



УДК 502.3  
ГРНТИ 87.33

## РЕГРЕССИОННО-КЛАССИФИКАЦИОННЫЕ МОДЕЛИ ПОКАЗАТЕЛЕЙ ЭКОЛОГИЧЕСКОГО СОСТОЯНИЯ АНТРОПОГЕННО-ИЗМЕНЕННЫХ ТЕРРИТОРИЙ

*В.М. УМЫВАКИН, доктор географических наук  
ВУНЦ ВВС «ВВА имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж)  
В.В. БОТАЛОВ, кандидат технических наук  
ВУНЦ ВВС «ВВА имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж)  
А.В. ПАХМЕЛКИН, кандидат географических наук  
ВУНЦ ВВС «ВВА имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж)  
А.В. ШВЕЦ  
ВУНЦ ВВС «ВВА имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж)*

В статье предлагаются регрессионно-классификационные (дискретно-непрерывные) модели анализа данных об экологическом состоянии антропогенно-измененных территорий.

*Ключевые слова:* антропогенно-измененная территория; экологическое состояние; природно-хозяйственный показатель; типологическая регрессия; дискретно-непрерывная модель.

### REGRESSION AND CLASSIFICATION MODELS OF INDICATORS ECOLOGICAL CONDITION OF ANTHROPOGENOUS CHANGED TERRITORIES

*V.M. UMYVAKIN, Doctor of Geographical Sciences  
MESC AF «N.E. Zhukovsky and Y.A. Gagarin Air Force Academy» (Voronezh)  
V.V. BOTALOV, Candidate of Technical Sciences  
MESC AF «N.E. Zhukovsky and Y.A. Gagarin Air Force Academy» (Voronezh)  
A.V. PAHMELKIN, Candidate of Geographical Sciences  
MESC AF «N.E. Zhukovsky and Y.A. Gagarin Air Force Academy» (Voronezh)  
A.V. SHVETS  
MESC AF «N.E. Zhukovsky and Y.A. Gagarin Air Force Academy» (Voronezh)*

The regression and classification (discrete and continuous) models of the analysis of the changed territories given about an ecological state anthropogenous are offered in this article.

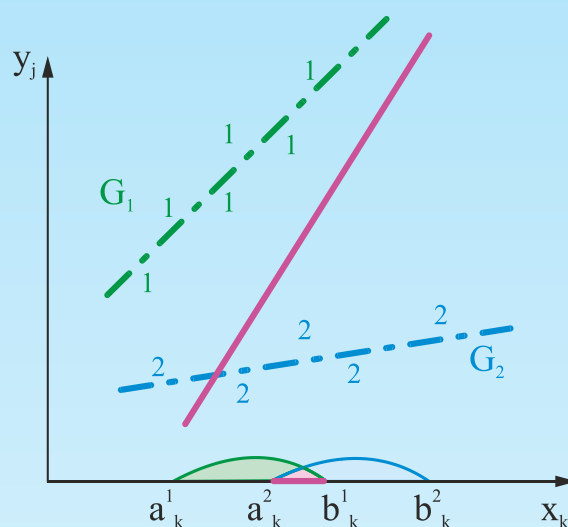
*Keywords:* the anthropogenous changed territory; an ecological state; a natural and economic indicator; typological regression; discrete and continuous model.

**Введение.** В настоящее время актуальной научной задачей является анализ экологической ситуации в регионах с интенсивной хозяйственной и иной деятельностью, результаты которого используются для обеспечения экологической безопасности территорий и населения.



Основной целью данной работы является разработка и применение классификационных моделей регрессионного анализа для прогнозирования экологического состояния антропогенно-измененных территорий (геообъектов).

**Методы построения дискретно-непрерывных моделей показателей экологического состояния геообъектов.** Для оценки экологического состояния антропогенно-измененных территорий широко применяются методы анализа многомерных природно-хозяйственных данных (методы корреляционно-регрессионного анализа, многомерной классификации, факторного анализа и др.) [1, 2]. Для повышения эффективности моделирования статистических взаимосвязей показателей качества окружающей среды геообъектов (например, муниципальных районов) предлагается методический подход, основанный на комплексировании процедур классификации и корреляционно-регрессионного анализа. Следует отметить, что основная роль методов классификации заключается в том, что их применение позволяет получить однородные классы геообъектов, «центрами» которых являются различные уравнения регрессии (рисунок 1). При этом повышается качество (достоверность) регрессионных моделей.



1 и 2 – объекты (точки) групп  $G_1$  и  $G_2$ ;

$[a_k^1, b_k^1]$  и  $[a_k^2, b_k^2]$  – диапазоны изменения параметра  $X_k$  соответственно в группах  $G_1$  и  $G_2$

Рисунок 1 – Классификация объектов по регрессионным моделям

Совместное применение методов классификации (типологической группировки) и корреляционно-регрессионного анализа называют методом дискретно-непрерывного моделирования показателей [3], а комплекс регрессионных моделей для всех классов (типов) – дискретно-непрерывной моделью (ДНМ) показателей. В работе О.В. Азаряна [4] построение дискретно-непрерывных моделей называется типологическим регрессионным анализом.

Рассмотрим методические подходы к построению ДНМ показателей экологического состояния антропогенно-измененных территорий. Первый подход, предложенный Б.Б. Розиным [3], заключается в первоначальном выделении (например, с помощью методов многомерной классификации [1]) по некоторым показателям однородных классов геообъектов, для которых затем строятся регрессионные уравнения [2]. Они должны существенно отличаться друг от друга.

Как правило, ставится цель построения ДНМ с линейными уравнениями регрессии (это наиболее распространенная постановка задачи).

В работе [3] задачу построения ДНМ (моделей типологической регрессии) показателей предлагается решать в рамках следующей схемы:



1) обоснование гипотезы о неоднородности совокупности анализируемых геообъектов и целесообразности использования дискретно-непрерывной модели;

2) построение дискретной части модели, включающее выбор способа классификации и набора показателей, по которым должно производиться разбиение совокупности на однородные классы, формирование математического описания классов и правила отнесения нового геообъекта к той или иной группе, оценку точности классификации;

3) построение непрерывной части модели, включающее все операции, присущие корреляционно-регрессионному анализу, но выполняемые для каждого класса геообъектов в отдельности;

4) оценка аппроксимирующей способности модели.

При этом особое значение имеют результаты проверки гипотезы о неоднородности совокупности анализируемых геообъектов и исследования вопроса соизмеримости показателей различных типов. Один из возможных способов проверки гипотезы о неоднородности совокупности заключается в использовании многомерной классификации с последующей оценкой существенности различий, полученных в результате разбиения классов.

Таким образом, если в результате разбиения получены существенно различные классы, то есть основание принять гипотезу о том, что исходная совокупность неоднородна и представляет собой смесь, композицию из нескольких более однородных подмножеств. Существенно различными классы считаются в том случае, если существенно различаются основные статистические характеристики каждого из показателей в отдельности (параметры кривых распределения) и характеристики их взаимосвязи (матрицы коэффициентов корреляции). Такой анализ может проводиться либо путем попарного сравнения классов между собой, либо путем сравнения каждого класса с общей совокупностью.

После того, как набор показателей экологического состояния для классификации геообъектов сформирован, можно переходить к выбору способа разбиения совокупности геообъектов на классы. Следует различать два принципиально различных способа классификации: метод последовательных логических делений по каждому показателю в отдельности (типологическую группировку) и метод формирования классов на основе набора репрезентативных показателей (многомерную классификацию), ни один из которых не является необходимым или достаточным условием принадлежности геообъекта к данному классу.

При использовании любого из методов многомерной классификации один из важных вопросов построения дискретной части ДНМ – определение рационального числа классов. С точки зрения способа определения числа классов все алгоритмы классификации можно подразделить на две группы: алгоритмы, где число классов заранее задается исследователем, и алгоритмы, где число классов в явной форме не задается, а находится в процессе разбиения. В этом подходе при построении ДНМ очень важна задача отыскания небольшого набора показателей, которые объясняли бы основную часть внутриклассовой вариации моделируемого критерия. Начальным этапом этого процесса является анализ структуры матриц парных коэффициентов корреляции и выделение взаимосвязанных групп показателей. Далее исключаются дублирующие и тесно коррелированные показатели, производится выбор существенных показателей для построения частных регрессий. Наиболее распространенные формальные способы отыскания существенных показателей базируются на последовательном исключении показателей, дающих наименьшее (наибольшее) изменение остаточной дисперсии по сравнению с исходным набором (например, с помощью метода главных компонент [1]).



Отметим, что в методе типологической регрессии [4] второй и третий этапы в приведенной выше схеме построения ДНМ находятся в соотношении итерационного взаимодействия, то есть непрерывно из очередной итерации второго этапа переходят в следующую итерацию третьего и обратно до тех пор, пока не будет осуществлена классификация и определены коэффициенты уравнения регрессии, удовлетворяющие некоторому критерию качества решения данной задачи.

В настоящей работе предлагается использовать оригинальный методический подход к построению ДНМ показателей экологического состояния геообъектов [5]. Из всех показателей выделим набор факторов-функций (критериев)  $y_1, y_2, \dots, y_m$  и набор факторов-аргументов (параметров)  $x_1, x_2, \dots, x_n$ . Обозначим через  $y_j^i$  и  $x_k^i$ , соответственно, значение анализируемого (моделируемого)  $j$ -го критерия и  $k$ -го параметра  $i$ -го геообъекта ( $j=1,2,\dots, m, k=1,2,\dots, n; i=1,2,\dots, N$ ). Требуется построить ДНМ критерия  $y_j$  от параметров  $x_1, x_2, \dots, x_n$ .

Предлагаемый подход основан на предположении, что исходная информация, необходимая для построения уравнений типологической регрессии, содержится в самом моделируемом показателе. Действительно, для геообъектов с близкими значениями параметров значения моделируемого критерия могут сильно отличаться (см. рисунок 1). Отсюда следует вывод о том, что измерения критерия сделаны на геообъектах из разных классов, которым соответствуют существенно различные уравнения регрессии – «центры» классов (см. рисунок 1). В основе используемого подхода к построению ДНМ критериев экологического состояния территорий лежит совместное применение двух методов статистического анализа данных – метода регрессионного анализа и метода главных компонент [1, 2]. Метод главных компонент позволяет проводить классификацию анализируемых геообъектов в пространстве новых некоррелированных переменных, линейно зависящих от исходных показателей экологического состояния геообъектов.

Как известно, важнейшими характеристиками качества (достоверности) уравнений регрессии являются выборочный коэффициент множественной корреляции  $R$  и дисперсионное отношение Фишера ( $F$  – критерий) [2]. Рассматриваемый метод построения ДНМ критериев экологического состояния территорий, по существу, направлен на повышение качества уравнений типологической регрессии в каждом классе геообъектов. Здесь можно рассмотреть вопрос о качестве типологической группировки, который решается с помощью проверки статистической гипотезы об идентичности уравнений регрессии [2]. Если гипотеза отвергается, то есть основания считать, что разбиение совокупности геообъектов на классы, «центрами» которых являются определенные регрессионные модели критериев, проведено верно.

С формальной точки зрения требуется разбить всю совокупность (множество)  $G$  анализируемых геообъектов на два класса  $G_1$  и  $G_2$ , в каждом из которых коэффициенты множественной корреляции  $R_1$  и  $R_2$  выше, чем коэффициент  $R$  для всей выборки. Приведем общее описание алгоритма метода построения ДНМ [5]. Вначале строится линейная регрессионная модель моделируемого критерия и определяются выборочные значения  $R$  и  $F$  для всей совокупности геообъектов. Если полученная модель является достоверной и статистические данные таковы, что  $R \geq R^*$ , где  $R^*$  – требуемое значение коэффициента множественной корреляции, разбиение множества  $G$  на классы  $G_1$  и  $G_2$  не производится. Предположим, что  $R < R^*$ . В этом случае метод построения ДНМ позволяет разделить всю совокупность геообъектов на два класса  $G_1$  и  $G_2$ , в которых соответствующие значения коэффициентов множественной корреляции  $R_1$  и  $R_2$  увеличиваются по сравнению с коэффициентом  $R$  для всей выборки. Если требование к качеству регрессионных моделей не выполняется, т.е.  $R_l \geq R^*$  ( $l=1,2$ ), классы  $G_1$  и  $G_2$  разбиваются на подклассы и т.д.



Применение дискретно-непрерывных моделей показателей для анализа медико-экологической ситуации в муниципальных районах Воронежской области. Матрица исходных данных для построения уравнений типологической регрессии (ДНМ) критерия «общая заболеваемость взрослого населения на 1000 человек» как фактор-функции от параметров (фактор-аргументов)  $x_1$  и  $x_2$  медико-экологической ситуации территорий муниципальных районов Воронежской области по состоянию на 2013 год приведена в таблице 1.

Таблица 1 – Медико-экологические данные для построения дискретно-непрерывных моделей критерия «общая заболеваемость взрослого населения»

№ п/п	Муниципальные районы и городские округа	Общая заболеваемость взрослого населения на 1000 человек ( $y_1$ )	Средняя концентрация фенола в воздухе населенных мест ( $x_1$ ), мг/м <sup>3</sup>	Средняя концентрация железа в питьевой воде ( $x_2$ ), мг/дм <sup>3</sup>
1	Аннинский	1148,56	0,004	0,26
2	Бобровский	1353,24	0,0037	0,59
3	Богучарский	1296,05	0,004	0,05
4	Борисоглебский	1032,63	0,004	0,223
5	Бутурлиновский	814,47	0,004	0,1
6	Верхнемамонский	1209,94	0,004	0,05
7	Верхнехавский	1285,30	0,0042	0,22
8	Воробьевский	1392,77	0,004	0,131
9	Грибановский	1042,26	0,0045	0,1032
10	Калачеевский	1234,49	0,004	0,13
11	Каменский	1199,12	0,0039	0,107
12	Кантемировский	971,44	0,0051	0,292
13	Каширский	858,01	0,004	0,16
14	Лискинский	1043,98	0,0057	0,09
15	Нижнедевицкий	1276,33	0,0070	0,14
16	Новоусманский	1272,64	0,004	0,15
17	Новохоперский	1317,11	0,004	0,4
18	Ольховатский	1219,40	0,0052	0,2127
19	Острогожский	1127,16	0,004	0,09
20	Павловский	1140,68	0,004	0,07
21	Панинский	1097,51	0,0058	0,133
22	Петропавловский	1382,15	0,004	0,11
23	Поворинский	1353,47	0,004	0,0595
24	Подгоренский	942,03	0,0053	0,16
25	Рамонский	1034,09	0,004	0,0075
26	Репьевский	1213,73	0,0043	0,054
27	Россошанский	1027,19	0,0049	0,107
28	Семилукский	1035,18	0,004	0,45
29	Таловский	1656,15	0,0041	0,84
30	Терновский	824,19	0,004	0,1536
31	Хохольский	1653,35	0,01567	0,34
32	Эртильский	1097,57	0,004	0,196





Как видно из таблицы 4, для всей совокупности 32 муниципальных образований регрессионная модель критерия  $y_1$  («общая заболеваемость населения») недостаточно достоверна ( $R=0.54$ ,  $F^{набл}=6.001$ ). Построение ДНМ анализируемого критерия позволило выделить два класса районов, в каждом из которых построено свое достоверное уравнение регрессии (см. таблицу 2).

Таблица 2 – Дискретно-непрерывные модели критерия «общая заболеваемость взрослого населения»

Класс муниципальных районов	Среднее значение	Корреляционная матрица	Уравнение регрессии и его достоверность
Вся выборка	$\bar{y}_1=1173,51$ $\bar{x}_1=0,005$ $\bar{x}_2=0,193$	1,00 0,37 0,43 0,37 1,00 0,10 0,43 0,10 1,00	$y_1=935,71+31258,69x_1+465,88x_2$ $R=0,54$ ; $F^{набл}=6,001$ ; $F^{кр}=3,33$
Первый класс (1,4-5,12-14,21,24,28,30-31)	$\bar{y}_1=1038,30$ $\bar{x}_1=0,006$ $\bar{x}_2=0,215$	1,00 0,90 0,48 0,90 1,00 0,29 0,48 0,29 1,00	$y_1=618,86+55825,59x_1+499,03x_2$ $R_1=0,92$ ; $F_1^{набл}=23,51$ ; $F_1^{кр}=4,46$
Второй класс (2-3,6-11,15-20,22-23,25-27,29,32)	$\bar{y}_1=1244,33$ $\bar{x}_1=0,004$ $\bar{x}_2=0,182$	1,00 -0,10 0,67 -0,10 1,00 -0,10 0,67 -0,10 1,00	$y_1=1189,35-8002,75x_1+490,58x_2$ $R_2=0,67$ ; $F_2^{набл}=7,49$ ; $F_2^{кр}=3,55$

Приложение:  $y_1$  – общая заболеваемость населения;  $x_1$  – концентрации фенола в воздухе;  $x_2$  – концентрация железа в питьевой воде;  $R$  – коэффициент множественной корреляции;  $F^{набл}$  и  $F^{кр}$  – соответственно наблюдаемое и критическое значения  $F$ -критерия Фишера.

При этом в первом классе коэффициент множественной корреляции  $R_1=0.92$ , критерий Фишера  $F_1^{набл}=23.51$ ,  $F_1^{кр}=4.46$ , а полученное уравнение регрессии является достоверным при уровне значимости 0.05.

В первом классе районов, расположенных в основном в северном и восточном секторах области преимущественно городского типа, заболеваемость населения имеет устойчивую положительную корреляцию с концентрацией фенола как индикатором повышенной автотранспортной нагрузки и фактором экологического риска в городских поселениях; во втором классе районов городского и, в большей степени, сельского типа более существенную роль в формировании общественного здоровья имеет качество питьевой воды (увеличение содержания железа – фактор риска повышения заболеваемости населения), что свидетельствует о различной обусловленности показателей общественного здоровья спецификой социально-экономической инфраструктуры.

**Выводы.** Таким образом, дискретно-непрерывная модель (ДНМ) показателей экологического состояния антропогенно-измененных территорий представляет собой комбинацию двух моделей – дискретной и непрерывной. При ее использовании совокупность анализируемых геообъектов моделируется на двух уровнях – уровне классов и уровне отдельных геообъектов. Модели верхнего уровня (уровень классов) – дискретные. В общем случае они строятся методами классификации. Модели нижнего уровня – непрерывные и представляют собой внутриклассовые уравнения регрессии. При этом исследуемая статистическая зависимость показателей экологического состояния геообъектов описывается комплексом типологических регрессионных уравнений, которые являются «центрами» различных классов.

Построенные ДНМ показателей качества окружающей среды и здоровья населения позволяют повысить качество и эффективность прогнозирования локального изменения критериев экологического состояния антропогенно-измененных территорий при варьировании параметров анализируемых геообъектов в задачах управления природопользованием и оценки экологических рисков.



## СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. М.: Статистика, 1974. 240 с.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей: справ. изд. М.: Финансы и статистика, 1989. 487 с.
3. Розин Б.Б. Теория распознавания образов в экономических исследованиях. М.: Статистика, 1973. 224 с.
4. Азарян О.В. Алгоритмы типологической регрессии в задачах автоматизации процесса проектирования (на примере создания и комбинирования проектных решений в гидростроительстве) // Статистика. Вероятность. Экономика: ученые записки по статистике. М.: Наука, 1985. Т. 49. С. 69–77.
5. Давнис В.В., Умывакин В.М., Климов А.В. Методические аспекты построения дискретно-непрерывных показателей эколого-экономических систем // Современная экономика: проблемы и решения. 2010. № 10 (10). С. 109–117.

## BIBLIOGRAPHY

1. Ayvazyan S.A., Bezhayeva Z.I., Staroverov O.V. Classification of multidimensional observations. M.: Statistics, 1974. 240 p.
2. Ayvazyan S.A., Enyukov I.S., Meshalkin L.D. Applied statistics. Research of dependences: reference book. M.: Finance and statistics, 1989. 487 p.
3. Rozin B.B. The theory of recognition of images in economic researches. M.: Statistics, 1973. 224 p.
4. Azaryan O.V. Algorithms of typological regression in designing process automation tasks (on the example of creation and a combination of project decisions in a hydroconstruction) // Statistics. Probability. Economy: scientific notes according to the statistics. M.: Science, 1985. Vol. 49. P. 69–77.
5. Davnis V.V., Umyvakina V.M., Klimov A.V. Methodical aspects of creation of discrete and continuous indicators of ecological-economic systems // Modern economy: problems and decisions. 2010. No. 10 (10). P. 109–117.

© Умывакин В.М., Боталов В.В., Пахмелкин А.В., Швец А.В., 2017

Умывакин Василий Митрофанович, доктор географических наук, старший научный сотрудник, ведущий научный сотрудник научно-исследовательского центра (проблем применения, обеспечения и управления авиацией Военно-воздушных сил), Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), Россия, 394064, г. Воронеж, ул. Старых Большевиков д. 54А, vaiu@mil.ru

Боталов Владимир Валерьевич, кандидат технических наук, начальник отдела – заместитель начальника управления научно-исследовательского центра (проблем применения, обеспечения и управления авиацией Военно-воздушных сил), Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), Россия, 394064, г. Воронеж, ул. Старых Большевиков д. 54А, vaiu@mil.ru

Пахмелкин Александр Васильевич, кандидат географических наук, старший научный сотрудник научно-исследовательского центра (проблем применения, обеспечения и управления авиацией Военно-воздушных сил), Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), Россия, 394064, г. Воронеж, ул. Старых Большевиков д. 54А, vaiu@mil.ru

Швец Алексей Владимирович, научный сотрудник научно-исследовательского центра (проблем применения, обеспечения и управления авиацией Военно-воздушных сил), Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина» (г. Воронеж), Россия, 394064, г. Воронеж, ул. Старых Большевиков д. 54А, vaiu@mil.ru