

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 7, №6 (2015) <http://naukovedenie.ru/index.php?p=vol7-6>

URL статьи: <http://naukovedenie.ru/PDF/74TVN615.pdf>

DOI: 10.15862/74TVN615 (<http://dx.doi.org/10.15862/74TVN615>)

УДК 519.67

Ботыгин Игорь Александрович

ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»

Россия, Томск¹

Доцент кафедры «Информатики и проектирования систем»

Кандидат технических наук

E-mail: bia@tpu.ru

Катаев Сергей Григорьевич

ФГБОУ ВО «Томский государственный педагогический университет»

Россия, Томск

Профессор кафедры «Общей физики»

Доктор технических наук

E-mail: sgkataev@sibmail.com

Шерстнёв Владислав Станиславович

ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»

Россия, Томск

Доцент кафедры «Вычислительной техники»

Кандидат технических наук

E-mail: vss@tpu.ru

Шерстнёва Анна Игоревна

ФГАОУ ВО «Национальный исследовательский Томский политехнический университет»

Россия, Томск

Доцент кафедры «Высшей математики»

Кандидат физико-математических наук

E-mail: sherstneva@tpu.ru

Методы классификации и анализа климатических полей

¹ 634034, Россия, г. Томск, ул. Советская, д. 84/3, оф. 406

Аннотация. В статье проведен сравнительный анализ методов нахождения закономерностей поведения во времени поля метеопараметра или иной характеристики на территориях разного масштаба. Показано, что выявление научно обоснованных глубинных тенденций изменения геофизического поля по исследуемой территории сопровождается резким увеличением объема данных, что требует не только использования современных информационных технологий хранения и обработки больших данных, но и проектирования новых подходов к их обработке. Отмечена перспективность методов и алгоритмов обнаружения закономерностей в экспериментальных данных на основе распределения объектов на группы. Вследствие многообразия систем признаков, по которым может определяться схожесть объектов, утверждается невозможность унификации процедур группировки объектов, так как все определяется спецификой решаемой задачи, авторским формированием классификационной системы критериев и их авторской смысловой нагрузкой. Основываясь на широко распространенном понимании выделенной группы объектов как подмножества, объединяющего объекты в некоторый единый абстрактный класс только на основе положения в пространстве признаков, сформулированы основные этапы, характеризующиеся индивидуальными задачами выбора и оптимизации, многомерной статистической процедуры, направленной на сбор данных о выборке объектов и упорядочения объектов в сравнительно однородные группы.

Ключевые слова: кластерный анализ; временной ряд; кластерная классификация; факторный анализ; кластеризация; группировка; геофизическое поле; экспериментальные данные.

Ссылка для цитирования этой статьи:

Ботыгин И.А., Катаев С.Г., Шерстнёв В.С., Шерстнёва А.И. Методы классификации и анализа климатических полей // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 7, №6 (2015) <http://naukovedenie.ru/PDF/74TVN615.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ. DOI: 10.15862/74TVN615

Статья опубликована 25.11.2015.

Интерес к изучению геофизических полей, к которым относят как естественные, так и искусственные физические поля планеты Земля, и, особенно, к их трансформации – не иссекаем. Но именно проблема изменения климата, обусловленная перестройкой геосистем, в том числе и антропогенным воздействием, в настоящее время очень актуальна [1-7].

Основная форма представления такого рода данных – это временные ряды, заданные в определенных пространственных точках. Все методы анализа многомерных геофизических данных направлены на достижения главной цели – поиска в этих данных тех или иных регулярностей, проявляющихся в существовании явных или неявных структур. К классическим методам многомерной статистики обычно относят: метод главных компонент, факторный анализ, корреляционный анализ, дискриминантный анализ, кластерный анализ, многомерное шкалирование. Выбор сочетания методов исследования зависит от целей исследования, природы данных и наличия априорной информации о возможных связях.

Анализ структуры климатических полей (климатическое районирование, выделение классов и т.п.), с одной стороны, направлено на анализ закономерностей формирования различных типов климата в глобальной климатической системе. С другой стороны, определение территориальных границ типов климата, различных по своим свойствам, позволяет организовывать огромное количество информации о климатических параметрах в гораздо меньшее число информационных структур с целью использования полученных результатов в хозяйственно-экономических мероприятиях и при моделировании климатических ситуаций. Таким образом, научная и практическая значимость любой климатической структуризации бесспорна. Чем больший региональный уклон имеют подобные исследования, тем более высокую социально-экономическую эффективность может нести полученная информация для конкретных отраслей. Если задача решается для исследования генезиса климата местности, основываясь на всем комплексе климатических условий соответствующих ландшафтных зон, то говорят о климатической классификации. Если выделение структур в полях элементов климата проводят для прикладных целей, то данную процедуру называют климатическим районированием.

Широко известным классическим классификациям климата присуща значительная доля субъективизма. При выделении климатических типов, зон, районов помимо, непосредственно, температурно-влажностных характеристик учитывается преобладание над территорией соответствующих типов воздушных масс по сезонам года и особенности их циркуляции, степень континентальности климата, характер подстилающей поверхности. Так, классы в одной из наиболее используемой в мире классификации климата В.П. Кеппена [8] выделяются на основе количественных критериев тепло-влажностного режима с учетом ландшафтных особенностей территории. Классификация климатов Л.С. Берга [3] основана на учете ландшафтно-географических зон суши. Границы климатических зон в генетической классификации Б.П. Алисова [3] определяются по среднему положению климатических фронтов, то есть в основе этой классификации учет условий формирования климатов в зависимости от типов воздушных масс и их циркуляции. На основе градаций характеристик тепло- и влагообеспеченности приземного воздуха и учета параметров теплового баланса деятельной поверхности построена классификация климатических режимов в работе [9]. Современные классификации предлагают более математизированные подходы, основанные, например: на учете вклада каждого влияющего фактора введением весов [10], на применении иерархического кластерного анализа данных температуры и осадков [11] и др.

В работах, посвященных нахождению закономерностей поведения во времени поля метеопараметра или иной характеристики на территориях разного масштаба, обычно исследуется поведение тренда в выбранном ареале, поскольку именно данная характеристика

дает возможность осуществлять прогнозирование. В этих задачах широко применяются кластерный и факторный анализы [12-18].

В ряде работ дополнительно исследуется структура временных рядов, под которой понимается ее характерные особенности, сформулированные в сжатом виде. Так, при исследовании периодических рядов среднемесячных температур (общего содержания озона, осадков и др.) [19-23] в качестве таких параметров выступали величины: мера разброса (дисперсия), среднее значение, фаза и др., характеризующие составляющие временного ряда: тренда, сезонной и нерегулярной. Используя этот набор данных можно с хорошей точностью, вернее, с той точностью, которую допускают данные, описать временные ряды и дать статистический прогноз его поведения.

Пространственно-временные связи между различными полями метеопараметров обычно изучают с использованием корреляционного анализа, который позволяет определять и временные лаги. Канонический корреляционный анализ в линейной и нелинейной формах широко используется в климатологии, в частности, для сезонного прогноза [24, 25], анализа структуры колебаний Эль-Ниньо [26], определения среднеширотного атмосферного отклика на вариации приповерхностной температуры Тихого океана [26, 27].

При решении задачи прогнозирования климата, также, используются разнообразные подходы, основанные на применении методов регрессионного, корреляционного анализа [28-34], метода главных компонент (эмпирических ортогональных функций) и др. [35-43].

Методы нахождения латентных структур временных рядов различного происхождения и поведения, включая хаотические ряды, развивается в разных направлениях. Популярность в последнее время приобрел метод сингулярного спектрального анализа [44, 45]. В этом методе исходный ряд представляется в виде взвешенной суммы членов, а последующий анализ позволяет понять, какому процессу соответствует то или иное слагаемое, весовой коэффициент указывает на вклад процесса. Этот метод дает: возможность выделять различные составляющие временных рядов, находить заранее неизвестные периодичности ряда, наилучшим образом определять компонент с заранее известным периодом. Если таковых не имеется, то мы имеем дело с хаотическими временными рядами, у которых также может иметься определенная структура. Большую популярность в последнее время получили изучающие такие ряды методы, основанные на фрактальной идее (самоподобие) [46, 47]. Так, например, фрактальные свойства временных рядов наблюдаются при измерениях таких естественных процессов как солнечная активность [48], уровень разливов рек [49], геофизическая и геомагнитная активность [50], физиологические характеристики организма человека [51]. При этом очень важно понять механизм возникновения фрактальности, особенно, в отличающихся естественных системах [52].

Среди методов многомерной статистики именно кластеризация чаще всего применяется для обработки данных во многих природно-климатических исследованиях. Заметим, что общепринятого устоявшегося определения понятий кластер или кластеризация нет, но вкладываемый в них смысл различными исследователями, примерно, один и тот же – разбиение множества объектов на классы, обладающих общим пространством признаков. Для каждого кластера вычисляются статистические и структурные характеристики. Именно структурные характеристики (параметры) дают возможность количественного сравнения различных кластеров друг с другом и сформулировать функционал качества всего разбиения.

Результат применения кластерного анализа – набор пространственных областей (кластеров), обладающих похожим поведением изучаемого параметра или целого набора параметров. В частности, в [53] иерархический кластерный анализ был использован для систематизации средних значений полей гидротермического коэффициента Селянинова и

было получено существенное сжатие данных (температур и атмосферных осадков) за период наблюдений с 1960 по 2010 годы. Иерархическая кластеризация использовалась также при решении задачи прогноза облачности [54]. Эксперименты проводились с разным числом кластеров и разными признаками. В результате было выделено 7 кластеров. В работе [55] иерархическая кластеризация использовалась для решения задачи климатического районирования. Исходные данные были взяты для 35 метеостанций по 22 климатическим показателям. В исследованиях по дистанционному зондированию Земли [56] также был успешно использован иерархический (дивизимный) алгоритм кластеризации.

Кластеризация по методу k -средних была применена для решения задачи пространственной декомпозиции метеорологических полей Северного полушария [57]. Для исследования были использованы данные из базы FAOCLIM-2. Полученные для кластеризации итоговые данные составили порядка 100000 значений годовых осадков и около 5000 среднегодовых температур.

В [58] для исследования был использован один из методов кластеризации – метод динамических ядер. Метод является разновидностью k -средних. С его помощью были выделены 4 кластера – географо-климатических варианта светлохвойных травяных лесов (подтайги). Другой алгоритм кластеризации «SNN» был использован в [59] для исследования окружающей природной среды. Полный объем использованного архива составил 2.9 Терабайта (по состоянию на август 2009 г.). Архив содержит более 80 различных переменных (включая атмосферное давление на уровне моря, влажность воздуха, солнечную радиацию) в нескольких координатных системах с 1948 г. по настоящее время. К ограничениям этого подхода можно отнести кластеризацию только по одной переменной.

Помимо перечисленных выше методов кластеризации, применённых для обработки природно-климатических данных, существуют и другие методы, использованные для решения задач рассматриваемой предметной области. Эти методы основаны на применении искусственных нейронных сетей. Например, нейронная самоорганизующаяся сеть Кохонена была использована для кластеризации при решении задачи классификации метеорологических сценариев [60]. С её помощью были выявлены базовые метеосценарии, использованные далее в задаче прогнозирования электропотребления. В работе [61] для кластеризации и классификации различных объектов пониженной интенсивности на изображениях Солнца были проанализированы данные о магнитном поле в основании областей пониженной интенсивности. Использовались деревья решений и нейронные сети Кохонена. Было выявлено, что сходство объектов обнаруженных кластеров может говорить о том, что данные области могут являться источниками медленного солнечного ветра со схожими характеристиками.

Таким образом, кластеризация является методом, интенсивно применяемым при анализе природно-климатических данных. Существует ещё много разновидностей методов кластеризации числовых данных, которые обладают как своими достоинствами, так и недостатками [62]. Тем не менее, у большинства из них есть общая черта – необходимость задавать количество кластеров до начала кластеризации, что существенно затрудняет обработку «сырых» данных и сужает возможности кластеризации. Поэтому для получения объективной информации требуются методы, в которых не нужно указывать количество кластеров. Такую возможность предоставляют иерархические методы (агломеративные и дивизимные), а из неиерархических – алгоритм CLOPE. Особенностью иерархических алгоритмов является обеспечение сравнительно высокого качества кластеризации. При этом не требуется предварительное задание количества кластеров. Сложность таких алгоритмов порядка $O(n^2)$ [63].

Основной проблемой иерархических методов кластеризации является сложность определения условия остановки. Остановиться нужно так, чтобы выделить «естественные» кластеры и не допустить их разбиения. Следующей проблемой этих методов является выбор точки разделения или слияния кластеров. Неверный выбор этой точки на каком-либо шаге приводит к некачественным результатам. И, главное, важно для обработки погодноклиматических данных, иерархические методы не применяются при обработке больших наборов данных [68], так как получение решения о разделении или слиянии кластеров приводит к большой вычислительной сложности метода.

Неиерархический алгоритм CLOPE предназначен для кластеризации больших наборов категориальных данных, а метеолюбные имеют числовой характер. Соответственно, этот метод неприменим для решения поставленных в исследовании задач.

Перспективен подход, основанный на применении для кластеризации самоорганизующихся нейронных сетей Кохонена. Обучающиеся модели, действительно, легко перенастраивать при использовании любого количества независимых и зависимых признаков, поступлении любого количества новых данных, обеспечения параллельного решения нескольких задач классификации или прогноза [67]. Тем не менее, сети Кохонена обладают тем же недостатком, что и все неиерархические алгоритмы кластеризации – необходимо знать количество кластеров до выполнения процедуры кластеризации.

Аналитика обработки временных рядов очень сильно зависит от временного интервала, по которому проводится их усреднение. Для построения обобщенной картины эволюции того или иного геофизического параметра, как правило, подходят достаточно длинные интервалы усреднения (например, среднемесячные значения). Однако в настоящее время появляются задачи, требующие детализации, связанной с переходом на уровень среднесуточных значений, а иногда и к более мелкому масштабу, так как необходимо выявить тонкие тенденции изменения геофизического поля по исследуемой территории. Такой переход сопровождается резким увеличением объема данных, что требует не только использования современных информационных технологий хранения и обработки больших данных, но и разработки новых подходов к их обработке.

И здесь, в поле зрения попадают, прежде всего, методы и алгоритмы многомерной статистики. Это и факторные, и дискриминантные, и кластерные методы и алгоритмы. Но кластерный анализ и кластерные методы классификации – самые востребованные. Сформулируем основные этапы многомерной статистической процедуры, направленной на сбор данных о выборке объектов, и упорядочения объектов в сравнительно однородные группы (кластерный анализ).

Принципы отбора системы признаков. Расширение пространства признаков позволяет получить и большее количество информации о структуре исследуемой системы. Но увеличение количества независимых факторов ведет и к увеличению классов реализации. Решение заключается в факторной оптимизации, т.е. в сведении до минимума количества значимых факторов (признаков) без потери информации об исследуемой системе. Так, например, в [68] для этих целей предложена трансформация исходных признаков в новые показатели, обладающие большей информативностью.

Выбор способа вычисления расстояния между объектами и группами. Мера близости (чаще всего выражается в виде скалярной величины) показывает количественную меру схожести, как правило, двух объектов. Чем меньше расстояние между объектами (в пространстве признаков), тем выше их схожесть. Заметим, что используемый исследователем метод кластеризации фактически определяет и способ вычисления расстояния между объектами. Расстояния могут быть и настоящими или некоторыми другими производными

мерами расстояния и задача исследователя заключается в выборе адекватного метода вычисления, зависящего от специфики применения и формы получаемых кластеров (гиперсфера, эллипс и т.д.).

Группировка объектов. Именно при группировке вся исследуемая совокупность объектов разделяется на группы по какому-то существенному признаку. При этом, разбиение исходного множества объектов может осуществляться по детерминированному, жестко заданному алгоритму или определяться вероятностной (нечеткой) характеристикой степени принадлежности каждого объекта к группам. Еще раз отметим, что результат группировки в значительной степени зависит от выбранной исследователем метрики в системе значимых факторов.

Представление результатов. Наглядное информативное описание результатов кластерного анализа очень важно хотя бы для самостоятельного использования в других статистических процедурах и методах. Помимо стандартных результатов кластерного анализа, например, в виде «типового объекта», который обладает значениями существенных признаков, близкими к фиктивному «среднему» объекту по группе, используется и представление полученных результатов в координатах специальных переменных, связанных с используемой системой признаков и позволяющих визуальную оценить полученную классификацию.

Определение качества полученной группировки. Сразу надо отметить, что наилучшего или универсального критерия качества полученной группировки (кластеризации) не существует. Таким образом, смысловая (семантическая) оценка качества выделения кластеров определяется субъективно исследователем [68]. Оценка качества, например, может базироваться на аналитике сравнительного анализа компактности классов.

Но существуют и формальные способы проверки качества кластеризации. Например, в [69] предлагается целый спектр формальных вычислений различных показателей качества. В частности, можно использовать индекс внутригруппового разброса, С-индекс, индекс Гудмана-Крускала и др. Алгоритмы базируются на нахождении вероятностей случайного образования групп, вычисляемых в рамках различных моделей распределения.

Таким образом, наличие субъективного влияния исследователя на выполнение каждого этапа кластерного анализа обуславливает неоднозначность полученного результата. Требуется и сравнение с аналогичными результатами, но в другом пространстве показателей, и сравнение с результатами анализа, но с другими инструментальными средствами. Стоит обратить внимание и на масштабирование показателей. В частности, на их нелинейное преобразование, повышающее, в общем случае, уровень корреляций между значимыми факторами.

В целом, аналитика кластерного анализа, включающая и семантическое осмысление данных, и решение проблемы больших данных (сокращение выборки до описания гипотетического среднего объекта), и выявление новых объектов, не вписывающихся в стандартные классы, не имеет общего универсального решения и определяется исключительно целями исследований и методами кластеризации.

ЛИТЕРАТУРА

1. Тимофеев Ю.М. Глобальная система мониторинга параметров атмосферы и поверхности. – СПб., 2009. – 129 с.
2. Мохов И.И. Диагностика структуры климатической системы. – СПб.: Гидрометеиздат, 1993. – 271 с.
3. Хромов С.П., Петросянц М.А. Метеорология и климатология. – М.: Изд-во МГУ, – 2004. – 582 с.
4. Груза Г.В., Рейтенбах Р.Г. Статистика и анализ гидрометеорологических данных. – Л.: Гидрометеиздат, 1982. – 215 с.
5. Предстоящие изменения климата / под ред. М.И. Будыко [и др.]. – Л.: Гидрометеиздат, 1991. – 272 с.
6. Груза Г.В., Ранькова Э.Я. Структура и изменчивость наблюдаемого климата. Температура воздуха Северного полушария. – Л.: Гидрометеиздат, 1980. – 72 с.
7. Винников К.Я. Современные изменения глобального климата // ВНИИГМИ – МЦД. Сер. метеорология. – 1985. – Вып. 8. – С. 5-20.
8. Köppen W. Das geographische System der Klimate. – Berlin: Verlag von Gebrüder Bornträger, 1936. – 44 s.
9. Григорьев А.А., Будыко М.И. Классификация климатов СССР // Изв. АН СССР. Серия геогр. – 1959. – №3. – С. 58-70.
10. Коробов В.Б., Васильев Л.Ю. Климатическое районирование территорий экспертно-статистическими методами. Постановка задачи // Метеорология и гидрология. – 2004. – №6. – С. 38-48.
11. Fovell R., Fovell M.-Y. Climate Zones of the Conterminous United States Defined Using Cluster Analysis // American Meteorological Society. – 1993. – №6. – PP. 2103-2135.
12. Armah Frederick Ato, Obiri Samuel et al, Mining and Heavy Metal Pollution: Assessment of Aquatic Environments in Tarkwa (Ghana) using Multivariate Statistical Analysis. Journal of Environmental Statistics, February 2010. – V. 1. – №4. – PP. 1-13.
13. Mrutu A., Luilo G.B. Data mining using multivariate statistical analysis The case of heavy metals in sediments of the Msimbazi Creek mangrove wetland // Environmental Skeptics and Critics. – 2013. – V. 2(4). – PP. 153-163.
14. Ryan P. Boyles*, Sethu Raman Analysis of climate trends in North Carolina (1949–1998) // Environment International. – 2003. – V. 29. – PP. 263 – 275.
15. T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. A local search approximation algorithm for k-means clustering. In Proceedings of ACM Symposium on Computational Geometry, pages 10–18, 2002.
16. Kalkstein, L.S., G. Tan, and J.A. Skindlov. An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification // Journal of Climate and Applied Meteorology. – 1987. – V. 26. – PP. 717-730.
17. Bretherton C.S., Smith C. and Wallace J.M. An intercomparison of methods for finding coupled patterns in climate data // J. Climate. – 1992. – V. 5. – PP. 541-560.

18. D'Andrea F.S., Tibaldi S., Blackburn M et al. Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979-1988 // *Clim. Dynamics*. – 1998. – V. 14. – PP. 385-407.
19. Катаев С.Г., Кусков А.И. Исследование озонных полей над территорией России и сопредельных государств. I. Составляющие полей озона и их // *Вестник ТГПУ. Сер.: Естественные и точные науки*. – 1998. – Вып. 5. – С. 3-9.
20. Катаев С.Г., Кусков А.И. Исследование озоновых полей над территорией России и сопредельных государств. II. Классификация составляющих полей // *Вестник ТГПУ. Сер.: Естественные и точные науки*. – 1998. – Вып. 5. – С. 10-17.
21. Кусков А.И., Катаев С.Г. Закономерности современных изменений теплового поля в приземном слое атмосферы Сибири и на Дальнем Востоке // *Изв. Вузов “Физика”*. – 2004. – №11. – С. 81-92.
22. Кусков А.И., Катаев С.Г. Структура и динамика приземного температурного поля над азиатской территорией России. – Томск: Изд-во ТГПУ, 2006. – 176 с.
23. Катаев С.Г. Алгоритмические и программные средства аналитического и неразрушающего контроля параметров природной среды на основе метода выделения структур.: дис. док. техн. наук: 05.11.13. защищена 12.02.2013; утв. 15.07.13. — Томск – 2013. — 352 с.
24. Wu, A.M., Hsieh, W.W., and Zwiers, F.W.: Nonlinear modes of North American winter climate variability derived from a general circulation model simulation, *J. Climate*, 16, 2325–2339, 2003.
25. Wu, A.M. and Hsieh, W.W.: The nonlinear northern hemisphere winter atmospheric response to ENSO, *Geophys. Res. Lett.*, 31, L02203, doi:10.1029/2003GL018885, 2004
26. Wu, A. and Hsieh, W.W.: Nonlinear interdecadal changes of the ElNino-Southern Oscillation, *Clim. Dynam.*, 21, 719–730, 2003.
27. Wu, A. and Hsieh, W.W.: Nonlinear canonical correlation analysis of the tropical Pacific wind stress and sea surface temperature, *Clim. Dynam.*, 19, 713–722, 2002.
28. T.P. Barnett and R. Preisendorfer. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, 115(9):1825–1850, 1987.
29. Roswintiarti O., Niyogi D.S., Raman S. Tele-connections between tropical Pacific sea surface temperature anomalies and North Carolina precipitation anomalies during El Niño events. *Geophys Res Lett* 1998; 25, P 4201 – 4204.
30. J. Sukharev, C. Wang, K.-L. Ma, A.T. Wittenberg, 2009, “Correlation Study of Time-Varying Multivariate Climate Data Sets”, *IEEE VGTC Pacific Visualization Symposium*, 161-168.
31. A.J. Cannon and W.W. Hsieh Robust nonlinear canonical correlation analysis: application to seasonal climate forecasting *Nonlin. Processes Geophys.*, 15, 221–232, 2008.
32. Livezey RE and Smith TM (1999) Considerations for use of the Barnett and Preisendorfer (1987) algorithm for canonical correlation analysis of climate variations. *J. Climate*, 12, 303–305.

33. Barnett, T.P. and Preisendorfer, R.: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis, *Mon. Weather Rev.*, 115, 1825–1850, 1987.
34. Barnston, A.G. and Ropelewski, C.F.: Prediction of ENSO episodes using canonical correlation-analysis, *J. Climate*, 5, 1316–1345, 1992.
35. Saha S., Nadiga .S, Thiaw C. et al. (2006) The NCEP Climate Forecast System. *J. Climate*, 19, 3483–3517.
36. A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk*, pp. 465-470 © 2007 Springer. Printed in the Netherlands.
37. Greenfield R.S. and Fisher G.M. (2003) Improving responses to climate predictions - An introduction. *Bull. Amer. Meteorol. Soc.*, 84, 1685–1685.
38. Hansen JW and Sivakumar MVK (2006) Advances in applying climate prediction to agriculture. *Climate Research* 33, 1–2.
39. Harrison MSJ (2005) The development of seasonal and interannual climate forecasting. *Climatic Change*, 70, 201–220.
40. Krishnamurti T.N., Kishtawal C.M., Zhang Z. et al. (2000) Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, 13, 4196–4216.
41. Kim K-Y and North GR (1999) EOF-based linear prediction algorithm: Examples. *J. Climate* 12, 2076–2092.
42. Kim K-Y and Wu Q (1999) A comparison study of EOF techniques: analysis of nonstationary data with periodic statistics. *J. Climate* 12, 185–199.
43. Tangang F.T., Tang B.Y., Monahan A.H. and Hsieh WW (1998) Forecasting ENSO events: A neural network extended EOF approach. *J. Climate*, 11, 29–41.
44. Elsner, J.B. and Tsonis, A.A. (1996): *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, Plenum Press.
45. Golyandina, N., and E. Osipov (2007) "The 'Caterpillar'-SSA method for analysis of time series with missing values", *J. Stat. Plan. Inference* 137(8), 2642–2653.
46. Mandelbrot B.B. *The Fractal Geometry of Nature*. // W.H. Freeman, Sun-Francisco, 1982.
47. Feder J. *Fractals*. // Plenum Press, New York, 1988.
48. Scafetta N. and B.J. West *Solar Flare Intermittency and Earth's Temperature Anomalies* // *Phys. Rev. Lett.*, 2003, v.90, № 24, pp. 248701.
49. Abarbanel H.D.I. *Analysis of Observed Chaotic Data*. // Springer, New York, 1996.
50. Yulmetyev R., Gafarov F., Hanggi P., Nigmatullin R. and Kayumov S., Possibility between earthquake and explosion seismogram differentiation by discrete stochastic non-Markov processes and local Hurst exponent analysis // *Phys. Rev. E*, 2001, №64, pp. 066132.
51. Costa M. Golderberger A.L. Peng C.K. *Multiscale Entropy Analysis of Complex Physiologic Time Series* // *Phys. Rev. Lett.*, 2002, V.89, №6.
52. Старченко Н.В. Индексы фрактальности и локальный анализ хаотических временных рядов.: дис. канд. физ-мат. наук: 05.13.18. защищена 15.02.06: утв.14.07.06. – М., – 2006. – 119 с.

53. Поляков Д.В., Кужевская И.В. Применение кластерного анализа для оценки температурно-влажностных условий в период активной вегетации на территории юга Западной Сибири и его связь с гидротермическим коэффициентом Т.Г. Селянинова // Вестн. Том. гос. ун-та. – 2012. – №360. – С. 188-192.
54. Ветрова Е.И., Скриптунова Е.Н., Шакина Н.П. Прогноз низкой облачности на аэродромах европейской территории бывшего СССР // Метеорология и гидрология. – 2013. – №1. – С. 12-31.
55. Овечкин С.В., Майнашева Г.М. Опыт использования кластерного анализа при климатическом районировании Московской области // Вестник МГПУ Серия «Естественные науки». – 2010. – №2 (6). – С. 65-74.
56. Вершовский Е.А. Разработка методов и алгоритмов кластеризации мультиспектральных данных дистанционного зондирования Земли // Автореферат диссертации на соискание ученой степени кандидата технических наук, 2010, 17 с.
57. Кирста Ю.Б., Курепина Н.Ю., Ловцкая О.В. Пространственная декомпозиция метеорологических полей Евразии: разделение воздействий растительности и антропогенной деятельности. // Фундаментальные исследования. – 2014. – №5. – С. 1030-1036.
58. Дробушевская О.В., Царегородцев В.Г. Географо-климатические варианты светлохвойных травяных лесов Сибири // Сибирский экологический журнал. – 2007. – №2. – С. 211-219.
59. Родригес Залепинос Р.А. Данные и методы интеллектуального анализа данных для исследования окружающей природной среды // Системный анализ и информационные технологии в науках о природе и обществе. – Донецк: ДонНТУ, 2011. – Вып. 1. – С. 94–107.
60. Вороненко Д.И. Классификация метеорологических сценариев для использования в предикторах энергопотребления жилых массивов города // Науковий вісник Чернівецького університету. – 2009. – Випуск 479. Комп'ютерні системи та компоненти. – С. 88-97.
61. Персианцев И.Г. Разработка нейросетевых методов обработки данных физического эксперимента. URL: <http://istina.msu.ru/projects/5667315/>.
62. Алгоритмы кластеризации на службе Data Mining // BaseGroup Labs [1995 - 2015]. URL: <http://www.basegroup.ru/library/analysis/clusterization/datamining/>.
63. Обзор алгоритмов кластеризации числовых пространств данных // Сайт habrahabr. URL: <http://habrahabr.ru/post/164417/>.
64. Макаров С.В., Кропотов Ю.А. Анализ и обработка экологической информации с помощью нейросетевых алгоритмов // Методы и устройства передачи и обработки информации. – 2007. – Вып. 9. – С. 93-99.
65. Нейский И.М. Классификация и сравнение методов кластеризации. URL: http://it-claim.ru/Persons/Neyskiy/Article2_Neyskiy.pdf.

66. Царегородцев В.Г., Погребная Н.А. Нейросетевые методы обработки информации в задачах прогноза климатических характеристик и лесорастительных свойств ландшафтных зон // Методы нейроинформатики. – Красноярск. –1998. – 205 с.
67. Бериков В.Б., Лбов Г.С. Современные тенденции в кластерном анализе. URL: <http://www.ict.edu.ru/ft/005638/62315e1-st02.pdf>.
68. Rand W. Objective criteria for the evaluation of clustering methods // Journal of American Statistical Association. –1971. – V.66. – PP. 846-850.

Рецензент: Статья рецензирована членами редколлегии журнала.

Botygin Igor' Aleksandrovich

National Research Tomsk Polytechnic University
Russia, Tomsk
E-mail: bia@tpu.ru

Kataev Sergey Grigor'evich

Tomsk State Pedagogical University
Russia, Tomsk
E-mail: sgkataev@sibmail.com

Sherstnev Vladislav Stanislavovich

National Research Tomsk Polytechnic University
Russia, Tomsk
E-mail: vss@tpu.ru

Sherstneva Anna Igorevna

National Research Tomsk Polytechnic University
Russia, Tomsk
E-mail: sherstneva@tpu.ru

Methods for classification and analysis of climatic fields

Abstract. The article features the comparative analysis of the methods for finding patterns of over time behavior for meteo-parameter field or any other characteristics in the territories of various sizes. It is shown that the identification of the science-proven underlying trends of the geophysical field changes on the researched area is accompanied by a sharp increase in the amount of stored data. This requires using modern information storage technologies and large data processing as well as designing new processing methods. The methods and algorithms can be effectively used for finding patterns in experimental data based on the distribution of objects into groups. Due to the diversity of systems of signs, which can be determined by the similarity of objects, unifying object grouping procedures is proven to be impossible, as all determined by the specific problem to be solved, the unique form of the criteria classification system and the author's meaning. The widespread understanding of the selected group of objects as a subset of combining objects into some single abstract class only based on the position in space of attributes, allows to specify the main stages with individual tasks of selection and optimization, multivariate statistical procedures aimed to collect data on the object selection and their arrangement into relatively homogenous groups.

Keywords: cluster analysis; time series; cluster classification; factor analysis; clustering; grouping; geophysical field; experimental data.

REFERENCES

1. Timofeev Yu.M. Global'naya sistema monitoringa parametrov atmosfery i poverkhnosti. – SPb., 2009. – 129 s.
2. Mokhov I.I. Diagnostika struktury klimaticheskoy sistemy. – SPb.: Gidrometeoizdat, 1993. – 271 s.
3. Khromov S.P., Petrosyants M.A. Meteorologiya i klimatologiya. – M.: Izd-vo MGU, – 2004. – 582 s.
4. Gruza G.V., Reytenbakh R.G. Statistika i analiz gidrometeorologicheskikh dannykh. – L.: Gidrometeoizdat, 1982. – 215 s.
5. Predstoyashchie izmeneniya klimata / pod red. M.I. Budyko [i dr.]. – L.: Gidrometeoizdat, 1991. – 272 s.
6. Gruza G.V., Ran'kova E.Ya. Struktura i izmenchivost' nablyudaemogo klimata. Temperatura vozdukha Severnogo polushariya. – L.: Gidrometeoizdat, 1980. – 72 s.
7. Vinnikov K.Ya. Sovremennye izmeneniya global'nogo klimata // VNIIGMI – MTsD. Ser. meteorologiya. – 1985. – Vyp. 8. – S. 5-20.
8. Köppen W. Das geographische System der Klimate. – Berlin: Verlag von Gebrüder Bornträger, 1936. – 44 s.
9. Grigor'ev A.A., Budyko M.I. Klassifikatsiya klimatov SSSR // Izv. AN SSSR. Seriya geogr. – 1959. – №3. – S. 58-70.
10. Korobov V.B., Vasil'ev L.Yu. Klimaticheskoe rayonirovanie territoriy ekspertno-statisticheskimi metodami. Postanovka zadachi // Meteorologiya i gidrologiya. – 2004. – №6. – S. 38-48.
11. Fovell R., Fovell M.-Y. Climate Zones of the Conterminous United States Defined Using Cluster Analysis // American Meteorological Society. – 1993. – №6. – PP. 2103-2135.
12. Armah Frederick Ato, Obiri Samuel et al, Mining and Heavy Metal Pollution: Assessment of Aquatic Environments in Tarkwa (Ghana) using Multivariate Statistical Analysis. Journal of Environmental Statistics, February 2010. – V. 1. – №4. – RR. 1-13.
13. Mrutu A., Luilo G.B. Data mining using multivariate statistical analysis The case of heavy metals in sediments of the Msimbazi Creek mangrove wetland // Environmental Skeptics and Critics. – 2013. – V. 2(4). – RR. 153-163.
14. Ryan P. Boyles*, Sethu Raman Analysis of climate trends in North Carolina (1949–1998) // Environment International. – 2003. – V. 29. – RP. 263 – 275.
15. T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, and A.Y. Wu. A local search approximation algorithm for k-means clustering. In Proceedings of ACM Symposium on Computational Geometry, pages 10–18, 2002.
16. Kalkstein, L.S., G. Tan, and J.A. Skindlov. An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification // Journal of Climate and Applied Meteorology. – 1987. – V. 26. – RR. 717-730.
17. Bretherton C.S., Smith C. and Wallace J.M. An intercomparison of methods for finding coupled patterns in climate data // J. Climate. – 1992. – V. 5. – RR. 541-560.

18. D'Andrea F.S., Tibaldi S., Blackburn M et al. Northern Hemisphere atmospheric block-ing as simulated by 15 atmospheric general circulation models in the period 1979-1988 // *Clim. Dynamics*. – 1998. – V. 14. – RR. 385-407.
19. Kataev S.G., Kuskov A.I. Issledovanie ozonnykh poley nad territoriei Rossii i sopredel'nykh gosudarstv. I. Sostavlyayushchie poley ozona i ikh // *Vestnik TGPU. Ser.: Estestvennye i tochnye nauki*. – 1998. – Vyp. 5. – С. 3-9.
20. Kataev S.G., Kuskov A.I. Issledovanie ozonovykh poley nad territoriei Rossii i sopredel'nykh gosudarstv. II. Klassifikatsiya sostavlyayushchikh poley // *Vestnik TGPU. Ser.: Estestvennye i tochnye nauki*. – 1998. – Vyp. 5. – S. 10-17.
21. Kuskov A.I., Kataev S.G. Zakonomernosti sovremennykh izmeneniy teplovogo polya v prizemnom sloe atmosfery Sibiri i na Dal'nem Vostoke // *Izv. Vuzov "Fizika"*. – 2004. – №11. – S. 81-92.
22. Kuskov A.I., Kataev S.G. Struktura i dinamika prizemnogo temperaturnogo polya nad aziatskoy territoriei Rossii. – Tomsk: Izd-vo TGPU, 2006. – 176 s.
23. Kataev S.G. Algoritmicheskie i programmnye sredstva analiticheskogo i nerazrushayushchego kontrolya parametrov prirodnoy sredy na osnove metoda vydeleniya struktur.: dis. dok. tekhn. nauk: 05.11.13. zashchishchena 12.02.2013: utv. 15.07.13. — Tomsk – 2013. — 352 s.
24. Wu, A.M., Hsieh, W.W., and Zwiers, F.W.: Nonlinear modes of North American winter climate variability derived from a general circulation model simulation, *J. Climate*, 16, 2325–2339, 2003.
25. Wu, A.M. and Hsieh, W.W.: The nonlinear northern hemisphere winter atmospheric response to ENSO, *Geophys. Res. Lett.*, 31, L02203, doi:10.1029/2003GL018885, 2004
26. Wu, A. and Hsieh, W.W.: Nonlinear interdecadal changes of the ElNino-Southern Oscillation, *Clim. Dynam.*, 21, 719–730, 2003.
27. Wu, A. and Hsieh, W.W.: Nonlinear canonical correlation analysis of the tropical Pacific wind stress and sea surface temperature, *Clim. Dynam.*, 19, 713–722, 2002.
28. T.P. Barnett and R. Preisendorfer. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, 115(9):1825–1850, 1987.
29. Roswintiarti O., Niyogi D.S., Raman S. Tele-connections between tropical Pacific sea surface temperature anomalies and North Carolina precipitation anomalies during El Nin^o events. *Geophys Res Lett* 1998; 25, P 4201 – 4204.
30. J. Sukharev, C. Wang, K.-L. Ma, A.T. Wittenberg, 2009, "Correlation Study of Time-Varying Multivariate Climate Data Sets", *IEEE VGTC Pacific Visualization Symposium*, 161-168.
31. A.J. Cannon and W.W. Hsieh Robust nonlinear canonical correlation analysis: application to seasonal climate forecasting *Nonlin. Processes Geophys.*, 15, 221–232, 2008.
32. Livezey RE and Smith TM (1999) Considerations for use of the Barnett and Preisendorfer (1987) algorithm for canonical correlation analysis of climate variations. *J. Climate*, 12, 303–305.

33. Barnett, T.P. and Preisendorfer, R.: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis, *Mon. Weather Rev.*, 115, 1825–1850, 1987.
34. Barnston, A.G. and Ropelewski, C.F.: Prediction of ENSO episodes using canonical correlation-analysis, *J. Climate*, 5, 1316–1345, 1992.
35. Saha S., Nadiga .S, Thiaw C. et al. (2006) The NCEP Climate Forecast System. *J. Climate*, 19, 3483–3517.
36. A. Troccoli et al. (eds.), *Seasonal Climate: Forecasting and Managing Risk*, pp. 465-470 © 2007 Springer. Printed in the Netherlands.
37. Greenfield R.S. and Fisher G.M. (2003) Improving responses to climate predictions - An introduction. *Bull. Amer. Meteorol. Soc.*, 84, 1685–1685.
38. Hansen JW and Sivakumar MVK (2006) Advances in applying climate prediction to agriculture. *Climate Research* 33, 1–2.
39. Harrison MSJ (2005) The development of seasonal and interannual climate forecasting. *Climatic Change*, 70, 201–220.
40. Krishnamurti T.N., Kishtawal C.M., Zhang Z. et al. (2000) Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, 13, 4196–4216.
41. Kim K-Y and North GR (1999) EOF-based linear prediction algorithm: Examples. *J. Climate* 12, 2076–2092.
42. Kim K-Y and Wu Q (1999) A comparison study of EOF techniques: analysis of nonstationary data with periodic statistics. *J. Climate* 12, 185–199.
43. Tangang F.T., Tang B.Y., Monahan A.H. and Hsieh WW (1998) Forecasting ENSO events: A neural network extended EOF approach. *J. Climate*, 11, 29–41.
44. Elsner, J.B. and Tsonis, A.A. (1996): *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, Plenum Press.
45. Golyandina, N., and E. Osipov (2007) "The 'Caterpillar'-SSA method for analysis of time series with missing values", *J. Stat. Plan. Inference* 137(8), 2642–2653.
46. Mandelbrot B.B. *The Fractal Geometry of Nature*. // W.H. Freeman, Sun-Francisco, 1982.
47. Feder J. *Fractals*. // Plenum Press, New York, 1988.
48. Scafetta N. and B.J. West *Solar Flare Intermittency and Earth's Temperature Anomalies* // *Phys. Rev. Lett.*, 2003, v.90, № 24, pp. 248701.
49. Abarbanel H.D.I. *Analysis of Observed Chaotic Data*. // Springer, New York, 1996.
50. Yulmetyev R., Gafarov F., Hanggi P., Nigmatullin R. and Kayumov S., Possibility between earthquake and explosion seismogram differentiation by discrete stochastic non-Markov processes and local Hurst exponent analysis // *Phys. Rev. E*, 2001, №64, pp. 066132.
51. Costa M. Golderberger A.L. Peng C.K. *Multiscale Entropy Analysis of Complex Physiologic Time Series* // *Phys. Rev. Lett.*, 2002, V.89, №6.
52. Starchenko N.V. *Индексы фрактальности и локальный анализ хаотических временных рядов*: дис. канд. физ-мат. наук: 05.13.18. заш.ч.ш.ч.ена 15.02.06: utv.14.07.06. – М., – 2006. – 119 с.

53. Polyakov D.V., Kuzhevskaya I.V. Primenenie klasterного анализа dlya otsenki temperaturno-vlazhnostnykh usloviy v period aktivnoy vegetatsii na territorii yuga Zapadnoy Sibiri i ego svyaz' s gidrotermicheskim koeffitsientom T.G. Selyaninova // Vestn. Tom. gos. un-ta. – 2012. – №360. – S. 188-192.
54. Vetrova E.I., Skriptunova E.N., Shakina N.P. Prognoz nizkoy oblachnosti na aerodromakh evropeyskoy territorii byvshego SSSR // Meteorologiya i gidrologiya. – 2013. – №1. – S. 12-31.
55. Ovechkin S.V., Maynasheva G.M. Opyt ispol'zovaniya klasterного анализа pri klimaticheskom rayonirovanii Moskovskoy oblasti // Vestnik MGPU Seriya «Estestvennye nauki». – 2010. – №2 (6). – S. 65-74.
56. Vershovskiy E.A. Razrabotka metodov i algoritmov klasterizatsii mul'tispektral'nykh dannykh distantsionnogo zondirovaniya Zemli // Avtoreferat dissertatsii na soiskanie uchenoy stepeni kandidata tekhnicheskikh nauk, 2010, 17 s.
57. Kirsta Yu.B., Kurepina N.Yu., Lovtskaya O.V. Prostranstvennaya dekompozitsiya meteorologicheskikh poley Evrazii: razdelenie vozdeystviy rastitel'nosti i antropogennoy deyatel'nosti. // Fundamental'nye issledovaniya. – 2014. – №5. – S. 1030-1036.
58. Drobushhevskaya O.V., Tsaregorodtsev V.G. Geografo-klimaticheskie varianty svetlokhvoynykh travyanykh lesov Sibiri // Sibirskiy ekologicheskiy zhurnal. – 2007. – №2. – S. 211-219.
59. Rodrigues Zalepinos R.A. Dannye i metody intellektual'nogo analiza dannykh dlya issledovaniya okruzhayushchey prirodnoy sredy // Sistemnyy analiz i informatsionnye tekhnologii v naukach o prirode i obshchestve. – Donetsk: DonNTU, 2011. – Vyp. 1. – S. 94–107.
60. Voronenko D.I. Klassifikatsiya meteorologicheskikh stsenariyev dlya ispol'zovaniya v prediktorakh energopotrebleniya zhilykh massivov goroda // Naukoviy visnik Chernivets'kogo universitetu. – 2009. – Vipusk 479. Komp'yuternisistemi ta komponenti. – S. 88-97.
61. Persiantsev I.G. Razrabotka neyrosetevykh metodov obrabotki dannykh fizicheskogo eksperimenta. URL: <http://istina.msu.ru/projects/5667315/>.
62. Algoritmy klasterizatsii na sluzhbe Data Mining // BaseGroup Labs [1995 - 2015]. URL: <http://www.basegroup.ru/library/analysis/clusterization/datamining/>.
63. Obzor algoritmov klasterizatsii chislovykh prostranstv dannykh // Sayt habrahabr. URL: <http://habrahabr.ru/post/164417/>.
64. Makarov S.V., Kropotov Yu.A. Analiz i obrabotka ekologicheskoy informatsii s pomoshch'yu neyrosetevykh algoritmov // Metody i ustroystva peredachi i obrabotki informatsii. – 2007. – Vyp. 9. – S. 93-99.
65. Neyskiy I.M. Klassifikatsiya i sravnenie metodov klasterizatsii. URL: http://it-claim.ru/Persons/Neyskiy/Article2_Neiskiy.pdf.
66. Tsaregorodtsev V.G., Pogrebnaya N.A. Neyrosetevye metody obrabotki informatsii v zadachakh prognoza klimaticheskikh kharakteristik i lesorastitel'nykh svoystv landshaftnykh zon // Metody neyroinformatiki. – Krasnoyarsk. –1998. – 205 s.
67. Berikov V.B., Lbov G.S. Sovremennye tendentsii v klasterном анализе. URL: <http://www.ict.edu.ru/ft/005638/62315e1-st02.pdf>.

68. Rand W. Objective criteria for the evaluation of clustering methods // Journal of American Statistical Association. –1971. – V.66. – RP. 846-850.