, а для y — $p_y = 5$. Для

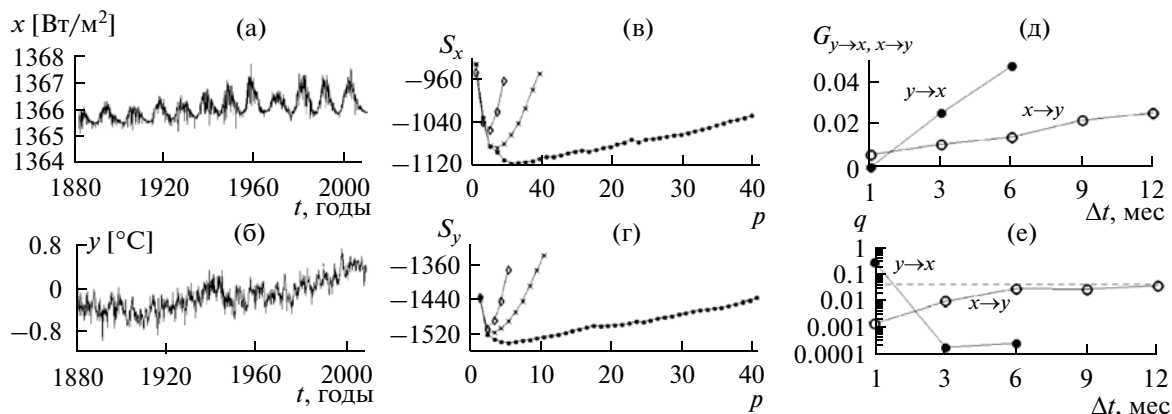


Рис. 3. Эффект редкой выборки при анализе связи вариаций потока солнечного излучения (солнечной постоянной) (а) и аномалий глобальной приповерхностной температуры (б): в, г – соответствующие зависимости критерия Шварца от порядка индивидуальной АР модели с многочленом степени 1 (кружки), 2 (крестики), 3 (ромбики); д, е – характеристики причинности по Грейнджеру и уровень значимости вывода об их положительности в зависимости от интервала выборки. Оптимальные значения АР порядков: $p_x = 6, 10, 5$ при $\Delta t = 1, 3, 6$ и $p_y = 5, 2, 4, 4$ при $\Delta t = 1, 3, 6, 9, 12$.

При $\Delta t = 9, 12$ оптимальные $p_x = 12, 9$ даже с $p_{xy} = 1$ дают число коэффициентов больше \sqrt{N} , так что оценки УП ненадежны (не показано). Пунктиром отмечен уровень $q = 0.05$.

оценки $G_{y \rightarrow x}$ по тому же критерию подбиралось оптимальное количество p_{xy} учтенных значений y при прогнозе x . Оно оказалось равным единице во всех случаях, кроме $\Delta t = 1$, где оно равно нулю. Аналогично, при расчете $G_{x \rightarrow y}$ оптимальное количество p_{yx} учтенных значений x при прогнозе y во всех случаях равно единице. При этом число оцениваемых коэффициентов АР-моделей всегда должно быть меньше \sqrt{N} для достаточной статистики [25], иначе оценке уровня значимости по F -тесту нельзя доверять.

Итак, при $\Delta t = 1$ величина $G_{y \rightarrow x}$ при любом ненулевом p_{xy} (например, $p_{xy} = 1$ на рис. 3д) статистически незначимо отлична от нуля (оценка уровня значимости q по F -тесту много больше 0.05 – см. рис. 3е), т.е. ложная связь $y \rightarrow x$ не отмечается. Что касается противоположного влияния, то величина $G_{x \rightarrow y} = 0.006$ значимо положительна на уровне $q < 0.002$, т.е. воздействие $x \rightarrow y$ корректно выявляется по ежемесячным данным. При $\Delta t > 1$ статистически значимые УП отмечаются в обоих направлениях (рис. 3д, 3е) – еще один пример эффекта редкой выборки. Таким образом, ложные выводы о ДС при анализе климатических рядов вполне возможны, и требуется специальная проверка.

5. ТЕСТ НА ДВУНАПРАВЛЕННУЮ СВЯЗЬ

Согласно полученным результатам при анализе временного ряда $\{x_n, y_n\}_{n=1}^N$ нельзя с достоверностью сделать вывод о наличии двунаправленной

связи, даже если обе оценки $G_{y \rightarrow x}$ и $G_{x \rightarrow y}$ статистически значимо отличны от нуля. Для такого вывода нужно провести специальную проверку “нулевой гипотезы” об ОС с ее подтверждением или опровержением. При этом вероятность ошибочного опровержения не должна превышать некоторой малой величины (уровня значимости). Для тестирования можно использовать, в частности, следующий подход [35]. Рассмотрим некоторый класс M моделей процесса (X, Y) с ОС и некоторым “внутренним” временным шагом, меньшим интервала выборки Δt . Проведем в M поиск модели, способной адекватно воспроизвести статистические характеристики наблюдаемых данных (x_n, y_n) . Если такая модель существует, то гипотеза ОС не может быть опровергнута (тест на двунаправленную связь дает отрицательный результат). Следует подчеркнуть, что тест проводится лишь на основе имеющихся данных с интервалом выборки Δt , поскольку данных с меньшим интервалом выборки (как в примере п. 4) обычно нет. Меньший временной шаг используется только в модели, “прореженные” характеристики которой сравниваются с их аналогами, оцененными по данным наблюдений.

Для реализации этого подхода на примере стационарных гауссовских процессов с дискретным временем рассмотрим класс M моделей вида

$$\begin{aligned} X(t) &= \sum_{k=1}^P A'_{X,k} X(t-k\tau) + \sum_{k=1}^S B'_{X,k} Y(t-k\tau) + \Xi'(t), \\ Y(t) &= \sum_{k=1}^Q A'_{Y,k} Y(t-k\tau) + \sum_{k=1}^R B'_{Y,k} X(t-k\tau) + \Psi'(t), \end{aligned} \quad (4)$$

где временной шаг $\tau = \Delta t/L$, т.е. τ в целое число раз L меньше Δt , $\Xi(t)$ и $\Psi'(t)$ – независимые гауссовские шумы с соответствующими дисперсиями σ_{Ξ}^2 и $\sigma_{\Psi'}^2$. Оба направления ОС $X \rightarrow Y$ и $Y \rightarrow X$ следует проверить по очереди. Для определенности проверим гипотезу ОС $X \rightarrow Y$, т.е. модель (4) с $S = 0$. При этом класс M характеризуется четырьмя параметрами P, Q, R, L . Свойства гауссовских процессов X и Y полностью определяются их авто- и взаимными ковариационными функциями (АКФ и ВКФ) [32]. Опираясь на это, можно найти закон распределения выборочных оценок АКФ и ВКФ для модели (4) и тем самым аналитический критерий статистического согласия между КФ модели и выборочных КФ анализируемых данных.

Для $(P + Q + R + 2)$ -мерного вектора параметров модели (4) можно ввести обозначение $\theta' = \left(\left\{ A'_{X,k} \right\}_{k=1}^P, \left\{ A'_{Y,k} \right\}_{k=1}^Q, \left\{ B'_{Y,k} \right\}_{k=1}^Q, \sigma_{\Xi}^2, \sigma_{\Psi'}^2 \right)$. При заданном значении θ' АКФ и ВКФ модели (3) могут быть точно найдены решением линейных алгебраических уравнений. Для D -мерного вектора значений АКФ и ВКФ модели ($D = 4K + 3$), где $\rho'_{XY}(l\Delta t) = M[X(t)Y(t + l\Delta t)]$, можно ввести обозначение $\rho' = \left(\left\{ \rho'_{XX}(l\Delta t) \right\}_{l=0}^K, \left\{ \rho'_{YY}(l\Delta t) \right\}_{l=0}^K, \left\{ \rho'_{XY}(l\Delta t) \right\}_{l=-K}^K \right)$, а также обозначение $\hat{\rho}$ для вектора выборочных оценок КФ, полученных для анализируемых временных рядов $\{x_n, y_n\}_{n=1}^N$, т.е. $\hat{\rho}_{XX}(l\Delta t) = (1/N) \sum_{n=1}^N x_n x_{n+l}$, $\hat{\rho}_{YY}(l\Delta t) = (1/N) \sum_{n=1}^N y_n y_{n+l}$, $\hat{\rho}_{XY}(l\Delta t) = (1/N) \times \sum_{n=1}^N x_n y_{n+l}$. Если анализируемый процесс принадлежит классу M (соответствует некоторому априори неизвестному значению $\theta' = \theta'_0$) и N велико, то для $\hat{\rho}$ распределение характеризуется D -мерным гауссовским законом со средним ρ' (также соответствующим $\theta' = \theta'_0$) и матрицей ковариаций C , выражаемой через ρ' по формуле Бартлетта [32]. Величина $\chi^2 = (\hat{\rho} - \rho')^T C (\hat{\rho} - \rho')$, где T означает транспонирование, распределена тогда по закону “хи-квадрат” с D степенями свободы [32]. Поскольку θ'_0 априори неизвестно, то с целью поиска наилучшей модели анализируемых процессов в классе M определим минимальное значение $\hat{\chi}_{\min}^2$ величины χ^2 как функции параметров модели θ' . Если анализируемый процесс принадлежит классу M , то величина $\hat{\chi}_{\min}^2$ распределена по закону “хи-квадрат”, но уже с $D - D'$ степенями свободы, так как по D' переменным проводилась оптимизация. Обозначим $(1 - q)$ -квантиль этого распределения

через χ_{1-q}^2 . Если $\hat{\chi}_{\min}^2 > \chi_{1-q}^2$, то гипотеза ОС опровергается на уровне значимости q (с вероятностью ошибки q). Далее при анализе используется обычное значение $q = 0.05$. Величина K принята равной 20, чтобы в рассматриваемых примерах вектор $\hat{\rho}$ включал в себя все существенно отличные от нуля значения АКФ и ВКФ.

Гипотеза может быть опровергнута ошибочно, если в классе M недостаточно велики значения P, Q, R или не должно L . Следует варьировать эти параметры в некотором диапазоне и повторить тест при различных P, Q, R, L . Если данные получены с усреднением (3), это учитывается при расчете значений ρ' в модели, других изменений вносить в процедуру тестирования не требуется.

Эффективность теста проверялась на примере осцилляторов (2) при $B_Y = 0.3$ и различных B_X . Для каждого набора значений параметров генерировались ансамбли из 100 рядов фиксированной длины $N = 1000, 2000, 3000$ с простым прореживанием при $\Delta t = 2$. Величина $B_X = 0$ соответствует ОС $X \rightarrow Y$ и положительному УП $G_{Y \rightarrow X}$ (см. рис. 2а при $T_Y = 5$). Оценки $G_{Y \rightarrow X}$ статистически значимо отличны от нуля (на уровне 0.05 согласно F -тесту) с вероятностью 0.22 при $N = 1000$, 0.53 при $N = 2000$ и 0.68 при $N = 3000$. Таким образом, ложный вывод о ДС только по оценкам УП весьма вероятен. Проверка гипотезы ОС $X \rightarrow Y$ проводилась для каждого временного ряда отдельно при $L = 2$ и различных P, Q, R в диапазоне от 1 до 5. Рассчитывалась относительная частота f опровержений нулевой гипотезы (доля временных рядов в ансамбле, для которых гипотеза оказалась опровергнутой). При $B_X = 0$ она характеризует частоту ошибочных выводов и, следовательно, должна быть не более заявленного уровня значимости $q = 0.05$, чтобы тест был корректным. Согласно рис. 4а–4г это справедливо, даже если P, Q, R заметно больше, чем порядок исходной системы (2).

При $B_X > 0$ частота f характеризует частоту правильных опровержений гипотезы ОС (чувствительность метода к ДС). Чем f больше, тем эффективнее тест. Она растет с ростом B_X и достигает больших значений, если P, Q, R не слишком велики (рис. 4а, 4б). Величина f уменьшается с уменьшением N и ростом P, Q, R (рис. 4г). Это ожидаемо – более широкий класс моделей дает большую вероятность найти модель с КФ, близкими к наблюдаемым выборочным КФ с точностью до ошибок их оценивания, которые растут с уменьшением N . Все результаты аналогичны для $\Delta t = 3$ и $\Delta t = 4$ (не показаны). Таким образом, тест работает корректно и с достаточно высокой чувствительностью.

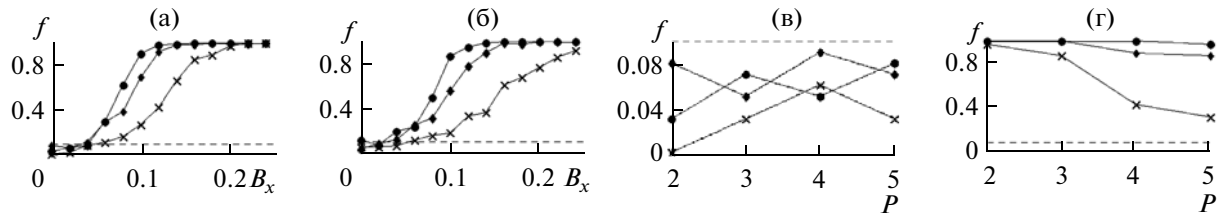


Рис. 4. Частота опровержения гипотезы ОС для системы (2) при $\Delta t = 2$: крестики для $N = 1000$, ромбики — $N = 2000$, кружки — $N = 3000$: а — $P = 2, Q = 2, R = 1$, б — $P = 3, Q = 3, R = 3$, в — $B_x = 0$, г — $B_x = 0.2$ ($P = Q, R = 1$ для в, г). При $B_x = 0$ f характеризует частоту ошибочных выводов. Пунктиром показана допустимая величина частоты ошибок при $q = 0.05$ и учете закона Бернулли для распределения частоты при ансамбле в 100 рядов. При $B_x > 0$ f характеризует чувствительность метода.

Подход может быть реализован и при описании процессов с помощью стохастических дифференциальных уравнений вместо разностных (4). При этом единственным отличием будет способ расчета модельных значений ρ' , которые найдутся путем решения соответствующей системы обыкновенных дифференциальных (а не алгебраических) уравнений. Он может быть расширен также на случай нелинейных систем [35].

6. ОЦЕНКА СВЯЗИ МЕЖДУ ЭНЮК И ИНДИЙСКИМ МУССОНОМ

Представленный тест был применен для проверки прежнего вывода о ДС [25, 26] между Эль-Ниньо — Южным колебанием и индийским муссоном — явлениями, с которыми связаны важнейшие процессы в Азиатско-Тихоокеанском регионе и имеющими глобальное значение [12, 25, 26, 39, 40]. Для сделанного вывода о ДС есть реальные основания с механизмами воздействий в обоих направлениях. В связи с формированием аномалий температуры поверхности океана в экваториальных широтах в Тихом океане во время Эль-Ниньо и Ла-Нинья с соответствующими изменениями конвективных процессов, зональной циркуляции Уокера и меридиональной Хэдли, смещением внутритропической зоны конвергенции проявляются значимые сезонные аномалии температуры и осадков во многих регионах, в том числе в области индийского муссона. В свою очередь, от режимов азиатского и австралийского муссонов зависят циркуляционные особенности, положение областей интенсивной конвекции и облачности в зоне формирования Эль-Ниньо и Ла-Нинья (см., например, [25, 26]).

На рис. 5 приведены временные ряды ежемесячных значений x_n для муссонного индекса (рис. 5а) — аномалий количества осадков над Индией (см., например, <http://climexp.knmi.nl/data/pALLIN.dat>) с удаленным годовым ходом и y_n для индекса ЭНЮК (рис. 5б) — аномалий температуры поверхности океана в области Ниньо-3 (см., например, [\[climexp.knmi.nl/data/iersst/_nino3a.dat\]\(http://climexp.knmi.nl/data/iersst/_nino3a.dat\)\) за период 1871–2006 гг. с длиной рядов \$N = 1632\$ при \$\Delta t = 1\$ мес. В \[25, 26\] на основе анализа этих рядов данных при использовании линейных АР моделей были получены положительные УП в обоих направлениях с \$G_{y \rightarrow x} = 0.020\$ и \$G_{x \rightarrow y} = 0.017\$, значительно отличными от нуля на уровне менее 0.001 согласно \$F\$ -тесту. Учет нелинейных членов мало менял результаты. При этом следует отметить, что подобные значения УП могут отмечаться и для систем с ОС, как в случае с модельной системой \(2\), особенно когда различаются времена релаксации двух процессов. АКФ для муссонного индекса спадает быстрее, чем АКФ для индекса ЭНЮК \(рис. 5а, 5б\), что дает добавочную мотивацию для более детальной проверки типа взаимосвязи двух процессов.](http://</p>
</div>
<div data-bbox=)

В связи с тем, что анализировались суммарное за месяц количество осадков над Индией и среднемесячное значение температуры поверхности океана в области проявления ЭНЮК, использовался тест с учетом осреднения (3). Применились только линейные АР модели (4), так как нелинейность играет малую роль.

Проверялась гипотеза об ОС для каждого направления при $L = 2$ (с 2-недельным “внутренним” временным шагом τ в динамике процессов), $L = 3$ (с 10-суточным шагом τ) и $L = 4$ (с недельным шагом τ). Временные масштабы порядка недели характерны для атмосферных процессов. Значения P, Q, R, S менялись в диапазоне от 1 до 5. Обе гипотезы об ОС оказались опровергнуты на уровне значимости $q < 0.05$ для всех анализируемых моделей. Наилучшее соответствие между модельными и наблюдаемыми КФ было достигнуто при $L = 2, P = Q = 3, R = 1$, т.е. для ОС “муссон \rightarrow ЭНЮК” (рис. 5в–5д). АКФ этой модели (крестики) достаточно близки к АКФ анализируемых данных (кружки) в отличие от ВКФ (рис. 5д). ВКФ анализируемых данных имеет максимум при временном лаге два месяца (“муссон ведет”), но ее значения при малых отрицательных лагах тоже достаточно велики. Оказыва-

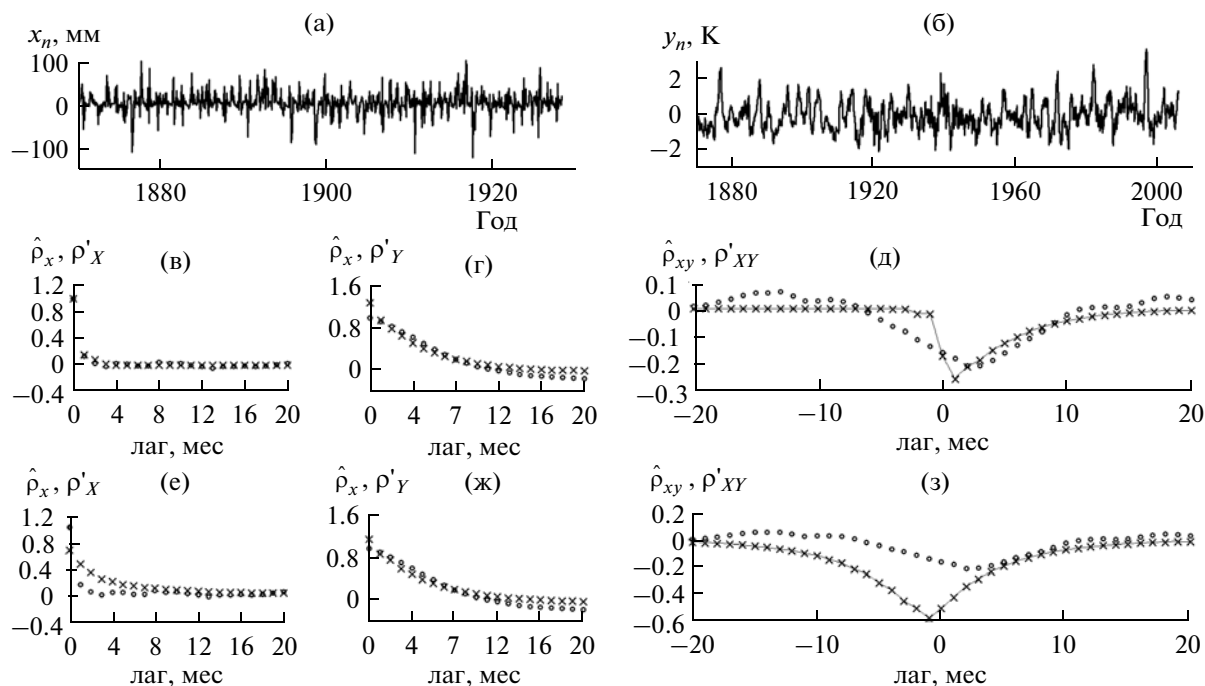


Рис. 5. Анализ взаимосвязи индийского муссона (а) и ЭНЮК (б): а, б – временные ряды; в–з – ковариационные функции наблюдаемых данных (кружки) и модели (крестики) с ОС в направлении “муссон → ЭНЮК” при $P = Q = 3$, $R = 1$, $L = 2$ (в–д) и в направлении “ЭНЮК → муссон” при $P = Q = 4$, $S = 1$, $L = 3$ (е–з).

ется, что любая модель (4) с ОС $X \rightarrow Y$ может иметь такую медленно меняющуюся в окрестности нуля ВКФ, только если при этом АКФ процесса X не является столь быстро спадающей, как на рис. 5в (кружки). Таким образом, эта модель может воспроизвести либо АКФ муссонного индекса, либо ВКФ, но не обе функции вместе, что свидетельствует о неадекватности модели с ОС “муссон → ЭНЮК”. Тест количественно оценивает меру этой неадекватности: для отмеченной наилучшей модели (см. рис. 5в–5д) $\hat{\chi}_{\min}^2 = 144 > \chi_{0.95}^2 = 95.1$. Модели с противоположным направлением ОС расходятся с анализируемыми данными еще сильнее: $\hat{\chi}_{\min}^2 = 705 > \chi_{0.95}^2 = 92.8$ для наилучшей модели с $L = 3$, $P = Q = 4$, $S = 1$ (см. рис. 5е–5з). Таким образом, вывод о ДС для ЭНЮК и индийского муссона прошел дополнительную проверку: он не является результатом редкой выборки, по крайней мере, если собственный временной шаг процессов τ лежит в диапазоне от недели до месяца.

7. ЗАКЛЮЧЕНИЕ

Показан эффект редкой выборки при анализе связей между климатическими процессами по временным рядам. В частности, при достаточно большом интервале выборки результаты анализа причинности “по Грейнджеру” могут для систем с

однонаправленной связью ошибочно свидетельствовать о двунаправленной связи. Этот эффект “ложных связей” из-за редкой выборки проявляется при анализе вариаций потока солнечного излучения и глобальной приповерхностной температуры. При использовании данных, прореженных и осредненных на интервалах 3 месяца или 6 месяцев, фиксируется значимое ненулевое “воздействие” ГПТ на солнечную активность, что не отмечается для ежемесячных данных. Представлен специальный тест для надежного выявления двунаправленной связи, который использован для анализа взаимодействия между ЭНЮК и индийским муссоном. Прежний вывод о взаимном воздействии этих процессов подтвержден этим тестом, по крайней мере, при анализированных временных шагах от недели до месяца.

Предложенный тест для подтверждения или опровержения предположений о типе взаимосвязи процессов может быть реализован при описании процессов с помощью дифференциальных уравнений, основанных на физических модельных соображениях, с достаточно широкой областью его применимости. Он представляется потенциально полезным в детальных исследованиях взаимодействия климатических процессов с проверкой выводов о двунаправленных связях аналогично проведенной в данной работе. Соответствующее более широкое рассмотрение причинности при наличии трех и более взаимодей-

ствующих процессов требует специальных исследований.

Работа выполнена в рамках проектов РФФИ (гранты №№ 11-02-00599, 11-05-01139, 12-05-91057-НЦНИ, 12-05-91323-СИГ), программ РАН, Министерства образования и науки РФ (ГК 11.519.11.5004) и при поддержке гранта Президента РФ НШ-5467.2012.5.

СПИСОК ЛИТЕРАТУРЫ

1. *Blanford H.F.* On the connection of the Himalaya snowfall with dry winds and seasons of drought in India // *Proc. Roy. Soc. London*. 1884. V. 37. P. 1–23.
2. *Walker G.T., Bliss E.W.* World weather V // *Mem. Roy. Meteor. Soc.* 1932. V. 4. № 36. P. 53–84.
3. *Kripalani R.H., Kulkarni A.* Rainfall variability over Southeast Asia: Connections with Indian monsoon and ENSO extremes: New perspectives // *Int. J. Climatol.* 1997. V. 17. № 11. P. 1155–1168.
4. *Huang J., Higuchi K., Shabbar A.* The relationship between the North Atlantic Oscillation and the ENSO // *Geophys. Res. Lett.* 1998. V. 25. P. 2707–2710.
5. *Арне К., Бенгтссон Л., Голицын Г.С. и др.* Анализ и моделирование изменений гидрологического режима в бассейне Каспийского моря // *ДАН*. 1999. Т. 366. № 2. С. 248–252.
6. *Груза Г.В., Ранькова Э.Я., Клещенко Л.К., Аристова Е.Н.* О связи климатических аномалий на территории России с явлением Эль-Ниньо–Южное колебание // *Метеорология и гидрология*. 1999. № 5. С. 32–51.
7. *Mokhov I.I., Bezverkhny V.A., Karpenko A.A.* Evolution of climatic characteristics and atmospheric components at Milankovitch scales from Vostok ice core // *Research Activities in Atmospheric and Oceanic Modelling* / Ed H. Ritchie. 2002. Rep. No. 32. WMO/TD-No. 1105. P. 2.17–2.18.
8. *Jevrejeva S., Moore J.C., Grinsted A.* Influence of the Arctic Oscillation and El Nino–Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach // *J. Geophys. Res.* 2003. V. 108. № D21. P. 4677, doi:10.1029/2003jd003417.
9. *Grinsted A., Moore J.C., Jevrejeva S.* Application of the cross wavelet transform and wavelet coherence to geophysical time series // *Nonlin. Proc. Geophys.* 2004. V. 11. P. 561–566.
10. *Wang W., Anderson B.T., Kaufmann R.K., Myneni R.B.* The relation between the North Atlantic Oscillation and SSTs in the North Atlantic basin // *J. Climate*. 2004. V. 17. № 24. P. 4752–4759.
11. *Мохов И.И., Безверхний В.А., Карпенко А.А.* Диагностика взаимных изменений содержания парниковых газов в атмосфере и температурного режима по палеорекострукциям для антарктической станции Восток // *Изв. РАН. Физика атмосферы и океана*. 2005. Т. 41. № 5. С. 579–592.
12. *Maraun D., Kurths J.* Epochs of phase coherence between El-Nino/Southern Oscillation and Indian monsoon // *Geophys. Res. Lett.* 2005. V. 32. P. L15709. doi:10.1029/2005GL023225.
13. *Мохов И.И., Смирнов Д.А.* Исследование взаимного влияния процессов Эль-Ниньо–Южное колебание и Северо-Атлантического и Арктического колебаний нелинейными методами // *Изв. РАН. Физика атмосферы и океана*. 2006. Т. 42. № 5. С. 650–667.
14. *Mokhov I.I., Smirnov D.A.* El-Nino Southern Oscillation drives North Atlantic Oscillation as revealed with nonlinear techniques from climatic indices // *Geophys. Res. Lett.* 2006. V. 33. P. L03708. doi:10.1029/2005GL024557.
15. *Mosedale T.J., Stephenson D.B., Collins M. et al.* Granger causality of coupled climate processes: Ocean feedback on the North Atlantic Oscillation // *J. Climate*. 2006. V. 19. № 7. P. 1182–1194.
16. *Palus M., Novotna D.* Quasi-biennial oscillations extracted from the monthly NAO index and temperature records are phase-synchronized // *Nonlin. Proc. Geophys.* 2006. V. 13. P. 287–296.
17. *Verdes P.F.* Global warming is driven by anthropogenic emissions: A time series analysis approach // *Phys. Rev. Lett.* 2007. V. 99. P. 048501.
18. *Мохов И.И., Смирнов Д.А.* Диагностика причинно-следственной связи солнечной активности и глобальной приповерхностной температуры Земли // *Изв. РАН. Физика атмосферы и океана*. 2008. Т. 44. № 3. С. 283–293.
19. *Козленко С.С., Мохов И.И., Смирнов Д.А.* Анализ причинно-следственных связей между Эль-Ниньо в Тихом океане и его аналогом в экваториальной Атлантике // *Изв. РАН. Физика атмосферы и океана*. 2009. Т. 42. № 6. С. 754–763.
20. *Bronnimann S.* Impact of El-Nino–Southern Oscillation on European climate // *Rev. Geosci.* 2007. V. 45. P. RG3003. doi:10.1029/2006 RG0001999.
21. *Мохов И.И., Смирнов Д.А.* Эмпирические оценки воздействия естественных и антропогенных факторов на глобальную приповерхностную температуру // *ДАН*. 2009. Т. 426. № 5. С. 679–684.
22. *Lu R., Chen W., Dong B.* How does a weakened Atlantic thermohaline circulation lead to an intensification of the ENSO–south Asian summer monsoon interaction? // *Geophys. Res. Lett.* 2008. V. 35. P. L08706. doi:10.1029/2008GL033394.
23. *Smirnov D.A., Mokhov I.I.* From Granger causality to long-term causality: Application to climatic data // *Phys. Rev. E*. 2009. V. 80. P. 016208. doi:10.1103/PhysRevE.80.016208.
24. *Бышев В.И., Нейман В.Г., Иванов Ю.А.* Природные факторы глобальной изменчивости современного климата // *Изв. РАН. Сер. геогр.* 2009. № 1. С. 7–13.
25. *Mokhov I.I., Smirnov D.A., Nakonechny P.I. et al.* Alternating mutual influence of El-Nino/Southern Oscillation and Indian monsoon // *Geophys. Res. Lett.* 2011. V. 38. P. L00F04. doi:10.1029/2010GL045932.
26. *Мохов И.И., Смирнов Д.А., Наконечный П.И. и др.* Взаимосвязь явлений Эль-Ниньо–Южное колебание и индийского муссона // *Изв. РАН. Физика атмосферы и океана*. 2012. Т. 48. № 1. С. 56–66.
27. *Chen W., Dong B., Lu R.* Impact of the Atlantic Ocean on the multidecadal fluctuation of El Nino–Southern Oscillation–South Asian monsoon relationship in a

- coupled general circulation model // *J. Geophys. Res.* 2010. V.115. P. D17109. doi:10.1029/2009JD013596.
28. Мохов И.И., Смирнов Д.А., Карпенко А.А. Оценки связи изменений глобальной приповерхностной температуры с разными естественными и антропогенными факторами на основе данных наблюдений // *ДАН.* 2012. Т. 443. № 2. С. 225–231.
 29. Granger C.W.J. Investigating causal relations by econometric models and cross-spectral methods // *Econometrica.* 1969. V. 37. № 3. P. 424–438.
 30. Sims C.A. Discrete approximations to continuous time distributed lags in econometrics // *Econometrica.* 1971. V. 39. P. 545–563.
 31. Marcellino M. Some consequences of temporal aggregation in empirical analysis // *J. Business and Economic Statistics.* 1999. V. 17. P. 129–136.
 32. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Ч. 1, 2. М.: Мир, 1974. 406 с.
 33. Schwarz G. Estimating the dimension of a model // *The Annals of Statistics.* 1978. V. 6. № 2. P. 461–464.
 34. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. 456 с.
 35. Smirnov D.A., Bezruchko B.P. Spurious causalities due to low temporal resolution: Towards detection of bidirectional coupling from time series // *Europhys. Lett.* 2012. V. 100. 10005./doi:10.1209/0295-5075/100/100005.
 36. Timmer J., Lauk M., Pflieger W., Deuschl G. Cross-spectral analysis of physiological tremor and muscle activity: I. Theory and application to unsynchronized electromyogram // *Biol. Cybern.* 1998. V. 78. P. 349–357.
 37. Lean J., Rottman G., Harder J., Kopp G. Source contributions to new understanding of global change and solar variability // *Solar Physics.* 2005. V. 230. № 1–2. P. 27–53.
 38. Hoyt D.V., Schatten K.H. The Role of the Sun in Climate Change. Oxford: Oxford Univ. Press, 1997. 279 p.
 39. Climate Change 2007: The Physical Science Basis. / Eds: S. Solomon et al. Cambridge: Cambridge Univ. Press, 2007. 996 p.
 40. Yamasaki K., Gozolchiani A., Havlin S. Climate networks around the globe are significantly affected by El Niño // *Phys. Rev. Lett.* 2008. V. 100. P. 228501.

Estimate of Interaction between Climatic Processes: Effect of Sparse Sample of Analyzed Data Series

D. A. Smirnov^a and I. I. Mokhov^b

^a *Kotelnikov Institute of Radio Engineering and Electronics, Saratov Department, ul. Zelenaya, Saratov, 410019 Russia*
e-mail: smirnovda@yandex.ru

^b *Obukhov Institute of Atmospheric Physics, Russian Academy of Sciences, per. Pyzhevskii, Moscow, 119017 Russia*
e-mail: mokhov@ifaran.ru

Received October 29, 2012; in final form, February 28, 2013

Abstract—It is shown that the approach based on the Granger causality may lead to erroneous conclusions about bidirectional coupling (BC) in the case of unidirectional coupling (UC) and quite a sparse sample of the data series analyzed. This effect was revealed in an analysis of coupling between variations in the solar irradiance and global near-surface temperature. We present a statistical test to confirm or reject speculations about the character (unidirectional or bidirectional) of coupling. The corresponding analysis of coupling between phenomena of El Niño and the Indian Monsoon confirmed the earlier conclusions about their mutual influence.

Keywords: time series, estimates of couplings, unidirectional coupling, bidirectional coupling, Granger causality, temperature, solar activity, El Niño, Southern Oscillation, Indian Monsoon.