

УДК 519.25, 330.46

DOI: 10.25559/SITITO.15.201902.441-455

## О корректности вычислительной задачи построения композитных индексов

Т. В. Жгун\*, А. В. Липатов, Г. А. Чалов

Новгородский государственный университет им. Ярослава Мудрого, г. Великий Новгород, Россия

173003, Россия, Новгородская обл., г. Великий Новгород, Большая Санкт-Петербургская ул., д. 41

\*Tatyana.Zhgun@novsu.ru

### Аннотация

Вычисление композитных индексов плохо формализованной системы, основанное на данных, содержащих ошибки, может рассматриваться как проблема выделения сигнала из шума. Сигнал в этом случае представляет собой весовые коэффициенты линейной свертки индикаторов в композитном индексе. Определяемые веса должны отражать структуру оцениваемой системы. Успешное применение метода главных компонент для описания структур различных систем позволяет предположить, что этот метод также даст адекватные результаты для описания социальных систем. Однако многомерный анализ определяют структуру основных компонентов и основных факторов по-разному для разных наблюдений. Причиной этого может быть наличие неизбежных ошибок в используемых данных. В качестве способа избежать этого предлагается модификация метода анализа главных компонент с учетом наличия ошибок в используемых данных. Решение проблемы требует детального понимания влияния ошибок входных данных на определяемые параметры. Поэтому важным является вопрос о корректности вычислительной задачи. Предложено уточнение понятия корректности задачи вычисления составного индекса изменения качества системы. Следствием устойчивости является в среднем небольшое изменение (приращение) ранга объектов для разных измерений. Этот прирост можно оценить апостериорно по ряду наблюдений с использованием предлагаемого дисперсионного критерия. Представлены результаты оценки устойчивости составных индексов по этому критерию. Интегральные показатели, рассчитанные по авторской методике, имеют хорошую устойчивость.

**Ключевые слова:** интегральные индикаторы, ошибки данных, метод главных компонент, корректность и устойчивость задачи, критерий надежности рангов.

**Для цитирования:** Жгун Т. В., Липатов А. В., Чалов Г. А. О корректности вычислительной задачи построения композитных индексов // Современные информационные технологии и ИТ-образование. 2019. Т. 15, № 2. С. 441-455. DOI: 10.25559/SITITO.15.201902.441-455

© Жгун Т. В., Липатов А. В., Чалов Г. А., 2019



Контент доступен под лицензией Creative Commons Attribution 4.0 License.  
The content is available under Creative Commons Attribution 4.0 License.



## On the Correctness of the Computational Problem of Composite Indices Construction

T. V. Zhgun\*, A. V. Lipatov, G. A. Chalov

Yaroslav-the-Wise Novgorod State University, Velikiy Novgorod, Russia

41 Bolshaya St. Petersburg St., Velikiy Novgorod 173003, Novgorod region, Russia

\*Tatyana.Zhgun@novsu.ru

### Abstract

The computation composite indices of a poorly formalized system, based on data containing errors, can be considered as a problem signal-to-noise discrimination. The signal in this case is the weight coefficients of the linear convolution of indicators. The weights to be determined should reflect the structure of the system being evaluated. The successful application of principal component analysis in different systems structure description allows us to suggest that the method will also provide adequate results to describe social systems. However, principal component analysis and factor analysis determine the structure of principal components and principal factors differently for different observations. The reason for this may be the presence of inevitable errors in the used data. As a method of avoiding this, a modification of the principal component analysis method is proposed, taking into account the presence of errors in the data used. A solution of the problem requires a detailed understanding of input data errors' influence on the calculated model's parameters. Therefore, the question of the problem correctness is essential. A clarification of the concept of computation a system's quality changes composite index problem correctness is proposed. The consequence of the stability is on average a slight change (increment) of objects Rank for different measurements. This increment can be estimated a posteriori using a number of observations of the proposed variance criterion. The results of different composite index evaluation stability according to this criterion are presented. The integral indicators computed using the author's method have a good stability.

**Keywords:** composite index, data error, principal component analysis, correctness and stability of composite index, rank robustness tests.

**For citation:** Zhgun T.V., Lipatov A.V., Chalov G.A. On the Correctness of the Computational Problem of Composite Indices Construction. *Sovremennye informacionnye tehnologii i IT-obrazovanie* = Modern Information Technologies and IT-Education. 2019; 15(2):441-455. DOI: 10.25559/SITI-TO.15.201902.441-455



## Введение

Задача определения комплексного показателя сложных систем возникают как при исследовании физических явлений, технических систем, так при решении проблем управления социально-экономическими системами. При описании стохастических динамических систем в задачах гидродинамики, магнитной гидродинамики, астрофизике, физике плазмы, радиофизике интегральные величины, характеризующие такие системы, являются их основными характеристиками. Например, все законы сохранения в механике и электродинамике сплошных сред записываются для интегральных величин. Интегральные характеристики описывают динамику в целом, позволяя отвлечься от побочных эффектов, связанных со случайностью показателей, искаженных помехами и являются ключом к пониманию структурообразования в стохастических динамических системах [1].

Построение интегрального индикатора вводит отношения порядка на многомерном множестве объектов и позволяет сравнивать качество объектов. Сопоставление интегральных индикаторов рассматриваемых объектов и их рейтингов, которые определяются интегральными характеристиками, позволяет судить о степени достижения цели управления. Международные организации, аналитические центры и социальные науки на рубеже тысячелетий существенно увеличили количество применяемых интегральных индикаторов для измерения разнообразных латентных характеристик социально-экономических систем: социального капитала, человеческого развития, качества жизни, качества управления и др. Согласно данным ООН, к 2011 году существовало 290 индексов, разработанных для ранжирования или комплексной оценки стран. Обзор феноменального роста количества применяемых для комплексной оценки стран сводных индексов, проведенного Бандура<sup>1,2,3</sup>, показал, что только 26 из них (9%) было сформировано до 1991 года, 132 (46%) — с 1991 года по 2006 год, а в период 2007–2011 годы было сформировано 130 сводных индексов. Огромное количество применяемых методик оценивания латентных характеристик социальных систем свидетельствует о неудовлетворенности получаемыми результатами и о необходимости дальнейших исследований в этой области.

Стремительное увеличение числа композитных индексов является явным признаком их значимости в политике и экономике в целом. Все основные международные организации, такие как Организация экономического сотрудничества и развития (OECD), Европейский Союз, Всемирный экономический форум или Международный Валютный Фонд конструируют

составные показатели в самых различных областях<sup>4</sup> [2]. Общей целью большинства из этих показателей является ранжирование объектов (стран) и их сравнительный анализ<sup>5,6</sup> [3, 4]. Использование для административных, управленческих, политических целей единого показателя, характеризующего сложные, плохо формализуемые процессы (качество жизни, финансовую устойчивость, рыночную политику и т.д.) является чрезвычайно привлекательной идеей. Следовательно, улучшение качества таких показателей является очень важным вопросом исследования и с теоретической, и с практической точек зрения.

Несмотря на огромное количество применяемых композитных индексов, нерешенные методические проблемы разработки интегральных индикаторов, проблемы неполноты и достоверности большинства статистических показателей, приводят к тому, что часто вычисленные композитные индексы вызывают больше вопросов, чем дают ответов. Обсуждение плюсов и минусов составных показателей приводится в работах [2, 3].

Организация экономического сотрудничества и развития (OECD) ведет непрерывную работу по совершенствованию методик построения композитных индексов<sup>7</sup> [2, 4, 5]. В 2008 году OECD совместно с Объединенным исследовательским центром (Joint Research Centre European Commission) с подготовлен справочник [6], явившийся итогом многолетних исследований в этой области<sup>8</sup> [2, 4, 5, 7], где изложен набор технических принципов формирования составных показателей. Основным методом агрегации данных авторы выбирают линейную свертку показателей, а основным инструментом построения сводных показателей — многомерный анализ.

Итак, для построения качественного интегрального показателя необходима, во-первых, тщательная проработка аспектов измеряемого явления, ибо «то, что плохо определено, вероятно, будет плохо измерено» [2], во-вторых, качественный выбор данных, так как качество составных индикаторов в значительной степени зависит от качества базовых показателей, и в-третьих, адекватный инструмент для работы с многомерными данными.

Принципиальным отличием определения композитных индексов для социально-экономических систем является неопределенность качества используемых данных, в отличие от характеристик технических и физических систем, для которых заранее известна погрешность измерения. Тем не менее, именно статистические данные, содержащие неустраняемые погрешности, в настоящее время представляют собой наилучшие оценки имеющихся реальных величин в социальных системах [5].

<sup>1</sup> Bandura R., del Campo C.M. A Survey of Composite Indices Measuring Country Performance: 2006 Update. A UNDP/ODS Working Paper. United Nations Development Programme, New York, 2006. [Электронный ресурс]. URL: [https://www.undp.org/content/dam/undp/library/corporate/Development%20Studies/measuring\\_country\\_performance\\_2006update.pdf](https://www.undp.org/content/dam/undp/library/corporate/Development%20Studies/measuring_country_performance_2006update.pdf) (дата обращения: 10.04.2019).

<sup>2</sup> Bandura R. A survey of composite indices measuring country performance: 2008 update. Technical report, Office of Development Studies, United Nations Development Programme (UNDP), New York, 2008.

<sup>3</sup> Bandura R. Composite Indicators and Rankings: Inventory 2011. Technical report, Office of Development Studies, United Nations Development Programme (UNDP), New York, 2011.

<sup>4</sup> Там же.

<sup>5</sup> Там же.

<sup>6</sup> Sharpe A. Literature Review of Frameworks for Macro-indicators // CSLS Research Reports 2004-03. Centre for the Study of Living Standards, 2004. URL: <http://www.csls.ca/reports/LitRevMacro-indicators.pdf> (дата обращения: 10.04.2019).

<sup>7</sup> Nicoletti G., Scarpetta S., Boylaud O. Summary indicators of product market regulation with an extension to employment protection legislation. Economics department working papers No. 226, ECO/WKP(99)18. 2000. URL: <https://www.oecd.org/eco/outlook/1880867.pdf> (дата обращения: 10.04.2019).

<sup>8</sup> Там же.



Инструментом работы с многомерными данными является многомерный статистический анализ данных, а именно факторный анализ и метод главных компонент (МГК). Применение этой методики к характеристике физических, биологических, технических систем дает содержательные и достоверные результаты. Поэтому является очевидным необходимость применения этого инструментария и для описания социальных систем. Впервые факторный анализ был использован для объединения множества показателей в единый индекс при разработке индекса здоровья в работе Хайтауэра в 1978 году [13], и сейчас широко используется для этих целей<sup>9</sup> [2, 7]. Метод главных компонент был выбран в качестве стандарта метода построения при вычислении Показателей социально-экономического статуса (*Socio-Economic Status Indices, SES*), и вычисляемый индекс определяла проекция на первую главную компоненту [9, 10]. Эту же методику использовали Линдман и Селин при создании Индекса экологической устойчивости [11], Сомарриба и Пена при измерении качества жизни в Европе [12]. Среди отечественных исследований следует отметить работы С.А. Айвазяна [13] по определению индекса качества жизни населения с использованием первой главной компоненты.

Однако первая главная компонента хорошо приближает моделируемую ситуацию, если максимальное собственное число ковариационной матрицы дает значительный вклад (не менее 70%) в сумму всех собственных чисел. В этом случае структура системы удовлетворительно описывается первой главной компонентой. Это условие выполняется, если система описывается небольшим количеством признаков (не более пяти), и при этом одно из свойств системы очевидно доминирует над остальными. При описании социально-экономических систем структура системы не допускает простого приближения. В качестве выхода из этой ситуации рассматривается согласно [13] понижение порога информативности до 55%, и деление исходной системы на подсистемы, описываемые меньшим количеством переменных. В рассмотренных выше исследованиях [8-10, 12] вклад наибольшего собственного числа колебался от 13% до 38%, за исключением работы [12], рассматривавшей модельный пример, где эта цифра составила 56%. Авторы следовали рекомендациям [12], утверждавшим, что первая главная компонента дает удовлетворительные весовые коэффициенты при вычислении интегрального показателя даже в тех случаях, когда наибольшее собственное число вносит небольшой вклад в сумму всех собственных чисел. Однако такое утверждение нельзя назвать бесспорным.

Исследователи Организации экономического сотрудничества и развития придерживаются иной точки зрения. Для формирования составного индикатора используется факторный анализ, где метод главных компонент применяется исключительно для извлечения факторов с тем, чтобы количество извлеченных факторов объясняло бы более 50% общей дисперсии. Значение композитного индекса в этом случае определяют только значимые нагрузки выбранных главных факторов после вращения. Де-факто именно эта методика становится стандартом при вычислении композитных индексов<sup>10</sup> [2, 7]. Хотя авторы [2] обращают внимание, что различные методы извлечения главных компонент и различные способы враще-

ния подразумевают различные значимые переменные, а значит, разные веса переменных при вычислении составного показателя, следовательно, и различные значения вычисляемого интегрального индикатора. Кроме того, факторный анализ предполагает наличие достаточной корреляции между исходными переменными, что в некотором смысле противоречит идее полного описания исследуемого явления совокупностью независимых величин.

Следует также отметить еще одно обстоятельство. Методика определения весовых коэффициентов с помощью факторного анализа (так же, как и метод главных компонент) не может применяться для сравнения характеристик описываемых объектов в динамике, так как даже при фиксированных методах извлечения факторов и способа вращения, факторный анализ для разных наблюдений системы по-разному определяет структуру главных факторов (значимые нагрузки), что делает бессмысленным сравнение объектов в динамике [14]. Причиной такого явления может являться недостаточное качество используемых данных, а именно, наличие неустранимых погрешностей используемых статистических данных, которые исследователь не может оценить. Влияние этих погрешностей может являться причиной неудовлетворительного решения поставленной задачи ранжирования объектов на основании вычисляемой интегральной характеристики.

Целью предлагаемой работы является рассмотрение корректности и устойчивости решения задачи вычисления интегральной характеристики системы, которая рассматривается как задача выделения сигнала на фоне помех, а именно:

обоснование целесообразности использования весовых коэффициентов линейной свертки, определяющей интегральную характеристику как характеристик, отражающих структуру системы;

обоснование возможности модификации метода главных компонент, применяемого для выявления структуры системы для зашумленного сигнала – ряда последовательных измерений; анализ специфики исследования корректности задач вычисления интегральной характеристики системы в условиях недостаточно качественной информации;

исследование устойчивости решения задачи вычисления интегрального индикатора количественными методами.

## 1 Постановка задачи

Рассмотрим построение интегральной оценки системы из  $m$  объектов, для которой известны таблицы описаний объектов для ряда наблюдений – матрицы размерностью  $m \times n$   $A^t = \{a_{ij}^t\}_{i,j=1}^{n,m}$ ,  $t = 1, \dots, T$ . Элемент матрицы  $a_{ij}^t$  – значение  $j$ -го показателя  $i$ -го объекта, вектор  $a_i^t = (a_{i1}^t, \dots, a_{im}^t)$  – описание  $i$ -го объекта в момент  $t$ . Для каждого момента  $t$  вектор интегральных показателей имеет вид

$$q^t = A^t \cdot w^t, \quad (1)$$

Или же, для  $i$ -го объекта в момент  $t$

$$q_i^t = \sum_{j=1}^n w_j^t \cdot a_{ij}^t \quad (2)$$

где  $q^t = \langle q_1^t, q_2^t, \dots, q_m^t \rangle^T$  – вектор интегральных индикато-

<sup>9</sup> Там же.

<sup>10</sup> Там же.



ров момента  $t$ ,  $w' = \langle w'_1, w'_2, \dots, w'_m \rangle^T$  – вектор весов показателей для момента  $t$ ,  $A^t$  – матрица преобразованных данных для момента  $t$ .

Численные характеристики системы предварительно подвергнуты унификации – приведению значений переменных на отрезок  $[0, 1]$  по принципу: «чем больше, тем лучше». Если исходный показатель связан с анализируемым интегральным свойством качества монотонной зависимостью, то при унификации исходных показателей исходные переменные  $x_j$  для каждого момента наблюдения преобразовываются по правилу:

$$a'_{ij} = s_j + (-1)^{s_j} \cdot \frac{x'_{ij} - m_j}{M_j - m_j},$$

где  $s_j = 0$ , если оптимальное значение  $j$ -го показателя максимально и  $s_j = 1$ , если оптимальное значение  $j$ -го показателя минимально,  $m_j$  – наименьшее значение  $j$ -го показателя по всей выборке (глобальный минимум),  $M_j$  – наибольшее значение  $j$ -го показателя по всей выборке (глобальный максимум). Если исходный показатель связан с анализируемым интегральным свойством качества немонотонной зависимостью (т.е. внутри диапазона изменения данного показателя существует значение  $x_j^{opt}$ , при котором достигается наивысшее качество), то значение соответствующего унифицированного показателя рассчитывается по формуле:

$$a'_{ij} = \left( 1 - \frac{|x'_{ij} - x_j^{opt}|}{\max((M_j - x_j^{opt}), (x_j^{opt} - m_j))} \right)$$

В дальнейших рассуждениях в качестве исходных данных выступают данные, к которым были применены необходимые преобразования. В частности, метод главных компонент предполагает, кроме унифицирования, также предварительное центрирование данных. Каждый элемент центрированной матрицы данных вычисляется по формуле:

$$z_{ij} = \frac{a_{ij} - \bar{a}_j}{s_j},$$

где  $\bar{a}_j = \frac{1}{m} \sum_{i=1}^m a_{ij}$  – выборочная средняя,  $s_j^2 = \frac{1}{m} \sum_{i=1}^m (a_{ij} - \bar{a}_j)^2$

– выборочная дисперсия. Центрированные данные используются для вычисления весовых коэффициентов  $w'$ .

Будем рассматривать оцениваемый объект как сложную (не поддающуюся удовлетворительной формализации), большую (число состояний выше современных вычислительных возможностей) систему. Такими системами являются как каждый биологический объект, так и любая социальная система. Система доступна для наблюдения, и известно конечное, достаточно большое число регистрируемых с некоторой точностью численных характеристик этой системы в разные моменты времени. Значимость регистрируемых показателей для функционирования системы в общем случае неизвестна. Для решения задачи управления требуется дать мотивированную оценку каждого наблюдаемого объекта на всем промежутке наблюдений, т.е. вычислить в динамике интегральную характеристику качества системы. Для построения искомого интегрального индикатора качества системы требуется найти веса пока-

зателей  $w'$  для каждого момента времени, адекватно отражающие свойства рассматриваемой системы. Т.е. определяемые весовые коэффициенты должны отражать **структуру** оцениваемой системы. Такая трактовка весовых показателей устраняет одну из главных неопределенностей при конструировании интегрального индикатора.

При использовании методов вычисления интегральных характеристик «с учителем» [15,16] весовые коэффициенты (1-2) назначаются экспертами и отражают значимость (или важность) переменных в рассматриваемой системе с точки зрения экспертов. Главным возражением против этого метода является то обстоятельство, что услуги экспертов являются товаром и весьма недешевы, следовательно, экспертная оценка неизбежно будет отражать мнение заказчика исследования. С другой стороны, важность переменной со стороны эксперта может совсем не совпадать со свойствами системы, которые отражают веса переменных.

Формальные методы построения интегральных индикаторов (методы «без учителя») не используют субъективных предпочтений для определения весовых коэффициентов. Одним из наиболее простых методов анализа структуры исследуемой системы является метод главных компонент (МГК). Пространство главных компонент оптимально для моделирования внутренней структуры данных.

В работе [17] МГК применяется для анализа структуры рельефа, в работе [18] для исследования структурных свойств карбонных нанотрубок, в [19] для анализа геометрических свойств химических соединений, в работе [20] для анализа спектра лекарственных средств, в работе [21] МГК применяется для анализа структур биомолекулярных объектов. Среди преимуществ метода главных компонент для анализа потока данных (рентгеновских изображений) авторы [21] отмечают высокую эффективность метода для задачи фильтрации шумов и для поиска наиболее характерных особенностей в данных. Успешное применение МГК для описания структуры систем самого различного вида позволяет предположить, что и для описания социальных систем метод будет тоже давать адекватные результаты.

При вычислении интегральных показателей методом главных компонент строятся проекции объектов на новую систему координат, называемую главными компонентами. При этом сумма квадратов расстояний от объектов до их проекций на первую главную компоненту минимальна. Рассмотрим ортогональную матрицу  $W$  в линейной комбинации  $Z^T = A^T \cdot W$  векторов строк матрицы  $A$  таких, что векторы столбцы матрицы  $Z$  имели бы максимальную сумму дисперсий

$$\sum_{j=1}^n \sigma^2(z_j) \rightarrow \max, \quad \text{где } \sigma^2(z) = \frac{1}{m} \sum_{i=1}^m (z_i - \bar{z})^2 \quad \text{и} \quad \bar{z} = \frac{1}{m} \sum_{i=1}^m z_i.$$

В [22, 23] показано, что векторы-столбцы матрицы  $W$  есть собственные вектора ковариационной матрицы данных  $\Sigma = A^T \cdot A$ . Эта матрица может быть найдена с помощью сингулярного разложения матрицы  $A^T \cdot A$ . Так как  $\Sigma = A^T \cdot A = W \Lambda^2 W^T$  (в сингулярном разложении матрицы ортогональны, а матрица  $\Lambda$  – диагональная с невозрастающими неотрицательными элементами), то  $\Sigma W = W \Lambda^2$  и  $W \Sigma = \Lambda^2 W$  является собственной системой матрицы  $\Sigma W$ .

Обычно вектор интегральных индикаторов  $q_{PCA} = A \cdot w$  вычисляется как проекция векторов – строк матрицы унифицированных данных  $A$  на первую главную компоненту, где соб-





ственный вектор  $w$  – первая строка матрицы  $W$ . Этот вектор соответствует максимальному собственному значению ковариационной матрицы. Этот метод используется для оценки статических систем в случае, если первая главная компонента хорошо приближает моделируемую ситуацию, т.е. максимальное собственное число ковариационной матрицы дает вклад не менее 70% в сумму всех собственных чисел. Такое соотношение выполняется, если рассматривается небольшое количество признаков.  $n$  невелико, как правило, не более пяти, и одно из свойств системы явно доминирует над остальными. При описании социально-экономических систем такое предположение не выполняется: число переменных значительно более пяти, и структура системы не допускает простого приближения. В качестве выхода из этой ситуации рассматривается согласно [24, 25] понижение порога информативности до 55% и деление исходной системы на блоки-подсистемы, описываемые меньшим количеством переменных, в которых выполняется необходимое условие информативности. В качестве критерия информативности используется дисперсионная информативность. Если проекцию на первую компоненту нельзя считать удачной оценкой, то ситуацию можно исправить, если вместо одной компоненты выбрать  $l$  таких компонент, чтобы относительная доля разброса  $\gamma_l$ , приходящаяся на первые  $l$  ( $l \leq n$ ) главных компонент:

$$\gamma_l = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_l}{\lambda_1 + \lambda_2 + \dots + \lambda_n} \geq \theta \quad (3)$$

была не менее определенной величины  $\theta$ . Выбранные компоненты являются столбцами матрицы  $W$ .

При использовании МГК выбранные главные компоненты формируют матрицу  $W$  и для каждого признака эффект воздействия выбранных компонент суммируем и таким образом определяют веса в (2):

$$w'_j = \sum_{i=1}^l w'_{ij}$$

При использовании факторного анализа выбранные компоненты подвергаются вращению, после чего в них определяются значимые и незначимые нагрузки. Обычно значимыми считаются нагрузки более 0,5. Незначимые нагрузки обнуляются. Критерием удовлетворительного решения обычно является возможность четкой содержательной интерпретации получившихся факторов. Полученные факторы также формируют матрицу  $W$  и для каждого признака эффект воздействия выбранных факторов суммируется таким же образом.

## Корректность задачи вычисления интегральной характеристики системы

Итак, вектор интегральных индикаторов вычисляется как взвешенная сумма, т.е. линейная комбинация столбцов матрицы данных  $A$  и вектора весов  $W$ , который определяется вычисляемыми по корреляционной матрице данных собственными числами и собственными векторами  $q = A \cdot W$ . Проблема этого подхода состоит в том, даже небольшое возмущение исходных данных может вызвать значительное изменение весовых коэффициентов при использовании методов многомерного анализа. При этом нарушится объективность интегрального показателя, характеризующего всю выборку.

Применение самых современных технологий при решении

прикладных задач требует детального понимания влияния степени погрешности наблюдаемых параметров на степень определенности вычисляемых параметров модели. Поэтому весьма существенным является вопрос о том, корректна ли решаемая задача. Большинство вычислительных задач может быть приведено к уравнению вида:

$$A \cdot x = y, \quad x \in X, \quad y \in Y, \quad (4)$$

в котором по заданному, не обязательно линейному оператору  $A$ , действующему из пространства  $X$  в пространство  $Y$  и по заданному элементу  $x \in X$  требуется определить решение  $y \in Y$ . Приведем определение согласно [26, 27].

**Определение.** Назовем задачу (4) *корректной* Адамара на множестве  $Y$ , если при любом фиксированном наборе начальных данных  $x \in X$  решение:

а) существует в пространстве  $Y$ ; б) единственно в  $Y$ ; в) устойчиво в  $Y$ .

Воспользовавшись нормами соответствующих пространств, можно записать условие устойчивости. Оно предполагает, что для любого  $\varepsilon > 0$  можно указать такое  $\delta(\varepsilon) > 0$ , что из неравенства  $\|\hat{x} - x\| \leq \delta(\varepsilon)$  следует неравенство  $\|\hat{y} - y\| \leq \varepsilon$ , где  $\hat{y}$  – приближенное решение, соответствующее приближенному набору входных данных  $\hat{x}$ ,  $y$  – неизвестное точное решение для точного набора данных  $x$ . Если хотя бы одно из условий не выполняется, то задачу называют *некорректной*.

Идея поиска решения не на всем множестве  $Y$ , а на некотором его подмножестве  $Y^* \subset Y$ , являющемся сужением исходного пространства решений, принадлежит А.Н. Тихонову. Сужение множества возможных решений производится на основе априорной информации о свойствах решения. Сознательно или интуитивно введение априорной информации (в той или иной форме) в алгоритм решения некорректных задач позволяет получить разумные результаты. Достоверность используемой априорной информации существенно влияет на точность получаемых решений.

Если существует непрерывная зависимость решения от начальных данных, то обратный оператор  $A^{-1}$  существует и он непрерывен, а, следовательно, и ограничен, и ограничено множество  $X^* = A^{-1} \cdot Y^*$ .

**Определение.** Назовем задачу (4) *корректной* по Тихонову [27, 28]. На множестве  $Y^* \subset Y$ , а само множество  $Y^*$  – множеством корректности (множеством допустимых решений), если для любого допустимого набора начальных данных  $x^* \in X^*$  из прообраза множества допустимых решений  $X^* = A^{-1} \cdot Y^*$ : а) существует решение  $y^* \in Y^*$ ; б)  $y^*$  единственно в  $Y^*$ ; в)  $y^*$  устойчиво в  $Y^*$ .

Смысл определения корректности по Тихонову заключается в том, что корректность может быть достигнута за счет сужения рассматриваемого множества решений до множества корректности

Рассмотрим проблему корректности задачи вычисления интегральной характеристики. При вычислении интегральной характеристики системы по формуле (1-2) существование и единственность решения очевидны. Открытым остается вопрос об устойчивости полученного решения. Если на рубеже тысячелетий внимание исследователей было сосредоточено на конструировании интегральных индикаторов [5-12], то в начале третьего тысячелетия в центре внимания исследователей находятся вопросы качества композитных индексов, их корректности и устойчивости [29-31].



При исследовании устойчивости интегральных индикаторов поведение моделей рассматривается на единичном  $n$ - мерном кубе  $[0, 1]^n \subset R^n$  для каждого объекта. Значения унифицированных переменных заполняют некоторое подмножество этого куба, однако неверно полагать, что для каждого объекта любая из унифицированных переменных пробегает все возможные значения на интервале  $[0, 1]$ . Так, например, при описании качества жизни субъектов Российской Федерации безосновательно полагать, что любая переменная, описывающая эту категорию, может для Мордовии и Тувы принимать такие же значения, как для Москвы или Калининграда.

Для однократного наблюдения допустимые значения  $j$ -ой переменной, характеризующей  $i$ -ый объект, лежат в некоторой окрестности от наблюдаемых значений  $a_{ij}$ . Если известны значения  $j$ -ой переменной, характеризующей  $i$ -ый объект для ряда наблюдений  $a'_{ij}$ ,  $t = 1, \dots, T$ , то значения этой переменной ограничены интервалом  $[m_{ij}, M_{ij}]$ , где  $m_{ij} = \min_t a'_{ij}$ ,  $M_{ij} = \max_t a'_{ij}$ . Тогда вся совокупность пере-

менных заполняет для объекта  $i$  подмножество единичного куба  $[m_{ij}, M_{ij}]^n = [m_{i1}, M_{i1}] \times \dots \times [m_{in}, M_{in}] \subset [0, 1]^n$ . Это подмножество является множеством допустимых значений для данных задачи (1-2)  $X^* = \bigcup_{i=1}^m [m_{ij}, M_{ij}]^n$ .

Так как оператор вычисления интегрального показателя непрерывен, можно говорить, что множеством корректности  $Y^*$  при вычислении интегрального показателя является образ множества допустимых решений  $Y^* = A(X^*) = A\left(\bigcup_{i=1}^m [m_{ij}, M_{ij}]^n\right)$ . Задача вычисления инте-

грального индикатора будет корректной, если её решение устойчиво на множестве допустимых входных данных, измеренных с некоторой погрешностью.

## Применение метода главных компонент для вычисления интегрального показателя изменения качества системы для ряда наблюдений

Любое измерение, в том числе и статистическое, связано с точностью измерительного прибора, поэтому результат измерения неизбежно содержит неустраняемую ошибку. Построение интегральной характеристики системы можно рассматривать как задачу выделения полезного сигнала на фоне шумов. Получение точных характеристик объекта (и весов композитных индексов в том числе) на основании однократного измерения, неизбежно содержащего неизвестную погрешность, не представляется возможным. Однако по серии таких измерений вычисление неизвестной характеристики вполне вероятно. Такую задачу, в частности, с успехом решает астрофотометрия, определяющая основные числовые параметры астрономических объектов не по однократному наблюдению (изображению), а по **серии** зашумленных изображений.

Задача отыскания наилучшего из способов распознавания сигнала при наличии помех является одной из основных и сложных проблем в измерительной технике, радиолокации, астрономии, оптической связи, локации, навигации, телевизионной автоматике и многих других областях науки и техники. Идеальный прием сигналов в условиях воздействия шума и по-

мех основывается на простых и глубоких идеях, изложенных в наиболее последовательной и ясной форме Ф.М. Вудвордом [32]. Следуя этой идее, задачей идеального устройства, на вход которого поступает смесь сигнала с шумом, является полное разрушение ненужной информации, содержащейся в смеси, и сохранение полезной информации о тех параметрах сигнала, которые представляют интерес для пользователя системы. В любой коммуникационной системе наиболее важным параметром, характеризующим уровень помех, является отношение уровня сигнала к уровню шума (S/N). Это величина наиболее полно описывает качество воспроизведения сигнала в телевизионных системах, в системах мобильной связи, в астрофотометрии. Приемлемое качество воспроизводимого сигнала определяется задаваемым стандартом пороговым значением отношением сигнал/шум.

Используя основные идеи, лежащие в основе астрофотометрии, можно рассматривать построение интегральной характеристики изменения качества сложной системы как решение задачи выделения полезного сигнала по серии наблюдений, содержащей описание неизвестного параметра (в многомерном массиве зашумленных данных) в условиях априорной неопределенности о свойствах полезного сигнала на основании задаваемого отношения сигнал/шум. Алгоритм приводится в [33, 34]. Определяемым сигналом в этом случае являются весовые коэффициенты, определяемые по набору матриц  $A^t = \{a'_{ij}\}_{i,j=1}^{n,m}$ . Эта задача аналогична задаче восстановления цифровых изображений, искаженных гауссовским шумом.

Методика многомерного анализа, прекрасно работающая для оценки технических систем, дает неудовлетворительный результат при построении интегральной характеристики социальных систем для ряда последовательных наблюдений. В частности, вычисляемые интегральные показатели чрезвычайно неустойчивы, рейтинги, определяемые этими показателями, имеют большой разброс. Выходом из этой ситуации будет модификация метода главных компонент, учитывающая наличие ошибок в используемых данных. Метод главных компонент (МГК) позволяет выделить структуру в зашумленном массиве данных и, в частности, с успехом применяется для шумоподавления при распознавании изображений, что позволяет надеяться на успешность применения этого метода для построения интегральных характеристик плохо формализуемых систем.

Если определяется интегральная характеристика системы для ряда последовательных наблюдений, то изменение данных во времени вызывается как изменением ситуации, так и случайными ошибками. Метод главных компонент на основании различных для разных моментов значений собственных векторов и собственных значений описывает неизменную структуру системы. Следовательно, именно неискаженные значения собственных чисел и собственных векторов будут характеризовать структуру рассматриваемой системы и будут тем сигналом, который нужно по имеющимся реализациям выделить из зашумленных данных. Усреднение работает на основе предположения об абсолютно случайной природе шума и дает стабильные результаты для технических систем. Например, именно усреднение значений используется в астрофотографии для подавления шумов в изображении.

Для собственных чисел таким сигналом будет усредненное значение по всем наблюдениям. Средние значения полученных величин назовем эмпирическими собственными числами

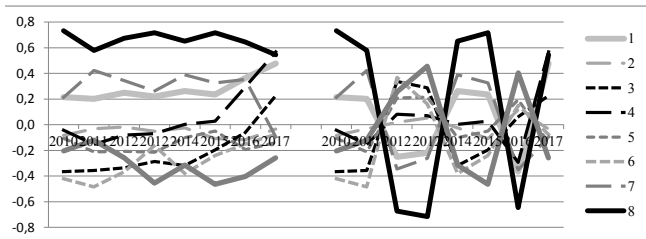


ми. Пример нахождения эмпирических собственных чисел по восьми наблюдениям приведен в таблице 1.

Т а б л и ц а 1. **Определение эмпирических собственных чисел**  
Table 1. **Determination of Empirical Eigenvalues**

Год	Собственные числа								
	1	2	3	4	5	6	7	8	9
2010	3.39	2.08	1.22	0.74	0.58	0.53	0.32	0.10	0.04
2011	3.27	2.21	1.33	0.70	0.61	0.41	0.31	0.10	0.05
2012	3.34	2.31	1.26	0.68	0.59	0.41	0.26	0.11	0.05
2012	3.31	2.33	1.28	0.68	0.58	0.43	0.24	0.11	0.04
2014	3.29	2.34	1.44	0.61	0.57	0.33	0.25	0.12	0.05
2015	3.35	2.25	1.46	0.66	0.49	0.37	0.26	0.14	0.04
2016	3.41	2.11	1.63	0.55	0.48	0.41	0.25	0.13	0.04
2017	3.27	2.47	1.52	0.59	0.48	0.29	0.21	0.13	0.03
Эмпирические собственные числа	<b>3.33</b>	<b>2.26</b>	<b>1.39</b>	<b>0.65</b>	<b>0.55</b>	<b>0.40</b>	<b>0.26</b>	<b>0.12</b>	<b>0.04</b>

Собственные векторы в МГК определяются с точностью до направления, в отличие от собственных чисел, определяемых однозначно. Среднее значение факторных нагрузок переменных зависит от выбранного направления главных компонент и не может однозначно характеризовать сигнал. На рис.1 представлен выбор направления собственных векторов для шестой главной компоненты. Изменения нагрузок переменных по годам не дает никакого представления об искомой эмпирической главной компоненте (на рисунке справа). Сплошные линии соответствуют тем переменным, нагрузки которых меняются незначительно – значение отношения сигнал/шум превышает пороговое значение, и переменная является значимой после согласования направления набора имеющихся собственных векторов (на рисунке слева).



Р и с. 1. Согласование направлений собственных векторов при определении шестой эмпирической главной компоненты. Справа – исходные направления собственных векторов, слева – выбор направления собственных векторов, обеспечивающего максимизацию уровня сигнала.

Fig. 1. Coordination of directions of eigenvectors in determining the sixth empirical principal component. On the right are the original directions of the eigenvectors, on the left is the choice of the direction of the eigenvectors, which ensures the maximum signal level.

На основе вычисленных для различных наблюдений собственных векторов (упорядоченных в порядке убывания собственных чисел), нужно распознать случайные и неслучайные компоненты этих векторов. Наличием неслучайного (значимого) вклада переменной в структуру главных компонент будем считать не большую величину факторной нагрузки после вращения, а **инвариантность** факторной нагрузки при возмущениях, признаком которой будет вычисляемое отношение сигнал/шум, определяемое отношением среднего значения фактор-

ной нагрузки (сигнал) к среднеквадратичному отклонению (шум). Если это отношение выше порогового значения, то такую переменную считаем неслучайной. Иначе переменная характеризует шумовую составляющую сигнала и не участвует в дальнейшем рассмотрении. Критерием выбора направления собственных векторов будет максимизация уровня сигнала – суммы вычисленных значений ОСШ у значимых переменных.

Т а б л и ц а 2. **Определение шестой эмпирической главной компоненты**  
Table 2. **Definition of the Sixth Empirical Principal Component**

Согласование знаков у переменных: 1, 2, 5, 6, 8, 9									
Год	Переменные								
	1	2	3	4	5	6	7	8	9
2010	0.21	-0.08	-0.37	-0.04	-0.11	-0.42	0.21	0.73	-0.20
2011	0.20	-0.03	-0.36	-0.15	-0.21	-0.48	0.42	0.58	-0.12
2012	0.25	-0.02	-0.34	-0.08	-0.21	-0.37	0.34	0.67	-0.26
2012	0.22	-0.05	-0.29	-0.07	-0.21	-0.16	0.26	0.72	-0.46
2014	0.26	-0.02	-0.32	0.00	-0.10	-0.38	0.39	0.65	-0.32
2015	0.23	-0.11	-0.20	0.03	-0.05	-0.24	0.32	0.72	-0.47
2016	0.37	-0.12	-0.06	0.29	-0.19	-0.15	0.35	0.64	-0.40
2017	0.48	-0.03	0.23	0.57	-0.14	-0.08	-0.10	0.55	-0.26
Среднее, m	<b>0.28</b>	-0.06	-0.21	0.07	-0.15	-0.29	0.28	<b>0.66</b>	<b>-0.31</b>
Среднеквадр. отклонение, s	0.10	0.04	0.20	0.24	0.06	0.15	0.16	0.07	0.12
ОСШ	<b>2.91</b>	1.50	1.04	0.28	2.43	1.93	1.68	<b>9.74</b>	<b>2.52</b>
Сумма ОСШ по строке									24.02
Сумма действующих ОСШ									15.16
Согласование знаков 3-ей и 7-ой переменной									
Год	Переменные								
	1	2	3	4	5	6	7	8	9
2010	0.21	-0.08	-0.37	-0.04	-0.11	-0.42	0.21	0.73	-0.20
2011	0.20	-0.03	-0.36	-0.15	-0.21	-0.48	0.42	0.58	-0.12
2012	0.25	-0.02	-0.34	-0.08	-0.21	-0.37	0.34	0.67	-0.26
2012	0.22	-0.05	-0.29	-0.07	-0.21	-0.16	0.26	0.72	-0.46
2014	0.26	-0.02	-0.32	0.00	-0.10	-0.38	0.39	0.65	-0.32
2015	0.23	-0.11	-0.20	0.03	-0.05	-0.24	0.32	0.72	-0.47
2016	0.37	-0.12	-0.06	0.29	-0.19	-0.15	0.35	0.64	-0.40
2017	-0.48	0.03	-0.23	-0.57	0.14	0.08	0.10	-0.55	0.26
Среднее, m	0.16	-0.05	<b>-0.27</b>	-0.07	-0.12	-0.27	<b>0.30</b>	0.52	-0.25
Среднеквадр. отклонение, s	0.26	0.05	0.10	0.24	0.12	0.18	0.11	0.43	0.24
ОСШ	0.60	1.00	<b>2.59</b>	0.31	0.97	1.45	<b>2.82</b>	1.20	1.04
Сумма ОСШ по строке									11.98
Сумма действующих ОСШ									5.41
Согласование знаков 4-ой переменной									
Год	Переменные								
	1	2	3	4	5	6	7	8	9
2010	0.21	-0.08	-0.37	-0.04	-0.11	-0.42	0.21	0.73	-0.20
2011	0.20	-0.03	-0.36	-0.15	-0.21	-0.48	0.42	0.58	-0.12
2012	0.25	-0.02	-0.34	-0.08	-0.21	-0.37	0.34	0.67	-0.26
2012	0.22	-0.05	-0.29	-0.07	-0.21	-0.16	0.26	0.72	-0.46
2014	-0.26	0.02	0.32	0.00	0.10	0.38	-0.39	-0.65	0.32
2015	-0.23	0.11	0.20	-0.03	0.05	0.24	-0.32	-0.72	0.47
2016	-0.37	0.12	0.06	-0.29	0.19	0.15	-0.35	-0.64	0.40
2017	-0.48	0.03	-0.23	-0.57	0.14	0.08	0.10	-0.55	0.26
Среднее, m	-0.06	0.01	-0.12	-0.15	-0.03	-0.07	0.03	0.02	0.05
Среднеквадр. отклонение, s	0.31	0.07	0.27	0.19	0.17	0.33	0.34	0.71	0.35
ОСШ	0.19	0.17	0.45	0.81	0.19	0.22	0.10	0.03	0.14
Сумма ОСШ по строке									2.3
Сумма действующих ОСШ									0





Полученную таким образом главную компоненту будем называть эмпирической главной компонентой (ЭГК). Значимые переменные ЭГК, как и в факторном анализе, будут участвовать в дальнейшем рассмотрении, а незначимые переменные обнуляются. В таблице 2 приведен пример определения значимых нагрузок ЭГК для набора главных компонент, представленных на рис.1. Максимальное вычисляемое отношение сигнал/шум для переменной достигается, если нагрузки этой переменной имеют один знак по рассматриваемой выборке. Значимые переменные нашлись в двух вариантах. Но в первом из вариантов сумма вычисленных значений ОСШ оказалась выше, что соответствует более высокому уровню выделяемого сигнала. В итоге первая, восьмая и девятая переменные оказались значимыми (вычисленное значение ОСШ выше порогового значения 2,2) и значение нагрузок у этих переменных будут ненулевыми нагрузками в этой ЭГК.

Качественное описание структуры системы требует либо всех главных компонент, либо их достаточно большого количества. Может оказаться, что ценная для конкретной задачи информация содержится как раз в направлениях с меньшей дисперсией. Например, при создании цифровой модели рельефа, которая строится по оцифрованным снимкам, искомым рельеф дают восьмая и девятая главные компоненты, а главные компоненты 12 и 13 в методе «Гусеница» свидетельствуют о наличии в анализируемых данных периоды с дробным периодом [35]. Подходы к оценке числа главных компонент по необходимой доле объясненной дисперсии формально применимы всегда, однако неявно они предполагают, что в данных нет разделения на «сигнал» и «шум», и любая заранее заданная точность имеет смысл. А если рассматриваются сигнал с наложенным шумом, задаваемая точность теряет смысл и требуется переопределить понятие информативности. Аналогично дисперсионной информативности согласно (3) можно определить  $SNR$ -информативность для выбранного числа эмпирических главных компонент  $N$ , определяемую задаваемым соотношением сигнал/шум (ОСШ,  $SNR$ ):

$$\gamma_{SNR} = \frac{S_{11} + S_{12} + \dots + S_{1N}}{S_{21} + S_{22} + \dots + S_{2N}}, \quad (5)$$

где  $S_{1k}$  — сумма величин ОСШ у выбранных действующих переменных  $k$ -ой ЭГК,  $S_{2k}$  — сумма ОСШ всех переменных этой ЭГК.  $SNR$ -информативность, характеризует уровень шума в сигнале (в главной компоненте), в отличие от дисперсионной информативности,  $SNR$ -информативность не может достигать 100% по логике построения. Такую информативность имеет сигнал при отсутствии шума, но в таком случае  $SNR$ -информативность не рассматривается. Полная информативность выбранной системы признаков определяется дисперсионной и  $SNR$ -информативностью:

$$\gamma = \gamma_{\sigma} \cdot \gamma_{SNR} \quad (6)$$

Число выбираемых ЭГК, участвующих в вычислении композитного индекса, должно максимизировать суммарную информативность решения, определяемую как традиционной накопленной дисперсионной информативностью, так и накопленной  $SNR$  – информативностью, характеризующей уровень сигнала ЭГК относительно уровня фона. Дисперсионная

информативность возрастает при увеличении числа используемых ЭГК. Однако  $SNR$ -информативность – уменьшается, так как младшие компоненты несут с себе больше шума.  $SNR$  – информативность ЭГК, определяемой в таблице 2, составляет  $15,16/24,02 = 0,63$ . Алгоритм, реализующий вышеизложенное, приведен в [34].

Оценим далее устойчивость интегральных индикаторов, получаемых с помощью этого алгоритма и с помощью других методов, анализируя особенности ранжирования объектов.

## Оценка устойчивости интегральных индикаторов

Основная область применения композитных индексов – ранжирование объектов. Исследователи, вычисляющие интегральные индикаторы для практического применения, часто свое исследование заканчивают на этапе вычисления и рассмотрения полученных рейтингов, не анализируя проблемы качества полученных результатов [7-13, 36, 37]. Именно положение объекта относительно других объектов является основанием и для привлечения общественного внимания, и для принятия политических решений. Если небольшие изменения входных данных при вычислении композитного индекса кардинально изменяют ранжирование объектов, произведенное на основе вычисленных интегральных показателей, то такой интегральный показатель не может считаться надежным. Необходимым признаком надежности композитного индекса является устойчивость относительно возмущений исходных данных рейтингов, определяемых интегральными характеристиками.

Зарубежные исследователи считают, что рейтинги являются функцией весов, и если существует неопределенность в отношении правильных весов, то существует и неопределенность в отношении достоверности этих рейтингов. Поэтому анализ устойчивости рангов, генерируемых композитными индексами с равными весами [38] или альтернативным весам [39].

Если же веса определены не произвольно, и нет сомнения в их достоверности, следует рассмотреть устойчивость композитных индексов относительно изменения характеристик объектов во времени. Наблюдаемые параметры изменений характеристик системы определяют множество допустимых значений переменных и можно исследовать поведение интегральных характеристик, когда изменение входных данных заведомо не покидает область допустимых значений и, следовательно, рассматриваемые решения не покидают области корректности. Следствием устойчивости интегральной характеристики, в частности, является незначительное (в среднем) изменение рейтинга объектов для разных моментов времени.

Например, ранги 25 стран членов Европейского Союза за 2009-2011 годы, выставленные на основании значений Индекса Развития Человеческого Потенциала (ИРПЧ), который вычисляет значения показателя ИРПЧ с помощью линейной свертки переменных с равными весами, дают среднее изменение рейтинга за год 7,7%<sup>11</sup>. Изменение рейтинга более 15% составляет 14% случаев, более 30% – 2%. Такое поведение рейтинга дает основание назвать устойчивой интегральный показатель ИРПЧ.

<sup>11</sup> Human Development Reports 1990-2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience. United Nations Development Programme, NY, USA, 2014. URL: <https://sustainabledevelopment.un.org/content/documents/1789hdr14-report-en-1.pdf> (дата обращения: 13.06.2018).



В исследовании, определяющем качество жизни населения Российской Федерации по методике, предложенной С.А. Айвазяном, автором метода [13, 25], рейтинги субъектам РФ выставлялись по двум методикам: в одном случае веса показателей в линейной свертке для каждого года определялись экспертами, в другом – определялись по предложенной автором методике. Случаев, когда изменение рейтинга составило более 15%: в случае экспертных весов 18, в случае вычисляемых весов – один. Среднее изменение рейтинга для первой методики 3,6%, для второй – 1,6%. Т.е. предложенная методика показала блестящее качество построенного композитного индекса.

Однако при тщательном применении этой методики другими авторами результаты оказываются совсем иными. Например, в работе [36] вычислялись интегральные характеристики качества жизни муниципальных образований Тюменской области за 2005-2008 годы. В этом случае изменение рейтинга более 15% от максимально возможного составляет 48,9% общего числа случаев. В 17,9% случаев эта величина превосходит 30%. Среднее изменение рейтинга составляет 16,9%. В работе [37] приведены значения рейтингов муниципальных образований Самарской области, вычисленные по той же методике. Одномоментное изменение рейтинга более 15% от максимально возможного составляет 45% общего числа случаев. В 21,6% случаев эта величина превосходит 30%. Среднее изменение рейтинга составляет 16,9%. В этом случае стоит говорить о неустойчивости интегрального индикатора.

Если небольшие изменения входных данных при вычислении композитного индекса кардинально изменяют ранжирование объектов, то такой интегральный показатель не может считаться надежным. Необходимым признаком надежности композитного индекса является устойчивость относительно возмущений исходных данных. В частности, следствием этого является незначительное (в среднем) изменение рейтинга объектов для разных измерений. Схемы определения весов с помощью факторного анализа или метода главных компонент таковым свойством не обладают. Определим далее меру для оценки малости изменения рейтинга объектов.

Пусть  $R_t = (r_{t1}, r_{t2}, \dots, r_{tm})$  – рейтинги  $m$  объектов для момента  $t$ . Известны значения наборов рейтингов  $R_t$  для моментов  $t = 1, \dots, T$ , которые представляют собой равномерно распределенные на отрезке  $[1, m]$  случайные величины, с числовыми характеристиками, соответствующими равномерному распределению:

$$M(R_t) = \frac{(m+1)}{2}, \quad D(R_t) = \frac{(m-1)^2}{12}, \quad t = 1, \dots, T \quad (7)$$

Значения математического ожидания и дисперсии величин  $R_t$ , определяемых (7), постоянны для всех наблюдений  $t$ . Заметим, что значения рейтингов для объекта  $i$  в последовательные моменты времени  $r_{ti}, r_{t+1,i}, r_{t+2,i}$  представляют численную реализацию сложной функциональной зависимости, отражающей свойства исследуемой системы. Найти формальное описание этой зависимости не представляется возможным. Степень линейной связи при этом между наборами рейтингов  $R_t, R_{t+1}$  высока, коэффициенты корреляции Пирсона и Спирмена близки к единице и не позволяют сделать выводов о качестве рейтингов.

Более информативны для описания качества рейтингов величины приращений рейтингов. Изменения рейтингов во времени диктуются в основном свойствами системы и – в меньшей мере – случайными факторами. Меру этой случайности можно оценить. Чем меньше доля случайности для величины разности рейтингов, тем лучше описана структура системы при выборе весовых коэффициентов и тем выше качество интегрального показателя.

Предложим метод оценки устойчивости композитного индекса на основе анализа разности рейтингов. Если рейтинги объектов для последовательных наблюдений никак не отражают свойства системы и выставляются абсолютно случайно, то случайные величины рейтингов  $R_t, R_{t+1}$  для последовательных моментов времени независимы, и дисперсия разности независимых случайных величин  $D(R_{t+1} - R_t)$  максимальна и равна сумме дисперсий величин  $R_t, R_{t+1}$ , равномерно распределенных на интервале  $[1, m]$ .

$$D_{\max} = D(R_{t+1} - R_t) = D(R_{t+1}) + D(R_t) = \frac{(m-1)^2}{6}, \quad (8)$$

где  $m$  – число объектов.

Можно оценить устойчивость интегральной характеристики, оценив хаотичность (случайность) разности выставленных рейтингов  $\Delta_t$ . Такой оценкой хаотичности является доля дисперсии реализации величины  $\Delta_t = R_{t+1} - R_t$  относительно дисперсии (8), которая достигает максимума, если рейтинги  $R_t, R_{t+1}$  для двух последовательных моментов времени абсолютно не зависят друг от друга и выставляются полностью случайно.

$$k = \frac{D(R_{t+1} - R_t)}{D_{\max}} \cdot 100\% \quad (9)$$

Поскольку случайность не является главной причиной изменения рейтинга, эта доля должна быть невелика. В таблице 3 приведены результаты оценки качества разных интегральных характеристик сравнением доли дисперсии разности рейтингов относительно максимально возможного значения такой дисперсии. В качестве условной единицы устойчивости можно рассматривать оценку интегрального показателя ИРПЧ, которая составляет около 7%. Величина оценки, сравнимая с этим значением, будет свидетельствовать о хорошей устойчивости интегрального индикатора по входным данным. Значения, которые значительно превосходят эту величину, характеризуют неустойчивость интегрального индикатора и, следовательно, его низкое качество. В приведенной таблице это исследование [36] и [37]. Интегральные индикаторы, рассчитанные по авторской методике, показывают хорошую устойчивость.

Т а б л и ц а 3. Сравнение показателей качества интегральных характеристик

Table 3. Comparison of Quality Indicators of Integrated Characteristics

Интегральный показатель	Источник	Период	Вес	Число		Оценка k
				Переменных $l$	объектов $m$	
ИРПЧ стран ЕС	Human Development Reports 1990-2014 <sup>12</sup>	2009-2011	expert	5	25	6.9
Качество жизни населения РФ	[13]	1997, 1999	expert	9	79	6.3
Качество жизни населения РФ	[13]	1997, 2000	expert	9	79	1.2

<sup>12</sup> Human Development Reports 1990-2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience. United Nations Development Programme, NY, USA, 2014. URL: <https://sustainabledevelopment.un.org/content/documents/1789hdr14-report-en-1.pdf> (дата обращения: 13.06.2018).



Качество жизни Самарской обл.	[37]	2002-2004	МГК	11	37	31.5
Качества жизни Тюменской обл.	[36]	2005-2008	МГК	17	26	22.7
Качество жизни населения РФ	[33]	2007-2014	МГК	37	83	1.7
Качество жизни населения РФ	(*)	2007-2016	МГК	37	85	1.6
Индекс устойчивости общества SSI. Благополучие человека (Human Wellbeing)	<a href="http://www.ssfindex.com/data-all-countries/">http://www.ssfindex.com/data-all-countries/</a>	2006-2016	expert	9	154	1
Индекс устойчивости общества SSI. Экологическое благополучие (Environmental Wellbeing)	<a href="http://www.ssfindex.com/data-all-countries/">http://www.ssfindex.com/data-all-countries/</a>	2006-2016	expert	8	154	3.5
Индекс устойчивости общества SSI. Экономическое благополучие (Economic Wellbeing)	<a href="http://www.ssfindex.com/data-all-countries/">http://www.ssfindex.com/data-all-countries/</a>	2006-2016	expert	5	154	4.5
Индекс экологической эффективности (Environmental Performance Index)	<a href="http://epi.yale.edu/reports/2016-report">http://epi.yale.edu/reports/2016-report</a> <a href="https://epi.envirocenter.yale.edu/downloads/epi2018policy-makerssummary01.pdf">https://epi.envirocenter.yale.edu/downloads/epi2018policy-makerssummary01.pdf</a>	2016, 2018	expert	19	176	11.5

## Заключение

Построение интегральной характеристики системы можно рассматривать как задачу выделения полезного сигнала на фоне шумов. Сигналом в этом случае являются весовые коэффициенты линейной свертки показателей. Определяемые весовые коэффициенты должны отражать структуру оцениваемой системы. Успешное применение МГК для описания структуры систем различного вида позволяет предположить, что и для описания социальных систем метод будет тоже давать адекватные результаты. Однако метод главных компонент и факторный анализ (даже при фиксированных методах извлечения факторов и способа вращения) по-разному определяют структуру главных компонент и главных факторов для разных наблюдений. Значит, методика определения весовых коэффициентов с помощью многомерного анализа не может применяться для сравнения характеристик объектов в динамике.

Причиной этого может являться наличие неустраняемых погрешностей используемых данных. Даже небольшое возмущение исходных данных может вызвать значительное изменение весовых коэффициентов при использовании методов многомерного анализа. В качестве выхода из ситуации предлагается модификация метода главных компонент, учитывающая наличие ошибок в используемых данных. Алгоритм использует новый подход к выбору числа главных компонент, к определению весов рассматриваемых подсистем и к определению информативности полученной характеристики на основании выбранного параметра отношения сигнал/шум.

Решение задачи требует детального понимания влияния погрешностей используемых данных на вычисляемые параметры модели. Поэтому существенным является вопрос о корректности задачи. Предлагается уточнение понятия корректности задачи вычисления интегральной характеристики

изменения качества системы. Корректность может быть достигнута за счет сужения рассматриваемого множества решений до множества корректности на основе априорной информации о свойствах решения. Задача вычисления интегрального индикатора будет корректной, если её решение устойчиво на множестве допустимых входных данных, определяемых по наблюдаемым измерениям. Следствием устойчивости является в среднем незначительное изменение (приращение) рейтинга объектов для разных измерений. Это приращение может быть апостериорно оценено по ряду наблюдений по предлагаемому дисперсионному критерию. Приведены результаты оценки устойчивости разных интегральных характеристик по этому критерию. Приведены оценки устойчивости разных интегральных характеристик по дисперсионному критерию. Интегральные индикаторы, рассчитанные с помощью предлагаемой модификации метода главных компонент, учитывающей наличие ошибок в используемых данных, показывают хорошую устойчивость к изменению входных данных.

## Список использованных источников

- [1] Кляцкин В. И. Интегральные характеристики – ключ к пониманию структурообразования в стохастических динамических системах // Успехи физических наук. 2011. Т. 181, № 5. С. 457-482. DOI : 10.3367/UFNr.0181.201105a.0457
- [2] Nardo M., Saisana M., Saltelli A., Tarantola S. (eds). Tools for composite indicators building. Ispra, Italy: Joint Research Centre – European Commission, 2005. 134 pp.
- [3] Foa R., Tanner J. C. Methodology of the Indices of Social Development // ISD Working Paper Series. 2012. Vol. 4. Pp. 1-66. URL: <http://hdl.handle.net/1765/50510> (дата обращения: 10.04.2019).
- [4] Saltelli A., Munda G., Nardo M. From complexity to multidimensionality: the role of composite indicators for advocacy of EU reform // Tijdschrift vor Economie en Management. 2006. Vol. L1, 3. Pp. 221-235.
- [5] Saltelli A. Composite Indicators between Analysis and Advocacy // Social Indicators Research. 2007. Vol. 81, Issue 1. Pp. 65-77. DOI: 10.1007/s11205-006-0024-9
- [6] Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publication. 2008. 162 pp.
- [7] Tarantola S., Saisana M., Saltelli A. Internal Market Index 2002: Technical details of the methodology. JRC European Commission. Institute for the Protection and Security of the Citizen Technological and Economic Risk Management Unit I-21020 Ispra (VA), Italy, 2002. 27 p. URL: [https://www.ec.europa.eu/internal\\_market/score/docs/score11/im-index-2002\\_en.pdf](https://www.ec.europa.eu/internal_market/score/docs/score11/im-index-2002_en.pdf) (дата обращения: 10.04.2019).
- [8] Hightower W. L. Development of an index of health utilizing factor analysis // Medical Care. 1978. Vol. 16, Issue 3. Pp. 245-255. DOI: 10.1097/00005650-197803000-00006
- [9] McKenzie D. J. Measuring inequality with asset indicators // Journal of Population Economics. 2005. Vol. 18, Issue 2. Pp. 229-260. DOI: 10.1007/s00148-005-0224-7
- [10] Vyas S., Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis // Health Policy and Planning. 2006. Vol. 21, Issue 6. Pp. 459-468. DOI: 10.1093/heapol/czl029
- [11] Lindman C., Sellin J. Measuring Human Development. The





- Use of Principal Component Analysis in Creating an Environmental Index. University essay from Uppsala universitet, Uppsala, 2011. 45 p. URL: <http://www.diva-portal.org/smash/get/diva2:464378/FULLTEXT03> (дата обращения: 10.04.2019).
- [12] Somarriba N., Pena B. Synthetic Indicators of Quality of Life in Europe // *Social Indicators Research*. 2009. Vol. 94, Issue 1. Pp. 115-133. DOI: 10.1007/s11205-008-9356-y
- [13] Айвазян С. А. Эмпирический анализ синтетических категорий качества жизни населения // *Экономика и математические методы*. 2003. Т. 39, № 3. С. 19-53. URL: <https://elibrary.ru/item.asp?id=17299169> (дата обращения: 10.04.2019).
- [14] Жгун Т. В. Построение интегральной характеристики изменения качества системы на основании статистических данных как решение задачи выделения сигнала в условиях априорной неопределенности // *Вестник Новгородского государственного университета им. Ярослава Мудрого*. 2014. № 81. С. 10-16. URL: <https://elibrary.ru/item.asp?id=22777430> (дата обращения: 10.04.2019).
- [15] Стрижов В. В., Шакин В. В. Согласование экспертных оценок // *Математические методы распознавания образов (ММРО-10), Доклады X всероссийской конференции. РАН, ВЦ, РФФИ, Москва, 2001. С. 137-138. URL: <http://strijov.com/papers/mmro10.pdf> (дата обращения: 10.04.2019).*
- [16] Кузнецов М. П., Стрижов В. В. Построение интегрального индикатора с использованием ранговой матрицы описаний // *Международная конференция «Интеллектуализация обработки информации» (ИОИ-9), Черногория, г. Будва, 16–22 сентября 2012. С. 130-132. URL: <http://strijov.com/papers/Kuznetsov2012IOI.pdf> (дата обращения: 10.04.2019).*
- [17] Пузаченко Ю. Г., Онуфреня И. А., Алещенко Г. М. Многомерный анализ структуры рельефа (метод главных компонент) // *Известия Российской академии наук. Серия географическая*. 2004. № 1. С. 26-36. URL: <https://elibrary.ru/item.asp?id=17636361> (дата обращения: 10.04.2019).
- [18] Oddershede J., Nielsen K., Stahl K. Using X-ray powder diffraction and principal component analysis to determine structural properties for bulk samples of multiwall carbon nanotubes // *Zeitschrift für Kristallographie - Crystal-Line Materials*. 2007. Vol. 222, Issue 3-4. Pp. 186-192. DOI: 10.1524/zkri.2007.222.3-4.186
- [19] Bosco J. P. da Silva, Ramos M. N. Principal Component Analysis of Molecular Geometries of Cis- and Trans-C<sub>2</sub>H<sub>2</sub>X<sub>2</sub> // *Journal of the Brazilian Chemical Society*. 2004. Vol. 15, No. 1. Pp. 43-49. DOI: 10.1590/S0103-50532004000100009
- [20] Li W, Zhong Y, Yu D, Qu D, Sun B, Li M, Liu J. Application of principal component analysis for identification of drugs packed in anthropomorphic phantom // *ARPN Journal of Engineering and Applied Sciences*. 2012. Vol. 7, No. 7. Pp. 915-921. URL: [https://arpnjournals.com/jeas/research\\_papers/grp\\_2012/jeas\\_0712\\_740.pdf](https://arpnjournals.com/jeas/research_papers/grp_2012/jeas_0712_740.pdf) (дата обращения: 10.04.2019).
- [21] Теслюк А. Б., Сенин Р. А., Ильин В. А. Применение метода главных компонент к анализу дифракционных изображений биомолекулярных объектов // *Математическая биология и биоинформатика*. 2013. Т. 8, № 2. С. 708-715. DOI: 10.17537/2013.8.708
- [22] Pao C. P. *Линейные статистические методы и их применения*. М., Наука, 1968. 548 с.
- [23] Isenmann A. J. *Modern Multivariate Statistical Techniques. Regression, Classification and Manifold Learning*. Springer, New York, NY, 2008. 733 p. DOI: 10.1007/978-0-387-78189-1
- [24] Айвазян С. А. Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении межрегиональных сопоставлениях. М.: ЦЭМИ РАН, 2000. 56 с. URL: <https://elibrary.ru/item.asp?id=26766108> 21104853 (дата обращения: 10.04.2019).
- [25] Айвазян С. А. К методологии измерения синтетических категорий качества жизни населения // *Экономика и математические методы*. 2003. Т. 39, № 2. С. 33-53. URL: <https://elibrary.ru/item.asp?id=17293598> (дата обращения: 10.04.2019).
- [26] Hadamard J. *Le probleme de Cauchy et les equations aux derives partielles lineaires hyperbolique*. Paris: Hermann, 1932.
- [27] Тихонов А. Н. *Методы решения некорректных задач* / А. Н. Тихонов, В. Я. Арсенин. М.: Наука, 1986. 285 с.
- [28] Тихонов А. Н. Регуляризирующие алгоритмы и априорная информация / А. Н. Тихонов, А. В. Гончарский, В. В. Степанов и др. М.: Наука, 1988. 198 с.
- [29] Saltelli A., Ratto M., Andres T., Campolongo F. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, Ltd, 2008. DOI: 10.1002/9780470725184
- [30] Huergo L., Münnich R., Saisana M. Robustness Assessment for Composite Indicators with R // *The second international conference useR! 2006, Vienna, Austria, 15-17 June 2006. URL: <https://www.r-project.org/conferences/useR-2006/Abstracts/Huergo+Munnich+Saisana.pdf> (дата обращения: 10.04.2019).*
- [31] Saltelli A., Annoni P., Azzini I., Campolongo F., Ratto M., Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index // *Computer Physics Communications*. 2010. Vol. 181, Issue 2. Pp. 259-270. DOI: 10.1016/j.cpc.2009.09.018
- [32] Вудворд Ф. М. *Теория вероятностей и теория информации с применениями в радиолокации*. М.: Издательство «Сов. радио», 1955. 128 с.
- [33] Жгун Т. В. Построение интегральной характеристики качества жизни субъектов Российской Федерации с помощью метода главных компонент // *Экономические и социальные перемены: факты, тенденции, прогноз*. 2017. Т. 10, № 2. С. 214-235. DOI: 10.15838/esc/2017.2.50.12
- [34] Жгун Т. В. Алгоритм построения интегрального индикатора качества сложной системы для ряда последовательных наблюдений // *Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика*. 2019. Т. 6, № 1. С. 5-25. DOI: 10.14529/cmse170101
- [35] Голяндина Н. Э., Усевич К. Д., Флоринский И. В. Анализ сингулярного спектра для фильтрации цифровых моделей // *Геодезия и картография*. 2008. № 5. С. 21-28. URL: <https://elibrary.ru/item.asp?id=11516739> (дата обращения: 10.04.2019).
- [36] Гайдамак И. В., Хохлов А. Г. Моделирование интеграль-





- ных показателей качества жизни населения юга Тюменской области // Вестник Тюменского государственного университета. Физико-математическое моделирование. Нефть, газ, энергетика. 2009. № 6. С. 176-186. URL: <https://elibrary.ru/item.asp?id=13287475> (дата обращения: 10.04.2019).
- [37] Айвазян С. А., Степанов В. С., Козлова М.И. Измерение синтетических категорий качества жизни населения региона и выявление ключевых направлений совершенствования социально-экономической политики (на примере Самарской области и ее муниципальных образований) // Прикладная Эконометрика. 2006. № 2(2). С. 18-84. URL: <https://elibrary.ru/item.asp?id=9482361> (дата обращения: 10.04.2019).
- [38] Foster J. E., McGillivray M., Seth S. Rank Robustness of Composite Indices: Dominance and Ambiguity // OPHI Working Papers 26b. University of Oxford, 2012. 21 pp. URL: [https://www.ophi.org.uk/wp-content/uploads/OPHI-wp26b\\_vs4.pdf](https://www.ophi.org.uk/wp-content/uploads/OPHI-wp26b_vs4.pdf) (дата обращения: 10.04.2019).
- [39] Seth S., McGillivray M. Composite Indices, Alternative Weights, and Comparison Robustness // OPHI Working Paper 106. University of Oxford, 2016. 24 pp. URL: <https://ophi.org.uk/wp-content/uploads/OPHIWP106.pdf> (дата обращения: 10.04.2019).
- Поступила 10.04.2019; принята к публикации 20.05.2019; опубликована онлайн 25.07.2019.
- Об авторах:**
- Жгун Татьяна Валентиновна**, доцент кафедры прикладной математики и информатики, Институт электронных и информационных систем, Новгородский государственный университет им. Ярослава Мудрого (173003, Россия, Новгородская обл., г. Великий Новгород, Большая Санкт-Петербургская ул., д. 41), кандидат физико-математических наук, доцент, ORCID: <http://orcid.org/0000-0002-7518-6925>, Tatyana.Zhgun@novsu.ru
- Липатов Александр Владимирович**, аспирант, Институт электронных и информационных систем, Новгородский государственный университет им. Ярослава Мудрого (173003, Россия, Новгородская обл., г. Великий Новгород, Большая Санкт-Петербургская ул., д. 41), ORCID: <http://orcid.org/0000-0001-9128-8975>, shmel9311@gmail.com
- Чалов Герман Александрович**, аспирант, Институт электронных и информационных систем, Новгородский государственный университет им. Ярослава Мудрого (173003, Россия, Новгородская обл., г. Великий Новгород, Большая Санкт-Петербургская ул., д. 41), ORCID: <http://orcid.org/0000-0003-1433-1778>, mr.germanmad@mail.ru
- Все авторы прочитали и одобрили окончательный вариант рукописи.
- References**
- [1] Klyatskin V.I. Integral characteristics: a key to understanding structure formation in stochastic dynamic systems. *Uspekhi Fizicheskikh Nauk = Physics-Uspekhi*. 2011; 54(5):441-464. (In Eng.) DOI: 10.3367/UFNe.0181.201105a.0457
- [2] Nardo M., Saisana M., Saltelli A., Tarantola S. (eds). *Tools for composite indicators building*. Ispra, Italy: Joint Research Centre – European Commission, 2005. 134 pp. (In Eng.)
- [3] Foa R., Tanner J.C. Methodology of the Indices of Social Development. *ISD Working Paper Series*. 2012; 4:1-66. Available at: <http://hdl.handle.net/1765/50510> (accessed 10.04.2019). (In Eng.)
- [4] Saltelli A., Munda G., Nardo M. From complexity to multidimensionality: the role of composite indicators for advocacy of EU reform. *Tijdschrift vor Economie en Management*. 2006; L1(3):221-235. (In Eng.)
- [5] Saltelli A. Composite Indicators between Analysis and Advocacy. *Social Indicators Research*. 2007; 81(1):65-77. (In Eng.) DOI: 10.1007/s11205-006-0024-9
- [6] Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publication. 2008. 162 pp. (In Eng.)
- [7] Tarantola S., Saisana M., Saltelli A. Internal Market Index 2002: Technical details of the methodology. JRC European Commission. Institute for the Protection and Security of the Citizen Technological and Economic Risk Management Unit I-21020 Ispra (VA), Italy, 2002. 17 pp. Available at: [https://www.ec.europa.eu/internal\\_market/score/docs/score11/im-index-2002\\_en.pdf](https://www.ec.europa.eu/internal_market/score/docs/score11/im-index-2002_en.pdf) (accessed 10.04.2019). (In Eng.)
- [8] Hightower W. L. Development of an index of health utilizing factor analysis. *Medical Care*. 1978; 16(3):245-255. (In Eng.) DOI: 10.1097/00005650-197803000-00006
- [9] McKenzie D.J. Measuring inequality with asset indicators. *Journal of Population Economics*. 2005; 18(2):229-260. (In Eng.) DOI: 10.1007/s00148-005-0224-7
- [10] Vyas S., Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. *Health Policy and Planning*. 2006; 21(6):459-468. (In Eng.) DOI: 10.1093/heapol/czl029
- [11] Lindman C., Sellin J. Measuring Human Development. The Use of Principal Component Analysis in Creating an Environmental Index. University essay from Uppsala universitet, Uppsala, 2011. 45 pp. Available at: <http://www.diva-portal.org/smash/get/diva2:464378/FULLTEXT03> (accessed 10.04.2019). (In Eng.)
- [12] Somarriba N., Pena B. Synthetic Indicators of Quality of Life in Europe. *Social Indicators Research*. 2009; 94(1):115-133. (In Eng.) DOI: 10.1007/s11205-008-9356-y
- [13] Ajvazjan S.A. Empirical Analysis of the Population's Life Quality Synthesized Categories. *Economics and Mathematical Methods*. 2003; 39(3):19-53. Available at: <https://elibrary.ru/item.asp?id=17299169> (accessed 10.04.2019). (In Russ., abstract in Eng.)
- [14] Zhgun T.V. Constructing the Integral Characteristic of the System Quality Changes on the Basis of Statistical Data as a Solution to the Problem of Signal Allocation Under Conditions of Prior Uncertainty. *Vestnik of Yaroslav the Wise Novgorod State University*. 2014; 81:10-16. Available at: <https://elibrary.ru/item.asp?id=22777430> (accessed 10.04.2019). (In Russ., abstract in Eng.)
- [15] Strizhov V.V., Shakin V.V. Coordination of expert assessments. *Proc. 10 th. Russ. Simp. Mathematical methods of pattern recognition*. Moscow, 2001, pp. 137-138. Available at: <http://strijov.com/papers/mmro10.pdf> (accessed 10.04.2019). (In Russ.)
- [16] Kuznetsov M.P., Strijov V.V. Integral indicator construction using rank-scaled design matrix. Proceedings of the Inter-



- national conference "Intelligent Information Processing" IIP-9. Montenegro, Budva, 2012; 130-132. Available at: <http://strijov.com/papers/Kuznetsov2012IOI.pdf> (accessed 10.04.2019). (In Russ.)
- [17] Puzachenko Yu.G., Onufrenia I.A., Aleshchenko G.M. Multi-dimensional Analyses of Relief Structure (Principle Components Metod. *Izvestiya RAN (Akad. Nauk SSSR). Seriya Geograficheskaya*. 2004; 1:26-36. Available at: <https://elibrary.ru/item.asp?id=17636361> (accessed 10.04.2019). (In Russ.)
- [18] Oddershede J., Nielsen K., Stahl K. Using X-ray powder diffraction and principal component analysis to determine structural properties for bulk samples of multiwall carbon nanotubes. *Zeitschrift für Kristallographie - Crystalline Materials*. 2007; 222(3-4):186-192. (In Eng.) DOI: 10.1524/zkri.2007.222.3-4.186
- [19] Bosco J.P. da Silva, Ramos M.N. Principal Component Analysis of Molecular Geometries of Cis- and Trans-C<sub>2</sub>H<sub>2</sub>X<sub>2</sub>. *Journal of the Brazilian Chemical Society*. 2004; 15(1):43-49. (In Eng.) DOI: 10.1590/S0103-50532004000100009
- [20] Li W., Zhong Y., Yu D., Qu D., Sun B., Li M., Liu J. Application of principal component analysis for identification of drugs packed in anthropomorphic phantom. *ARPN Journal of Engineering and Applied Sciences*. 2012; 7(7):915-921. Available at: [https://arpnjournals.com/jeas/research\\_papers/rp\\_2012/jeas\\_0712\\_740.pdf](https://arpnjournals.com/jeas/research_papers/rp_2012/jeas_0712_740.pdf) (accessed 10.04.2019). (In Eng.)
- [21] Teslyuk A.B., Senin R.A., Ilyin V.A. Applying Principal Component Analysis for Macromolecular Objects Diffraction Images Sorting. *Mathematical Biology and Bioinformatics*. 2013; 8(2):708-715. (In Russ., abstract in Eng.) DOI: 10.17537/2013.8.708
- [22] Rao C.R. Linear Statistical Inference and its Applications. Second Edition. New York, John Wiley & Sons, 1973. 625 pp. (In Eng.)
- [23] Izenmann A.J. Modern Multivariate Statistical Techniques. Regression, Classification and Manifold Learning. Springer, New York, NY, 2008. 733 pp. (In Eng.) DOI: 10.1007/978-0-387-78189-1
- [24] Ajvazjan S.A. *Integral'nye indikatory kachestva zhizni nasele-niya: ih postroenie i ispol'zovanie v social'no-jekonomicheskom upravlenii I mezhregional'nyh sopostavleniyah* [Integrated indicators of the quality of life: their construction and use in social and economic management and interregional comparisons]. Moscow: CJEI RAN, 2000. 56 p. (In Russ.)
- [25] Ajvazjan S.A. Towards a Methodology of Measuring of the Population's Life Quality Synthesized Categories. *Economics and Mathematical Methods*. 2003; 39(2):33-53. Available at: <https://elibrary.ru/item.asp?id=17293598> (accessed 10.04.2019). (In Russ., abstract in Eng.)
- [26] Hadamard J. Le probleme de Cauchy et les equations aux derives particlee lineaires hyperbolique. Paris: Hermann, 1932. (In Eng.)
- [27] Tikhonov A.N., Arsenin V.Ya. *Methods for Solving Ill-Posed Problems*. Nauka, Moscow, 1986. 285 pp. (In Russ.)
- [28] Tikhonov A.N., Goncharsky A.V., Stepanov V.V., Yagola A.G. *Regularizing Algorithms and A Priori Information*. Nauka, Moscow, 1983. 198 pp. (In Russ.)
- [29] Saltelli A., Ratto M., Andres T., Campolongo F. Global Sensitivity Analysis: The Primer. John Wiley & Sons, Ltd, 2008. (In Eng.) DOI: 10.1002/9780470725184
- [30] Huergo L., Münnich R., Saisana M. Robustness Assessment for Composite Indicators with R. *The second international conference useR!* Vienna, Austria, 2006. Available at: <https://www.r-project.org/conferences/useR-2006/Abstracts/Huergo+Munnich+Saisana.pdf> (accessed 10.04.2019). (In Eng.)
- [31] Saltelli A., Annoni P., Azzini I., Campolongo F., Ratto M., Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*. 2010; 181(2):259-270. (In Eng.) DOI: 10.1016/j.cpc.2009.09.018
- [32] Woodward P.M. Probability and Information Theory with Applications to Radar. London, Pergamon Press, 1955. 128 pp. (In Eng.)
- [33] Zhgun T.V. Building an Integral Measure of the Quality of Life of Constituent Entities of the Russian Federation Using the Principal Component Analysis. *Economic and Social Changes: Facts, Trends, Forecast*. 2017; 10(2):214-235. (In Russ., abstract in Eng.) DOI: 10.15838/esc/2017.2.50.12
- [34] Zhgun T.V. An algorithm for constructing integral quality indicator of complex systems for a sequence of observations. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*. 2017; 6(1):5-25. (In Russ.) DOI: 10.14529/cmse170101
- [35] Golyandina N.E., Usevich K.D., Florinsky I. V. Singular spectrum analysis for filtering of digital terrain models. *Geodesy and Cartography*. 2008; 5:21-28. Available at: <https://elibrary.ru/item.asp?id=11516739> (accessed 10.04.2019). (In Russ., abstract in Eng.)
- [36] Gaydamak I.V., Khokhlov A.G. Modelling of Integral Indexes of Life Quality for the South of Tyumen Region. *Tyumen State University Herald. Physical and Mathematical Modeling. Oil, Gas, Energy*. 2009; 6:176-186. Available at: <https://elibrary.ru/item.asp?id=13287475> (accessed 10.04.2019). (In Russ., abstract in Eng.)
- [37] Ajvazjan S.A., Stepanov V.S., Kozlova M.I. *Izmerenie sinteticheskikh kategorij kachestva zhizni naselenija regiona i vyjavlenie kljuchevykh napravlenij sovershenstvovaniya social'no-jekonomicheskoy politiki (na primere Samarskoj oblasti i ee municipal'nyh obrazovanij)* [Measurement of synthetic categories of quality of life of the population of the region and identify key areas of improvement in socio-economic policy (in the Samara region and its municipalities an example)]. *Prikladnaja Ekonometrika = Applied Econometrics*. 2006; 2(2):18-84. Available at: <https://elibrary.ru/item.asp?id=9482361> (accessed 10.04.2019). (In Russ., abstract in Eng.)
- [38] Foster J.E., McGillivray M., Seth S. Rank Robustness of Composite Indices: Dominance and Ambiguity. *OPHI Working Papers* 26b. University of Oxford, 2012. 21 pp. Available at: [https://www.ophi.org.uk/wp-content/uploads/OPHI-wp26b\\_vs4.pdf](https://www.ophi.org.uk/wp-content/uploads/OPHI-wp26b_vs4.pdf) (accessed 10.04.2019). (In Eng.)
- [39] Seth S., McGillivray M. Composite Indices, Alternative Weights, and Comparison Robustness. *OPHI Working Paper 106*. University of Oxford, 2016. 24 pp. Available at: <https://ophi.org.uk/wp-content/uploads/OPHIWP106.pdf> (accessed 10.04.2019). (In Eng.)

Submitted 10.04.2019; revised 20.05.2019;  
published online 25.07.2019.



**About the authors:**

**Tatyana V. Zhgun**, Associate Professor of the Department of Applied Mathematics and Computer Science, Institute of Electronic and Information Systems, Yaroslav-the-Wise Novgorod State University (41 Bolshaya St. Petersburg St., Velikiy Novgorod 173003, Novgorod region, Russia), Ph.D. (Phys.-Math.), Associate Professor, ORCID: <http://orcid.org/0000-0002-7518-6925>, [Tatyana.Zhgun@novsu.ru](mailto:Tatyana.Zhgun@novsu.ru)

**Aleksandr V. Lipatov**, postgraduate student, Institute of Electronic and Information Systems, Yaroslav-the-Wise Novgorod State University (41 Bolshaya St. Petersburg St., Velikiy Novgorod 173003, Novgorod region, Russia), ORCID: <http://orcid.org/0000-0001-9128-8975>, [shmel9311@gmail.com](mailto:shmel9311@gmail.com)

**German A. Chalov**, postgraduate student, Institute of Electronic and Information Systems, Yaroslav-the-Wise Novgorod State University (41 Bolshaya St. Petersburg St., Velikiy Novgorod 173003, Novgorod region, Russia), ORCID: <http://orcid.org/0000-0003-1433-1778>, [mr.germanmad@mail.ru](mailto:mr.germanmad@mail.ru)

*All authors have read and approved the final manuscript.*

