

Сравнение временных рядов в прикладных исследованиях: близость, синхронность и корреляция

О.В. Максимова^{1,2)}

¹⁾ФГБУ «Институт глобального климата и экологии имени академика Ю.А. Израэля»,
Россия, 107258, Москва, ул. Глебовская, 20Б

²⁾Университет МИСИС,
Россия, 119049, Москва, Ленинский пр-кт, 4

*Адрес для переписки: *o-maximova@yandex.ru*

Реферат. Задача сравнения временных рядов возникает во множестве прикладных исследований. Сложность такого сравнения обусловлена тем, что с течением времени вероятность, что все данные описывают первоначальный процесс, уменьшается, и некоторые краткосрочные изменения могут быть вызваны специальными причинами, которые не повторяются. В работе рассматриваются три возможные задачи, связанные со сравнением временных рядов в прикладных исследованиях: поточечная близость, синхронность изменения во времени и корреляция уровней рядов. Предложены подходы к решению каждой задачи в отдельности, и выявлены их особенности, которые необходимо учитывать при постановке задачи сравнения массивов данных во времени. Установлено и продемонстрировано: синхронно изменяющиеся временные ряды могут быть не близки поточечно и иметь слабую корреляционную связь, высокая корреляция может не обеспечивать синхронность временных рядов и давать ложные выводы о взаимосвязи показателей во времени, и т.п. Выводы, полученные при решении только одной задачи, нельзя переносить как на выводы другой, так и на связь временных рядов в целом.

Продемонстрированные в работе модельные примеры и примеры прикладных исследований дают понимание, как можно сформулировать задачу сравнения массивов данных во времени, и подойти к ее решению и интерпретации. Показано, что возможность установления наличия причинно-следственных связей временных рядов в целом зависит от качественного анализа содержания рассматриваемых процессов.

Ключевые слова. Временной ряд, корреляция временных рядов, коэффициент синхронности, ложные корреляции, климат, приросты сосны.

Comparison of time series in applied research: proximity, synchronism and correlation

O.V. Maksimova^{1, 2)}

¹⁾Yu.A. Izrael Institute of Global Climate and Ecology,
20B, Glebovskaya str., 107058, Moscow, Russian Federation

Abstract. The problem of comparing time series arises in many applied studies. The complexity of such a comparison is due to the fact that over time the probability that all data describe the initial process decreases, and some short-term changes can be caused by special causes that do not repeat. The paper considers three possible problems associated with comparing time series in applied studies: point-by-point proximity, synchronicity of change over time, and correlation of series levels. Approaches to solving each problem separately are proposed, and their features are identified that must be taken into account are identified while comparing data arrays over time. It has been established and demonstrated that synchronously changing time series may not be close point-by-point and have a weak correlation; high correlation cannot ensure synchronicity of time series and give false conclusions about the relationship of indicators over time, etc. Conclusions obtained when solving only one problem cannot be transferred either to the conclusions of another, or to the relationship of time series as a whole. The model examples and applied research examples demonstrated in the work provide an understanding of how the problem of comparing data arrays over time can be formulated and how to approach its solution and interpretation. It is shown that the possibility of establishing the presence of cause-and-effect relationships in time series as a whole depends on a qualitative analysis of the content of the processes under consideration.

Keywords. Time series, time series correlation, synchronicity coefficient, spurious correlations, climate, pine growths.

*В сущности, все модели неверны,
но некоторые из них полезны*

Вох Г.

Введение

Очень часто при анализе процессов приходится сталкиваться с данными, измеренными во времени (временной ряд). Такие данные можно встретить при наблюдениях за природными процессами и явлениями, производственными и экономическими процессами, и, конечно, при наблюдениях за динамикой различных медицинских показателей состояния здоровья человека. Принципиальной характеристикой таких данных является факт, что номер наблюдения (т.е. время) имеет значение: важны не только сами показатели, но и момент времени, соответствующий их измерению. Сложность анализа и прогноза множится за счет следующих причин:

— чем дальше мы смотрим назад во времени, тем меньше вероятность, что все данные описывают один и тот же первоначальный процесс;

– некоторые краткосрочные изменения могут быть вызваны специальными причинами, которые не повторяются и, как отмечал в своей книге Николас Талеб, наоборот, способствуют пониманию процесса (Taleb, 2001).

При сравнении же двух наборов временных данных исследователь, как правило, начинает с подсчета коэффициента корреляции. Если коэффициент принимает высокие абсолютные значения, то можно выдвинуть гипотезу о причинно-следственной связи при отсутствии оснований считать эту связь заведомо ложной. Однако если коэффициент корреляции невысок и при этом есть основания для поиска причинно-следственной связи изменчивости показателей во времени, исследование можно продолжить, выходя за рамки простого расчета коэффициента корреляции. Тогда возникают смежные задачи, которые могут решаться как по-отдельности, так и одновременно. *Первая задача* – изучить попарную близость массивов данных во времени. Важно при этом, чтобы имелись одинаковые единицы измерения данных, формирующих временные ряды. Примером такой задачи в биологии может служить сравнение динамик приростов сосны в разных местообитаниях (Максимова, Кухта, 2024), в климатологии – подбор модели, наиболее точно воспроизводящей региональный климат (Максимова, Гинзбург, 2020; Шерстюков, 2011; Лбов, Герасимов, 2010), в экономических исследованиях – анализ близости обменных курсов валют за несколько лет (Бурнаев, Оленев, 2006), и т.п. *Вторая задача* – изучить синхронность изменения массивов данных во времени. Такая задача может возникать, когда нас интересуют не величины массивов показателей, а направленность их изменений за один или несколько временных периодов. Сходство синхронности показателей исследуется в биологических задачах, например, для параметров состояния крон деревьев (Феклистов и др., 2020; Кухта и др., 2024; Шерстнева и др., 2024), в экологии – при оценке изменений различных загрязняющих веществ в каких-либо средах (Асфандиярова и др., 2013; Розенталь, Шпер, 2023), в изучении изменений климата – при сравнении трендов приземной температуры воздуха различных климатических моделей (Израэль и др., 2006; Бардин и др., 2020; Максимова, Кухта, 2022а), в экономических исследованиях – при анализе динамики сезонности цен на различные товары, ВВП и налоговые сборы (Сушко, 2017; Шрамко, 2012), в музыке – при изучении схожести вокальных исполнений (Катаева, Якимук, 2019), и т.д. Актуальность такой постановки возрастает при сравнении массивов с разноименными единицами измерений. *Третья задача* – определить взаимосвязь самих наборов показателей. В таком контексте задача сравнения рядов наиболее привлекательна для исследователя, поскольку высокие значения коэффициента корреляции могут дать возможность численному прогнозу показателей одного признака по соответствующим значениям другого. Поэтому корреляционный анализ широко применяется в самых различных областях (Кухта и др., 2024; Сушко, 2017; Лебедев, Спесивцев, 2015; Розенталь, Шпер, 2023; Хамитова и др., 2023; Aronov et al., 2020).

Интересно, что выявление сходств временных рядов в контексте решения одной из этих задач не гарантирует сходств в контексте других. Именно поэтому проблема сравнения двух временных рядов лежит далеко за пределами простого подсчета коэффициента корреляции.

Цель исследования – рассмотреть подходы к сравнению временных данных в контексте решений задач исследования их поточечной близости, синхронности и корреляции и продемонстрировать их решение на примерах как прикладных исследований, так и модельных.

Методы и материалы

Реализация модельных примеров проводилась в программах *Statistica 15* и *Excel*. Для решения поставленных задач использовались следующие понятия и инструменты: стандартное отклонение, евклидово расстояние, коэффициент вариации, коэффициент корреляции Пирсона r , линейная регрессия, критерий G-знаков, критерий χ^2 -Пирсона (Кобзарь, 2012).

Помимо построенных модельных примеров использовались результаты предыдущих исследований автора и данные открытого доступа с указанием соответствующих ссылок.

Результаты и обсуждение

Что понимать под взаимосвязью временных рядов? Интуитивно понятно, что массивы рядов связаны, если по значениям одного можно сделать прогноз значений другого в конкретный разрез времени. Однако при сравнении массивов не всегда возможно решить такую задачу, так как их значения (уровни временных рядов) часто формируются под воздействием множества также меняющихся во времени факторов. Поэтому в прикладных исследованиях могут возникать и другие смежные задачи, которые выделены во введении: точечная близость уровней рядов, синхронность динамики за один или несколько временных периодов и корреляция временных рядов.

Но насколько допустимо переносить выводы о тесноте связи между колебаниями на синхронность и близость рядов? Очевидно, что совпадение синхронности во времени не обеспечивает поточечную близость рядов (на рис. 1а, к примеру, точки 14 и 15), и обратно, поточечная близость не всегда задает синхронность изменчивости (на рис. 2а, например, пары точек с номерами 14 и 15). Однако остается вопрос. Что подразумевать под поточечной близостью рядов? Эта проблема будет рассмотрена далее при описании первой задачи. Наиболее интересна ситуация возможности переноса вывода о корреляции рядов на выводы о близости и синхронности. На рис. 1а,б представлены модельные данные рядов X и Y , для которых $r = 0.3$, но синхронность наблюдается повсюду за исключением двух выделенных промежутков. А на рис. 2а,б, напротив, прослеживается асинхронность¹⁾ на всех промежутках, кроме отмеченных четырех, однако рассчитанный коэффициент корреляции между показателями рядов высокий и составляет $r = 0.9$.

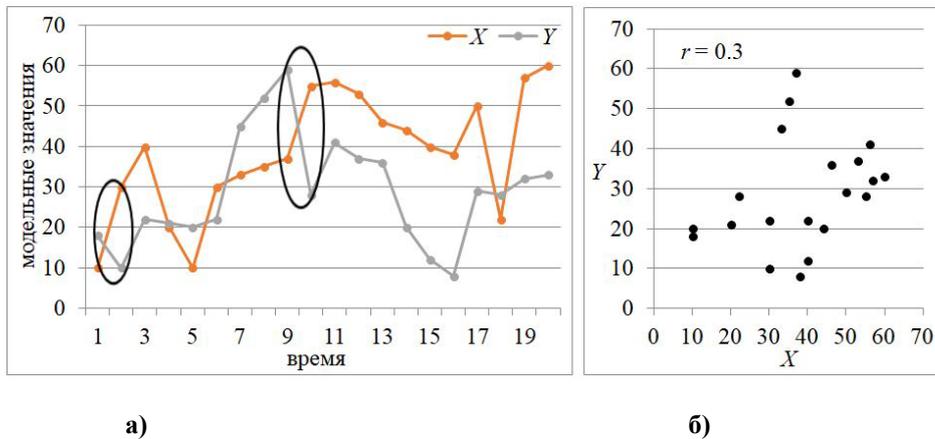


Рисунок 1. Модельные значения временных рядов X и Y : а) во времени, 2 несовпадения направлений динамики из 19, б) диаграмма рассеяния, коэффициент корреляции $r = 0.3$

Figure 1. Model values of time series X and Y : а) in time, 2 mismatches in the directions of dynamics out of 19, б) scatter diagram, correlation $r = 0.3$

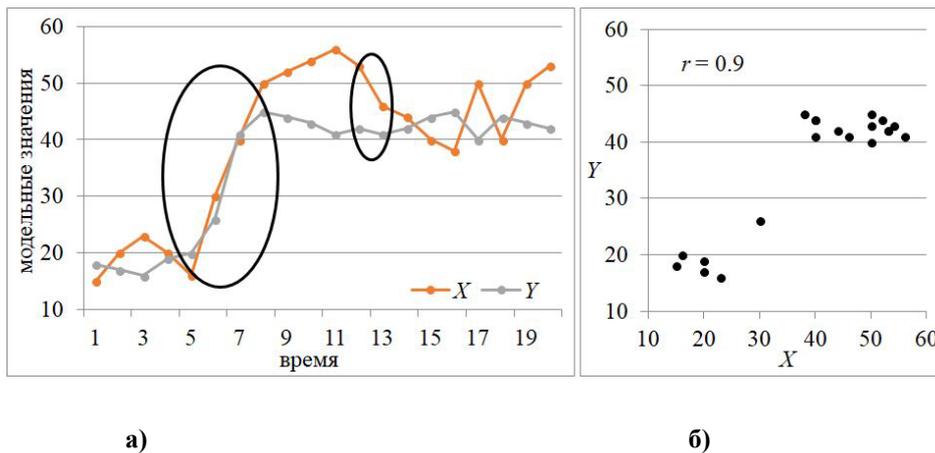


Рисунок 2. Модельные значения временных рядов X и Y : а) во времени, 4 совпадения направлений динамики из 19, б) диаграмма рассеяния, коэффициент корреляции $r = 0.9$

Figure 2. Model values of time series X and Y : а) in time, 4 coincidences of the directions of dynamics out of 19, б) scatter diagram, correlation $r = 0.9$

Очевидно, что могут быть и ситуации обратные: синхронность на большинстве временных интервалов будет сопровождаться высоким значением корреляции, а асинхронность – близким к нулю. Таким образом, выводы о тесноте корреляционной связи между колебаниями рядов в общем случае не экстраполируются на выводы об их поточечной близости и синхронности (равно как неверно в общем случае и обратное суждение).

1) Асинхронность – несовпадение монотонности на рассматриваемом промежутке.

Рассмотрим каждую задачу отдельно.

Первая задача. Поточечная близость уровней временных рядов может рассматриваться через евклидово расстояние, если сравнение производится среди трех и более рядов и необходимо выбрать наиболее близкие из них:

$$\rho_{jk} = \sqrt{\frac{\sum_{i=1}^n (y_{ij} - y_{ik})^2}{n}}, \quad (1)$$

где y_{ij} и y_{ik} – i -е значения j -го и k -го динамического рядов, $i = 1, \dots, n$. Такая задача решалась, к примеру, авторами при выборе глобальной модели, наиболее точно воспроизводящей климат Арктического региона (Максимова, Гинзбург, 2020). Однако если нужно провести оценку близости между двумя рядами, то возникает вопрос: с чем сравнивать? И тогда обычная евклидова метрика не подходит (Бурнаев, Оленев, 2006; Максимова, Кухта, 2022б). С этой целью авторы исследования (Максимова, Кухта, 2022б; Максимова, Кухта, 2023) для сравнения параметров изменчивости приростов сосны в разных местообитаниях разработали и верифицировали на нескольких наборах натуральных данных коэффициент:

$$V_{\text{ско}_{\text{связи}}} = \frac{\rho_{jk}}{\text{med}_{jk}}, \quad (2)$$

где med_{jk} – медиана данных для j -го и k -го рядов биотопов, а ρ_{jk} рассчитывается по формуле (1). Коэффициент (2) показывает, во сколько раз расстояние между данными отличается от их медианного значения. В качестве порогового значения для оценки близости выбрано 0.35 как для похожего по структуре часто применяемого в статистике коэффициента вариации оценки однородности данных. Действительно, если $V_{\text{ско}_{\text{связи}}}$ меньше 0.35, то это означает, что средняя разность²⁾ между точками меньше примерно в три и более раз их медианного значения. Однако, следует отметить, что выбор такого порогового значения на данный момент обосновано лишь эмпирическими исследованиями и в дальнейшем требует теоретической проверки.

У других авторов разработан подход оценки близости на основе вейвлет-коэффициентов: предложена методика расчета мер близости между временными рядами на основе характеристик этих рядов, получаемых с помощью дискретного вейвлет-преобразования (Бурнаев, Оленев, 2006). На экспериментах с финансовыми данными этот подход подтвердил эффективность, но, несмотря на некоторые достоинства, следует отметить, что он отли-

²⁾ Под средней разностью понимается корень квадратный из средней суммы квадратов разностей между точками (согласно формуле (2))

чается гораздо большей технической сложностью, чем подход, описанный выше.

Иногда используются для сравнения временных рядов критерии проверки параметрических гипотез и критерии согласия, например χ^2 -Пирсона, на принадлежность выборок (соответствующим массивам двух временных рядов) одному распределению. Но даже если при таком подходе получены положительные результаты о сходстве, невозможно сделать вывод о близости рядов, так как сравнение данных в пространстве не учитывает порядок следования точек, который играет принципиально важную роль во временных рядах.

Как было отмечено во введении, в такой постановке задачи важно, чтобы имелись одинаковые единицы измерения данных, формирующих временные ряды.

Вторая задача. Синхронность временных рядов возникает во множестве прикладных исследований: в биологии, например, при определении сходства дендрохронологических данных образцов, взятых на разных участках (Феклистов и др., 2020), при оценке направленности отклика воздействия трансграничного загрязнения или климатических факторов на параметры роста древостоя (Кухта и др., 2024; Шерстнева и др., 2024). Для оценки синхронности временных рядов вводится коэффициент, равный доле совпадающих направлений изменчивости уровней рядов среди всех направлений (Феклистов и др., 2020; Шерстнева и др., 2024). Авторами работы (Феклистов и др., 2020) предложена граница коэффициента синхронности в 50%, начиная с которой можно судить о неслучайном совпадении, но такая оценка лишь приближенная. Поэтому для принятия решения о значимом числе совпадений рекомендуется применять критерий G-знаков (Максимова и др., 2023; Шерстнева и др., 2024; Максимова, Гинзбург, 2020). Приведем пример из работы (Максимова и др., 2023), в которой исследовалась сонаправленность динамик индексов линейных приростов сосны во влажных биотопах и средних температур прошлогоднего вегетационного сезона заповедника «Кивач» (рис. 3). Для установления статистической значимости их сонаправленности используется критерий G-знаков, в котором выдвигается основная гипотеза о случайном совпадении направлений, а конкурирующая – не случайном. Число несовпадений из возможных 22 направлений составляет $G_{расч} = 5$, а критическое значение для уровня ошибки 5% составляет $G_{крит} = 6$. Так как $G_{расч} < G_{крит}$, то синхронность рядов признается неслучайной (статистически значимой).

Для модельного примера на рис. 1 коэффициент $V_{ско_{связи}} = 0.58 > 0.35$, что фиксирует большую поточечную разницу рядов, но при этом ряды ведут себя синхронно (синхронность составляет 89.5%). Для примера на рис. 2 $V_{ско_{связи}} = 0.17$, что дает хорошую поточечную близость, но синхронность составляет лишь 21%. Это дополнительно демонстрирует содержательную разницу постановок задач сравнения поточечной близости и синхронности временных рядов.



Рисунок 3³⁾. График хода средних температур (апрель-сентябрь прошлого года, °C) и индексов линейных приростов влажных биотопов

Figure 3. Average temperatures graph (April-September last year, °C) and linear increments indices of wet biotopes

Третья задача. Расчет коэффициента корреляции между признаками. Эта задача имеет широкий спектр применений в разных областях. Известно, что коэффициент корреляции Пирсона служит мощным статистическим инструментом, который может помочь исследователям обнаружить скрытые закономерности и взаимосвязь между переменными (Максимова, 2024; Благовещенский, 2009). Однако у этого инструмента есть ограничения как при расчете, так и интерпретации в условиях прикладной задачи. Применение альтернативных коэффициентов корреляции, например, Спирмена и Кендалла, также не обеспечивает точность выводов в задаче сравнения временных рядов, так как можно прийти к неверным умозаключениям, вызываемым ложной корреляцией. Это связано с тем, что все показатели меняются со временем t , которое может рассматриваться в качестве общей причины для всех временных рядов. При этом могут встречаться две ситуации:

- 1) связь временных рядов имеет практическое обоснование, но величина коэффициента корреляции это не подтверждает,
- 2) коэффициент корреляции между временными рядами показывает наличие синхронности, однако, связь ложная.

Далее будут рассмотрены ситуации и приведены примеры в условиях расчета коэффициента корреляции Пирсона.

Для ситуации 1) рассмотрим диаграмму рассеяния на рис. 4, построенную для данных исследования, представленных на рис. 3. Обнаруженная синхронность динамик, как отмечалось, значимая, однако коэффициент корреляции Пирсона между признаками невысок для такого небольшого объ-

³⁾ Расчеты и рисунок выполнен автором и опубликован в исследовании (Максимова и др., 2023)

ема данных и составляет $r = 0.4$ (что объясняется наличием воздействия множества других факторов). Известно из натуральных наблюдений, что такая связь оправдана и подтверждает лимитирующую роль температур на этапе формирования почек возобновления для сосняков в сфагновых биотопах, характеризующихся холодным субстратом (Максимова и др., 2023). В качестве математического подтверждения наличия причинно-следственной связи рекомендуется в этом случае продолжить ряд наблюдений по мере накопления данных.

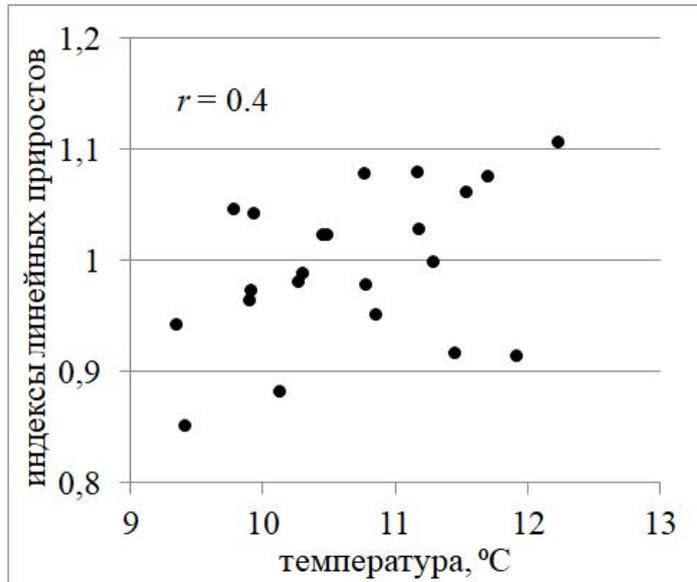


Рисунок 4. Диаграмма рассеяния для данных, представленных на рис. 3

Figure 4. Scatter plot for the data presented in Picture 3

Приведем еще один пример, данные для которого заимствованы из (Елисеева, Юзбашев, 2006). Рассмотрим изменения двух признаков $X = \{\text{урожайность картофеля, ц/га}\}$, $Y = \{\text{стоимость картофеля за ц}\}$ за 13 лет с 1977 по 1989 гг. (рис. 5). По рис. 5б видно, что коэффициент корреляции Пирсона близок к нулю. Но по законам экономики при пространственной корреляции связь урожайности и себестоимости отрицательная сильная: чем выше урожайность, тем ниже в среднем формируется себестоимость. Наиболее вероятно, наблюдается ложная слабая корреляция. Графически на рис. 5а можно видеть асинхронность рядов во все временные промежутки, кроме одного. Для математического подтверждения связи между уровнями урожайности и себестоимости можно рекомендовать рассмотреть корреляцию между колебаниями без учета трендов (так как причиной ложной корреляции может быть однонаправленность трендов обоих признаков) или рассмотреть цены на картофель, приведенные с помощью дефляторов к одному базисному году. В частности, корреляция между колеблемостью признаков после устранения трендовой составляющей в каждом ряду (т.е. корреляция отклонений от трен-

дов), составляет $r = -0.94$, что может служить обоснованием выдвинутой гипотезы об обратной связи между признаками.

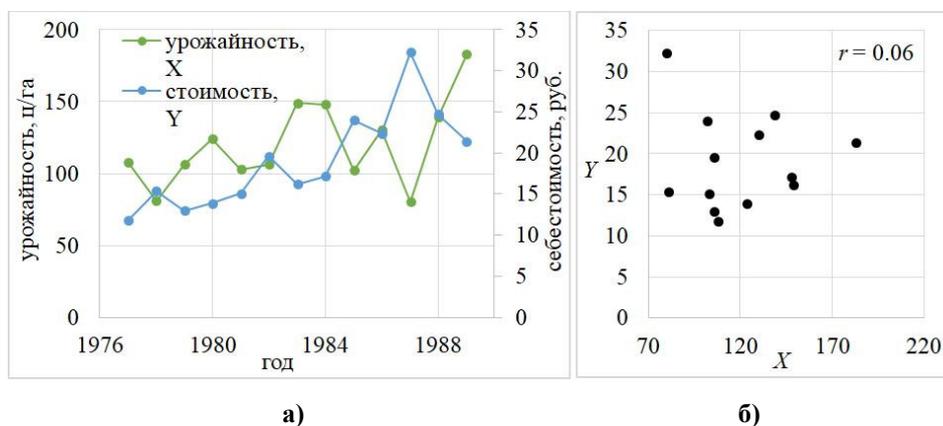


Рисунок 5. Данные по урожайности картофеля X и его стоимости Y с 1977 по 1989 гг.: а) во времени, б) диаграмма рассеяния

Figure 5. Data on potato yield X and its cost Y from 1977 to 1989: a) in time, b) scatter diagram

Ситуация 2), когда коэффициент корреляции дает высокое абсолютное значение, а связь является ложной, может иметь разные причины:

– выбран случайным образом отрезок времени, в течение которого признаки варьируют по совпадению случайных причин (пример такой ситуации – случайная связь между температурой приземного воздуха и осадками за выбранный недлительный период, описанной в работе (Максимова и др., 2023)),

– неоднородность данных в рядах: наличие выбросов или периодические ошибки измерений (известно, к примеру, что коэффициент корреляции Пирсона чувствителен к выделяющимся наблюдениям (Максимова, 2024)),

– наличие сезонных волн колебаний одинаковой тенденции или наличие влияния третьего фактора.

Приведем пример для последней ситуации. Рассмотрим данные⁴⁾ для США по потреблению сыра моцарелла в год на человека, X (фунтов), и по числу защищенных диссертаций в строительстве, Y (шт.), с 2000 по 2009 гг. (рис. 6а). Коэффициент корреляции между признаками X и Y равен $r = 0.96$, который, очевидно, высокий даже для такого небольшого объема данных (рис. 6б). Кажется бы, что можно сделать вывод, что увеличение съедаемого сыра моцарелла способствует увеличению числа диссертаций в строительстве, или чем больше докторских степеней будет получено, тем больше сыра моцарелла будет потребляться. Однако анализ за представленные годы говорит о приросте численности населения⁵⁾, которое как третий фактор наиболее

⁴⁾данные восстановлены из интернет-ресурса: <http://www.tylervigen.com/spurious-correlations> (дата обращения 24.01.2024)

⁵⁾ https://ru.wikipedia.org/wiki/Население_США (дата обращения 28.01.2025)

вероятно оказывает влияние на признак Y (рассчитанная корреляция между числом защищенных диссертаций в строительстве и общей численностью населения составила $r = 0.95$). Также с ростом населения могут суммарно варьироваться вкусовые предпочтения людей в сторону увеличения потребления сыра. Возникает тем самым однонаправленность временных трендов для X и Y (рис. 6а). Действительно, полученная первоначально высокая связь не является причинно-следственной и наиболее вероятно вызвана действием третьего фактора – приростом населения⁴). Рассматривать такую взаимосвязь для практических выводов не уместно.

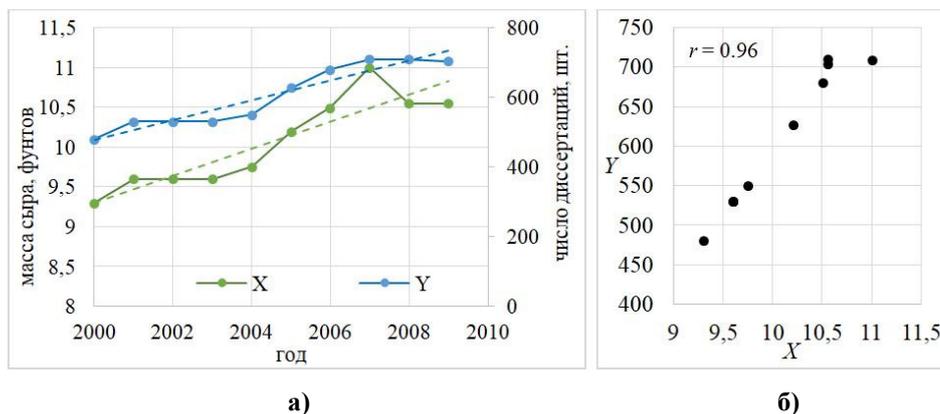


Рисунок 6. Данные⁵) по годовому потреблению сыра моцарелла на душу населения в США (X, фунты) и числу защищенных диссертаций в строительстве (Y, шт.) с 2000 по 2009 гг.: а) во времени, б) диаграмма рассеяния

Figure 6. Data in the United States on annual per capita consumption of mozzarella cheese (X, pounds) and the number of dissertations defended in construction (Y, units) from 2000 to 2009: a) over time, b) scatter plot

Таким образом, коэффициент корреляции при наличии сонаправленных трендов может быть необоснованно высоким по абсолютной величине. И в одном, и в другом ряду уровни более поздних лет будут либо больше, либо меньше, уровней более ранних периодов. Коэффициент корреляции окажется положительным. Но одинаковая направленность трендов вовсе не влечет причинной зависимости.

Иногда при наличии трендов для установления взаимосвязи рядов строят уравнение регрессии, описывающее зависимость значений одного ряда от значений другого при включении переменной времени t в уравнение в качестве объясняющей. Этот прием широко используется в анализе временных рядов, когда тенденция фиксируется через включение фактора времени в модель. Данный подход позволяет учесть всю информацию, содержащуюся в исходных данных, и не приводит к потере числа наблюдений. Основная проблема такого подхода – интерпретация коэффициента при факторе времени, поэтому этот прием нельзя считать универсальным при сравнении временных рядов (Елисеева, Юзбашев, 2006).

Следствием изложенных подходов служит тот факт, что ответ на вопрос, насколько допустимо переносить выводы о тесноте связи между колебаниями

на связь динамических рядов в целом, зависит от качественного содержания процесса и причинного механизма связи.

Иногда исследователей при сравнении временных рядов интересуют все три рассмотренные в настоящей статье задачи, поэтому в таком случае проводится комплексный анализ с разработкой подходов, присущих специфике данных (Максимова, Гинзбург, 2020; Бардин и др., 2020; Sun et al., 2015). Например, в работе автора, посвященной выбору глобальной климатической модели, наиболее точно воспроизводящей региональные проекции климата, исследуются как близость числовых показателей подготовленных массивов, так и попарная близость уровней рядов, и синхронность трендов, и корреляция (Максимова, Гинзбург, 2020; Максимова, Кухта, 2022а). На рис. 7 приведен результат такого подбора модели для Арктического региона (Архангельская обл., Ненецкий АО, республика Коми) после предварительного сглаживания данных с окном $l = 10$ лет (Максимова, Кухта, 2022а).

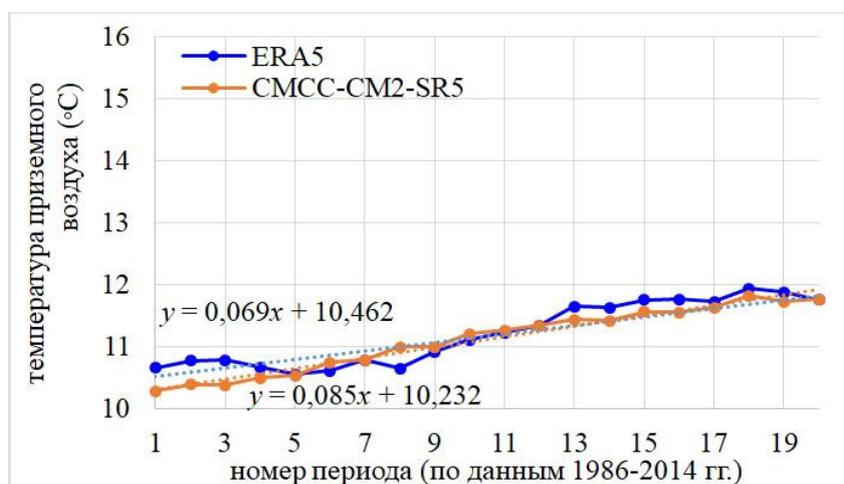


Рисунок 7⁶⁾. Средние температуры приземного воздуха за июль-сентябрь с окном сглаживания $l = 10$ лет модели CMCC-CM2-SR5 по Арктическому региону (Архангельская обл., Ненецкий АО, респ. Коми) для исторического периода в сравнении с данными реанализа ERA5 с указанием трендов

Figure 7. Average surface air temperatures for July-September with a smoothing period $l = 10$ years of the CMCC-CM2-SR5 model for the Arctic region (Arkhangelsk Region, Nenets Autonomous District, Rep. Komi) with trends: for the historical period in comparison with the ERA5 reanalysis data

С целью выбора модели была проведена верификация глобальных климатических моделей проекта CMIP6⁷⁾ по главному члену ансамбля r1i1p1 с

⁶⁾ Расчеты и рисунок выполнен автором и опубликован в исследовании (Максимова, Кухта, 2022а)

⁷⁾ CMIP – проект сравнения объединенных глобальных климатических моделей, CMIP 6 – 6-я фаза этого международного проекта: <https://esgf-index1.ceda.ac.uk/projects/cmip6-ceda/> (дата обращения 03.06.2021)

данными реанализа ERA5 по температуре приземного воздуха за период с 1986 по 2014 гг. для вегетационного сезона июль-сентябрь. Наиболее близкой в этих условиях к данным реанализа оказалась модель CMCC-CM2-SR5⁸⁾ с коэффициентом корреляции $r = 0.93$, близкими трендами и поточечной близостью. Это дало возможность признать ее наиболее точной в оценке воспроизводимости климата по температуре приземного воздуха для Арктического региона и использовать для дальнейшего прогноза изменения температуры.

Выводы и заключение

При сравнении временных рядов в работе рассмотрены следующие задачи: поточечная близость, синхронность изменения во времени и корреляция уровней рядов. Предложены подходы решения каждой задачи в-отдельности, продемонстрированные как на модельных примерах, так и на задачах прикладных исследований. Выявлены особенности каждого подхода, которые необходимо учитывать при постановке задачи о сравнении массивов данных во времени.

Установлено, что выводы, полученные при решении только одной задачи, нельзя переносить как на выводы другой, так и на связь временных рядов в целом. Например, синхронность изменения показателей рядов не обязательно влечет поточечную близость и корреляцию их уровней, а высокая корреляция может не обеспечивать синхронность и давать ложные выводы о взаимосвязи показателей во времени, и т.п.

Возможность получения выводов о наличии причинно-следственных связей временных рядов в целом зависит от качественного анализа содержания рассматриваемых процессов. Это проблема, выходящая далеко за пределы статистической науки, и чаще всего лежит в понимании описываемых процессов.

Благодарности

Автор признателен Кухта А.Е. (к.б.н., в.н.с. Института глобального климата и экологии имени академика Ю.А. Израэля) за идеологический вклад в понимание и интерпретацию биологических выводов.

Исследование выполнено в соответствии с планом НИТР РОСГИДРОМЕТА: "Развитие методов и технологий мониторинга и оценки загрязнения природной среды вследствие трансграничного атмосферного переноса загрязняющих веществ и кислотных выпадений (ЕЭК ООН: ЕМЕП, МСП КМ; ЕАНЕТ). регистрационный номер 125031703847-1".

⁸⁾ <https://cmip-publications.llnl.gov/view/CMIP6/?type=model&option=CMCC-CM2-SR5>
(дата обращения 29.01.2025)

Список литературы

- Асфандиярова, Л.Р., Асфандияров, Р.Н., Рашидова, А.Р., Юнусова, Г.В. (2013) Анализ сезонных изменений концентраций загрязняющих веществ в р. Белая, *Башкирский химический журнал*, т. 20 (4), с. 126-131.
- Бардин, М.Ю., Ранькова, Э.Я., Платова, Т.В., Самохина, О.Ф., Корнева, И.А. (2020) Современные изменения приземного климата по результатам регулярного мониторинга, *Метеорология и гидрология*, № 5, с. 29-45.
- Благовещенский, Ю.Н. (2009) *Тайны корреляционных связи в статистике*, Серия «Библиотека Солев», М., Научно-издательский центр ИНФРА-М, Научная книга, 158 с.
- Бурнаев, Е.В., Оленев, Н.Н. (2006) Меры близости на основе вейвлет коэффициентов для сравнения статистических и расчетных временных рядов, *Межвузовский сборник научных и научно-методических трудов за 2005 год (Десятый выпуск)*, Киров, Изд-во Вятский государственный университет, с. 41-51.
- Елисеева, И.И., Юзбашев, М.М. (2006) *Общая теория статистики: учебник для студентов вузов, обучающихся по направлению и специальности "Статистика"*, М., Финансы и статистика, 654 с.
- Израэль, Ю.А., Павлов, А.В., Анохин, Ю.А., Мяч, Л.Т., Шерстюков, А.Б. (2006) Статистические оценки изменения элементов климата в районах вечной мерзлоты на территории Российской Федерации, *Метеорология и гидрология*, т. 5, с. 27-38
- Катаева, Е.С., Якимук, А.Ю. (2019) Применение метода выделения синхронности при оценке сходства вокальных исполнений, *Доклады Томского государственного университета систем управления и радиоэлектроники*, т. 22 (3), с. 49-54.
- Кобзарь, А.И. (2012) *Прикладная математическая статистика. Для инженеров и научных работников*, М., ФИЗМАТЛИТ, 816 с.
- Кухта, А.Е., Максимова, О.В., Махрова, Т.Г., Громов, С.А. (2024) Отклик приростов сосны обыкновенной *Pinus sylvestris* бореальных фитоценозов на трансграничное загрязнение, *Проблемы региональной экологии*, № 6, с. 26-34, doi: 10.24412/1728-323X-2024-2-26-34.
- Лбов, Г.С., Герасимов, М.К. (2010) Прогнозирование экстремальных событий на основе анализа многомерных разнотипных временных рядов, *Вычислительные технологии*, т. 15 (5), с. 81-91.
- Лебедев, В.А., Спесивцев, Б.И. (2015) Возможность прогнозирования кризисных явлений на мировом рынке цен на нефть и газ на основе корреляционного анализа, *Записки Горного института*, т. 213, с. 94-99.
- Максимова, О.В. (2024) Об ошибках обращения с коэффициентом корреляции Пирсона, *Экологический мониторинг и моделирование экосистем*, т. XXXV, № 3-4, с. 88-102.
-

Максимова, О.В., Гинзбург, В.А. (2020) Изменения приземной температуры воздуха в континентальной части Арктического региона (Архангельская обл., респ. Коми, Ненецкий АО) – анализ и выбор модели прогноза, *Мониторинг состояния и загрязнения окружающей среды. Экосистемы и климат Арктической зоны*, Расширенные тезисы докладов, Институт глобального климата и экологии имени академика Ю.А. Израэля, Институт географии РАН, с. 79-82.

Максимова, О.В., Кухта, А.Е. (2022а) Оценка изменения линейных приростов сосны обыкновенной Печоро-Илычского заповедника на основе климатического прогноза температуры приземного воздуха в российской Арктике, *Арктика: экология и экономика*, т. 12, № 1, с. 77-86, doi: 10.25283/2223-4594-2022-1-77-86.

Максимова, О.В., Кухта, А.Е. (2022б) Вариабельность линейных и радиальных приростов сосны обыкновенной побережья Белого моря в зависимости от условий произрастания, *Экологический мониторинг и моделирование экосистем*, т. XXXIII, № 20-36, с. 20-36, doi: 10.21513/0207-2564-2022-3-4-20-36.

Максимова, О.В., Кухта, А.Е. (2023) Сравнительный анализ изменчивости линейных и радиальных приростов сосны обыкновенной в разных типах местообитаний государственного природного заповедника «Кивач», *Экологический мониторинг и моделирование экосистем*, т. 34, № 3-4, с. 45-61, doi: 10.21513/0207-2564-2023-3-4-45-61.

Максимова, О.В., Кухта, А.Е. (2024) Об особенностях сравнения динамических рядов параметров хода роста сосны обыкновенной в разных типах местообитаний, *Стандарты и качество*, № 13, с. 55-61.

Максимова, О.В., Кухта, А.Е., Коротков, С.А. (2023) Воздействие черного углерода и других климатических факторов на линейные приросты сосны обыкновенной (*Pinus Sylvestris* L.) на территории заповедника «Кивач», *Лесной вестник / Forestry Bulletin*, т. 27, № 5, с. 48-59, doi: 10.18698/2542-1468-2023-5-48-59.

Розенталь, О.М., Шпер, В.Л. (2023) Разведочный анализ характеристик природных вод, *Экологические системы и приборы*, № 3, с. 21-29, doi: 10.25791/esip.3.2023.1358.

Сушко, О.П. (2017) Особенности динамики цен сельскохозяйственной продукции северных стран, *Арктика и Север*, т. 26, с. 38-52.

Феклистов, П.А., Бызова, Н.М., Пашкевич, А.И., Сазанова, Е.В., Соболев, А.Н. (2020) Дендрохронологическое исследование древесины в исторически значимых арктических объектах, *Известия высших учебных заведений. Лесной журнал*, 5(377), с. 106-118.

Хамитова, Р.Я., Лоскутов, Д.В., Валеева, Э.Р., Степанова, Н.В. (2023) Анализ взаимосвязей заболеваемости взрослого населения инсулиннезависимым сахарным диабетом и алкогольных показателей, *Наука молодых – Eruditio Juvenium*, т. 11(3), с. 303-317.

Шерстюков, Б.Г. (2011) *Изменения, изменчивость и колебания климата*, Обнинск, ФГБУ "ВНИИГМИ-МЦД", 293 с.

Шрамко, О.Г. (2012) Влияние синхронности макроэкономических показателей на оценку эффективности экономической политики государства, *Региональная экономика и управление: электронный научный журнал*, т. 30, с. 43-49.

Aronov, I.Z., Maksimova, O.V., Galkina, N.M. (2020) COVID-19 Highest Incidence Forecast in Russia Based on Regression Model, *International Journal of Mathematical, Engineering and Management Sciences*, vol. 5, no. 5, pp. 812-819, doi.org/10.33889/IJMEMS.2020.5.5.063.

Sun, Q., Miao, C., Duan, Q. (2015) Comparative analysis of CMIP3 and CMIP5 global climate models for simulating the daily mean, maximum, and minimum temperatures and daily precipitation over China, *J. Geophys. Res. Atmos.*, vol. 120, pp. 4806-4824, doi: 10.1002/2014JD022994.

Taleb, N.N. (2001) *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*, Random House, US, 316 p.

Reference

Asfandiyarova, L.R., Asfandiyarov, R.N., Rashidova, A.R., Yunusova, G.V. (2013) Analiz sezonnykh izmeneniy kontsentratsiy zagryaznyayushchikh veshchestv v R. Belaya [Analysis of seasonal changes in pollutant concentrations in the Belaya River], *Bashkirskiy khimicheskiy zhurnal*, vol. 20 (4), pp. 126-131.

Bardin, M.Yu., Ran'kova, E.Ya., Platova, T.V., Samokhina, O.F., Korneva, I.A. (2020) Sovremennyye izmeneniya prizemnogo klimata po rezul'tatam regul'yarnogo monitoring [Modern changes in surface climate based on the results of regular monitoring], *Meteorologiya i gidrologiya*, vol. 5, pp. 29-45.

Blagoveshchenskiy, Yu.N. (2009) *Tajny korrelyacionnykh svyazi v statistike* [Secrets of correlation in statistics], Seriya «Biblioteka Soley», «Nauchno-izdatel'skiy centr INFRA-M», «Nauchnaya kniga», Moscow, Russia, 158 p.

Burnayev, Ye.V., Olenov, N.N. (2006) Mery blizosti na osnove veyvlet koeffitsiyentov dlya sravneniya statisticheskikh i raschetnykh vremennykh ryadov [Measures of Proximity Based on Wavelet Coefficients for Comparison of Statistical and Estimated Time Series], *Mezhvuzovskiy sbornik nauchnykh i nauchno-metodicheskikh trudov za 2005 god (Desyatyy vypusk)* [Interuniversity collection of scientific and scientific-methodical works for 2005 (Tenth issue)], Izd-vo Vyatskiy gosudarstvennyy universitet, Kirov, Russia, pp. 41-51.

Yeliseyeva, I.I., Yuzbashev, M.M. (2006) *Obshchaya teoriya statistiki: ucheb. dlya studentov vuzov, obuchayushchikhsya po napravleniyu i spetsial'nosti "Statistika"* [General Theory of Statistics: a textbook for university students majoring in Statistics], Finansy i statistika, Moscow, Russia, 654 p.

Izrael', Yu.A., Pavlov, A.V., Anokhin, Yu.A., Myach, L.T., Sherstyukov, A.B. (2006) Statisticheskiye otsenki izmeneniya elementov klimata v rayonakh vechnoy

merzloty na territorii Rossiyskoy Federatsii [Statistical estimates of climate change in permafrost regions on the territory of the Russian Federation], *Meteorology and hydrology*, vol. 5, pp. 27-38.

Katayeva, Ye.S., Yakimuk, A.Yu. (2019) Primeneniye metoda vydeleniya sinkhronnosti pri otsenke skhodstva vokal'nykh ispolneniy [Application of the synchronicity detection method in assessing the similarity of vocal performances], *Doklady Tomskogo gosudarstvennogo universiteta sistem upravleniya i radioelektroniki*, vol. 22 (3), pp. 49-54.

Kobzar', A.I. (2012) *Prikladnaya matematicheskaya statistika. Dlya inzhenerov I nauchnykh rabotnikov* [Applied mathematical statistics. For engineers and scientists], Fizmatlit, Moscow, Russia, 816 p.

Kukhta, A.Ye., Maksimova, O.V., Makhrova, T.G., Gromov, S.A. (2024) Otklik prirostov sosny obyknovennoy *Pinus sylvestris* boreal'nykh fitotsenozov na transgranichnoye zagryazneniye [Response of Scots pine *Pinus sylvestris* growth in boreal phytocenoses to transboundary pollution], *Problemy regional'noy ekologii*, no. 6, pp. 26-34, doi: 10.24412/1728-323X-2024-2-26-34.

Lbov, G.S., Gerasimov, M.K. (2010) Prognozirovaniye ekstremal'nykh sobyitiy na osnove analiza mnogomernykh raznotipnykh vremennykh ryadov [Forecasting extreme events based on the analysis of multivariate heterogeneous time series], *Vychislitel'nyye tekhnologii*, vol. 15 (5), pp. 81-91.

Lebedev, V.A., Spesivtsev, B.I. (2015) Vozmozhnost' prognozirovaniya krizisnykh yavleniy na mirovom rynke tsen na neft' i gaz na osnove korrelyatsionnogo analiza [The Possibility of Forecasting Crisis Phenomena in the World Market of Oil and Gas Prices Based on Correlation Analysis], *Zapiski Gornogo instituta*, no. 213, pp. 94-99.

Maksimova, O.V. (2024) Ob oshibkakh obrashcheniya s koeffitsiyentom korrelyatsii Pirsona [The Possibility of Forecasting Crisis Phenomena in the World Oil and Gas Price Market Based on Correlation Analysis], *Ekologicheskii monitoring i modelirovaniye ekosistem*, vol. XXXV, no. 3-4, pp. 88-102.

Maksimova, O.V., Ginzburg, V.A. (2020) Izmeneniya prizemnoy temperatury vozdukha v kontinental'noy chasti Arkticheskogo regiona (Arkhangel'skaya obl., resp. Komi, Nenetskiy AO) – analiz i vybor modeli prognoza [Changes in Surface Air Temperature in the Continental Part of the Arctic Region (Arkhangelsk Region, Komi Republic, Nenets Autonomous Okrug) – Analysis and Selection of a Forecast Model], *Monitoring sostoyaniya i zagryazneniya okruzhayushchey sredy. Ekosistemy i klimat Arkticheskoy zony* [Monitoring of the state and pollution of the environment. Ecosystems and climate of the Arctic zone], Extended abstracts of reports, Yu. A. Izrael Institute of Global Climate and Ecology, Institute of Geography of the Russian Academy of Sciences, pp. 79-82.

Maksimova, O.V., Kukhta, A.Ye. (2022a) Otsenka izmeneniya lineynykh prirostov sosny obyknovennoy Pechoro-Ilychskogo zapovednika na osnove klimaticheskogo prognoza temperatury prizemnogo vozdukha v rossiyskoy Arktike

[Assessment of changes in linear increments of Scots pine of the Pechora-Ilychsky Reserve based on the climate forecast of surface air temperature in the Russian Arctic], *Arktika: ekologiya i ekonomika*, vol. 12, no. 1, pp. 77-86, doi: 10.25283/2223-4594-2022-1-77-86.

Maksimova, O.V., Kukhta, A.Ye. (2022b) Variabel'nost' lineynykh i radial'nykh prirostov sosny obyknovennoy poberezh'ya Belogo morya v zavisimosti ot usloviy proizrastaniya [Variability of linear and radial increments of Scots pine on the White Sea coast depending on growing conditions], *Ekologicheskij monitoring i modelirovaniye ekosistem*, vol. XXXIII, vol. 20-36, pp. 20-36, doi: 10.21513/0207-2564-2022-3-4-20-36.

Maksimova, O.V., Kukhta, A.Ye. (2023) Sravnitel'nyy analiz izmenchivosti lineynykh i radial'nykh prirostov sosny obyknovennoy v raznykh tipakh mestoobitaniy gosudarstvennogo prirodnogo zapovednika «Kivach» [Comparative analysis of variability of linear and radial increments of Scots pine in different types of habitats of the Kivach State Nature Reserve], *Ekologicheskij monitoring i modelirovaniye ekosistem*, vol. 34, № 3-4, pp. 45-61, doi: 10.21513/0207-2564-2023-3-4-45-61.

Maksimova, O.V., Kuhta, A.E. (2024) Ob osobennostyakh sravneniya dinamicheskikh ryadov parametrov hoda rosta sosny obyknovennoy v raznykh tipakh mestoobitaniy [On the peculiarities of comparison of dynamic series of parameters of the growth course of Scots pine in different types of habitats], *Standarty I kachestvo*, no.13, pp. 55-61.

Maksimova, O.V., Kuhta, A.E., Korotkov, S.A. (2023) Vozdejstvie chernogo ugleroda I drugih klimaticheskikh faktorov na linejnye prirosty sosny obyknovennoy (*Pinus Sylvestris* L.) na territorii zapovednika «Kivach» [Impact of black carbon and other climatic factors on linear growth of Scots pine (*Pinus Sylvestris* L.) in the territory of the Kivach nature reserve], *Lesnoj vestnik*, vol. 27, no. 5. pp. 48-59, doi: 10.18698/2542-1468-2023-5-48-59.

Rozental', O.M., Shper, V.L. (2023) Razvedochnyy analiz kharakteristik prirodnykh vod [Exploratory analysis of natural water characteristics], *Ekologicheskije sistemy i pribory*, no. 3, pp. 21-29, doi: 10.25791/esip.3.2023.1358.

Sushko, O.P. (2017) Osobennosti dinamiki tsen sel'skokhozyaystvennoy produktsii severnykh stran [Features of the dynamics of prices for agricultural products of northern countries], *Arktika i Sever*, vol. 26, pp. 38-52.

Feklistov, P.A., Byzova, N.M., Pashkevich, A.I., Sazanova, Ye.V., Sobolev, A.N. (2020) Dendrokronologicheskoye issledovaniye drevesiny v istoricheskikh znachimykh arkticheskikh ob'yektakh [Dendrochronological study of wood in historically significant Arctic objects], *Izvestiya vysshikh uchebnykh zavedeniy. Lesnoy zhurnal*, 5 (377), pp. 106-118.

Khamitova, R.Ya., Loskutov, D.V., Valeyeva, E.R., Stepanova, N.V. (2023) Analiz vzaimosvyazey zaboilevayemosti vzroslogo naseleniya insulinnezavisimym

sakharnym diabetom i alkogol'nykh pokazateley [Analysis of the relationships between the incidence of non-insulin-dependent diabetes mellitus in the adult population and alcohol indicators], *Nauka molodykh – Eruditio Juvenium*, vol. 11 (3), pp. 303-317.

Sherstyukov, B.G. (2011) *Izmeneniya, izmenchivost' i kolebaniya klimata* [Climate Changes, Variability and Fluctuations], "VNIIGMI-MTSD", Obninsk, Russia, 293 p.

Shramko, O.G. (2012) Vliyaniye sinkhronnosti makroekonomicheskikh pokazateley na otsenku effektivnosti ekonomicheskoy politiki gosudarstva [The Impact of Synchronicity of Macroeconomic Indicators on the Assessment of the Effectiveness of State Economic Policy], *Regional'naya ekonomika i upravleniye: elektronnyy nauchnyy zhurnal*, vol. 30, pp. 43-49

Aronov, I.Z., Maksimova, O.V., Galkina, N.M. (2020) COVID-19 Highest Incidence Forecast in Russia Based on Regression Model, *International Journal of Mathematical, Engineering and Management Sciences*, vol. 5, no. 5, pp. 812-819, doi.org/10.33889/IJMEMS.2020.5.5.063.

Sun, Q., Miao, C., Duan, Q. (2015) Comparative analysis of CMIP3 and CMIP5 global climate models for simulating the daily mean, maximum, and minimum temperatures and daily precipitation over China, *J. Geophys. Res. Atmos.*, vol. 120, pp. 4806-4824, doi: 10.1002/2014JD022994.

Taleb, N.N. (2001) *Foiled by Randomness: The Hidden Role of Chance in Life and in the Markets*, Random House, US, 316 p.

Статья поступила в редакцию (Received): 31.01.2025.

Принята к публикации (Accepted): 21.04.2025.

Для цитирования / For citation:

Максимова, О.В. (2025) Сравнение временных рядов в прикладных исследованиях: близость, синхронность и корреляция, *Экологический мониторинг и моделирование экосистем*, т. XXXVI, № 1-2, с. 46-64, doi:10.24412/2782-3237-2025-1-2-46-64.

Maksimova, O.V. (2025) Comparison of time series in applied research: proximity, synchronism and correlation, *Environmental Monitoring and Ecosystem Modelling*, vol. XXXVI, no. 1-2, pp. 46-64, doi:10.24412/2782-3237-2025-1-2-46-64.