

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

УДК 519.2

В. Е. ГВОЗДЕВ, Д. В. СЕМЕНЕНКО

**ИНФОРМАЦИОННАЯ ПОДДЕРЖКА
АНАЛИЗА СОСТОЯНИЯ ТЕРРИТОРИАЛЬНЫХ СИСТЕМ
ПО РАЗНОТИПНЫМ ПРИЗНАКАМ**

Настоящая статья посвящена вопросам комплексного оценивания состояния территориальных систем на основе статистических индексов. Расчет значений индексов осуществляется на основе типологической классификации территорий по разнотипным признакам состояния. Описана процедура нахождения границ пересекающихся классов на основе анализа законов распределения порядковых статистик. *Территориальная система ; статистические индексы ; классификационная шкала ; закон распределения непрерывной случайной величины ; порядковая статистика*

ВВЕДЕНИЕ

Одной из задач, связанных с анализом состояния территориальных систем (ТС), является обоснованный выбор методов и процедур оценивания текущего состояния ТС и территориально-временной изменчивости таких систем в контексте рассматриваемых признаков состояния. Важной характеристикой методов анализа является формализация этих подходов. Анализ текущего и будущего состояния ТС является неотъемлемой частью информационного обеспечения управления и контроля состояния территориальных систем.

Важнейшим принципом управления состоянием территориальных систем является принцип комплексности. Согласно ему, управление ТС должно осуществляться с учетом единства природной, техногенной, социальной, экономической компонент.

Анализируя изложенные в литературе подходы к комплексному исследованию ТС, можно выделить два направления исследований:

1) использование методов многомерного статистического анализа. В рамках этого подхода наибольшее распространение получили фак-

торные модели (компонентный и факторный анализ), таксономические модели, методы кластерного анализа и др.;

2) разработка интегральных показателей состояния окружающей среды. В рамках этого подхода разработан набор интегральных показателей, характеризующих состояние природных компонентов по содержанию в них различных химических соединений, демографическим показателям, экономическим показателям и другие.

**ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ
ИНДЕКСОВ ДЛЯ ОЦЕНКИ СОСТОЯНИЯ
ТЕРРИТОРИАЛЬНЫХ СИСТЕМ**

В настоящей работе для комплексной оценки состояния ТС в качестве подхода предлагается использование описанных в литературе статистических индексов. Обоснованием целесообразности использования индексов является следующее:

- возможность количественно охарактеризовать изменение состояния ТС;
- возможность выбирать базовый период (месяц, год), по сравнению с которым оценивается изменение рассматриваемого показателя состояния ТС;
- обеспечивать сопоставимость данных при изменении административного устройства изучаемой территории или изменении состава рассматриваемых признаков.

Анализ состояния ТС должен содержать в себе ответы на два основных вопроса: стабильно ли состояние рассматриваемой ТС и каким образом изменяется состояние ТС. Для ответа

Контактная информация: (347)272-82-80

Результаты научных исследований поддержаны грантом РФФИ 08-07-97011 «Фундаментальные основы краткосрочного прогнозирования паводковой ситуации на территории Республики Башкортостан по малому числу статистических данных».

на эти вопросы в данной работе предлагается использование двух видов индексов.

В основе конструирования индексов, характеризующих стабильность состояния ТС, лежат результаты типологической классификации ТС. Расчет индексов по результатам классификации производится на основе соотношения:

$$Q_k^{(g)} = \sum_{i=1}^{n_k} \frac{S_i^{(g,k)}}{S} \ln \frac{S_i^{(g,k)}}{S}, \quad (1)$$

где $S_i^{(g,k)}$ – суммарная площадь территорий, отнесенных к i -му классу состояния в g -м временном срезе; S – общая площадь исследуемой территории; n_k – число классов состояния k -го признака.

По форме записи данное выражение близко к энтропии Шеннона и может выступать как интегрированная характеристика различий состояний территорий.

Охарактеризовать стабильность ТС по результатам расчета такого типа индексов можно следующим образом: чем ближе к нулевому значению индекс, тем стабильнее система. Это объясняется тем, что если все объекты ТС попадают в один класс состояния, то выражение $\frac{S_i^{(g,k)}}{S}$ принимает единичное значение, а выра-

жение $\ln \frac{S_i^{(g,k)}}{S}$ принимает нулевое значение,

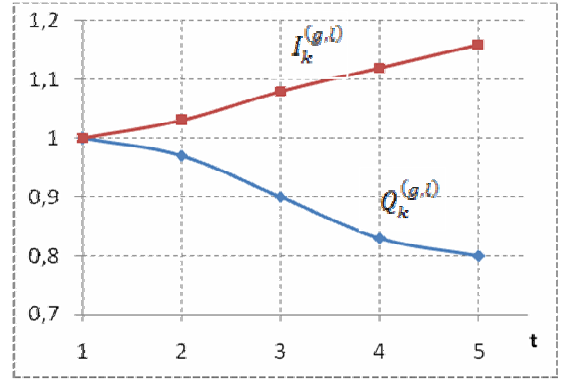
что характеризует минимальную степень различия объектов территориальной системы в контексте рассматриваемого признака и, следовательно, стабильность системы.

Для оценки направления изменения состояния ТС предлагается использовать индекс, характеризующий среднее значение рассматриваемого признака:

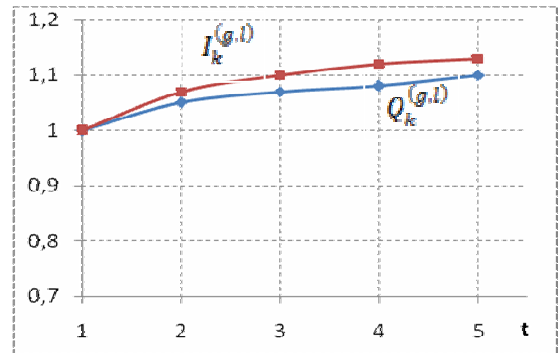
$$I_k^{(g)} = \frac{\sum_{i=1}^N \frac{\Pi_i^{(g,k)}}{S_i^{(g,k)}}}{N}, \quad (2)$$

где $S_i^{(g,k)}$ – площадь i -го участка территории, входящего в ТС в g -м временном срезе; $\Pi_i^{(g,k)}$ – значение рассматриваемого k -го признака i -го участка территории в g -м временном срезе; N – общее число участков территорий, входящих в ТС.

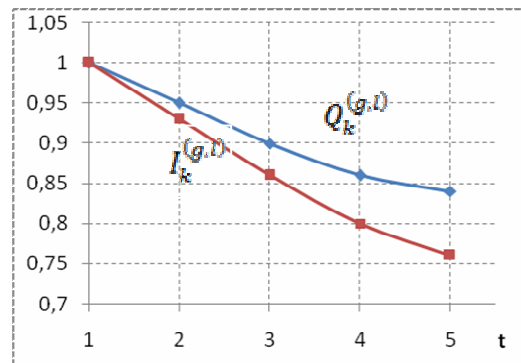
Одновременное использование двух индексов позволяет охарактеризовать стабильность изменчивости ТС и определить направление изменения состояния (улучшение или ухудшение).



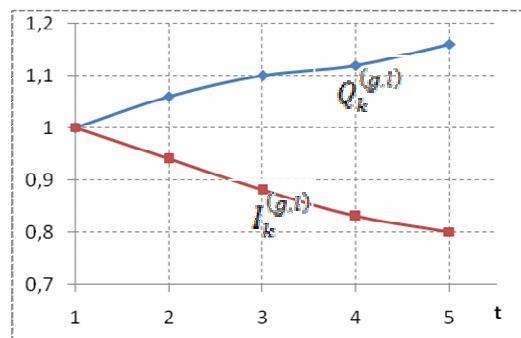
а



б



в



г

Рис. 1. Варианты изменения состояния ТС для признака, характеризующего негативное влияние:

- а – стабильно ухудшающееся состояние;
- б – нестабильно ухудшающееся состояние;
- в – стабильно улучшающееся состояние;
- г – нестабильно улучшающееся состояние

На рис. 1 приведены различные варианты изменения состояния ТС для признака, характеризующего негативное воздействие на ТС.

Для получения зависимостей, приведенных на рисунке, использовались следующие соотношения, основанные на (1) и (2):

$$Q_k^{(g,l)} = \frac{Q_k^{(g)}}{Q_k^{(l)}}; \quad I_k^{(g,l)} = \frac{I_k^{(g)}}{I_k^{(l)}} \quad (3)$$

В случае, если признак отражает положительное воздействие на состояние ТС, изменение значений $I_k^{(g,l)}$ на данном рисунке будет трактоваться противоположно приведенному, трактовка изменения значений $Q_k^{(g,l)}$ останется прежней.

ТИПОЛОГИЧЕСКАЯ КЛАССИФИКАЦИЯ ТЕРРИТОРИЙ КАК ОСНОВА ПОСТРОЕНИЯ СТАТИСТИЧЕСКИХ ИНДЕКСОВ

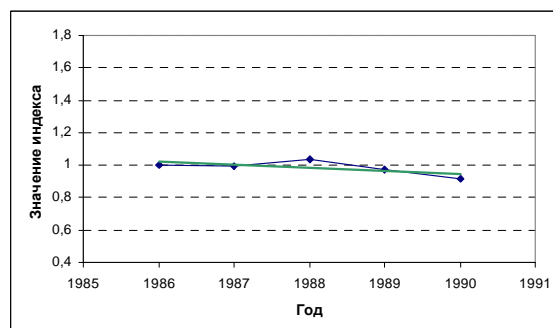
В настоящей статье предлагается подход к построению статистических индексов, основанный на приемах типологической классификации территорий, позволяющий учесть, что состав территориальных единиц, входящих в ТС, может меняться.

Целью классификации является разработка процедуры отнесения участков ТС к тому или иному классу состояния в контексте анализируемых характеристик. В силу того, что территориальный объект является сложной системой, для оценки его состояния по какому-либо признаку, в зависимости от целей исследования, могут быть построены различные классификационные шкалы.

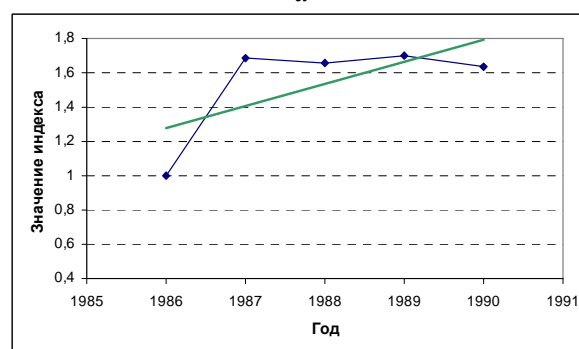
Следует подчеркнуть, что получаемые значения индексов стабильности зависят от способа построения классификационной шкалы. На рис. 2 представлены результаты, полученные с использованием различных подходов к построению классификационных шкал, при обработке одних и тех же данных о заболеваемости ревматизмом на территории Республики Башкортостан в период с 1986 по 1991 гг. В первом случае (рис. 2, а) был применен метод естественного разбиения на классы, встроенный в пакет ArcGIS 9.X; во втором случае (рис. 2, б) – метод стандартного отклонения, также встроенный в пакет ArcGIS 9.X.

Из рисунка видно, что на основе одних и тех же данных, в зависимости от выбора классификационной шкалы, можно получить совершенно разные оценки стабильности состояния ТС по рассматриваемому признаку. Таким образом,

задача построения классификационных шкал является важной для комплексной оценки состояния ТС и может быть выделена в отдельную задачу.



а



б

Рис. 2. Влияние подходов к формированию классификационных шкал на оценку состояния ТС: а – классификация с использованием метода естественного разбиения; б – классификация с использованием метода стандартного отклонения

В литературе [2, 3] описаны подходы к построению классификационных шкал с непересекающимися классами с учетом статистических особенностей исходных данных в случае использования одномерного характеристического признака. Недостатком этих подходов является то, что возможна ситуация, когда объекты с весьма незначительными отличиями характеристического признака относятся к разным классам. Использование пересекающихся классов позволяет получить области пересечения классов, в которые попадают объекты, не обладающие явно выраженной принадлежностью к тому или иному классу состояния. Кроме того, преимущество данного метода заключается в том, что число классов задается заранее и выбирается исходя из целей классификации и удобства трактовки получаемого результата. При заданном числе классов состояния n количество интервалов группирования составит $b = 2n - 1$.

Сравнительный анализ результатов классификации объектов различными методами [1]

позволяет заключить, что классификация на пересекающиеся классы является более объективной с позиции оценки вероятности попадания объектов разных классов в один и тот же интервал группирования. В упомянутой работе описана процедура построения классификационных шкал с пересекающимися классами на основе анализа законов распределения порядковых статистик (законов распределения случайных величин внутри классов). В основе этой процедуры лежит известное соотношение, устанавливающее связь законов распределения случайных величин $F(x)$ и плотностей распределения порядковых статистик [4]:

$$\varphi_n(x_m) = n \cdot C_{n-1}^{m-1} [F(x)]^{m-1} [1 - F(x)]^{n-m} f(x), \quad (4)$$

где $f(x)$ – плотность распределения (дифференциальная функция распределения) случайной величины, соответствующая $F(x)$; n – число классов группирования; m – рассматриваемый класс ($m = \overline{1, n}$).

Аналитически функцию плотности распределения порядковой статистики можно построить только для ограниченного набора $F(x)$, однако использование численных методов снимает это ограничение.

Опишем процедуру определения интервалов группирования на примере двух классов:

Шаг 1. Выбор критерия классификации. В качестве критерия могут быть предложены оценка эффективности использования информации, содержащейся в исходных данных, или оценка близости начальных моментов, но, учитывая, что $F(x)$ – исчерпывающая характеристика случайной величины, в качестве критерия качества классификации целесообразно предложить следующий:

$$D = \max_{\beta} |\widehat{F}(x) - F(x)| \rightarrow \min, \quad (5)$$

где $F(x)$ – теоретический закон распределения случайной величины; $\widehat{F}(x)$ – оценка закона распределения случайной величины, представленная в виде ступенчатой функции, соответствующая выбранному способу назначения интервалов группирования; β – заранее выбранное значение вероятности, с которым случайная величина принадлежит тому или иному классу состояния

Шаг 2. Определение границ классов. Для интегральных функций распределения $\Psi_2(x_1)$, $\Psi_2(x_2)$ случайных величин внутри первого и второго классов задается значение $x_1^{гп}$, $X \in 0 \div 8\sigma$, такое, которому соответствует выбранное значение β , т. е. вероятность того, что

первая порядковая статистика лежит в интервале от 0 до заданного $x_1^{гп}$. Далее задается значение $x_2^{гп}$, такое, чтобы вторая порядковая статистика лежала в интервале от $x_2^{гп}$ до 8σ с той же вероятностью β :

$$\begin{aligned} P(0 < x_1 < x_1^{гп}) &= \beta, \\ P(x_2^{гп} < x_2 < 8\sigma) &= \beta. \end{aligned}$$

После определения $x_1^{гп}$ и $x_2^{гп}$ определяются границы классов: от 0 до $x_1^{гп}$ – границы первого класса, от $x_2^{гп}$ до 8σ – границы второго класса. Таким образом, для двух классов получено три интервала группирования. В области пересечения классов могут находиться объекты как первого, так и второго класса. На рис. 3 приведены иллюстрации формирования границ классов для различных интегральных и дифференциальных законов распределения случайной величины.

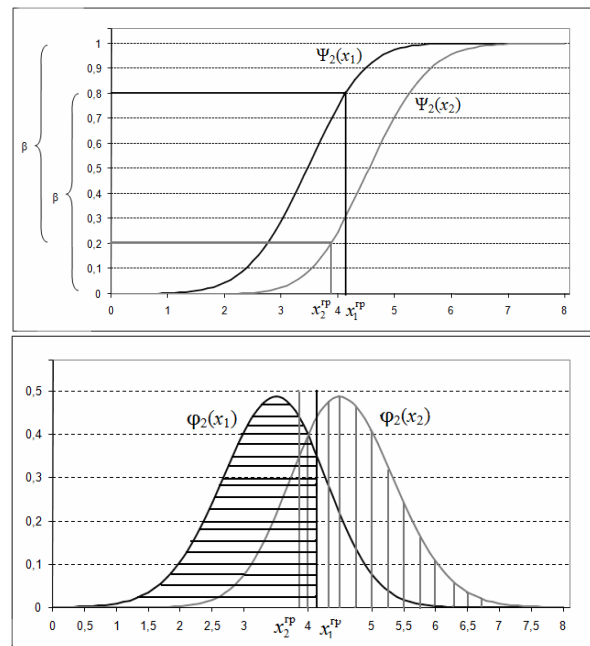


Рис. 3. Формирование $x_2^{гп}$ и $x_1^{гп}$

Шаг 3. Для полученных интервалов группирования строится ступенчатая функция и оценивается ее соответствие исходному распределению по критерию (5).

Шаг 4. Целенаправленно изменяется значение β до тех пор, пока разница между исходной $F(x)$ и аппроксимирующей ступенчатой $\widehat{F}(x)$ функциями не будет минимальна.

Процедура определения границ классов по описанной выше методике не вызывает трудностей, если заранее известны интегральный и дифференциальный законы распределения вы-

борочных данных. На практике же возникает задача оценивания закона распределения непрерывной случайной величины на основе выборочных данных. Таким образом, разработка методов построения классификационных шкал при заранее неизвестном законе распределения случайной величины является актуальной и может быть выделена в качестве самостоятельной задачи исследования.

В ранее проводившихся исследованиях [3] установлено, что при известной границе диапазона возможных значений случайной величины для оценивания закона распределения случайной величины достаточно учитывать математическое ожидание $M[x]$ и среднее квадратическое отклонение σ . При этом характеристикой формы закона распределения является значение $h=M[x]/\sigma$, а характеристикой масштаба – значение σ .

В данной работе для построения оценки закона распределения случайной величины по выборочным данным предлагается унифицированная параметрическая модель, представляющая собой таблицу, включающую известные законы распределения, у которых $0 < x < \infty$. На рис. 4 приведен фрагмент таблицы эталонных параметрических законов распределения случайных величин, содержащий типы и параметры (формы α , масштаба β , $\lambda=1/\beta$) законов распределения случайных величин, рассчитанных для единичного σ (нормированные законы распределения). Записи упорядочены по первичному ключу $h=M[x]/\sigma$.

$h=M[x]/\sigma$	Тип закона распределения	Параметры распределения	
0,8	Гамма	$\alpha=0,64$	$\beta=1,25$
0,9	Гамма	$\alpha=0,81$	$\beta=1,1111$
1	Экспоненциальный	$\lambda=1$	
1,0986	Вейбулла	$\alpha=1,1$	$\beta=1,1386$
1,1949	Вейбулла	$\alpha=1,2$	$\beta=1,2703$

Рис 4. Фрагмент унифицированной параметрической модели

На основе данной таблицы по схеме, представленной на рис. 5, строится закон распределения случайной величины по выборочным данным.

В ходе исследования предлагаемой унифицированной параметрической модели было установлено, что характеристика формы закона распределения $h=M[x]/\sigma$ и характеристика масштаба порядковых статистик функционально зависят от характеристик формы и масштаба исходного закона распределения. Примеры зависимостей для двух порядковых статистик приведены на рис. 6, 7. На рисунках обозначены: n – общее число классов (порядковых статистик); m – номер класса; ($m = \overline{1, n}$); $h_m^{(n)}$ – значение формы закона распределения m -й порядковой статистики; $\sigma_m^{(n)}$ – среднее квадратическое отклонение m -й порядковой статистики.

Полученные результаты позволяют разработать достаточно простую процедуру оценки изменчивости ТС (рис. 8).

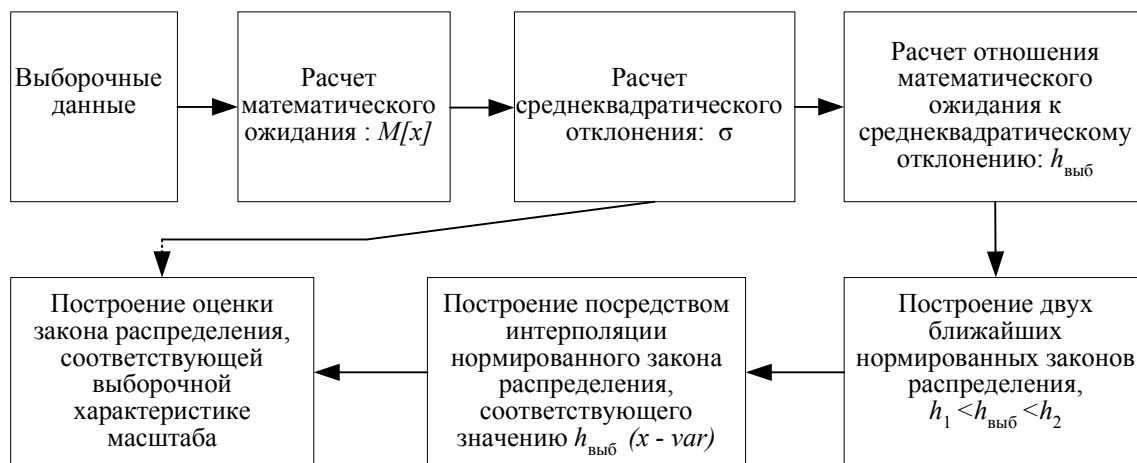


Рис. 5. Схема построения закона распределения случайной величины

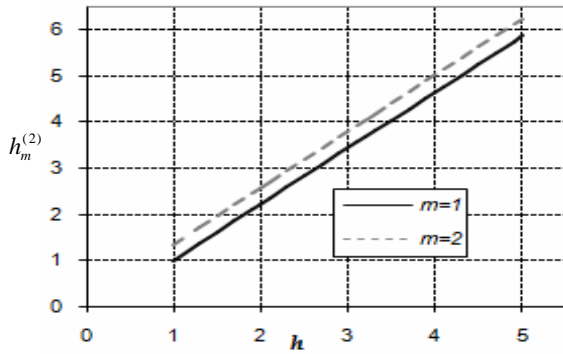


Рис. 6. Зависимость характеристик формы порядковых статистик от характеристик формы h исходного закона распределения

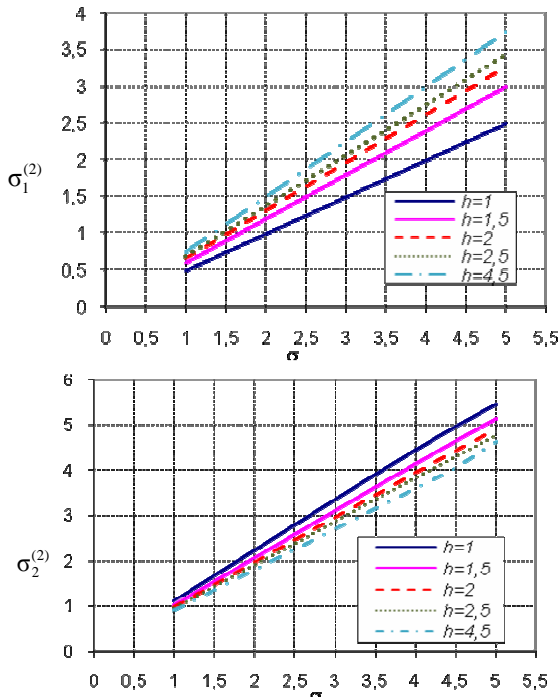


Рис. 7. Зависимость характеристик масштаба порядковых статистик от характеристик масштаба исходного закона распределения для различных форм исходного закона распределения

Как видно из схемы, такая процедура позволяет избежать построения законов распределения непрерывной случайной величины на основе выборочных данных для оценки законов распределения порядковых статистик по формуле (4) и дает возможность непосредственно перейти к построению законов распределения порядковых статистик с определенными по графикам параметрами на основе унифицированного параметрического метода, что позволяет уменьшить погрешность расчетов при восстановлении законов распределения порядковых статистик и увеличить точность определения границ классов.

Для данной выборки были рассчитаны значения математического ожидания и среднеквадратического отклонения: $M[x] = 12,92$; $\sigma = 4,58$; $M[x]/\sigma = 2,82$

Для восстановления закона распределения случайной величины использовалась унифицированная параметрическая модель, которая позволила определить, что искомым законом распределения лежит между следующими:

$M[x]/\sigma$	Тип закона распределения	Параметры распределения	
		α	β
2,751436	Вейбулла	$\alpha=3$	$\beta=3,081186$
2,828427	Гамма	$\alpha=8$	$\beta=0,353553$

На шаге определения границ классов на основе соотношения (5) были получены следующие зависимости критерия качества классификации D от вероятности принадлежности случайной величины β к выбранному классу состояния (рис. 10).

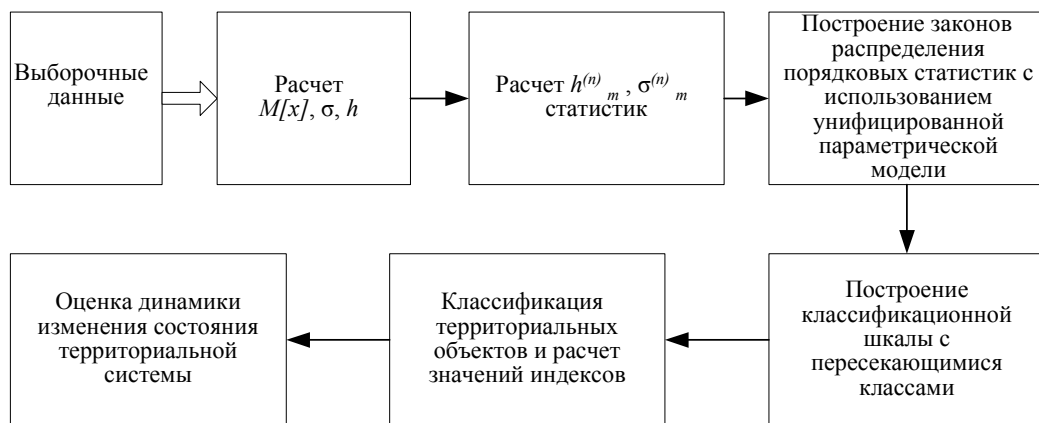


Рис. 8. Процедура типологической классификации на основе унифицированной параметрической модели

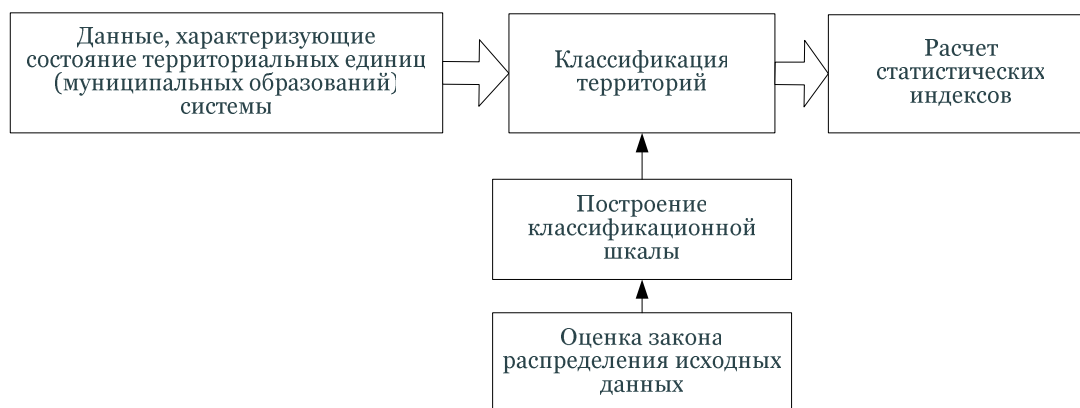


Рис. 9. Схема решения задачи

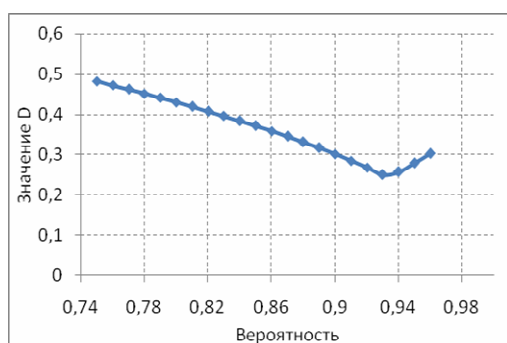


Рис. 10. Выбор оптимального значения β

Для оптимального β были определены границы интервалов группирования:
 интервал 1: 0 – 10,02
 интервал 2: 10,02 – 15,92
 интервал 3: 15,92 – 35

На рис. 11 приведены дифференциальные функции распределения порядковых статистик и отмечены границы интервалов группирования. На рис. 12 приведены графики изменения значений индексов.

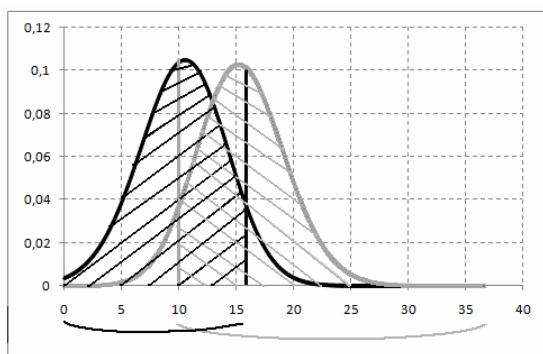


Рис. 11. Границы интервалов группирования

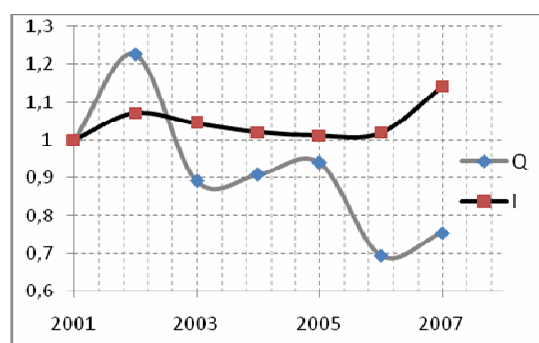


Рис. 12. Изменения значений индексов по отношению к индексу 2001 г.

На основе полученных результатов делаются выводы о стабильности состояния ТС и направлении изменения состояния. Следует учесть, что анализ тенденций изменения состояния ТС по малому числу данных – предмет отдельных исследований.

ЗАКЛЮЧЕНИЕ

Предлагаемый подход к оценке состояния ТС на основе статистических индексов позволяет оценить стабильность изменения состояния ТС и дает возможность охарактеризовать направление изменения состояния рассматриваемой системы. Данный подход основан на решении задачи классификации и отличается от известных использованием унифицированной параметрической модели для оценивания законов распределения случайных величин и использованием законов распределения порядковых статистик для нахождения границ классов.

СПИСОК ЛИТЕРАТУРЫ

1. **Gvozdev, V. E.** Definition of the casual value grouping intervals by the account of serial statistics properties / V. E. Gvozdev, O. Y. Bejaeva, D. Y. Semenenko // Proc. of 9th International Workshop on Computer Science and Information Technologies (CSIT'2007), Vol. 2. USATU, Ufa, Russia, 2007. P.98–102.

2. **Гвоздев, В. Е.** Зонирование территории Республики Башкортостан по степени опасности для населения/ В. Е. Гвоздев, Е. В. Заяц, С. В. Павлов и др // Башкирский экологический вестник : науч. журнал. 2000. № 1(8).

3. **Гвоздев, В. Е.** Непараметрическое оценивание функциональных зависимостей по эмпирическим данным / В. Е. Гвоздев, А. Е. Колоденкова // Мехатроника, автоматизация, управление : теор. и прикл. науч.-техн. журнал. 2005. № 8. С. 12–18.

4. **Ефимов, А. Н.** Порядковые статистики – их свойства и приложения / А. Н. Ефимов. М. : Знание, 1980. 62 с.

ОБ АВТОРАХ

Гвоздев Владимир Ефимович, зав. каф. автоматиз. проектир. информ. систем. Дипл. инж. электрон. тех-ки (УАИ, 1978). Д-р техн. наук по АСУ (УГАТУ, 2000). Иссл. в обл. АСУ, открытых информ. систем, прикл. статистики, теории надежности, контроля и управ. сост. окр. среды, управл. прогн. проектами.



Семеновна Дарья Викторовна, асп. той же кафедры. Дипл. инж. по инф. системам и технологиям (УГАТУ, 2006). Готовит дис. в области анализа и управл. террит. системами.