

УДК 614.87

## О подходе к кластеризации термоточек по результатам многолетних климатических наблюдений

ISSN 1996-8493

DOI:10.54234/CST.19968493.2022.19.4.74

© Технологии гражданской безопасности, 2022

**Е.В. Иванов, А.В. Дмитриев, В.Л. Шимитило, Д.С. Федоровичев**

### Аннотация

Статья посвящена исследованию структуры погодных условий, связанных с термоточками пожаров, методами кластерного анализа и является частью научного исследования авторов, посвященного предсказанию вероятности возникновения природных пожаров. Основным методом исследования является кластерный анализ. Выделение самой структуры происходит из непространственных признаков климатических наблюдений.

**Ключевые слова:** природный пожар; термоточки; кластерный анализ; модель гауссовой смеси; квадратичная энтропия Рао.

## On the Approach to the Thermal Points Clustering Based on the Results of Long-Term Climatic Observations

ISSN 1996-8493

DOI:10.54234/CST.19968493.2022.19.4.74

© Civil Security Technology, 2022

**E. Ivanov, A. Dmitriev, V. Shimitilo, D. Fedoroviche**

### Abstract

The article is devoted to the study of the weather conditions structure connected with the fires thermal points by cluster analysis methods and is the part of the authors' scientific research devoted to predicting the probability of wildfires. The main research method is cluster analysis. Isolation of the structure comes from non-spatial features of climatic observations.

**Key words:** natural fire; heat points; cluster analysis; Gaussian mixture model; quadratic entropy of Rao.

24.04.2022

В настоящее время вектор государственной политики в области обеспечения защиты населения и территорий от чрезвычайных ситуаций смещен в сторону снижения последствий чрезвычайных ситуаций за счет реализации превентивных, предупредительных мероприятий [1, 2]. Это в полной мере касается и вопросов минимизации ущерба от лесных (природных) пожаров, которые по размерам ущерба ежегодно в нашей стране становятся лидерами [3].

С точки зрения проведения предупредительных мероприятий для лесных (природных) пожаров важную роль будет играть определение состояния лесопожарной обстановки, которая может характеризоваться рядом количественных и качественных показателей [4].

В нашей статье предлагаются к рассмотрению результаты исследования влияния погодных условий на вероятность возникновения лесных (природных) пожаров на основе обработки статистических данных методом кластерного анализа.

Для нахождения такой зависимости предложено определение структуры непространственных признаков, характеризующих природный пожар, и выявление на основе полученных зависимостей их новой классификации.

В качестве субъекта Российской Федерации, для которого осуществлялось построение модели, выбран Красноярский край.

Подготовка исходных данных и их преобработка осуществлялись с применением стандартных инструментов Jupyter Notebook. Источником информации о положении и дате фиксации термоточек являлось «Озеро данных МЧС России».

Источником данных по метеопараметрам являлся набор данных, предоставляемый NCEP-DOE (National Centers for Environmental Prediction — Национальный центр климатических, океанических и атмосферных наблюдений). Это постоянно обновляемый (с 1948 г. по настоящее время) набор данных с привязкой к глобальной сетке, который представляет состояние атмосферы Земли, включая результаты наблюдений и модели численного прогнозирования погоды с 1948 года.

В исследовании для построения моделей отбирались только подтвержденные как лесные и природные пожары (категория) термоточки. С учетом довольно «слабого» уровня сети метеостанций для рассматриваемого субъекта погодные данные для участков, не охваченных сетью метеоконтроля, получены бикубической интерполяцией из набора данных NCEP-DOE Reanalysis 2, содержащего ежедневные метеорологические наблюдения, наложенные на регулярную сетку, начиная с 1979 года. Таким образом, был отобран набор данных, сопоставляющих в зафиксированной термоточке в 2020 году приближенное значение погодных параметров в ее окрестностях в день фиксации.

Из 56 признаков набора данных NCEP-DOE отобрано 8:

- 1) *skt* — температура поверхности;
- 2) *weasd* — водный эквивалент накопленной высоты снежного покрова (кг/м<sup>2</sup>);
- 3) *cprat* — интенсивность осадков конвекции;
- 4) *srfp* — атмосферное давление;

5) *soilw* — среднемесячная объемная влажность почвы (кг/м<sup>2</sup>);

6) *shum* — удельная влажность;

7) *runof* — накопленный среднесуточный поверхностный сток воды за продолжительный период;

8) *srfpt* — потенциальная температура поверхности.

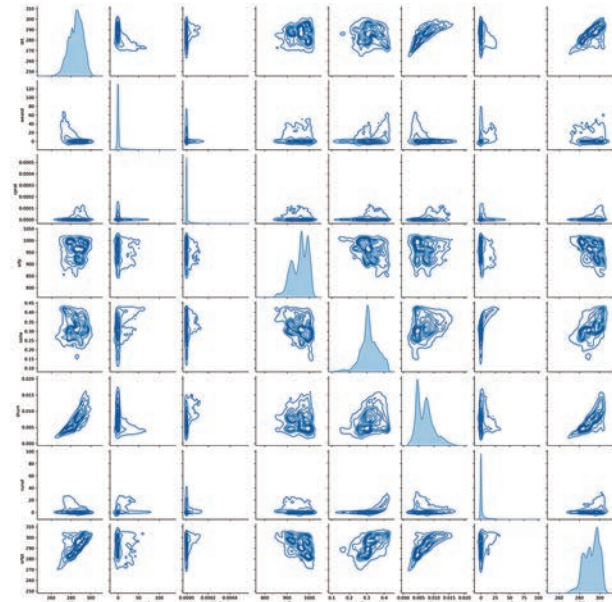


Рис. 1. Плотность распределений пар погодных признаков в используемом наборе данных. Признаки слева направо (сверху вниз): *skt*, *weasd*, *cprat*, *srfp*, *soilw*, *shum*, *runof*, *srfpt*

Полученный набор данных содержит 83 833 наблюдения, оценка плотности вероятности для пар признаков представлена на рис. 1.

Разнородность используемых данных накладывает ряд ограничений на допустимые методы кластерного анализа. Чем выше размерность данных, тем сложнее применение к ним методов, основанных на расстоянии [5, 6]; поэтому для достижения наилучшего разбиения следует понизить размерность признакового представления наблюдений, исключив признаки, не вносящие дополнительной информации для разделения (имеют единственный пик на графике плотностей вероятности на диагонали (рис. 1). Дополнительные сложности создает различие в масштабе признаков, для решения проблемы следует применить нормировку признаков либо использовать плотностные или параметрические методы кластеризации. По этой причине для проведения кластеризации был выбран метод Гауссовской модели смесей, основанный на оценке параметров многомерного нормального распределения.

Следующей задачей для анализа является линейная неразделимость исходных данных. Проблема возникает ввиду неочевидности формирования кластеров по фиксируемой плотности наблюдений в них. Из рис. 1 видно, что такие сравнения признаков друг с другом содержат пики плотности распределения различной формы. Решением проблемы стало построение кластеров на данных, из которых исключены области с наименьшей плотностью. Для выделения пиков проведена оценка функции плотности вероятности наблюдений для каждого

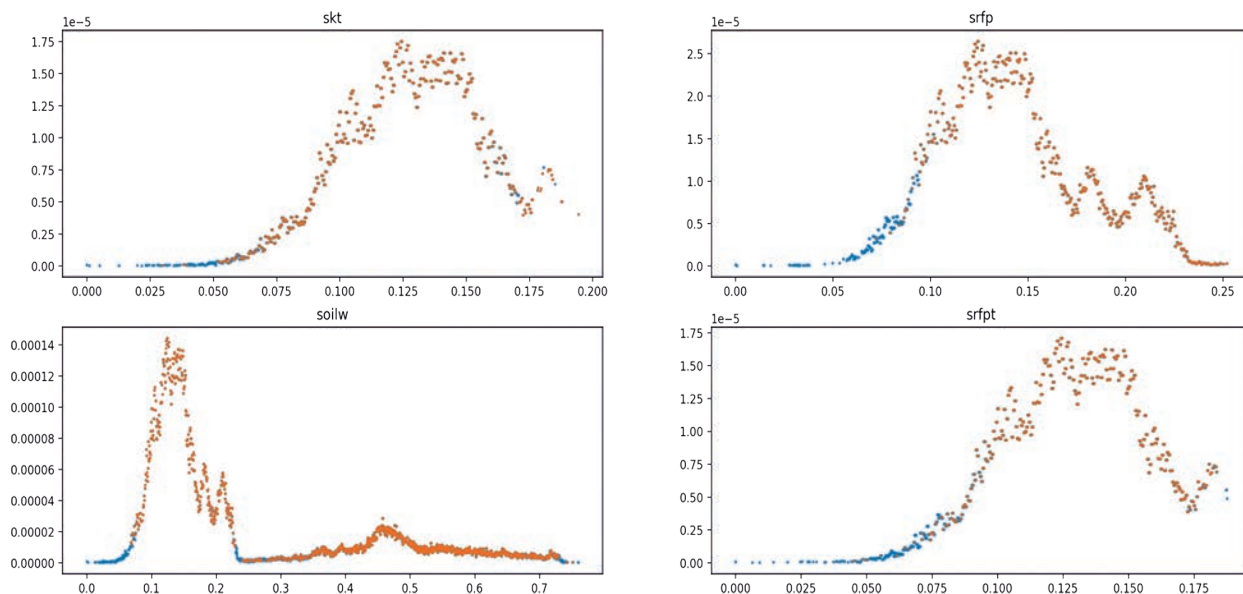


Рис. 2. Выбор наблюдений по усредненному значению оценки плотности вероятности их нормированных значений

признака, а при усредненных значениях — для признака в точке наблюдения. Наблюдения, среднее значение плотности вероятности которых по всем признакам ниже критического уровня  $1,2 \cdot 10^{-5}$ , в дальнейшем анализе не использовались. Выбор наблюдений по плотности проиллюстрирован на рис. 2. Исключению из выборки подверглись также те данные, которые имели достаточно большой разброс в значениях [7]. Для оценки плотности проведена нормировка значений параметров. После исключения областей с наименьшей плотностью набор данных содержит информацию о 38 269 наблюдениях.

Исходя из вида распределения результатов наблюдений (рис. 2), для четырех отобранных параметров (*skt* — температура поверхности; *soilw* — среднемесячная объемная влажность почвы; *srfp* — атмосферное давление; *srfpt* — потенциальная температура поверхности) можно сделать вывод о том, что они представляют собой смесь нормальных Гауссовских распределений.

То есть в дальнейшем будет рассматриваться статистическая модель для представления нормально распределенных подмножеств (двумерные пространства на рис. 1) внутри общего множества  $n$ -мерного пространства признаков.

Оптимальное разделение на кластеры признакового пространства осуществляется на основе минимизации функции энтропии. По сути решается задача минимизации «расстояний» между точками одного кластера с максимизацией «расстояний» между центрами соседних кластеров.

Зная количество нормальных распределений в смеси, пользуясь аппаратом математической статистики, определяем параметры распределения случайных величин (дисперсия, математическое ожидание). На этом принципе основывается вероятностная модель кластеризации — Гауссовская модель смесей (GMM — англ. Gaussian Mixture Model). Регулирующим параметром для поиска оптимального разбиения (гиперпараметра) в GMM является количество кластеров, описываемых  $n$ -мерными нормальными распределениями.

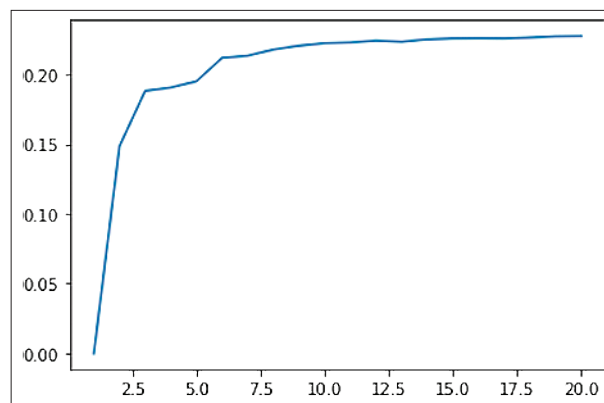


Рис. 3. Оценка энтропии для построенной модели

Для выбора оптимального значения этого гиперпараметра алгоритма производится линейный поиск с помощью квадратичной энтропии Рао [8, 9] для оценки качества итоговой кластеризации (формула 1). Для каждого значения числа кластеров в диапазоне от 1 до 20 строится разбиение и по нему оценивается значение энтропии. Эта метрика позволяет выбрать значение, дающее сбалансированные кластеры, при которых ни один кластер не будет преобладать над остальными, с минимальным внутрикластерным и максимальным межкластерным расстоянием [10]. Полученная оценка энтропии представлена на рис. 3. В результате поиска будет получено количество кластеров, на которое следует разбить наблюдения, начиная с которого дальнейшее разбиение будет избыточным.

$$E = \sum_{i=1}^{S-1} \sum_{j=i+1}^S d_{ij} p_i p_j.$$

Формула 1. Квадратичная энтропия Рао, где  $S$  — количество кластеров;  $d_{ij}$  — метрика расстояния между  $i$ -ым и  $j$ -ым кластерами;  $p_i$  — априорная вероятность принадлежности произвольного наблюдения к  $i$ -ому кластеру в разбиении.

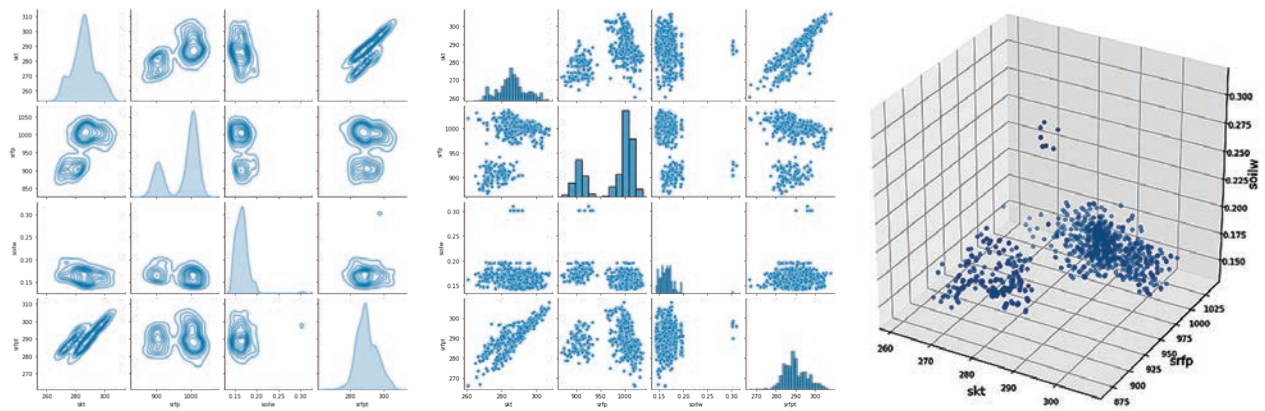


Рис. 4. (а) График оценки плотности распределений; (б) график пар наблюдений признаков; (в) график наблюдения в пространстве признаков skt, soilw, srfp

На графике результат изменений прекращает резкий рост энтропии после разбиения на три кластера. Это значение в дальнейшем выбирается в качестве оптимального для проведения кластеризации. Вывод об избыточности разбиения более чем на три кластера согласуется с интуитивным выводом, полученным при изучении модифицированного набора данных, представленных на рис. 4.

Итоговое разбиение, полученное в Гауссовской модели смесей для трех кластеров, представлено на рис. 5.

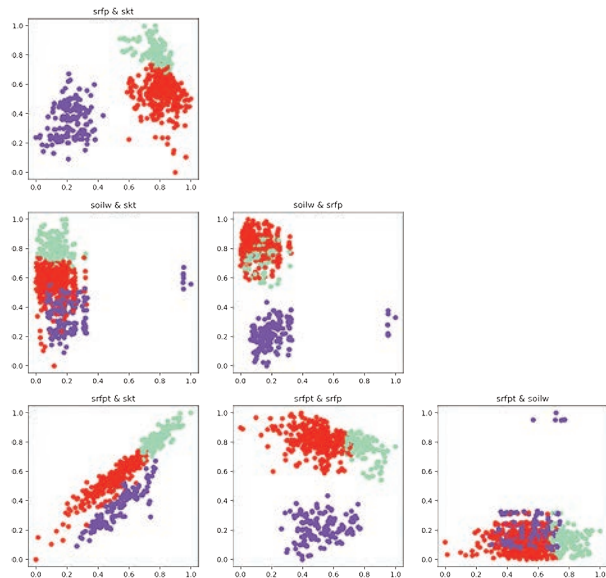


Рис. 5. Визуальное представление разбиения: кластеры выделены цветами на графиках пар значений параметров

Полученная кластеризация показывает возможность использования отобранных погодных признаков для группировки природных пожаров. Кроме представленного примера, раскрываемый в работе алгоритм может найти применение для решения других задач факторного анализа в исследуемой области. В частности, перспективным выглядит направление оценки влияния на вероятность возникновения природных пожаров факторов антропогенной нагрузки.

Таким образом, проведенный отбор факторов по основным группам, влияющим на пожарную обстановку,

перечисленным в работе [4], позволит сформировать математическую модель оценки вероятности возникновения природных пожаров, обладающую достаточной точностью.

Результаты могут быть использованы в дальнейших разработках по классификации пожаров, а значит и для разработки сценариев их предотвращения и реагирования на них.

### Литература

1. Волокитина А. В. Защита населенных пунктов от природных пожаров / А. В. Волокитина, Т. М. Софронова // Пожаровзрывобезопасность. 2011. Т. 20. № 3. С. 22–31.
2. Андреев Ю. А. Планирование и проведение мероприятий по профилактике природных пожаров в Алтае-Саянском экорегионе на основе оценки уровня антропогенной пожарной опасности / Ю. А. Андреев, А. Ю. Андреев, А. В. Брюханов // Сибирский пожарно-спасательный вестник. 2018. № 3(10). С. 38–46.
3. Арефьева Е. В. О подходе к определению расчетных индексов подверженности природным пожарам территорий пилотного субъекта РФ в рамках технологии дистанционной оценки риска / Е. В. Арефьева, В. В. Артюхин, А. С. Котосонов // Моделирование сложных процессов и систем: Сб. трудов секции № 10 XXXII Международной научно-практической конференции. Химки, 1 марта 2022 года. Химки: АГЗ МЧС России, 2022. С. 74–80.
4. Мазаник А. И. Модель оценки влияния метеорологических параметров на уровень природной пожарной опасности / А. И. Мазаник, А. В. Рыбаков, Р. Л. Белоусов, А. И. Араштаев // Научные и образовательные проблемы гражданской защиты. 2020. № 2(45). С. 12–24.
5. Рыбаков А. В. Разработка научно обоснованных моделей прогнозирования природных пожаров на основе данных дистанционного зондирования земли / А. В. Рыбаков, А. В. Дмитриев, А. Н. Тимарин, Е. В. Иванов, Э. К. Фукс // Сб. трудов Конкурса научно-исследовательских работ (Конкурса НИР). Материалы Молодежной программы 25-й Международной специализированной выставки и Форума «Безопасность и охрана труда» БИОТ-2021. М.: 2021. С. 109–112.
6. A Survey of Distance Metrics in Clustering Data Mining Techniques / M. Mercioni, S. Holban. DOI 10.1145/3338472.3338490 // ICGSP '19: Proceedings of the 2019 3rd International Conference on Graphics and Signal Processing. 2019. P. 44–47.
7. Рыбаков А. В. Проблема применения больших данных в интересах выполнения задач, стоящих перед МЧС России / А. В. Рыбаков, Е. В. Иванов, А. В. Дмитриев, В. С. Сидоров // Сибирский пожарно-спасательный вестник. 2021. № 2. С. 54–57.
8. Rai P. A Survey of Clustering Techniques / Pradeep Rai, Singh Shubha. DOI 10.5120/1326–1808 // International Journal of Computer Applications. 2018. № 12. P. 1–5.
9. Xu D. A Comprehensive Survey of Clustering Algorithms / D. Xu, Y. Tian. DOI 10.1007/s40745–015–0040–1 // Ann. Data. Sci. 2015. Vol 2. P. 165–193.
10. Kingrani S. Estimating the number of clusters using diversity / Suneel Kingrani, Mark Levene, Dell Zhang. DOI 10.5430/air.v7n1p15 // Artificial Intelligence Research. 2017.

**Сведения об авторах**

**Иванов Евгений Вячеславович:** к.т.н., АГЗ МЧС России, доц. каф. аварийно-спасат. работ. Химки, Россия. SPIN-код: 5470-1533.

**Дмитриев Антон Викторович:** АГЗ МЧС России, н. с. науч.-исслед. отд. (по проблемам ГО и ЧС). Химки, Россия. SPIN-код: 7743-7402.

**Шимитило Виктория Леонидовна:** к.т.н., АГЗ МЧС России, доц. каф. высш. математ. Химки, Россия. SPIN-код: 4525-8387.

**Федоровичев Дмитрий Станиславович:** АГЗ МЧС России, оператор науч.-исслед. отд. (по пробл. ГО и ЧС). Химки, Россия.

**Information about authors**

**Ivanov Evgeny V.:** PhD (Technical Sc.), Civil Defence Academy EMERCOM of Russia, Associate Professor of the Department of Emergency and Rescue Works. Khimki, Russia. SPIN-scientific: 5470-1533.

**Dmitriev Anton V.:** Civil Defence Academy EMERCOM of Russia, Researcher of the Research Department (for Civil Defense and Emergency Situations). Khimki, Russia. SPIN-scientific: 7743-7402.

**Shimitilo Victoria L.:** PhD (Technical Sc.), Civil Defence Academy EMERCOM of Russia, Associate Professor of the Department of Higher Mathematics. Khimki, Russia. SPIN-scientific: 4525-8387.

**Fedorovichев Dmitry S.:** Civil Defence Academy EMERCOM of Russia, Operator of the Research Department (for Civil Defense and Emergency Situations). Khimki, Russia.

**Издания ФГБУ ВНИИ ГОЧС (ФЦ)**

<b>Авторы, название</b>	<b>URL</b>
Наука и технологии обеспечения безопасности в чрезвычайных ситуациях: Материалы мероприятий деловой части программы, организатором которых выступил ФГБУ ВНИИ ГОЧС (ФЦ), Часть I.	<a href="https://elibrary.ru/item.asp?id=47150447">https://elibrary.ru/item.asp?id=47150447</a>
Наука и технологии обеспечения безопасности в чрезвычайных ситуациях: Материалы мероприятий деловой части программы, организатором которых выступил ФГБУ ВНИИ ГОЧС (ФЦ), Часть II.	<a href="https://elibrary.ru/item.asp?id=47233683">https://elibrary.ru/item.asp?id=47233683</a>
Совершенствование защиты населения, материальных и культурных ценностей от опасностей, возникающих при военных конфликтах или вследствие этих конфликтов с учётом современных угроз: Материалы научно-практической конференции.	<a href="https://elibrary.ru/item.asp?id=47152118">https://elibrary.ru/item.asp?id=47152118</a>
<i>Мошков В.Б. и др.</i> Добровольная сертификация как инструмент повышения качества аварийно-спасательных средств.	<a href="https://elibrary.ru/item.asp?id=47276629">https://elibrary.ru/item.asp?id=47276629</a>
<i>Акимов В.А. и др.</i> Исследование чрезвычайных ситуаций природного, техногенного и биолого-социального характера современными научными методами.	<a href="https://elibrary.ru/item.asp?id=47429574">https://elibrary.ru/item.asp?id=47429574</a>
<i>Азанов С.Н. и др.</i> Перспективы развития единой государственной системы предупреждения и ликвидации чрезвычайных ситуаций.	<a href="https://elibrary.ru/item.asp?id=47436936">https://elibrary.ru/item.asp?id=47436936</a>
<i>Мошков В.Б. и др.</i> I Международная научно-практическая конференция по развитию робототехники в области обеспечения безопасности жизнедеятельности «RoboEmercom»: Сборник материалов конференции/	<a href="https://elibrary.ru/item.asp?id=47781245">https://elibrary.ru/item.asp?id=47781245</a>
<i>Виноградов О.В. и др.</i> Основы прогнозирования параметров опасных зон для населения, проживающего вблизи районов размещения полигонов твердых бытовых отходов.	<a href="https://elibrary.ru/item.asp?id=48170922">https://elibrary.ru/item.asp?id=48170922</a>
<i>Козлов И.А. и др.</i> Индикация в различных природных средах агрессивных и радиоактивных химических веществ и методы их утилизации и конверсии.	<a href="https://elibrary.ru/item.asp?id=48176202">https://elibrary.ru/item.asp?id=48176202</a>
<i>Мануйло О.Л.</i> Творчество юных во имя спасения: Литературно-художественный альманах, т. Вып. № 6.	<a href="https://elibrary.ru/item.asp?id=48506682">https://elibrary.ru/item.asp?id=48506682</a>
<i>Качанов С.А. и др.</i> Глобальная и национальные стратегии управления рисками катастроф и стихийных бедствий: Материалы конгресса.	<a href="https://elibrary.ru/item.asp?id=48698036">https://elibrary.ru/item.asp?id=48698036</a>