

С. Г. Катаев, А. И. Кусков

ИССЛЕДОВАНИЕ ОЗОННЫХ ПОЛЕЙ НАД ТЕРРИТОРИЕЙ РОССИИ И СОПРЕДЕЛЬНЫХ ГОСУДАРСТВ. II. КЛАССИФИКАЦИЯ СОСТАВЛЯЮЩИХ ПОЛЕЙ ОЗОНА

Проблема выявления закономерностей эволюции озона состоит в том, что данные измерений озона формируются во временные ряды, где каждый член ряда представляет собой среднemesячное содержание озона в слое над станцией (ОСО). Поля ОСО, как следует из предыдущей работы, характеризуются значительной пространственной изменчивостью, которая обусловлена как общими, так и региональными процессами, а также неоднородностью антропогенных воздействий. Все это побуждает провести классификацию полей озона, которая позволила бы оценить региональный характер, определить компактные районы с квазиоднородными характеристиками поля ОСО.

В данной статье решается задача определения внутренней структуры поля общего содержания озона изучаемого региона (т. е. нахождения областей с похожей зависимостью). С этой целью был проведен кластерный анализ, в котором в качестве объектов выступали 45 озонметрических станций, а в качестве признаков – средние месячные значения озона $T_{ik}, i = 1, \dots, N_s; k = 1, \dots, N_p$ за 16 лет ($N_s = 45, N_p = 192$). Список озонметрических станций с соответствующей нумерацией приведен в табл. 1.

Таблица 1

Нумерованный список станций

1	Абастумани	16	Караганда	31	П.-Камчатский
2	Алма-Ата	17	Киев	32	Печера
3	Аральское море	18	о. Котельный	33	Рига
4	Архангельск	19	Красноярск	34	Екатеринбург
5	Ашхабад	20	Куйбышев	35	Семипалатинск
6	Большая Елань	21	Санкт-Петербург	36	Сковородино
7	Витим	22	Львов	37	Тбилиси
8	Владивосток	23	Марково	38	Тикси
9	Воронеж	24	Москва	39	Тура
10	Гурьев	25	Мурманск	40	Феодосия
11	о. Диксон	26	Нагаево	41	Ханты-Мансийск
12	Душанбе	27	Ник.-на-Амуре	42	о. Хейса
13	Игарка	28	Одесса	43	Цимлянск
14	Иркутск	29	Оленек	44	Чарджоу
15	Капустин Яр	30	Омск	45	Якутск

Классификация составляющих поля ОСО по какому-либо одному признаку является проблематичной. Это обусловлено тем, что ни одна составляющая поля не описывается одним признаком. Поле долговременного тренда определяется двумя признаками: относительным трендом и константой, которая изменяется от пункта к пункту. Поле сезонной составляющей обусловлено также двумя признаками – амплитудой годового хода и фазовым смещением. Поле аномалий характеризуется только дисперсией со средним значением, равным нулю. Ясно, что одинаковые средние и одинаковые дисперсии не являются признаком коррелированности рядов. В этой связи классификация должна проводиться с учетом временной связности рядов.

Основную задачу таксономии в приложении к этой проблеме можно сформулировать следующим обра-

зом: необходимо разделить N_s объектов (озонметрических станций) на небольшое число групп, руководствуясь при этом некоторым критерием, по которому конкретный объект будет отнесен к той или иной группе. В качестве такого критерия обычно выбирают "близость" объектов в N_p -мерном пространстве признаков R_p . В этом пространстве каждый объект номера i , ($i = 1, \dots, N_s$) будет отображаться в точку

с координатами T_{ik} , $k = 1, \dots, N_p$. Тогда похожие по свойствам объекты будут локализоваться, образуя своеобразное "сгущение", называемое таксоном или классом, а значительно отличающиеся друг от друга объекты попадут в разные сгустки. Степень близости

может быть описана количественно – введением в R_T метрики. В качестве меры длины между этими точка-

ми используют различные метрики: [1]

$$\text{Евклидова мера: } D_{ij}^E = \left\{ \sum_{k=1}^{N_p} (T_{ik} - T_{jk})^2 \right\}^{1/2}$$

$$\text{Мера Минковского: } D_{ij}^M = \left\{ \sum_{k=1}^{N_p} |T_{ik} - T_{jk}|^l \right\}^{1/l}$$

$$\text{Мера Хемминга: } D_{ij}^H = \sum_{k=1}^{N_p} |T_{ik} - T_{jk}|$$

Процедура таксономии проводилась с двумя алгоритмами $A_{\text{сф}}$, $A_{\text{св}}$; для вычисления расстояний между станциями использовались меры Евклида и Хемминга.

В первом алгоритме $A_{\text{сф}}$ используются идеи, изложенные в [2]. Этот алгоритм позволяет находить таксоны сферической формы, причем радиусы получаемых кластеров и их число находятся в процессе вычисления. При этом допускается, чтобы некоторое число объектов не попало ни в один из классов. Второй алгоритм $A_{\text{св}}$ выделяет среди всех объектов группы, в которых элементы связаны друг с другом сильнее, нежели со всеми остальными. Идея, реализуемая в этом алгоритме, восходит к агломеративным методам [1] и наиболее просто может быть описана на языке теории графов следующим образом. Соотнесем каждой станции вершину x_i некоторого графа G , а стоимости ребра (x_i, x_j) , связывающего вершины x_i и x_j , припишем величину, равную расстоянию D^H (D^E) между соответствующими станциями. Если теперь последовательно удалить из графа ребра стоимостью D , большей некоторого порогового значения D_p , то, начиная с $D = D_{\text{max}}$, исходный граф станет несвязным. Дальнейшее уменьшение порога приводит к увеличению числа компонент связности, среди которых будут подграфы с различным числом вершин и изолированные вершины. Каждый такой k -тый подграф можно охарактеризовать двумя пороговыми значениями: $D : D_{\text{max}}^k$ – значение порога, при котором этот подграф отделяется от остального графа; и D_{min}^k – значение порога, при котором из этого подграфа выходит одна вершина или сразу несколько вершин. Таким образом, величина D_{min}^k в какой-то степени определяет степень компактности k -го подграфа; чем меньше эта величина, тем сильнее связь между всеми элементами этого кластера. D_{max}^k для k -го кластера показывает насколько он связан с остальной частью графа, а максимальная величина этого параметра $D_{\text{max}} = \max(D^k), k = 1, N_d$, т. е. значение порога, при котором из графа выделяется первый подграф, характеризует степень связности всего графа. По раз-

ности $\eta = D_{\text{max}}^k - D_{\text{min}}^k$ можно судить о степени изолированности этой структуры от остальной части графа (свособразное "время жизни" k -го таксона).

Выбор меры длины зависит от вида данных и желания исследователя, причем нет однозначного соответствия между типом экспериментальных данных и применяемой метрикой D_{ij} . Результат классификации определяется тем, как точки группируются в пространстве признаков, а это зависит от используемой при этом метрики и алгоритма. Соответственно и сама процедура в общем случае не является однозначной. Выявление логики в расположении точек в кластерах составляет задачу нахождения скрытых эмпирических закономерностей в массиве данных измерений.

Проведенный таксономический анализ показал, что использование расстояний по Евклиду D^E и по Хеммингу D^H приводит к качественно подобным результатам, поэтому все данные, помещенные в таблицах, даются только для расчетов с D^H . В расчетах матрица D_{ij} нормировалась, т. е. максимальное расстояние между станциями принималось за единицу.

В табл. 2–5 приведены результаты кластерного анализа полей ОСО и его составляющих с использованием алгоритма "связи" АВ. В таблицах, наряду с уже упомянутыми величинами, помещены и другие параметры, характеризующие как отдельный класс, так и все разбиение в целом. N – номер класса, N_{st} – число станций в этом классе, $T_{\text{ср}}$ – среднее значение, а S_q – среднееквадратичное отклонение, S_i – среднее расстояние между станциями в i -ом классе, D_{sum} – сумма средних расстояний S_i ; $D_{\text{sum}} = \sum_{i=1}^{N_d} S_i$, N_d – число классов. Величина D_{sum} характеризует компактность разбиения объектов на классы, чем меньше D_{sum} , тем лучше проведена процедура таксономии, и поэтому ее часто используют в роли функционала качества, т. е. критерия, по которому среди всевозможных разбиений выбирается оптимальное. Результаты таксономии, полученные для обоих алгоритмов, качественно совпадают. И это свидетельствует об устойчивости выяв-

ленных классов. По критерию качества D_{sum} лучшая классификация получилась для алгоритма A_B , поэто-

му дальнейшее обсуждение будем вести для результатов, полученных именно с этим алгоритмом.

Суммарный озон

При постепенном уменьшении величины порога D_p , с некоторого значения все объекты разделяются на 2 класса. Причем граница раздела проходит по территории Западной Сибири. Дальнейшее уменьшение D_p приводит к возникновению 5 классов. На рис. 1 показаны пространственные области, относящиеся к различным классам с указанием номера класса и средней величины озона для него. Изолированные, т. е. не попавшие ни в один из классов станции, помечены треугольником и могут рассматриваться как самостоятельные классы. Параметры, характеризующие разбиение, приведены в табл. 2. Используя среднее значение длины связи S в качестве критерия компактности таксона, можно говорить, что "все

классы достаточно компактны ($0,193 < S < 0,265$). Следует также заметить, что третий класс, охватывающий большую часть Европейской территории России и сопредельных государств, является наиболее "долгоживущим", время жизни для него составляет $\eta = 0,039$, а второй класс – наиболее изолирован от остальных, поскольку он выделяется первым с максимальным пороговым значением ($D_{max} = 0,263$) и существует достаточно долго $\eta = 0,030$. Кроме того, видно, что пространственная компактность классов и их размер хорошо коррелируют со средней длиной связи: чем больше S , тем большую территорию занимает класс.

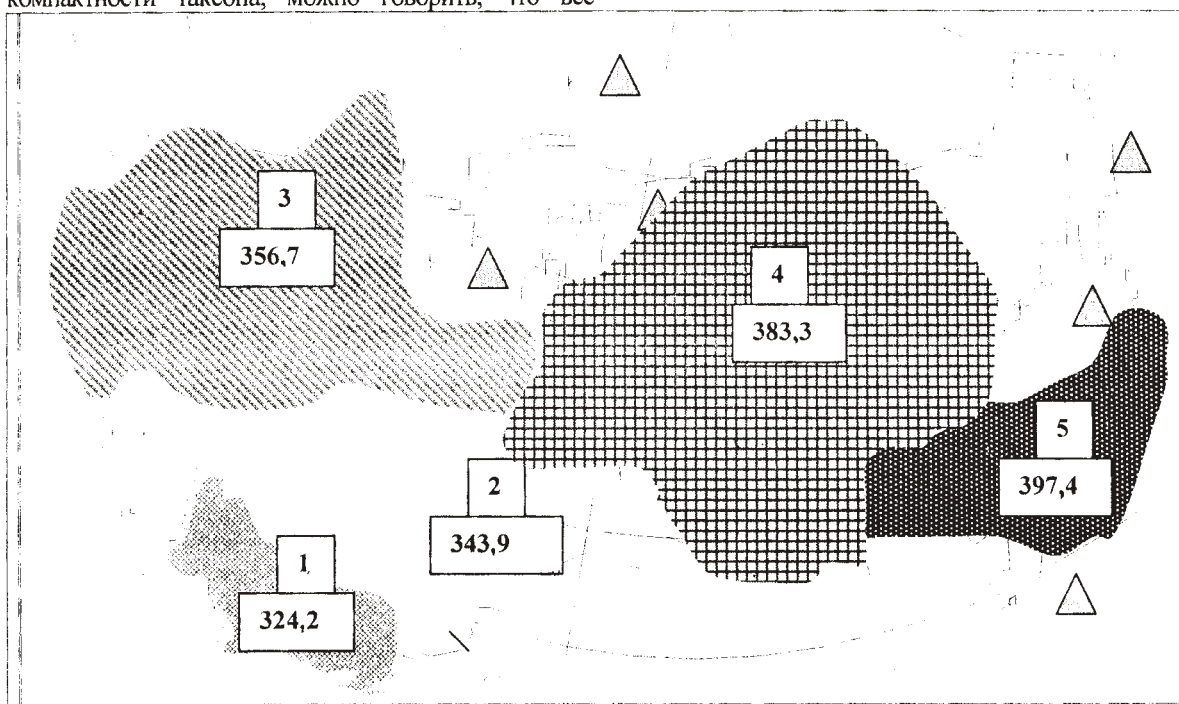


Рис. 1. Распределение станций по таксонам для исходного поля ОСО на территории России и сопредельных государств

Таблица 2

Суммарные данные по полному озону

N	N_{st}	Номера станций	T_{cp}	S_q	D_{max}	D_{min}	η	S	D_{sum}	
1	3	5, 12, 44	324,2	23,30	0,238	0,223	0,015	0,193	1,171	
2	7	1, 2, 3, 10, 16, 35, 37	343,8	29,72	0,263	0,233	0,030	0,241		
3	16	4, 9, 15, 17, 20, 21, 22, 24, 25, 28, 30, 33, 34, 40, 41, 43	356,7	34,93	0,241	0,202	0,039	0,265		
4	8	7, 13, 14, 18, 29, 38, 39, 45	383,3	51,67	0,235	0,211	0,024	0,264		
5	4	6, 27, 31, 36	397,4	50,32	0,236	0,199	0,011	0,208		
	7	8, 11, 19, 23, 26, 32, 42	Изолированные станции							

Полученное разбиение подтверждает, что средние значения не могут являться определяющими признаками при классификации ОСО. На самом деле, как видно из табл. 2, в третий класс попадают станции с номерами 40 и 41 (Феодосия и Ханты-Мансийск), у которых среднее значение ОСО, T_{cp} составляет соот-

ветственно 350,48 и 357,38. В то же время на станциях с номерами 10 (Гурьев), 35 (Семипалатинск), принадлежащих другому – второму – классу, средние значения находятся внутри интервала, характерного для 3-его класса и составляют соответственно величины 351,37 и 354,32.

Тренд

Долговременный тренд ведет себя при уменьшении порога немного иначе, чем ОСО. Первыми при $D_p = 0,097$ из графа выходят 3 станции: Абастумани, Чарджоу, Душанбе, которые как класс существуют достаточно долго и входят затем в результирующую классификацию. И только после этого оставшийся подграф раскалывается по Западной Сибири на 2 час-

ти. Оптимальное разбиение поля тренда содержит 5 классов (см. рис. 2), параметры таксонов приведены в табл. 3. Обращает на себя внимание тот факт, что выделенные классы имеют очень хорошую локализацию. Средняя длина связи для всех таксонов тренда на порядок меньше, чем для полного озона.

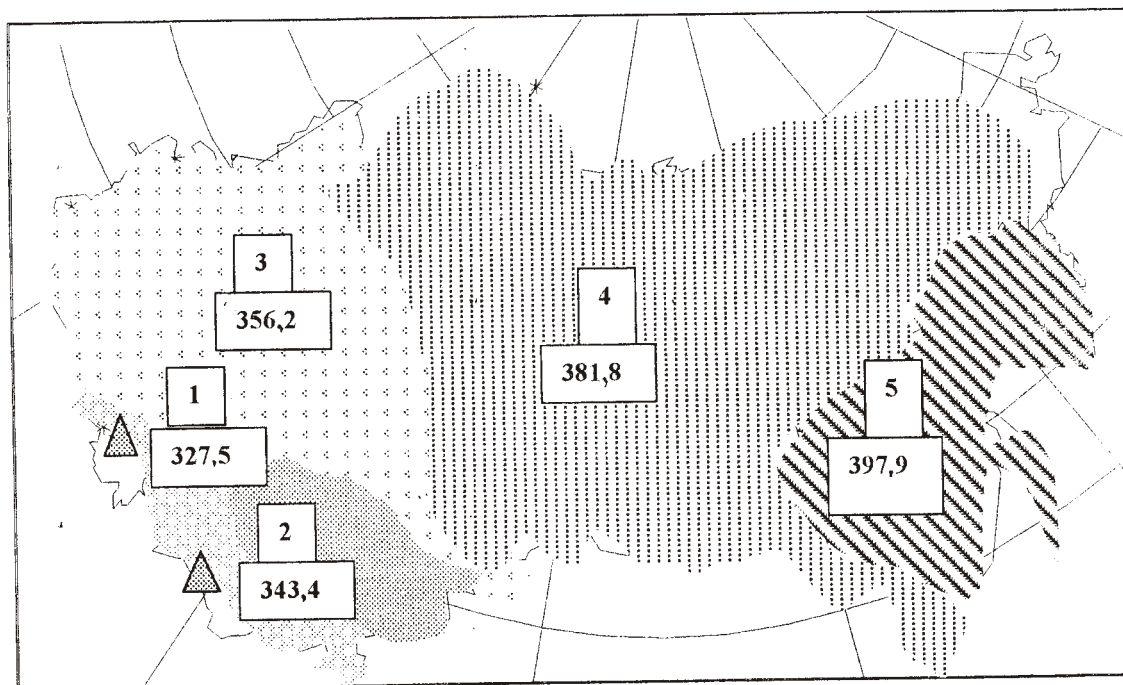


Рис. 2. Распределение станций по таксонам для долговременного тренда ОСО на территории России и сопредельных государств. (Цифрами помечены номера классов и средние значения для них.)

Таблица 3

Суммарные данные по тренду

N	N _{st}	Номера станций	T_{cp}	Sq	D_{max}	D_{min}	η	S	D_{sum}	
1	3	1, 12, 44	327,5	1,84	0,097	0,027	0,07	0,027	0,244	
2	2	2, 3	343,4	3,37	0,083	0,007	0,076	0,007		
3	20	4,9,10,15,16,17,20,21, 22,24,25,28,30,32,33,3 4,35,40,41,43	356,3	5,97	0,083	0,037	0,046	0,068		
4	13	7,8,11,13,14,18,19,23, 29,38,39,42,45	381,8	4,74	0,065	0,052	0,013	0,091		
5	5	6, 26, 27, 31, 36	397,9	6,80	0,065	0,051	0,014	0,051		
	2	5, 37	Изолированные станции							

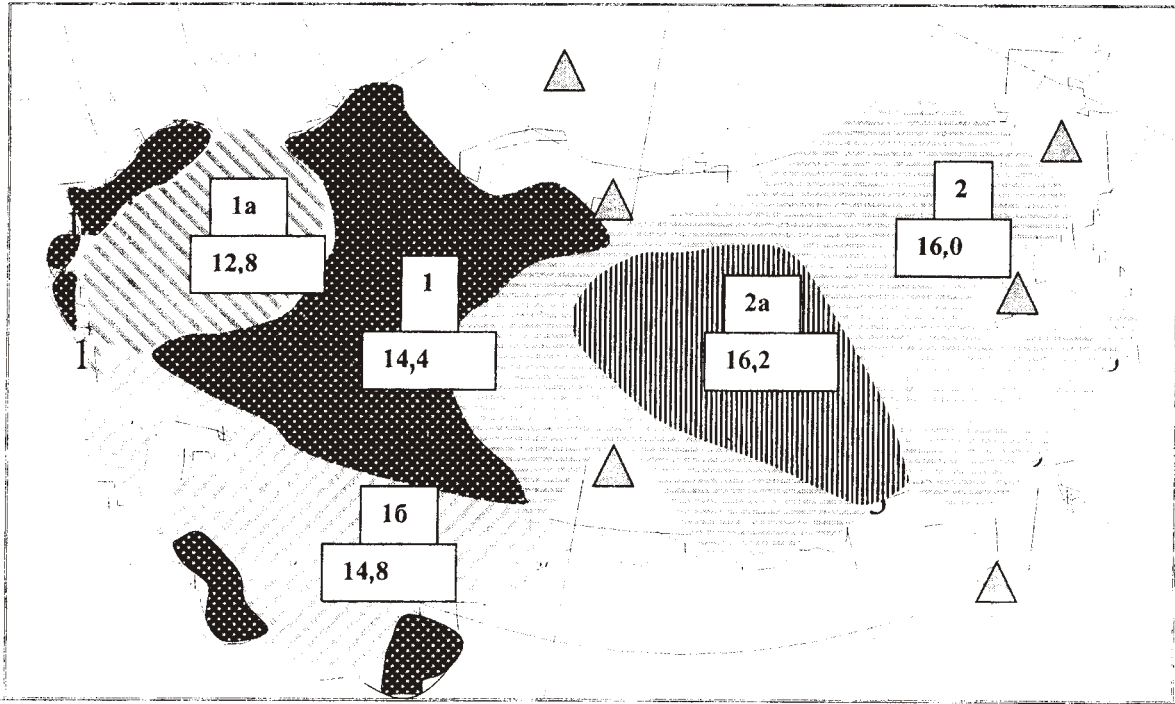


Рис. 3. Распределение станций по таксонам для аномалий ОСО на территории России и сопредельных государств. (Цифрами помечены номера классов и значения среднеквадратичного отклонения для них)

Аномалии

Аномальная составляющая, так же как и тренд, при значении порога 0,544 делится на два таксона.

Первый класс локализуется над Европейской территорией России (ЕТР) и захватывает часть Западной Сибири. Расстояние внутри класса 0,637. Второй – располагается от 85° в.д. и охватывает весь Дальний Восток. Расстояние внутри класса примерно такое же 0,641. Более детальное рассмотрение позволяет выделить внутри каждого класса небольшие, но хорошо локализованные в реальном про-

странстве таксоны. В первом классе содержится два таксона 1а и 1б. Один располагается вблизи Москвы и включает в себя Воронеж, Киев, Санкт-Петербург, Феодосию и Одессу. Другой охватывает районы Аральского моря, Средней Азии, Казахстана и Кавказа (см. рис. 3 и табл. 4). Классы для аномальной составляющей очень сильно размыты, средняя длина связи почти у всех таксонов больше 0,5. Это говорит о сильном влиянии регионального фактора на формирование аномалий.

Таблица 4

Суммарные данные по аномалии

N	N _{st}	Номера станций	T _{cp}	Sq	D _{max}	D _{min}	η	S	D _{sum}
1	26	1,2,3,4,5,9,10,12, 15,16,17,20,21,22, 24,25,28,30,32,33, 34,35,37,40,43,44	0,	14,4	0,544	0,520	0,024	0,637	1,278
2	13	6,7,13,14,18,27,29, 31,36,38,39,41,45	0,	16,0	0,544	0,526	0,018	0,641	
1a	8	1,3,10,12,16,35, 37,44	0,	12,8	0,458	0,442	0,016	0,496	1,512
16	6	9,17,21,24,28,40	0,	13,8	0,458	0,432	0,026	0,487	
2a	5	7,13,29,36,39	0,	16,2	0,474	0,444	0,030	0,529	

Сезонный ход

Процесс выделения классов для сезонного хода аналогичен тренду и аномалиям: при пороге, равном 0,544 выделяется 2 больших класса, по 26 и 17 элементов соответственно. Первый класс содержит 3 таксона: 1а, 1б и 1в, а во втором есть только один хорошо изолированный таксон 2а. Распределение по классам изображено на рис. 5. Из рисунка видно, что про-

странственные области, попадающие в таксоны первого класса, примерно одинаковы по площади и имеют преимущественно широтную ориентацию. Напротив, таксон 2а на Дальнем Востоке обладает ярко выраженным меридиональным характером.

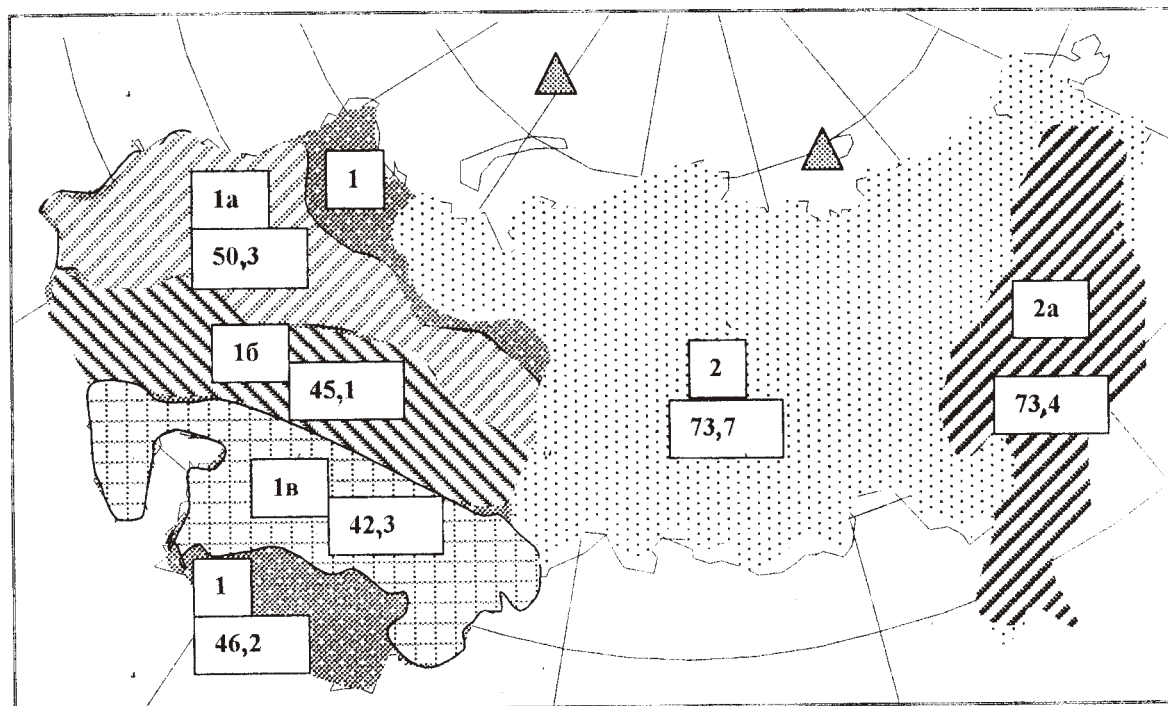


Рис. 4. Распределение станций по таксонам для сезонного хода ОСО на территории России и сопредельных государств. (Цифрами помечены номера классов и значения амплитуды для него)

Таблица 5

Суммарные данные по сезонному ходу

N	N _{ст}	Номера станций	T _{ср}	Sq	D _{max}	D _{min}	η	S	D _{sum}
1	26	1,2,3,4,5,9,10,12, 15,16,17,20,21,22, 24,25,28,30,33,34, 35,37,40,41,43,44	0,	14,4	0,150	0,120	0,030	0,247	0,506
2	17	6,7,8,11,13,14, 19,23,26,27,29, 31,32,36,38,39,45	0,	16,0	0,150	0,110	0,040	0,259	
1а	6	17,21,22,24,33,41	0,	32,8	0,065	0,055	0,010	0,095	0,433
1б	8	9,15,20,28, 30,34,40,43	0,	28,9	0,065	0,055	0,010	0,088	
1в	7	1,2,3,10,16,35,37	0,	27,1	0,075	0,055	0,020	0,083	
2а	6	6,8,23,26,27,31	0,	47,1	0,125	0,05	0,075	0,057	

В табл. 6 приведены расстояния между классами для ОСО и каждой составляющей. Под S_{kl} мы понимаем среднее расстояние между всевозможными

парами объектов, принадлежащих классам k и l за вычетом расстояний между объектами, находящимися в одном классе.

Таблица 6

Расстояния между классами для исследуемых параметров

Тренд		Аномалия		Сезонность		Полная	
классы	S_{kl}	классы	S_{kl}	классы	S_{kl}	классы	S_{kl}
1-2	0,193	1-2	0,742	1-2	0,339	1-2	0,335
1-3	0,349	1a-1б	0,610	1a-1б	0,179	1-3	0,333
1-4	0,858	1a-2a	0,710	1a-1в	0,339	1-4	0,708
1-5	0,662	1б-2a	0,728	1б-1в	0,189	1-5	0,592
2-3	0,156			1a-2a	0,694	2-3	0,509
2-4	0,664			1б-2a	0,616	2-4	0,632
2-5	0,468			1в-2a	0,515	2-5	0,500
3-4	0,508					3-4	0,916
3-5	0,312					3-5	0,808
4-5	0,197					4-5	0,306

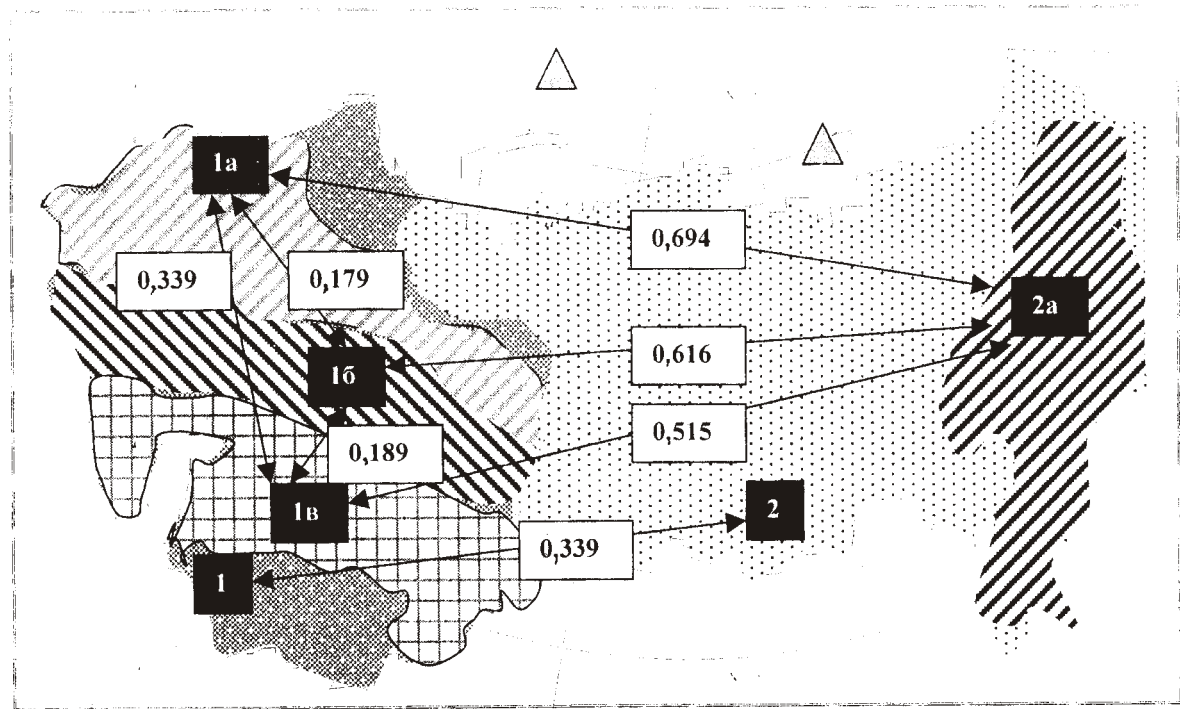


Рис. 5. Иллюстрация связи расстояний между классами и их реальным положением на местности для сезонного хода. (Цифры на темном фоне помечают номера классов, а числа на стрелках указывают расстояния между таксонами)

Видно, что расстояния в пространстве признаков между классами согласуются с расстояниями на местности между группами станций, принадлежащих различным таксонам.

Это факт иллюстрируется на рис. 5 для сезонного хода ОСО. Кроме того, отмечается корреляция расстояний между таксонами S_{kl} и разностью

$\Delta T_{kl} = T_k - T_l$ их средних значений ОСО: чем больше ΔT_{kl} , тем больше расстояние S_{kl} . Однако это только тенденция. Например, расстояния между первым и пятым классами S_{15} и S_{24} для тренда при-

близительно одинаковы: 0,662 и 0,664 соответственно, а разность между средними значениями отличается почти в 2 раза: $\Delta T_{15} = 70,4$; $\Delta T_{24} = 38,4$.

Заключение

В результате проведенного исследования было установлено:

выделение составляющих разного временного масштаба и последующая классификация дают возможность с любой степенью подробности (вплоть до отдельной станции) выявлять районы с репрезентативным поведением озона и, как следствие, выявлять факторы, формирующие это поведение. Так, мы видим, что все составляющие ОСО в первом приближении образуют 2 пространственные области, с границей раздела по Западной Сибири. Но тонкая структура каждой области для каждой составляющей – своя.

Компактная пространственная локализация таксонов, охватывающая значительные территории для раз-

ных составляющих ОСО, свидетельствует, во-первых, о существовании общего для рассматриваемого региона механизма формирования ОСО и, во-вторых, о влиянии на него региональных факторов.

Независимость результатов классификации от выбора метода таксономии и типа метрики позволяет говорить о существовании устойчивых классов. Это утверждение справедливо как для исходного поля, так и каждой из его составляющих.

Полученные результаты могут служить основой для последующего моделирования эволюции озонового слоя с учетом его региональных особенностей.

Литература

1. Дюк В.А. Компьютерная психодиагностика. – С.-Пб.: Братство, 1994. – 364 с.
2. Загоруйко Н.Г., Елкина В.Н., Лбов Г.С. Алгоритмы обнаружения эмпирических закономерностей. – Новосибирск: Наука, 1985. – 105 с.

УДК 539.2.001.5:548.313/.4

Т.Я. Голубенко

ТЕОРЕТИКО-ГРУППОВОЙ АНАЛИЗ ДОМЕННОЙ СТРУКТУРЫ ПРИ УПОРЯДОЧЕНИИ АТОМОВ В ГЦК ТВЕРДОМ РАСТВОРЕ

Образование сверхструктуры при упорядочении атомов в твердом растворе в большинстве случаев связано с появлением сложной доменной структуры [1]. Симметричные причины этого явления обсуждаются в настоящей статье.

В фазовом переходе типа упорядочения происходит потеря части элементов симметрии по сравнению со средней решеткой неупорядоченного сплава [2]. В соответствии с принципом Кюри [3] система стремится «восстановить» утраченную при переходе симметрию с помощью нового типа дефектов, которыми в упорядоченном сплаве являются антифазные границы (АФГ) и ориентационные двойники превращения. Теоретико-групповой анализ способов вложения низкосимметричной группы G_0 в высокосимметричную G проводился многими авторами [4–8] и уже давно стал традиционным при исследовании фазовых переходов. Однако для сверхструктур, образующихся при упорядочении в ГЦК твердом растворе, такой анализ отсутствует в литературе. Исключение составляет лишь

сверхструктура $D1_a$, рассмотренная в работе [8]. К тому же, в отличие от традиционных подходов, здесь проведена современная классификация доменной структуры путем построения варианто-образующей группы (VGG) [8].

С теоретико-групповой точки зрения равновероятно образуются все варианты упорядоченной фазы (домены), если они связаны операцией симметрии разупорядоченной матрицы, не являющейся элементом симметрии упорядоченной. Операторы, которые производят все варианты направлений (ориентационные домены) исходя из первоначально данного, образуют группу VGG. Для построения VGG в настоящей работе необходимо было разложить точечную группу O_h разупорядоченной структуры на смежные классы по подгруппе H_i (точечная группа i -го варианта упорядоченной фазы):