

УДК 005:332.1:631.145

UDC 005:332.1:631.145

**КЛАСТЕРНЫЙ АНАЛИЗ СОСТОЯНИЯ
ОКРУЖАЮЩЕЙ СРЕДЫ СУБЪЕКТОВ
ФЕДЕРАЦИИ ЮГА РОССИИ**

**CLUSTER ANALYSIS OF ENVIRONMENTAL
SITUATION OF SUBJECTS OF SOUTHERN
RUSSIA**

Сахнюк Татьяна Ивановна
к.э.н., доцент

Sakhnyuk Tatiana Ivanovna
Cand.Econ.Sci., associate professor

Сахнюк Павел Анатольевич
к.т.н., доцент
*Ставропольский государственный аграрный
университет, Ставрополь, Россия*

Sakhnyuk Pavel Anatolevich
Cand.Tech.Sci., associate professor
Stavropol State Agrarian University Stavropol, Russia

Проведен кластерный анализ состояния окружающей среды субъектов регионов юга России. Предлагаются меры по улучшению состояния окружающей среды для регионов каждого кластера

In the article we have performed a cluster analysis of the environmental situation in the regions of southern Russia. The article proposes the measures to improve the environment for the regions of each cluster

Ключевые слова: КЛАСТЕРНЫЙ АНАЛИЗ,
СОСТОЯНИЕ ОКРУЖАЮЩЕЙ СРЕДЫ

Keywords: CLUSTER ANALYSIS, ENVIRONMENT

Целью работы является разбиение территорий субъектов Южного и Северокавказского федеральных округов на кластеры по комплексу существенных показателей, влияющих на состояние окружающей среды и отнесение их к наиболее, наименее благополучных и территориям, занимающим срединные позиции.

Общая картина состояния окружающей среды может быть составлена на основании анализа всех основных характеризующих ее показателей, а также анализа наличия и использования важнейших природных ресурсов. При проведении анализа были использованы последние данные статистического сборника федеральной службы государственной статистики за 2010 год [1], личные наблюдения авторов [4]. Полученная объективная информация может быть использована для выработки стратегии и принятия управленческих решений относительно охраны окружающей среды субъектов территорий федеральных округов юга России.

Для проведения анализа применяется рабочее место аналитика *Deductor Studio*, которое входит в состав аналитической платформы *Deductor* [2]. Данное приложение содержит набор механизмов импорта,

обработки, визуализации и экспорта данных для быстрого и эффективного анализа информации. Будем использовать инструменты проведения кластерного анализа, имеющиеся в *Deductor Studio*, а именно алгоритмы *k-means* и самоорганизующаяся карта Кохонена.

Анализируемые показатели представим относительными величинами интенсивности в зависимости либо от площади территории, либо от числа городского населения исследуемых субъектов РФ. Такой перевод является важным для объективной оценки территорий по уровню их загрязненности. Для удобства анализа преобразуем данные таким образом, чтобы наилучшим показателям состояния окружающей среды соответствовали наименьшие значения, а наихудшим – наибольшие значения.

1. Кластерный анализ территорию юга России по уровню загрязненности

Для анализа будем использовать следующие наиболее показатели: 1) сброс загрязненных сточных вод в поверхностные водные объекты, млн. м³; 2) число источников воды не отвечающих санитарным нормам и правилам; 3) удельный вес проб воды 2 категории не соответствующих гигиеническим нормативам, % по микробиологическим показателям; 4) выбросы в атмосферу загрязняющих веществ - всего, тыс. т; 5) выбросы в атмосферу от автомобильного транспорта, тыс. т; 6) выбросы в атмосферу от стационарных источников по различным видам экономической деятельности, тыс. т; 7) удельный вес выбросов стационарных источников в общем объеме выбросов, %; 8) выбросы загрязняющих атмосферу веществ, отходящих от стационарных источников от сжигания топлива, тыс. т; 9) уловлено и обезврежено загрязняющих веществ, тыс. т; 10) уловлено и обезврежено загрязняющих веществ в % от общего количества отходящих загрязняющих веществ от стационарных источников; 11) уловлено и обезврежено загрязняющих веществ из общего количества

уловленных использовано загрязняющих веществ, тыс. т; 12) уловлено и обезврежено загрязняющих веществ в % от общего количества уловленных и обезвреженных загрязняющих веществ; 13) выбросы основных загрязняющих атмосферу веществ, отходящих от стационарных источников, от сжигания топлива, тыс. т.

Анализ других показателей затрудняется отсутствием числовых или категориальных значений по отдельным субъектам РФ.

Кластерный анализ указанных показателей будем проводить с помощью алгоритма *k-means*, основанного на оптимизации целевой функции, определяющей оптимальное в определенном смысле разбиение множества объектов на кластеры. В качестве целевой функции используется сумма квадратов взвешенных отклонений координат объектов от центров искомых кластеров. Кластеры ищутся сферической либо эллипсоидной формы. Алгоритм оптимизации целевой функции носит итеративный характер, и на каждой итерации требуется рассчитывать матрицу расстояний между объектами. Вычислительная сложность *i*-й итерации алгоритма *k-means* оценивается как $O(kmn)$, где *k*, *m*, *n* – количество кластеров, атрибутов и объектов соответственно [3].

Таблица 1 – Результаты кластеризации территорий субъектов юга РФ по уровню загрязненности алгоритмом *k-means*

Субъект	Номер кластера
▶ Республика Дагестан	0
Республика Калмыкия	0
Ставропольский край	0
Астраханская область	0
Волгоградская область	0
Республика Адыгея	1
Республика Ингушетия	1
Кабардино-Балкарская Республика	1
Карачаево-Черкесская Республика	1
Республика Северная Осетия-Алания	1
Краснодарский край	1
Чеченская Республика	2
Ростовская область	2

Однако, подход связанный с идеей поиска кластеров сферической или эллипсоидной формы в *k-means* имеет ряд недостатков. Подход хорошо работает, когда данные в пространстве образуют компактные сгустки, хорошо отличимые друг от друга. А если данные имеют вложенную форму, то ни один из алгоритмов семейства *k-means* никогда не справится с такой задачей. Также алгоритм плохо работает в случае, когда один кластер значительно больше остальных, и они находятся близко друг от друга – возникает эффект "расщепления" большого кластера. Поэтому для оценки объективности полученных результатов проведем кластерный анализ с использованием нейронной сети Кохонена.

Самоорганизующийся слой Кохонена – это однослойная нейронная сеть с конкурирующей передаточной функцией, которая анализирует выходные значения нейронов слоя и выдаёт в качестве результата наибольшее из этих значений (значение нейрона-победителя). Алгоритм функционирования самообучающихся карт (Self Organizing Maps – SOM) представляет собой один из вариантов кластеризации многомерных векторов. Важным отличием алгоритма SOM является то, что в нем все нейроны (центры классов) упорядочены в некоторую структуру (обычно двумерную сетку). При этом в ходе обучения модифицируется не только нейрон-победитель, но и его соседи, но в меньшей степени. За счет этого SOM можно считать одним из методов проецирования многомерного пространства в пространство с более низкой размерностью. Поэтому обучение по Кохонену напоминает натягивание эластичной сетки прототипов на массив данных из обучающей выборки. При использовании этого алгоритма вектора, схожие в исходном пространстве, оказываются рядом и на полученной карте.

Результатом кластеризации является самоорганизующаяся карта Кохонена (рис. 2) и выходная таблица с номерами кластеров (табл. 2).

Таблица 2 – Результаты кластеризации территорий субъектов юга РФ по уровню загрязненности методом карт Кохонена

Субъект	Номер кластера
Чеченская Республика	0
Ростовская область	0
Республика Адыгея	1
Республика Ингушетия	1
Кабардино-Балкарская Республика	1
Карачаево-Черкесская Республика	1
Республика Северная Осетия-Алания	1
Краснодарский край	1
Ставропольский край	1
Астраханская область	1
Волгоградская область	1
Республика Дагестан	2
Республика Калмыкия	2

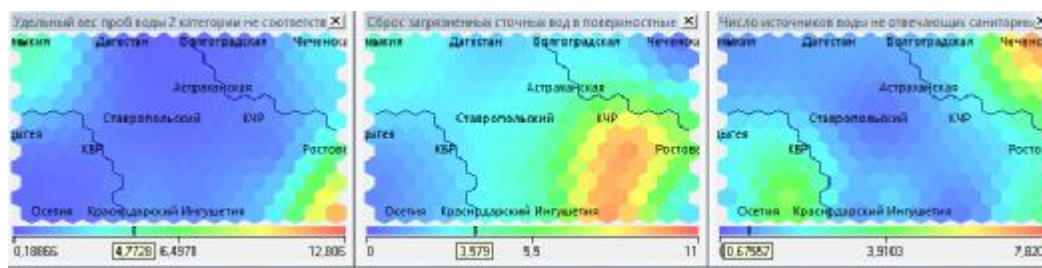


Рисунок 2 – Карты Кохонена показателей уровня загрязненности

Сравнивая результаты кластеризации территорий субъектов юга РФ по уровню загрязненности алгоритмом *k-means* и методом карт Кохонена, необходимо отметить, что отличия в результатах наблюдается для Краснодарского, Ставропольского края и Астраханской области, которые в первом случае оказались вместе с Дагестаном и Калмыкией, а во втором – отнесены к другому кластеру. Отличия в результатах объясняются тем, что самый большой кластер №1, сформированный методом карт Кохонена и содержащий 9 субъектов был "расщеплен" алгоритмом *k-means*.

Результаты по сформированным кластерам наиболее удобно рассматриваются с помощью визуализатора "Куб", в котором встроена кросс-диаграмма, изображающая полученные кластеры в графическом виде, что существенно упрощает анализ.

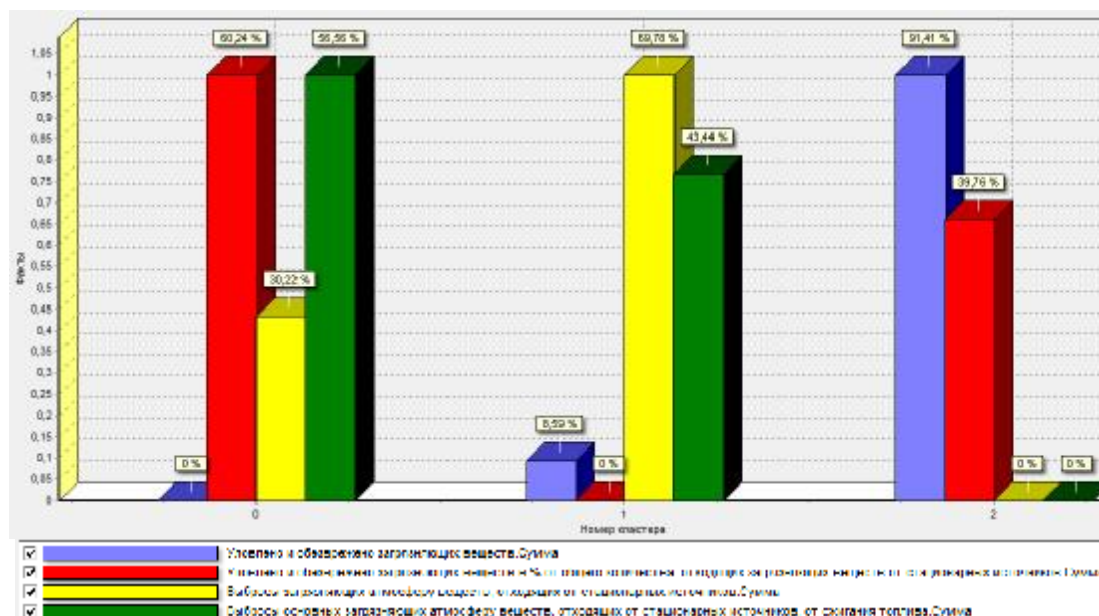


Рисунок 3 – Нормализованная столбчатая диаграмма кластеров по показателям *уровня загрязненности территорий* имеющих наибольшую значимость

Из рисунка 3 видно, что наилучшими (наименьшие по значению) показателями уровня загрязненности обладает кластер №2 (Дагестан и Калмыкия). Несколько более худшее положение занимает кластер №1 (республики Адыгея, Ингушетия, КБР, КЧР, Осетия, Краснодарский, Ставропольский край, Астраханская обл.). Причем хотя интенсивность выбросов во втором кластере низкая, однако уловлено и обезврежено из них крайне мало. Противоположная ситуация в 1-ом кластере, где интенсивность выбросов велика, однако уловлено и обезврежено из них достаточное количество.

Аутсайдером по уровню загрязненности являются Чеченская республика и Ростовская область.

2. Кластерный анализ состояния городской среды

Для исследования состояния городской среды использовались следующие показатели: 1) использование мощности водопроводов в городах и поселках городского типа, тыс. м³ в сутки; 2) использование

мощности водопроводов в % от общей установленной производственной мощности; 3) среднесуточный отпуск воды населению, литров; 4) использование мощности очистных сооружений канализаций, тыс. м³ в сутки; 5) использование мощности очистных сооружений канализаций в % от общей мощности очистных сооружений; 6) площадь зеленых массивов и насаждений в городах, тыс. га; 7) площадь зеленых массивов и насаждений, в том числе в расчете на одного городского жителя, м²; 8) площадь, убираемая механизированным способом, тыс. м²; 9) вывезено бытового мусора, тыс. м³; 10) удельный вес населения, обеспеченного недоброкачественной питьевой водой в общей численности населения.

Результаты кластеризации алгоритмом *k-mean* приведены в таблице 3.

Как и в предыдущем случае для оценки объективности полученных результатов проведем кластерный анализ состояния городской среды с использованием нейронной сети Кохонена.

Таблица 3 - Результаты кластеризации субъектов юга РФ по состоянию *городской среды* алгоритмом *k-means*

Субъект	Номер кластера
Краснодарский край	0
Ставропольский край	0
Волгоградская область	0
Ростовская область	0
Республика Адыгея	1
Республика Дагестан	1
Кабардино-Балкарская Республика	1
Республика Калмыкия	1
Карачаево-Черкесская Республика	1
Республика Северная Осетия-Алания	1
Чеченская Республика	1
Астраханская область	1
Республика Ингушетия	2

Таблица 4 - Результаты кластеризации состояния *городской среды* субъектов юга РФ методом карт Кохонена

Субъект	Номер кластера
▶ Республика Ингушетия	0
Республика Адыгея	1
Республика Дагестан	1
Кабардино-Балкарская Республика	1
Республика Калмыкия	1
Карачаево-Черкесская Республика	1
Республика Северная Осетия-Алания	1
Астраханская область	1
Чеченская Республика	2
Краснодарский край	2
Ставропольский край	2
Волгоградская область	2
Ростовская область	2

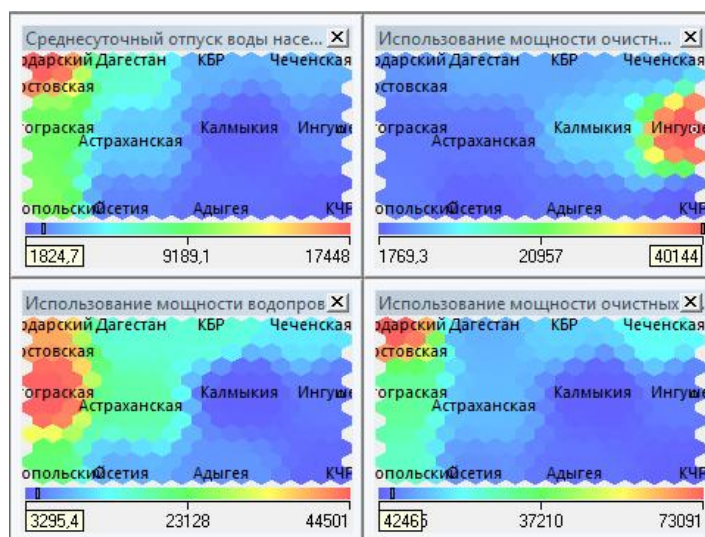


Рисунок 4 – Карты Кохонена показателей состояния *городской среды*

Результаты кластеризации различными методами не совпадают только для Чеченской республики: алгоритм *k-means* отнес ее к одному кластеру вместе с республиками Адыгея, КБР, КЧР, Осетия, Калмыкией, Астраханской обл., в то время как методом карт Кохонена Чеченская республика оказалась в одном кластере с Краснодарским, Ставропольским краем, Волгоградской и Ростовской областями.

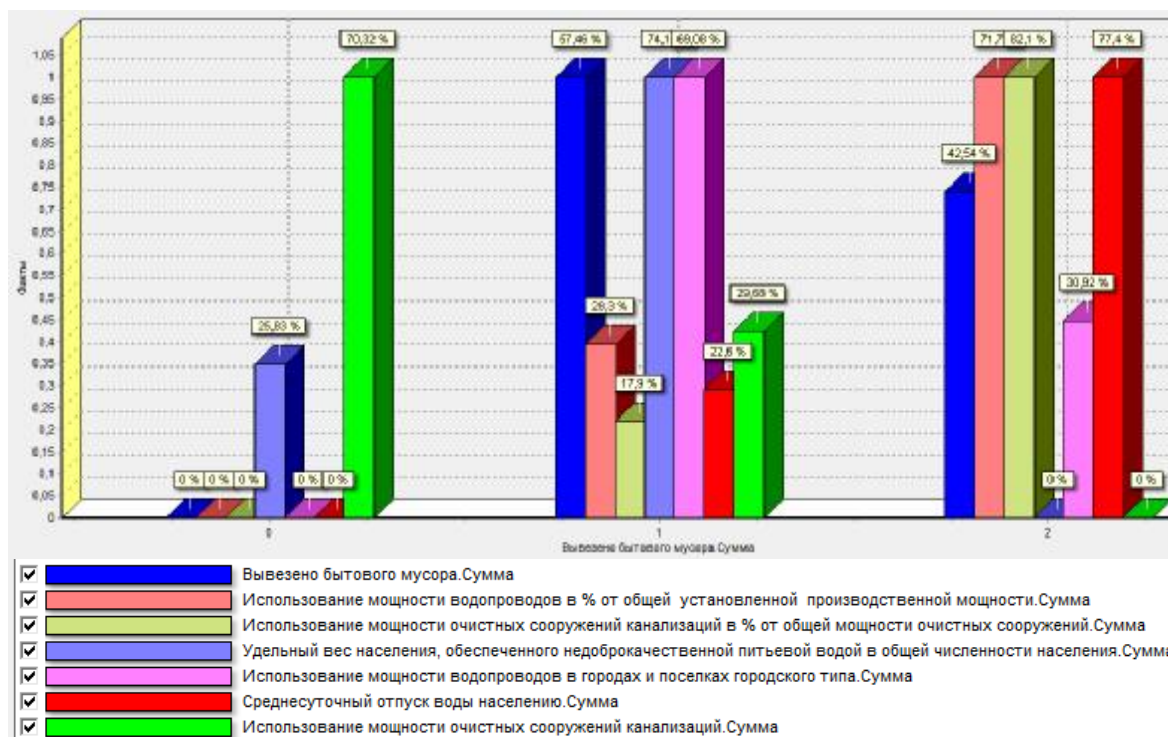


Рисунок 5 - Нормализованная столбчатая диаграмма кластеров по показателям состояния *городской среды* имеющих наибольшую значимость

Здесь явным фаворитом по состоянию городской среды оказалась республика Ингушетия - кластер №0, единственной проблемой в котором является использование мощности очистных сооружений канализации. Остальные субъекты юга РФ имеют намного более худшие показатели по состоянию городской среды, разделение их на разные кластеры связано с состоянием водных ресурсов, а именно: в 1-ом кластере среднему уровню среднесуточного отпуска воды населению соответствует средний уровень использования мощности очистных сооружений канализаций, при наихудшем уровне удельной обеспеченности недоброкачественной питьевой водой населения и плохом использовании мощности водопроводов; во 2-ом кластере еще более худшему уровню среднесуточного отпуска воды населению соответствует наилучшие уровни использования мощности очистных сооружений канализаций и удельной обеспеченности недоброкачественной питьевой водой населения, при еще худшем уровне использования мощности водопроводов.

3. Кластерный анализ общего состояния окружающей среды субъектов юга РФ

Для исследования общего состояния окружающей среды проведем кластерный анализ по 34 показателям, добавив к 23 ранее использованным для оценки уровня загрязненности территорий и состояния городской среды следующие 11: 1) инвестиции в основной капитал, направленные на охрану окружающей среды, млн. руб.; 2) текущие затраты на охрану окружающей среды, млн. руб.; 3) общая площадь земель лесного фонда и земель иных категорий, на которых расположены леса, тыс. га; 4) лесистость территории, %; 5) общий запас древесины на корню, млн. м³; 6) площадь земель лесного фонда, покрытая лесной растительностью, тыс. га; 7) искусственное лесовосстановление, га; 8) ввод молодняков в категорию хозяйственно ценных лесных насаждений, га; 9) удельный вес населения, обеспеченного недоброкачественной питьевой водой в общей численности населения; 10) число источников воды; 11) число источников воды не соответствующих гигиеническим нормам %, по микробиологическим показателям.

Результаты кластеризации алгоритмом *k-means* представлены в таблице №5.

Проведем кластерный анализ общего состояния окружающей среды с использованием нейронной сети Кохонена.

Результаты кластеризации различными методами не совпадают только для Астраханской области: алгоритм *k-means* отнес ее к к одному кластеру вместе с Дагестаном, Калмыкией и Ростовской областью, в то время как методом карт Кохонена Астраханская область отнесена к 1-му кластеру, содержащему республики Адыгея, КБР, Осетию, Чеченскую республику, а также Краснодарский, Ставропольский край и Волгоградскую область.

Таблица 5 - Результаты кластеризации субъектов юга РФ по *общему* состоянию окружающей среды алгоритмом *k-means*

Субъект	Номер кластера
▶ Республика Адыгея	0
Кабардино-Балкарская Республика	0
Республика Северная Осетия-Алани	0
Чеченская Республика	0
Краснодарский край	0
Ставропольский край	0
Волгоградская область	0
Республика Дагестан	1
Республика Калмыкия	1
Астраханская область	1
Ростовская область	1
Республика Ингушетия	2
Карачаево-Черкесская Республика	2

Таблица 6 - Результаты кластеризации субъектов юга РФ по *общему* состоянию окружающей среды методом карт Кохонена

Субъект	Номер кластера
▶ Республика Ингушетия	0
Карачаево-Черкесская Республика	0
Республика Адыгея	1
Кабардино-Балкарская Республика	1
Республика Северная Осетия-Алания	1
Чеченская Республика	1
Краснодарский край	1
Ставропольский край	1
Астраханская область	1
Волгоградская область	1
Республика Дагестан	2
Республика Калмыкия	2
Ростовская область	2



Рисунок 6 – Карты Кохонена показателей *общего* состояния окружающей среды

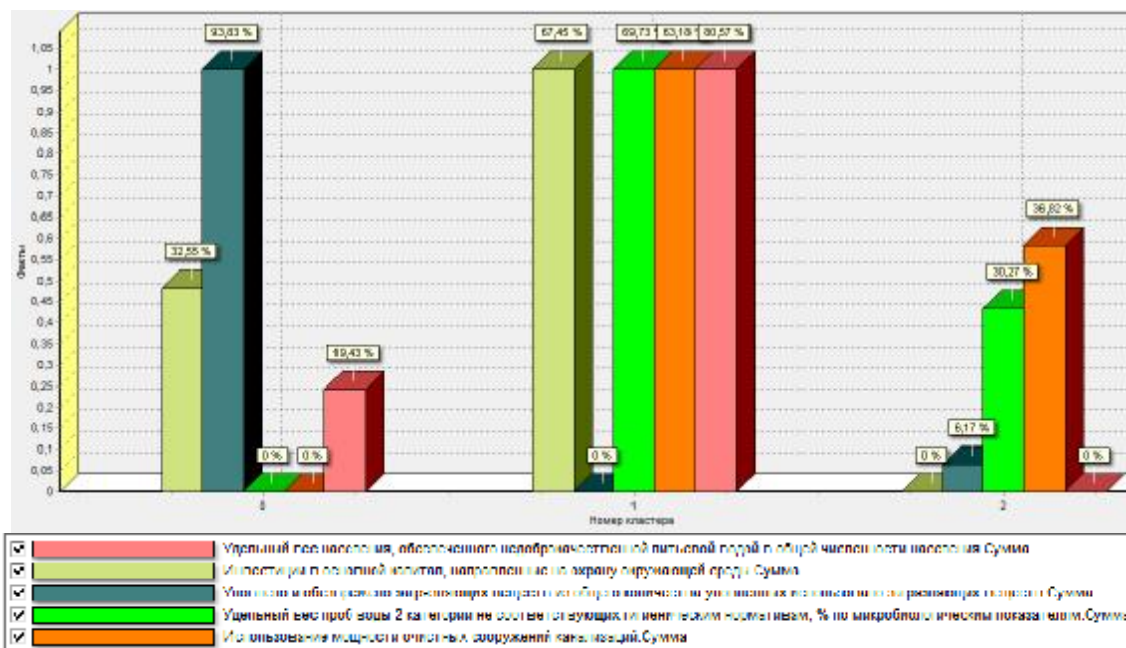


Рисунок 7 – Нормализованная столбчатая диаграмма кластеров по показателям *общего* состояния окружающей среды имеющих наибольшую значимость



Рисунок 8 – Кластеризация субъектов юга РФ по состоянию окружающей среды

Выводы:

1. Основными, определяющими состав кластеров факторами, являются:

- состояние их водных ресурсов: уровни обеспеченности недоброкачественной водой населения, качество воды 2-ой категории, использования мощности очистных сооружений;
- уровень уловленных и обезвреженных загрязняющих веществ;
- инвестиции в основной капитал, направленный на охрану окружающей среды.

2. Наиболее благополучными территориями по состоянию окружающей среды среди субъектов юга РФ являются представители кластера №2: республики Адыгея, Кабардино-Балкарская, Чеченская, республика Северная Осетия-Алания, Астраханской и Волгоградская области, Краснодарский и Ставропольским край.

На основании анализа данных по состоянию окружающей среды для закрепления лидирующих позиций предлагается: в Волгоградской и Астраханской области, Краснодарском крае значительно уменьшить уровень обеспеченности недоброкачественной водой населения; в Кабардино-Балкарской, Чеченской и республике Северная Осетия-Алания улучшить качество воды 2-ой категории; в Кабардино-Балкарской, Чеченской республике и Астраханской области значительно увеличить уровень уловленных и обезвреженных загрязняющих веществ; в Кабардино-Балкарской и Чеченской республиках увеличить уровень использования мощности очистных сооружений; в Волгоградской области, Краснодарском и Ставропольском крае увеличить инвестиции в основной капитал, направленный на охрану окружающей среды.

Таблица 7 - данные по кластеру №2

Субъект	Удельный вес населения, обеспеченного недоброкачественной питьевой водой в общей численности населения	Удельный вес проб воды 2 категории не соответствующих гигиеническим нормативам, % по микробиологическим показателям	Уловлено и обезврежено загрязняющих веществ из общего количества уловленных использовано загрязняющих веществ	Инвестиции в основной капитал, направленные на охрану окружающей среды	Использование мощности очистных сооружений канализаций
▶ Республика Северная Осетия-Алания	0,1	1,6	0,0398	0,01782	1928,26143
Республика Адыгея	0,3	0,62821	2	0,01847	3052,71895
Кабардино-Балкарская Республика	0,3	1,744	15,625	0,07711	6109,23077
Чеченская Республика	1,4	3,46795	15,6	0,07529	6954,11765
Ставропольский край	4,5	0,34894	1,50455	0,14823	3567,93383
Волгоградская область	8,6	0,18866	1,16392	0,33363	3257,45467
Краснодарский край	9	0,21722	0,10649	0,11046	4938,60299
Астраханская область	11,5	0,22245	8,59649	0,06567	3033,88812

3. Полученные результаты кластеризации состояния объектов окружающей среды могут служить основой для разработки прогнозных сценариев развития экономических систем региона [1].

4. Неплохие показатели по состоянию окружающей среды у представителей кластера №0: республики Дагестан и Калмыкия, имеющие наилучшие показатели по уровню загрязненности и Ростовской области, хотя и является аутсайдером по уровню загрязненности.

Таблица 8 - данные по кластеру №0

Субъект	Удельный вес населения, обеспеченного недоброкачественной питьевой водой в общей численности населения	Удельный вес проб воды 2 категории не соответствующих гигиеническим нормативам, % по микробиологическим показателям	Уловлено и обезврежено загрязняющих веществ из общего количества уловленных использовано загрязняющих веществ	Инвестиции в основной капитал, направленные на охрану окружающей среды
▶ Ростовская область	14,7	0,4099	1,98039	0,23494
Республика Калмыкия	20	0,33333	373,5	2,02989
Республика Дагестан	23	0,41352	251,5	0,18298

Для улучшения состояния окружающей среды, прежде всего, необходимо провести следующие мероприятия: в Дагестане и Калмыкии значительно увеличить уровень уловленных и обезвреженных загрязняющих веществ; во всех трех субъектах значительно уменьшить уровень обеспеченности недоброкачественной водой населения и увеличить инвестиции в основной капитал, направленный на охрану окружающей среды.

5. Плохие показатели по состоянию окружающей среды у представителей кластера №1: КЧР и республики Ингушетия.

Таблица 9 - данные по кластеру №1

Субъект	Удельный вес населения, обеспеченного недоброкачественной питьевой водой в общей численности населения	Удельный вес проб воды 2 категории не соответствующих гигиеническим нормативам, % по микробиологическим показателям	Уловлено и обезврежено загрязняющих веществ из общего количества уловленных использовано загрязняющих веществ	Инвестиции в основной капитал, направленные на охрану окружающей среды	Использование мощностей очистных сооружений канализаций
▶ Карачаево-Черкесская Республика	61,4	5,07692	0,05417	0,56522	1769,3346
Республика Ингушетия	65,5	12,80556	3,6	3,6	40144,36364

Для исправления положения необходимо принять следующие меры: значительно уменьшить уровень обеспеченности недоброкачественной водой населения; улучшить качество воды 2-ой категории; увеличить уровень использования мощности очистных сооружений; увеличить инвестиции в основной капитал, направленный на охрану окружающей среды.

6. Качество проведения кластерного анализа значительно увеличится в случае наличия большего числа показателей, по которым есть числовые или категориальные значения по всем анализируемым субъектам РФ. Однако наибольшее влияние на полученные результаты оказывает квалификация и интуиция эксперта-аналитика. Именно он решает, на сколько кластеров необходимо разбить исследуемый набор данных, и какие свойства будут основными при построении кластера.

Литература:

1. Байдаков А.Н., Назаренко А.В. Сценарное прогнозирование в управлении аграрными экономическими системами/А.Н. Байдаков, А.В. Назаренко//Вестник Института дружбы народов Кавказа, №4(20), 2011 С. 21-26
2. Охрана окружающей среды в России. 2010: Стат. сб./Росстат. - М., 2010. – 303с.
3. Паклин Н.Б., Орешков В.И. Бизнес аналитика: от данных к знаниям: Учебное пособие 2-е изд., перераб. и доп. – СПб.: Петербург, 2010.
4. Сахнюк Т.И., Сахнюк П.А., Левушкина С.В. Управление природоохранной деятельностью в региональном АПК на основе методов экономико-математического моделирования/Т.И. Сахнюк, П.А. Сахнюк, С.В. Левушкина//Научный журнал КубГАУ [Электронный ресурс]. -Краснодар: КубГАУ, 2011. -№72(08). -Шифр Информрегистра: 0421100012\0319. -Режим доступа: <http://ej.kubagro.ru/2011/08/pdf/40.pdf>