

# **ТИПОЛОГИЗАЦИИ ТЕРРИТОРИЙ ВЬЕТНАМА ПО ХАРАКТЕРИСТИКАМ ПОЖАРНОЙ ОПАСНОСТИ В СЕКТОРЕ ХОЗЯЙСТВУЮЩИХ СУБЪЕКТОВ**

**В.А. Минаев, доктор технических наук, профессор,  
заслуженный работник высшей школы Российской Федерации.  
Московский государственный технический университет им. Н.Э. Баумана.  
Н.Г. Топольский, доктор технических наук, профессор;  
Дао Ань Туан.  
Академия ГПС МЧС России.  
Нгуен Туан Ань, кандидат технических наук.  
Институт пожарной безопасности Вьетнама**

Рассматриваются многомерные процессы и явления, определяющие пожарные риски на различных территориях. Дана постановка задачи типологизации территорий Вьетнама на основе методов кластерного анализа. Показана методика и результаты снижения размерности пространства исходных признаков. Описан общий алгоритм типологизации территорий Вьетнама по пожарным рискам. Рассматриваются результаты решения задачи многомерной кластеризации территорий Вьетнама по состоянию пожарной опасности в секторе хозяйствующих субъектов.

*Ключевые слова:* Вьетнам, типологизация, кластерный анализ, алгоритм, пожарная безопасность, пожарные риски, сектор хозяйствующих субъектов

## **TIPOLOGIZATION OF THE PROVINCES ACCORDING TO CHARACTERISTICS OF FIRE DANGER FOR BUSINESS OBJECTS SECTOR IN VIETNAM**

V.A. Minaev. Bauman Moscow state technical university.  
N.G. Topolsky; Dao Anh Tuan. Academy of State fire service of EMERCOM of Russia.  
Nguyen Tuan Anh. Institute of fire safety of Vietnam

Multidimensional processes and phenomena that determine fire risks in different areas are considered. Statement of the problem of Vietnam's territories typology on the basis of cluster analysis methods is given. Characteristics of the fire situation in Vietnam at the residential sector and at the business objects sector defined in the result of expert procedures are given. The methodology and results of reducing space of initial indicators are shown. The general of Vietnam's territories typology for fire risks is described. Results of the multidimensional Vietnam's territories clustering for business objects sector on fire danger are discussed.

*Keywords:* Vietnam, typology, cluster analysis, algorithm, fire safety, fire risks, business objects sector

Типологизация территорий по пожарной опасности должна учитывать многие факторы: объемы и класс горючих материалов, находящихся на них, особенности климата и рельефа местности; текущие условия погоды, частоту возникновения источников зажигания, наличие и качество противопожарных средств [1]. Чтобы учесть территориальные аспекты пожарной опасности, относящиеся к жилым массивам, промышленным и сельскохозяйственным объектам в регионах Вьетнама, необходимо включить в рассмотрение такие системные факторы, как социально-экономические,

демографические, организационно-управленческие и материально-технические (относящиеся к деятельности противопожарной службы) и др.

Методы сложной многомерной классификации, предназначенные для разделения некоторой совокупности объектов (провинций Вьетнама) на однородные группы в смысле схожести условий по пожарным рискам, как правило, включают кластерный анализ. Наличие множества исходных признаков, характеризующих пожарные риски, заставляет отбирать из них наиболее существенные и изучать меньший набор признаков, то есть снижать размерность детерминант пожарного риска. Как правило, при этом исходное поле детерминант подвергается некоторому преобразованию, обеспечивающему минимальную потерю информации. Что дает возможность лаконичного и более простого объяснения многомерных структур, связанных с пожарными рисками.

Задача типологизации территорий по пожарным рискам. Пусть  $X$  – множество территорий (провинций) Вьетнама;  $Y$  – множество кластеров (типов территорий по пожарному риску). Задана некоторая функция расстояния между территориями  $\rho(x_i, x_j)$ , где  $i, j$  – индексы территорий;  $i=1, \dots, n$ ;  $j=1, \dots, n$ ;  $n$  – общее количество территорий во множестве  $X=\{x_i\}$ .

Требуется разбить множество территорий  $X$  на непересекающиеся  $Y^1 \cup Y^2 \cup \dots \cup Y^m = X$  подмножества, называемые кластерами  $Y^m$ , ( $m=1, \dots, M$ ), так, чтобы каждый кластер (тип территорий) состоял из объектов, близких по метрике  $\rho$ , а объекты разных кластеров существенно отличались по той же метрике. При этом каждому объекту  $x_i \in X$  приписывается еще и номер кластера  $x_i m$ ;  $m=1, \dots, M$ .

Алгоритм кластеризации – это функция  $\Psi(X) \rightarrow Y$ , которая любому объекту  $x_i \in X$  ставит в соответствие номер кластера  $Y^m$ . Цель реализации алгоритма – определить оптимальное число кластеров с точки зрения некоторого критерия качества кластеризации, отражающего многомерное представление пожарных рисков для территории страны.

Решение задачи кластеризации, как правило, неоднозначно, обуславливается тремя основными причинами:

- критерий качества кластеризации является эвристическим, завися от представлений экспертов о предметной области, оценивающих разумность выделения кластеров;
- число кластеров в общем случае априори неизвестно, устанавливаясь в соответствии с некоторым субъективным критерием меры близости оцениваемых объектов;
- результат кластеризации существенно зависит от метрики, выбор которой определяется экспертами, оценивающими поле пожарных рисков.

С учетом того, что множество рисков отличается, как правило, очень большой размерностью, задача типологизации является своеобразной «сверткой» исходных информационных таблиц. При этом число выделяемых типов объектов или явлений всегда меньше, чем уникальных единиц.

В результате получается лаконичное, логичное и наглядное представление типов территорий по пожарной обстановке в пространстве существенно меньшей размерности, позволяющее более точно обосновывать и принимать решения о предупреждении и эффективном тушении пожаров в зависимости от типа территории.

Задача кластерного анализа состоит в изучении по эмпирическим данным, каким образом территории «связываются» в «скопления» – кластеры, при этом никаких априорных предположений о структуре и количестве типов не производится. Таким образом, решается задача разбиения на типы с целью выделения групп однородных территорий, сходных между собой по характеристикам пожарной опасности, при существенном отличии этих групп друг от друга.

Большинство методов кластеризации [2] основывается на анализе квадратной и симметричной относительно главной диагонали матрицы коэффициентов сходства (расстояния, корреляции и т.д.) между объектами исходной матрицы наблюдений. Мерами дистанции могут служить: мера Минковского, мера расстояния по Евклиду, супремум-норма

или расстояние Чебышева, меры сходства Жаккара и Сьеренсена, коэффициент корреляции Пирсона и многие другие меры [3].

Кластерный анализ включает набор алгоритмов типологизации, группирующих данные в наглядные структуры-таксоны: иерархическая древовидная кластеризация, двухходовое объединение, метод  $K$  – средних и др. [4]. Наиболее часто используется иерархический алгоритм «Дендрограмма», версии которого отличаются правилами вычисления расстояний между кластерами.

Критерием корректности типологизации является устойчивость результата относительно выбора алгоритма кластерного анализа. Проверяют устойчивость, применяя несколько различающихся алгоритмов. Если результаты содержательно близки, то полученная типология корректна. В ином случае необходимо предположить, что задача кластерного анализа не имеет решения, и в реальности корректной типологии не существует.

Кроме иерархических методов типологизации получили распространение итерационные процедуры, с помощью которых пытаются найти наилучшее разбиение, ориентируясь на заданный критерий оптимизации, не строя при этом полного дерева (метод  $K$ -средних, алгоритмы «Форель», «Медиана», «Краб» и т.д.). В этом случае итерационный процесс начинается, как правило, со случайно выбранных кластеров, а затем путем вариации принадлежности объектов к различным кластерам решается двухаспектная задача:

- минимизации изменчивости внутри кластеров;
- максимизации изменчивости между кластерами.

В задачах анализа и прогнозирования пожарной обстановки весьма перспективно сочетание кластерного анализа с другими количественными методами (например, с корреляционным, регрессионным и факторным).

Уточняя формальную постановку задачи кластерного анализа провинций Вьетнама по пожарной опасности, отметим, что практически она заключается в том, чтобы на основании нормализованных данных, содержащихся в двумерной матрице размером  $n \times f$  (где  $n$  – число провинций,  $f$  – число социально-экономических, климатических характеристик, оперативно-служебных параметров противопожарных служб и иных показателей, связанных с пожарным риском), разбить множество провинций  $n$  на  $m$  кластеров (достаточно однородных подгрупп)  $Q_1, Q_2, \dots, Q_m$ , так, чтобы каждый объект принадлежал одной и только одной подгруппе разбиения. При этом провинции, принадлежащие одному и тому же кластеру, должны быть сходными, в то время как провинции, принадлежащие разным кластерам – разными.

Кластер имеет следующие математические характеристики: центр, радиус, среднеквадратическое отклонение, размер кластера. Определим их для четкого понимания методических аспектов решения задачи классификации провинций по пожарным рискам.

Центр кластера – это среднее геометрическое место точек в пространстве переменных.

Радиус кластера – максимальное расстояние точек от центра кластера.

Размер кластера может быть определен либо по радиусу кластера, либо по среднеквадратичному отклонению провинций для этого кластера.

Принято следующее правило – провинция относится к кластеру, если расстояние от объекта до центра кластера меньше радиуса кластера. Если это условие выполняется для двух и более кластеров, то объект является спорным. Очевидно, что такая неопределенность может быть устранена экспертом или аналитиком, хорошо разбирающимся в практических аспектах пожарных рисков в стране.

Наряду с необходимостью решения задачи предварительной стандартизации переменных, то есть приведения значений всех характеристик к единому диапазону значений, нередко решается задача определения коэффициента важности или веса, который отражает значимость соответствующей характеристики провинции. Как правило, в качестве весов выступают экспертные оценки, полученные в ходе опроса специалистов предметной области.

Авторы оперировали равнозначными коэффициентами по каждой из характеристик, а их стандартизация осуществлялась по формуле:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_j)^2}}, \quad (1)$$

где  $X_{ij}$  – представляет собой значение  $j$  характеристики в  $i$  провинции;  $I$  – число провинций, равное 63;  $J$  – общее число показателей, характеризующих состояние пожарной опасности;  $\bar{X}_j$  – среднее значение  $j$  характеристики пожарной опасности.

### Отбор показателей для типологизации провинций Вьетнама

Исходя из результатов анализа факторов, определяющих состояние пожарных рисков на территориях Вьетнама, для решения задачи их типологизации по состоянию пожарной опасности авторы опирались на следующую схему, отражающую факторный комплекс детерминации пожаров в стране (рис. 1).

В обоснование факторного комплекса легли результаты экспертных процедур по отбору практиками и научными работниками показателей, наиболее полно характеризующих пожарные риски в провинциях Вьетнама, а также результаты работ [5, 6].



Рис. 1. Факторный комплекс детерминации пожаров во Вьетнаме

Исследования авторов с использованием корреляционного анализа показали, что характеристики, детерминирующие пожарные риски в жилом секторе и секторе хозяйствующих субъектов различаются, пересекаясь в некоторой своей части. При этом рассмотрена динамика характеристик с 2006 по 2015 гг., а также их усредненный показатель за те же годы, по которому и производилась типологизация.

Применительно к жилому сектору при решении задачи типологизации методом кластеризации рассматривалась матрица размером 63 провинции, 27 характеристик; при решении той же задачи применительно к сектору хозяйствующих субъектов – матрица 63 провинции, 18 характеристик (табл. 1). Учитывая, что факторы, определяющие пожарную

опасность в названных секторах, во многом различаются, задача типологизации решалась раздельно применительно к каждому из них, затем решалась агрегированная задача.

Таблица 1. Показатели факторного комплекса пожарных рисков в секторе хозяйствующих субъектов

Номер показателя	Усредненные показатели по жилому сектору (2006–2015 гг.)
1	Средний ущерб на один пожар, \$ тыс.
2	Частота пожаров на один объект надзора, %
3	Доля объектов первого класса пожарной опасности, %
4	Доля объектов второго класса пожарной опасности, %
5	Доля объектов третьего класса пожарной опасности, %
6	Ежемесячный средний доход в госсекторе, \$
7	Доля городского населения, %
8	Количество погибших в среднем на один пожар, чел.
9	Средняя температура января, °С
10	Средняя температура июля, °С
11	Среднее количество осадков в январе, мм
12	Среднее количество осадков в июле, мм
13	Средний размер штрафа на объект надзора I класса, \$ тыс.
14	Средний размер штрафа на объект надзора II класса, \$ тыс.
15	Средний размер штрафа на объект надзора III класса, \$ тыс.
16	Среднее количество штрафов на объект надзора I класса
17	Среднее количество штрафов на объект надзора II класса
18	Среднее количество штрафов на объект надзора III класса

Рассмотрим сначала результаты типологизации территорий Вьетнама применительно к сектору хозяйствующих субъектов.

Общий алгоритм решения задачи типологизации состоял из девяти последовательных этапов (рис. 2), каждый из которых связан с предыдущими этапами обратными связями, позволяющими на любом из них производить корректировку алгоритма. При существовании приемлемой типологизации территорий по пожарной обстановке указанные на рис. 2 этапы могут повторяться необходимое число раз (с учетом мнений экспертов) для корректировки как меры расстояний между кластерами, методов кластеризации провинций, так и системы исходных показателей и, в приемлемых границах (объясненных с аналитической и практической точек зрения), конечного содержания кластеров.

Отметим, что в процессе реализации методов кластерного анализа в современной науке применяются два основных алгоритма обработки входных данных: сравнение объектов, исходя из признаков,  $Q$ -тип анализа; сравнение признаков, исходя из характеристик объектов –  $R$ -тип анализа. Авторами при построении типологии территорий Вьетнама применен гибридный тип анализа ( $RQ$ -анализ).

На этапе отбора показателей для устранения сильно связанных из них (дублирующих друг друга), характеризующих состояние пожарной опасности территорий Вьетнама, использован  $R$ -тип анализа. А именно, методами кластерного анализа в табл. 1 были найдены сильно связанные показатели (абсолютная величина коэффициента корреляции больше 0,8).

На рис. 3 применительно к табл. 1 показана метрика расстояний Чебышева между показателями с использованием метода медиан или взвешенного центроидного метода (weighted pair-group centroid method), позволяющего достаточно отчетливо выявить сильные связи.

Из дендрограммы, приведенной на рис. 3, следует, что показатели под номерами 16–18, а также показатели 13–15 являются сильно связанными (в табл. 1 эти две группы выделены цветом). Оставляя из каждой группы по одному показателю, размерность пространства, характеризующего пожарную опасность территорий для хозяйствующих субъектов, может быть снижена с 18 до 14. Экспертами из указанных сильно связанных

шести показателей были выбраны два – 14 – «Средний размер штрафа на объект надзора первого класса», «Среднее количество штрафов на объект надзора первого класса». В результате типологизация в окончательном виде производилась в пространстве практически не связанных, а также слабо связанных признаков (абсолютные значения коэффициентов корреляции не выше 0,4).



Рис. 2. Алгоритм решения задачи типологизации провинций Вьетнама по комплексным характеристикам пожарной опасности

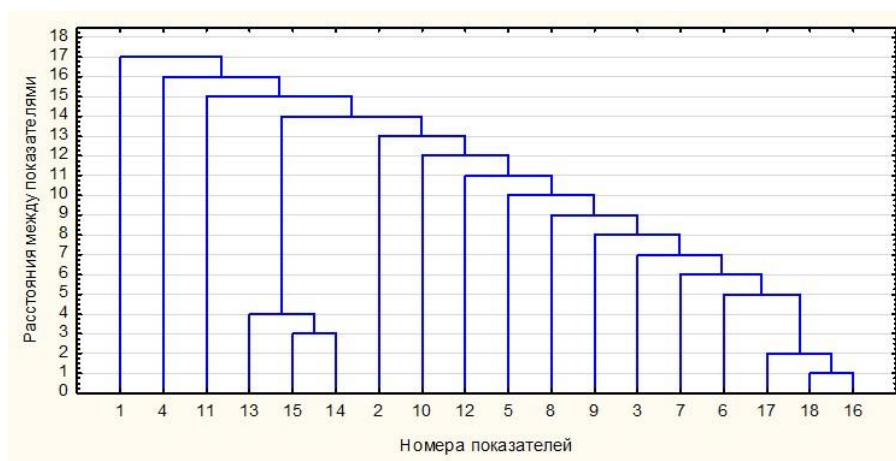


Рис. 3. Дендрограмма связи показателей пожарной опасности в провинциях Вьетнама по сектору хозяйствующих субъектов (взвешенный центроидный метод, метрика расстояния Чебышева)

В результате применительно к сектору хозяйствующих субъектов при решении задачи типологизации методом кластеризации рассматривалась матрица размером 63 провинции, 14 характеристик, являющаяся частью табл. 1.

Выбор меры расстояния между кластерами и метода кластеризации провинций осуществлялся путем перебора всех вариантов, предусмотренных в статистическом программном пакете Statistica 12. Затем с привлечением экспертов проводился анализ вариантов, среди которых выбирался тот, где выделенные кластеры подчинялись трем условиям:

- компактностью расположения на территории Вьетнама;
- схожестью социально-экономических условий и состояния оперативной обстановки по линии пожарной охраны, характеризующейся похожими характеристиками внешней среды её функционирования (плотность и демографические характеристики населения, климат, пожароопасность хозяйствующих субъектов и т.п. факторами);
- схожестью параметров сил и средств противопожарной службы.

Среди метрик расстояния между кластерами исследовались шесть включенных в указанный программный пакет вариантов:

- евклидово расстояние;
- квадрат евклидова расстояния;
- манхэттенское расстояние;
- расстояние Чебышева;
- расстояние Минковского;
- коэффициент корреляции Пирсона.

В качестве методов кластеризации применялись включенные в пакет иерархические (древовидные) процедуры кластерного анализа:

- правило одиночной связи (ближайшего соседа);
- правило полных связей (наиболее удаленных соседей);
- правило невзвешенного попарного среднего;
- правило взвешенного попарного среднего;
- невзвешенный центроидный метод;
- взвешенный центроидный метод;
- правило Варда (Ward).

Экспертам были представлены 15 вариантов, отражающих различные сочетания метрик расстояний и методов кластеризации. При этом явно нелогичные результаты типологизации территорий Вьетнама, с очевидностью неподчиняющиеся вышеназванным условиям экспертного анализа, были исключены из рассмотрения.

Наиболее логичной типологизацией территорий Вьетнама экспертами выбрана приведенная в табл. 2 и изображенная на рис. 4, 5.

**Таблица 2. Результаты типологизации территорий Вьетнама по состоянию пожарной опасности в секторе хозяйствующих субъектов**

Номер кластера (количество территорий)	Территории, вошедшие в кластер
1 (11)	29–39
2 (15)	1, 3, 5, 7, 12, 14–17, 20–25
3 (13)	2, 4, 6, 8, 9–11, 13, 18, 19, 26–28
4 (15)	45, 46, 51–59, 60–63
5 (9)	40–44, 47–50

Выбранная типология территорий осуществлена с помощью метода иерархического кластерного анализа – метода взвешенного попарного среднего (Weighted Pair-Group Method Using Arithmetic Averages – WPGMA), при этом метрикой выступала обратная величина коэффициента Пирсона, которую целесообразно применять для большого количества переменных.

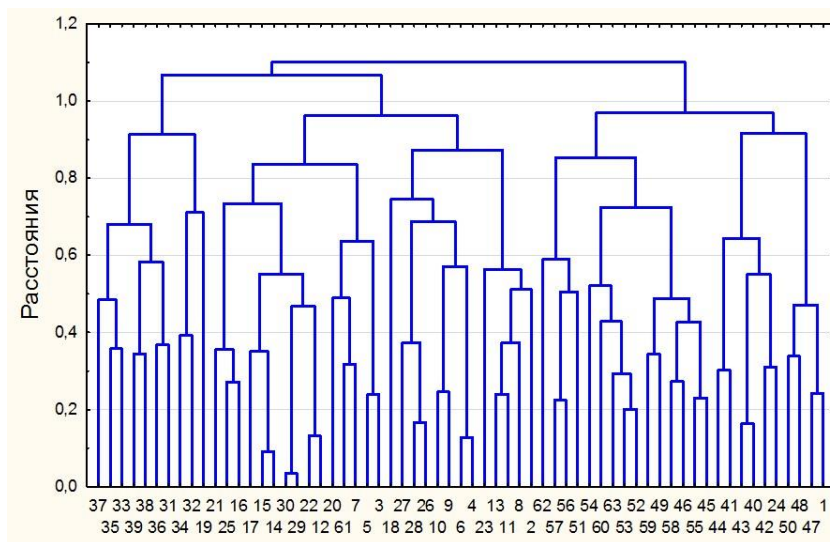


Рис. 4. Дендрограмма территорий Вьетнама по состоянию пожарной опасности в секторе хозяйствующих субъектов

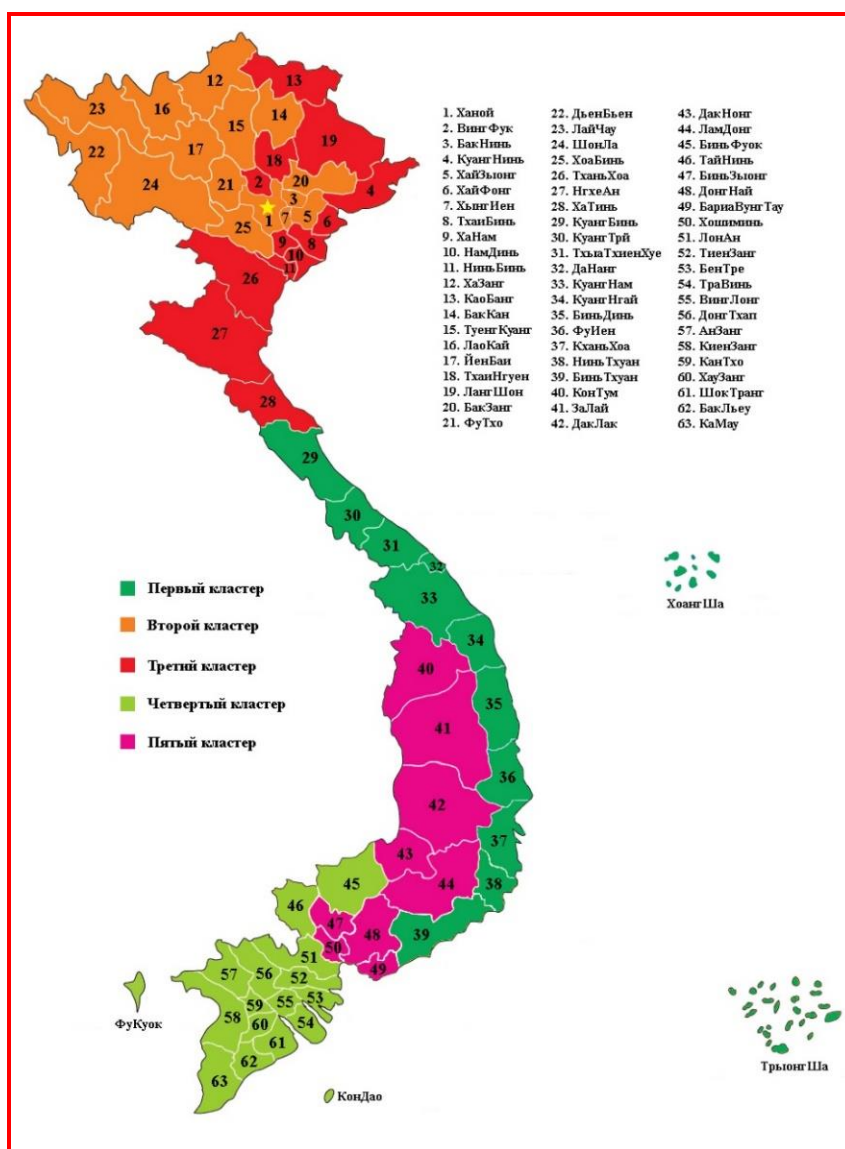


Рис. 5. Типология территорий Вьетнама по показателям пожарной опасности в секторе хозяйствующих субъектов



Забегаая вперед, укажем, что отобранная в настоящей статье система показателей применительно к сектору хозяйствующих субъектов оказалась весьма информативной, позволив содержательно и логически эффективно описать типологию пожарной опасности территорий Вьетнама.

### **Литература**

1. Минаев В.А., Топольский Н.Г., Дао Ань Туан. Проблемы и основные факторы оценки пожарных рисков во Вьетнаме // Интернет-журнал «Технологии техносферной безопасности». 2016. Вып. № 1 (65). 9 с.
2. Дюран Б., Оделл П. Кластерный анализ. М.: Статистика, 1977. 128 с.
3. Григорьев А.А. Меры сходства в кластеризации // Электронный научный журнал «Известия РЭУ им. Г.В. Плеханова». 2013. Вып. 11.
4. Многомерные статистические методы / Н.И. Гришакина [и др.]. Ч. IV: Кластерный анализ: учеб.-метод. пособие. В. Новгород: Изд-во Новгородского гос. ун-та им. Ярослава Мудрого, 2005. 54 с.
5. Минаев В.А., Тростянский С.Н., Чу Куок Минь. Оценка вероятности возникновения пожаров при нарушении требований пожарной безопасности // Интернет-журнал «Технологии техносферной безопасности». 2013. Вып. 5 (51).
6. Минаев В.А., Фаддеев А.О. «Медленные» катастрофы, здоровье и безопасность населения // Системы безопасности: материалы XV Науч.-техн. конф. М., 2006. С. 14–17.

### **References**

1. Minaev V.A., Topol'skij N.G., Dao An' Tuan. Problemy i osnovnye faktory ocenki pozharnyh riskov vo V'etname // Internet-zhurnal «Tekhnologii tekhnosfernoj bezopasnosti». 2016. Vyp. № 1 (65). 9 s.
2. Dyuran B., Odell P. Klasternyj analiz. M.: Statistika, 1977. 128 s.
3. Grigor'ev A.A. Mery skhodstva v klasterizacii // Ehlektronnyj nauchnyj zhurnal «Izvestiya REHU im. G.V. Plekhanova». 2013. Vyp. 11.
4. Mnogomernye statisticheskie metody / N.I. Grishakina [i dr.]. Ch. IV: Klasternyj analiz: ucheb.-metod. posobie. V. Novgorod: Izd-vo Novgorodskogo gos. un-ta im. Yaroslava Mudrogo, 2005. 54 s.
5. Minaev V.A., Trostyanskij S.N., Chu Kuok Min'. Ocenka veroyatnosti vznikoveniya pozharov pri narushenii trebovanij pozharnoj bezopasnosti // Internet-zhurnal «Tekhnologii tekhnosfernoj bezopasnosti». 2013. Vyp. 5 (51).
6. Minaev V.A., Faddeev A.O. «Medlennye» katastrofy, zdorov'e i bezopasnost' naseleniya // Sistemy bezopasnosti: materialy XV Nauch.-tekhn. konf. M., 2006. S. 14–17.