

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

Для решения задач оценки территории была проведена автоматическая систематизация средних значений полей гидротермического коэффициента Селянинова с применением иерархического кластерного анализа. Проведена статистическая оценка полученных классов с применением критерия качества, которая в последующем была подтверждена физическим смыслом классификации.

Ключевые слова: тепловлагообеспеченность территории; кластерный анализ.

Научные подразделения Росгидромета считают, что продовольственная безопасность России в ближайшие десятилетия будет зависеть от темпов и направленности усиливающегося процесса глобальных изменений климата. В данной связи действующая парадигма постоянства климата стала явно нереалистичной. Температурно-влажностный режим территории юга Западной Сибири является неотъемлемой частью общего климатического комплекса, анализ которого представляется не только актуальным, но и стратегически важными для аграрного комплекса страны. Для России оценка состояния климатических ресурсов была выполнена в ряде работ разными авторами с нескольких разных точек зрения [1. С. 247]. Одним из основных выводов отдельных работ [2. С. 183] является заключение о том, что сельское хозяйство России имеет большие возможности для развития благодаря корректному использованию своего биоклиматического потенциала. Агроклиматический потенциал весьма полно учитывается в сельскохозяйственном производстве ведущих стран, однако для нашей страны представляет собой нереализованный резерв развития (по некоторым оценкам, до 30%).

Данное исследование позволяет рассмотреть тематику районирования территории на более современном уровне, основа которой опирается на автоматическую систематизацию полученных данных и служит следующим этапом развития методики оценки территории [3. С. 136].

Климатическое районирование территории – очень сложная и неоднозначная задача исследования, в которую входит принцип определения и построения зон охвата. В условиях современного изменения климата аграрные предприятия Западной Сибири заинтересованы в поиске современной методики оценки агроклиматических условий и решения проблем районирования, с последующим оптимизированием используемых агроклиматических ресурсов.

Методика кластерного анализа активно разрабатывается и используется для анализа данных уже давно. Существует литература, описывающая различные варианты методов кластерного анализа (например, [4. С. 119]). Однако большие массивы данных стало возможным обрабатывать и, следовательно, получать статистически значимые результаты только с появлением быстродействующих и современных ЭВМ. Методика кластерного анализа позволяет выделить из множества объектов такие, которые имели бы какие-либо близкие свойства, поэтому кла-

стерный анализ носит также название классификации. Для решения приведенных целей ставится задача исследования температурно-влажностных условий территории с помощью кластерного анализа, который позволяет выполнить объективную систематизацию.

Кластерный анализ, а именно иерархическая классификация, получил широкое применение в развитии автоматических классификаций метеорологических полей и синоптических процессов [5. С. 7]. Расчетная часть анализа уже не является трудоёмкой, потому как производится в автоматическом режиме [7. С. 38]. На первое место выходит проблема корректной интерпретации полученных результатов [6. С. 44].

В данной работе классификации подверглись поля средних значений ГТК (гидротермический коэффициент Селянинова) над территорией юга Западной Сибири как индикатор температурно-влажностного режима и изменчивости климата. Была проведена статистическая оценка полученных классов с применением критерия качества, которая в последующем отвечала физическому смыслу классификации ГТК, представленного как критерий оценки температурно-влажностных условий территории исследования в период активной вегетации.

Исходные материалы. Для исследования была сформирована база данных компонентов индекса ГТК, а именно сумм активных температур и сумм атмосферных осадков в фазу активной вегетации ($T \geq 10^\circ\text{C}$) за 51-летний период наблюдений (1960–2010 гг.). Полученные компоненты были применены для расчета индекса ГТК с последующим осреднением по исследуемым станциям. Для устранения влияния пространственной неоднородности местоположения метеорологических станций проводилась пространственная интерполяция рассчитанных средних годовых значений величин на регулярной сетке $1 \times 1^\circ$ по алгоритму Крайгинга. Пространственные значения индекса ГТК подверглись кластеризации с последующим получением вариантов классификации. После завершения процедур классификации была проведена статистическая оценка значимости полученных классов с использованием меры качества классификации, которую принято называть критерием качества. Критерием качества классов в данной работе выступала суммарная внутриклассовая дисперсия (F_3).

Расчет проводился по данным станций международного обмена (ВНИИГМИ-МЦД): Омск, Барабинск, Барнаул, Рубцовск, Красноярск, Минусинск, Томск,

Колпашево, Тобольск, Екатеринбург, Курган, Енисейск, Тара, Петропавловск, Костанай, Астана и Иртышск (рис. 1).

Обзор территории исследования. Территория исследования представляется в виде сетки от центрального Предуралья до Енисея (50–60° с.ш. и 60–95° в.д.). Привлечение станций Урала и Казахстана использовались для улучшения точности результатов интерполяции, проводимой в данном исследовании (т.е. в качестве граничных условий).

Географически данный регион, по большому счету, представляет собой юг Западной Сибири, который по краям граничит: на юге с Казахским мелкосопочником (200–500 м над ур. м.), на востоке и юго-востоке с Кузнецким Алатау (1000–2000 м над ур. м.), а на западе с горами Урала (500–1000 м над ур. м.). Юг Западной Сибири представляет собой плоскую равнину, которая поднимается над уровнем моря всего на 50–250 м, но значительно удалена от ближайшего океана – более чем на полторы тысячи километров.

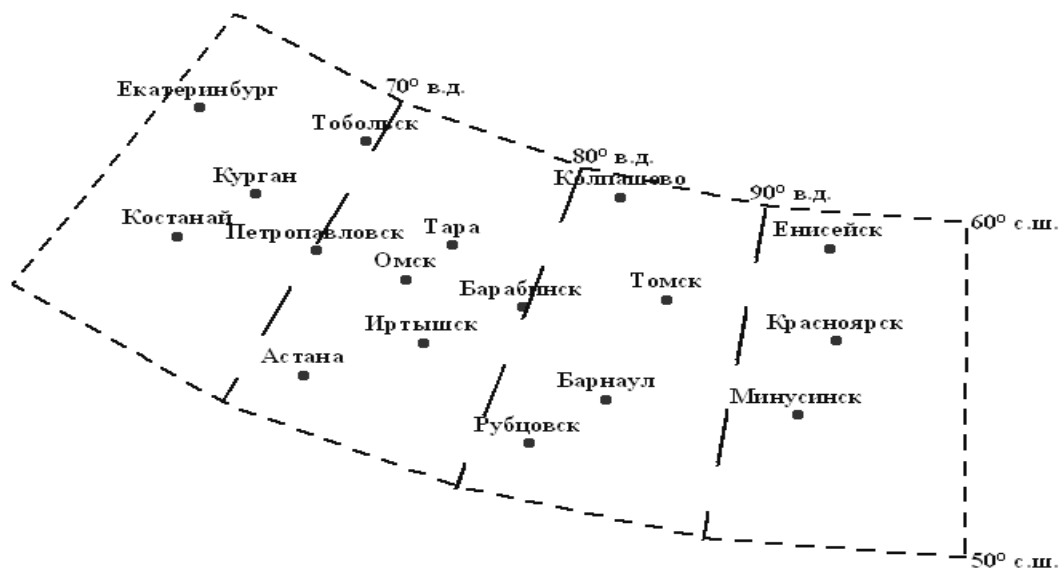


Рис. 1. Исследуемая территория юга Западной Сибири и северо-востока Казахстана

Методика кластерного анализа. В [8. С. 117] предложено использовать для кластеризации метод акреции (bump hunting) и k-средних (k-mean). Иными словами, проводится поиск такого разбиения на классы, при котором дисперсия разбиения максимальна для заданного числа классов. Излагать эти методы здесь не имеет смысла, потому как они имеют небольшие недостатки, связанные с тем, что приходится задавать число классов, а на данном этапе исследования мы не знаем точного числа выделяемых изначально классов. Кроме того, с помощью данных методов мы не можем получить структурированные данные от более к менее общим особенностям. Поэтому для анализа данных был применен другой метод. Одним из наиболее гибких и информативных методов кластеризации является метод восходящей иерархической классификации. Подробно иерархическая классификация излагается в [4. С. 77], ниже приводится краткое изложение модифицированного алгоритма, используемого в данной работе.

Алгоритм строится из последовательных шагов:

1. На начальном этапе в имеющейся пространственной выборке, состоящей из полей средних значений индекса ГТК по исследуемым станциям, отождествим каждую из них в отдельный класс. Такие классы, состоящие из единственного поля, называются начальными. Значения поля можно рассматривать как векторы в евклидовом пространстве. Общее число рассматриваемых векторов равно количеству полей в выборке N , т.е. количеству рассматриваемых станций (17 станций).

2. Для проведения классификации получим квадратную матрицу взаимных расстояний между всеми

начальными классами с размерностью $N \times N$. Это сделать легко, поскольку каждый начальный класс содержит единственный вектор. Существует несколько способов получения расстояний между векторами в зависимости от задач классификации и анализируемых данных. Представляется естественным использование формулы евклидовой метрики

$$d_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2}, \quad (1)$$

где d_{ij} – расстояние между i -м и j -м объектом; x_{ik} , x_{jk} – значение k -й переменной у i -го и j -го объектов.

3. Идея кластерного анализа заключается в последовательном объединении близких по каким-либо признакам кластеров, при котором строится бинарное «перевернутое дерево» иерархии классов. Критерием близости классов в данной работе, как и в работе [6. С. 40], был использован метод Уорда (Ward's method), в некоторых источниках он встречается под названием метода Варда. Из практического применения статистики (см., например, [4. С. 119]) известно, что сумма дисперсий зависимых случайных величин, под которыми в нашем случае понимается евклидова метрика, необходимо на каждом шаге объединять такие два класса, которые дают наименьшее увеличение дисперсии разбиения на каждом шаге. Данный метод предполагает, что на первом шаге каждый кластер состоит из одного объекта, и первоначально объединяются два ближайших кластера. Для них определяются средние значения каждого признака и рассчитывается сумма квадратов отклонений V_k :

$$V_k = \sum_{i=1}^{mk} \sum_{j=1}^p (x_{ij} - \bar{x}_{jk})^2, \quad (2)$$

где k – номер кластера; i – номер объекта; j – номер признака; p – количество признаков, характеризующих каждый объект; m_k – количество объектов в k -м кластере.

Таким образом, поля средних значений индекса ГТК будут близки между собой и по структуре, поскольку будут иметь большие значения пространственной корреляции. В дальнейшем на каждом шаге работы алгоритма объединяются те объекты или кластеры, которые дают наименьшее приращение величины V_k . Метод Уорда приводит к образованию кластеров приблизительно равных размеров с минимальной внутриклассовой вариацией. В итоге все объекты оказываются объединенными в один кластер.

4. На данном этапе получаем информацию о количестве полей, попавших в каждый класс. Эту информацию нетрудно получить, используя спуск по полученному бинарному дереву (дендрограмме). Более того, приемники каждого класса позволяют получить информацию о том, какие именно поля составляют данный класс.

Заслуживает внимания проблема интерпретация полученных результатов. Самый верхний класс, завершающий иерархию, будет идентичен среднему всей выборки, что ясно из процедуры его получения.

При спуске по ветвям дендрограммы будут выделяться все более тонкие и детальные структуры, свойственные климатическим особенностям.

Однако начиная с некоторого уровня данные структуры уже не будут значимыми как физически, так и статистически.

По дендрограмме визуально можно определить разницу между уровнями объединения. Уровнем будем считать очередной шаг алгоритма, на котором происходит объединение кластеров. Наибольшая разница по оси расстояний между соседними уровнями указывает на предпочтительное число классов (соответствующее уровню, от которого осуществляется переход к последующему). Этот подход возможно использовать в условиях данного исследования, потому как имеется небольшое количество объектов и выражается структура данных, но для корректности воспользуемся проверкой на качество.

Результаты кластерного анализа. Напомним, что основной задачей построения классификации являлось определение классов [9. С. 50]. В ходе визуальной оценки было выяснено, что с помощью иерархического кластерного анализа предпочтительней использовать деление на три класса. Результаты деления на кластеры для территории юга Западной Сибири представлены на рис. 2 в виде дендрограмм.

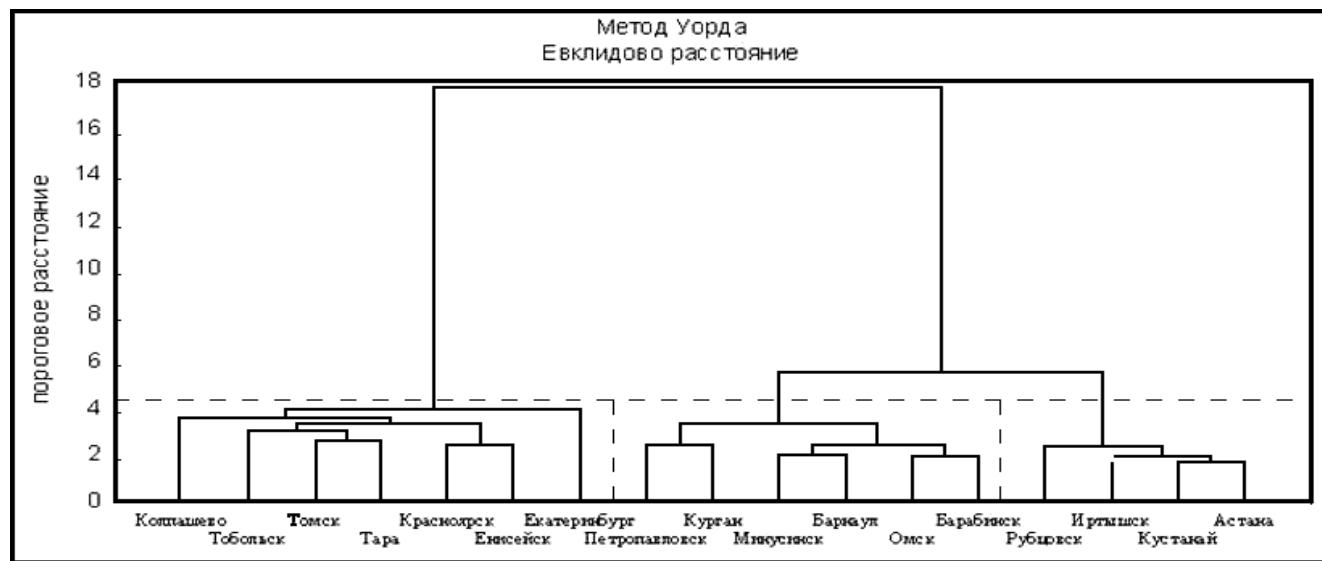


Рис. 2. Дендрограммы иерархического кластерного анализа территории юга Западной Сибири на основе метода Уорда за период с 1960 по 2010 г.

При анализе рис. 2 можно сказать, что у бинарного дерева классов выделяются три основные ветви. Еще на нижних ярусах они вобрали в себя значительную часть объектов. В первую группу попадают станции: Колпашино, Тобольск, Томск, Тара, Красноярск, Енисейск и Екатеринбург; во вторую – Петропавловск, Курган, Минусинск, Барнаул, Омск и Барабинск, а в третью – Рубцовск, Иртышск, Кустанай и Астана.

Сравним полученную классификацию со средними значениями ГТК по станциям за период исследования, где за физический смысл обоснования используется шкала классификации уровней тепловлагообеспеченности по значениям ГТК. Результаты сравнения фор-

мальной классификации и классификации уровней тепловлагообеспеченности по значениям ГТК представлены в таблице.

Рассматривая таблицу, можно сказать, что визуально наблюдается абсолютное совпадение классификации кластерного анализа с классификацией уровней тепловлагообеспеченности по значениям ГТК. Используя шкалу классификации уровней влагообеспеченности по значению ГТК, полученные классы можно интерпретировать следующим образом.

К первому классу относятся условия с повышенными значениями влагообеспеченности, второй класс выражает условия оптимального влагосодержания, а тре-

тий – недостаточного. Представим полученные классы в пространстве и сравним их с распределением ГТК.

Результаты формальной классификации представлены на рис. 3.

Сравнение классификации кластерного анализа (класс) и классификации уровней тепловлагообеспеченности по средним значениям ГТК, приведенной для территории юга Западной Сибири

Станции	ГТК _{сред}	Класс	Станции	ГТК _{сред}	Класс
Барабинск	1,01	2	Красноярск	1,53	1
Барнаул	1,02	2	Тара	1,46	1
Минусинск	1,22	2	Тобольск	1,57	1
Омск	1,13	2	Томск	1,52	1
Курган	1,04	2	Астана	0,71	3
Петропавловск	1,04	2	Иртышск	0,72	3
Екатеринбург	1,63	1	Кустанай	0,84	3
Енисейск	1,51	1			
Колпашево	1,82	1	Рубцовск	0,83	3

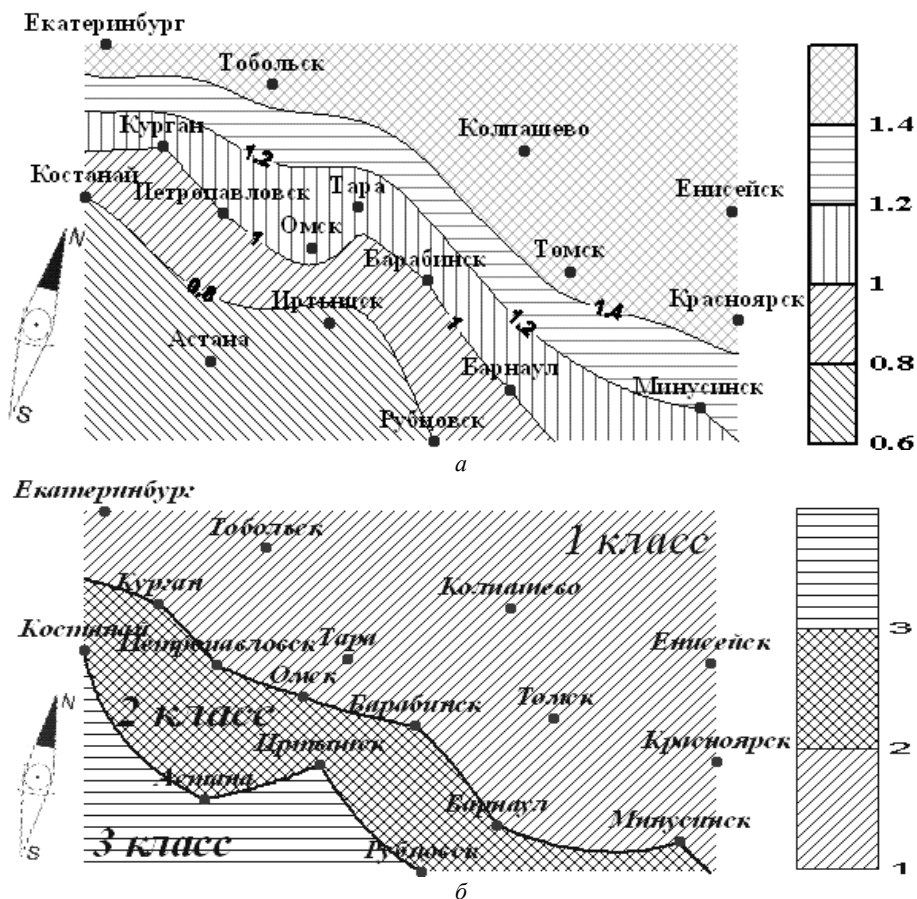


Рис. 3. Результаты формальной классификации (1960–2010 гг.): а – пространственное изменение индекса ГТК по территории юга Западной Сибири; б – пространственное распределение классов по территории юга Западной Сибири

При анализе рис. 3 можно сказать, что на территории исследования прослеживается зональное распределение индекса ГТК (см. рис. 3, а). В среднем за период исследования наибольшие значения индекса ГТК приходятся на районы ст. Екатеринбург, Тобольск, Колпашево, Томск, Красноярск и Енисейск (1,4–1,8). Наименьшие значения ГТК (слабой засушливости) за вегетационный период приходятся на районы ст. Астана и Иртышск ($< 0,75$). Наиболее благоприятные условия приходятся на районы ст. Курган, Омск, Барабинск, Барнаул и Минусинск (1–1,4).

Из рис. 3, б видно, что в пространстве классы получили широтное распределение без каких-либо «островов», данная картина соответствует распределению индекса ГТК по территории.

При рассмотрении рис. 3, а можно сказать, что в современных климатических условиях на территории юга Западной Сибири за период исследования пространственное распределение ГТК отражает увеличение континентальности и засушливости с северо-востока на юго-запад. В целом за период активной вегетации недостаток увлажнения ($\text{ГТК} < 1$) отмечается южнее «пояса» Барнаул – Барабинск – Курган.

Нормальное увлажнение ($1,0 < \text{ГТК} \leq 1,4$) на территории исследования реже всего отмечается на юго-западе. По мере продвижения на северо-восток условия оптимального тепловлагообеспечения занимают территорию в виде «пояса» Минусинск – Барнаул – Омск – Курган. Условия переувлажнения ($\text{ГТК} < 1,4$) наблюдаются на севере и северо-востоке юга Западной Сибири.

Засушливые условия ($ГТК < 0,8$) отмечаются на юго-западе территории (Рубцовск – Иртышск – Костанай). Этот район считается самым засушливым на территории юга Западной Сибири, где на фоне высоких температур количество осадков незначительно.

В целом для исследуемой территории Западной Сибири среднее значение индекса ГТК равно 1,2. Это говорит о том, что большая часть территории юга Западной Сибири находится в оптимально благоприятных условиях за время активной вегетации.

В результате проведенных исследований с помощью иерархического кластерного (формального) анализа объективным способом удалось получить научно обоснованную классификацию, хорошо согласующуюся с уже известной, физически обоснованной шкалой классификации ГТК. Кроме того, представленный метод кластерного анализа позволил нам сэкономить время расчета и сжать информацию, что является важным фактором в условиях использования больших массивов данных. Кластерный анализ также подтверждает то, что шкала классификации ГТК по тепловлагообеспеченности имеет статистическую значимость и может применяться для объективного районирования территории.

Важность учета погодно-климатических факторов в сельскохозяйственном производстве однозначна и она еще больше возрастает из-за сокращения числа зернопроизводящих экономических районов в России в целом и на юге Западной Сибири в частности. По сравнению с СССР продовольственная безопасность страны снизилась в два раза [10. С. 161]. В этом отношении высокая степень зависимости урожая от погодно-климатических условий свойственна как зерновым, так и техническим и овощным культурам, сенокосным травам.

При рассмотрении большинства сценариев изменения климата отмечается продолжение потепления (от 4°C на конец столетия при жестком сценарии А2, до $1,8^{\circ}\text{C}$ в щадящем сценарии В1). В тех же моделях неопределенность поведения осадков существенно больше, но большинство сценариев дает увеличение осадков жидкой фазы для умеренных широт, что может неоднозначно отразиться на температурно-влажностном режиме в этих широтах. Комплексный показатель ГТК возможно самостоятельно использовать для моделирования аномалий индекса в будущем, используя, соответственно, надежную метеорологическую информацию. В целом по результатам прогностических оценок изменения ГТК можно предположить, что границы агроклиматических поясов, выделяемых по совокупности тепла и влаги, будут претерпевать незначительные изменения. Вероятнее, что наиболее комфортными в отношении колебаний ГТК будут северные территории исследования с оптимальными и избыточными условиями тепловлагообеспечения. При этом на территориях с недостаточными и засушливыми значениями ГТК возрастет опасность повторяемости засух различных интенсивностей.

Как это отразится на условиях произрастания сельскохозяйственных культур, сказать трудно. Условия произрастания культур изменяются и методы агротехники интенсивно развиваются. Сложившаяся зона рискованного земледелия, очевидно, расширится к северу. Если судить только по ГТК, то площади комфортного произрастания теплолюбивых культур также расширятся на север. Таким образом, в условиях сохранения современных тенденций изменения климата юг Западной Сибири в ближайшие 100–200 лет имеет все предпосылки к повышению агроклиматического потенциала за счет северных и восточных территорий.

ЛИТЕРАТУРА

1. Гордеев А.В. и др. Биоклиматический потенциал России: теория и практика. М. : Товарищество научных изданий КМК, 2006. 508 с.
2. Гридасов В.Ф. Оценка влагообеспеченности сельского хозяйства с помощью агрогидрологических свойств почв // Труды ВНИИСХМ. 2006. № 35. С. 178–184.
3. Зойдзе Е.К., Хомякова Т.В. Оценка засушливых явлений в Российской Федерации // Труды ВНИИСХМ. 1993. № 33. С. 133–144.
4. Жамбю М. Иерархический кластерный анализ и соответствия. М. : Финансы и статистика, 1988. 344 с.
5. Груза Г.В. О типизации форм циркуляции в умеренных широтах Северного полушария по положению осевой изогипсы ВФЗ на поверхности 500 гПа // Метеорология и гидрология. 1996. № 2. С. 5–12.
6. Эзау И.Н. Кластерный анализ данных наблюдений и результатов численных экспериментов с моделью ОЦА // Метеорология и гидрология. 1995. № 12. С. 40–53.
7. Золотокрылин А.Н., Коняев К.В., Эзау И.Н. Сравнение синоптической и формальной классификаций крупномасштабной циркуляции атмосферы Северного полушария // Метеорология и гидрология. 1988. № 12. С. 34–44.
8. Гмурман В.Е. Теория вероятностей и математическая статистика. М. : Высшая школа, 1977. 479 с.
9. Зойдзе Е.К. Основы оперативной системы оценки развития засух и опыт ее экспериментальной эксплуатации // Труды ВНИИСХМ. 2002. № 34. С. 48–66.
10. Кислов А.В. и др. Прогноз климатической ресурсообеспеченности Восточно-Европейской равнины в условиях потепления XXI века. М. : МАКС Пресс, 2008. 292 с.

Статья представлена научной редакцией «Науки о Земле» 20 апреля 2012 г.