

В.Т. Калайда, В.И. Хамарин

МЕТОД ОБЪЕКТИВНОГО РАСПОЗНАВАНИЯ ПОЛЯРНЫХ И ТРОПИЧЕСКИХ ПРОФИЛЕЙ ОЗОНА

Исследуются возможности метода построения оптимальной разделяющей гиперплоскости («обобщенный портрет») для классификации вертикальных профилей парциального давления озона. Показано, что этот метод может быть использован для решения данной задачи. Анализируются оптимальные признаки в задаче классификации и восстановления вертикальных профилей озона.

Необходимость классификации вертикальных профилей озона, как и других метеорологических элементов, обусловлена прикладными задачами анализа и прогноза погоды, дистанционного зондирования атмосферы и подстилающей поверхности, изучением климата Земли.

В настоящей работе рассматриваются возможности использования метода «обобщенный портрет» [1] для объективной классификации вертикальных профилей парциального давления озона. Существенной особенностью данного метода распознавания образов является его независимость от вида закона распределения вероятностей исследуемой величины. Как показано в [2], распределение вероятностей парциального давления озона в основном не соответствует нормальному закону, что ограничивает возможности использования методов автоматического распознавания образов. Кроме того, метод «обобщенный портрет» позволяет рассматривать векторы случайных величин, а не поля отдельных точек, что также очень важно в задачах классификации характеристик воздушных масс.

Как отмечалось в [3], для вертикального распределения озона характерны четыре типа: А – тропический, В – умеренный, С – полярный, Д – комбинированный. Наиболее четко проявляются эти типы при анализе средних профилей озона. Однако до настоящего времени отсутствуют количественные критерии классификации вертикальных профилей озона. Для решения данной задачи мы использовали метод «обобщенный портрет», основанный на построении оптимальной гиперплоскости.

Кратко остановимся на описании метода «обобщенный портрет», следуя [1]. Рассмотрим два конечных множества случайных векторов, а именно: данные многолетнего вертикального озонзондирования на экваториальной станции Кантон (множество $X = x_1, \dots, x_n$) и на полярной станции Гус-Бэй (множество $Y = y_1, \dots, y_m$). Выбор этих данных для классификации был обусловлен существенными отличиями вертикальных профилей парциального давления озона в тропических широтах от полярных. Данные озонзондирования на станции Кантон представляли первый класс, а на станции Гус-Бэй – второй. Множества X и Y разделимы ориентированной гиперплоскостью, если для некоторого $k < 1$ существует такой вектор φ , что выполняется неравенство:

$$\begin{aligned} x_i^T \varphi &\geq 1, \quad i = 1, \dots, n; \\ y_j^T \varphi &\leq 1, \quad j = 1, \dots, m. \end{aligned} \quad (1)$$

Если существует вектор φ , для которого выполняются неравенства (1), то существует и множество векторов φ , удовлетворяющих неравенствам (1). Оптимальный по модулю среди этих векторов называется «обобщенным портретом» [1].

Среди однопараметрического (по параметру $k < 1$) семейства векторов «обобщенного портрета» существует вектор φ_0 , определяющий такое направление, на котором проекции множества X и Y наиболее отстоят друг от друга:

$$\varphi_0 = \operatorname{argmax}_{\varphi} [\min_{x_i \in X} x_i^T \varphi - \max_{y_j \in Y} y_j^T \varphi]. \quad (2)$$

Этот вектор называется оптимальным, а полученная с его помощью разделяющая гиперплоскость

$$X\varphi_0 = C_0,$$

где

$$C_0 = \frac{\min_{x_i \in X} x_i^T \varphi_0 + \max_{y_j \in Y} y_j^T \varphi_0}{2} \quad (3)$$

является оптимальной разделяющей гиперплоскостью. Оптимальная гиперплоскость отделяет точки множества X (для этих точек $X_{\Phi_0} > C_0$) от точек множества Y (для этих точек $Y_{\Phi_0} < C_0$) и наиболее удалена от элементов объединенного множества $X \cup Y$.

Для классификации вертикальных профилей парциального давления озона был использован комплекс программ «FOP» [1]. На первом этапе численных экспериментов были отобраны 4 вектора наблюдений на станции Кантон, которые, на наш взгляд, наиболее близко подходят к эталонному. Эти векторы были использованы в обучающей выборке первого класса. Остальные 26 векторов станции Кантон были подвергнуты испытаниям на принадлежность к первому классу. Каждый вектор был задан 27 признаками, в качестве которых использовались значения парциального давления озона на 27 уровнях атмосферы: Земля, 975, 950, 925, 900, 875, 850, 825, 800, 750, 700, 650, 600, 500, 400, 300, 250, 225, 200, 175, 150, 100, 70, 50, 30, 20 и 10 гПа. Комплекс программ «FOP» при стандартных значениях параметров RK [1] произвел вычисление «обобщенного портрета» по 4 векторам и расклассифицировал остальные 26 векторов. Все 26 векторов были отнесены к первому классу, т. е. к тропическому типу. Такой результат вполне закономерен, так как вертикальное распределение озона в тропиках имеет довольно устойчивую стратификацию и относительно небольшую изменчивость.

На втором шаге к 30 векторам станции Кантон были добавлены 89 векторов полярной станции Гус–Бэй. В качестве обучающейся выборки первого класса были использованы все 30 векторов станции Кантон, а для второго класса – 5 наиболее характерных профилей полярного типа. В результате работы комплекса «FOP» 81 вектор станции Гус–Бэй был отнесен ко второму классу и 4 вектора этой выборки попали в первый класс. Как показал анализ, эти 4 вектора имели грубые ошибки измерений.

Таким образом, первые численные эксперименты по классифицированию вертикальных профилей озона показали, что комплекс «FOP» может удовлетворительно классифицировать два типа профилей: тропические и полярные.

С целью повышения надежности вычисления характеристик «обобщенного портрета» был проведен ряд экспериментов по наращиванию объема выборки первого класса. Эксперименты предусматривали следующие шаги. На первом шаге к экзаменуемой (рабочей) выборке добавлялись новые векторы, относящиеся к той или иной станции озонзондирования. Затем производилась их классификация по обучающимся выборкам станций Кантон и Гус–Бэй. После классификации выбранные векторы первого класса включались в выборку первого класса в качестве обучающихся. Таким образом производилось постепенное наращивание объема выборки первого класса. Всего в нашем распоряжении были 4 станции, наблюдения на которых можно было отнести к тропическому типу. Это станции: Тривандрум (8° с.ш., 77° в.д.), Нью-Дели (28° с.ш., 77° в.д.), Кагосима (32° с.ш., 131° в.д.), станция Кантон (3° ю.ш., 172° з.д.). Результаты классификации наблюдений на станции Тривандрум показали, что из 29 векторов лишь 5 не были отнесены к тропическому типу. На станции Нью-Дели количество таких профилей возросло до 24 из общего количества 92, и на станции Кагосима – 4 из 94.

Проведенный анализ вертикальных профилей озона, полученных на тропических станциях Кантон, Тривандрум, Нью–Дели и Кагосима, показал, что все они имеют общие закономерности, которые позволяют использовать их для автоматической классификации. В табл. 1 приведены параметры «обобщенного портрета», полученные при различных объемах выборки первого класса.

Анализ табл. 1 показывает, что с увеличением объема выборки растет количество градаций и их вклад в «обобщенный портрет». Кроме того, происходит перераспределение вклада градаций. Заслуживает особого рассмотрения вопрос о количестве признаков, их информативности и их вкладе в «обобщенный портрет». Как показывают данные табл. 1, этот вклад может значительно изменяться от выборки.

В прикладных задачах очень важно получать желаемый результат с минимальным объемом исходной информации. С целью минимизации пространства признаков, в качестве которых использовались результаты озонзондирования на 27 уровнях, был произведен выбор оптимальной совокупности уровней, обеспечивающих задачу классификации. Для этой цели был использован алгоритм NMIN [1]. Алгоритм NMIN предназначен для выделения в исходном пространстве признаков такого подпространства, в котором построенная гиперплоскость обладает минимальной гарантированной вероятностью ошибок квалификации. Поиск оптимального минимума осуществляется по критерию

$$P = \nu + \frac{d \left(\ln \frac{l}{d} + 1 \right) + \ln C_n^h - \ln \eta}{2l} \times \left(1 + \sqrt{1 + \frac{4\nu l}{d \left(\ln \frac{l}{d} + 1 \right) + \ln C_n^h - \ln \eta}} \right),$$

где P – вероятность ошибочной классификации с помощью построенной гиперплоскости; ν – частота ошибочной классификации; $d = \min\left(n^{-h}, \left[\frac{D^2}{\rho^2}\right] + 1, r\right)$ – внутренняя размерность задачи; n – размерность пространства; D – диаметр множества;

$$\rho = \frac{1}{\sqrt{2W(a)}};$$

r – число информативных векторов; l – количество векторов обучающей последовательности; h – минимальное число векторов обучающей последовательности, по которым раскладывается вектор направляющих косинусов разделяющей гиперплоскости.

Т а б л и ц а 1

П а р а м е т р ы о б о б щ е н н о г о п о р т р е т а

Кантон, Гус-Бэй				Кантон, Гус-Бэй, Тривандрум, Нью-Дели, Кагосима			
$ISP=1$	$LR=4, W=0,5327$			$ISP=16$	$LR=13$	$W=1,0538$	
Номер признака	Решающее правило с порогом $=-3,73$			Решающее правило с порогом $=1,37$			
	Градации			Градации			
	1	2	3	1	2	3	4
2	0,26	-3,27	3,02	-1,85	-4,22	5,71	0,36
3	3,27	-3,27	0	-1,85	-6,47	3,97	4,35
4	3,27	-3,27	0	12,99	-6,47	-6,89	0,36
5	3,27	-3,27	0	-4,09	6,42	-2,69	0,36
6	3,27	-3,27	0	-1,85	0,40	1,09	0,36
7	0,77	-0,77	0	-1,85	0,40	1,09	0,36
8	0,77	-0,77	0	-1,29	4,03	-4,20	1,09
9	0,77	-0,77	0	-1,85	0,40	1,09	0,36
10	0,77	-0,77	0	-1,85	0,40	1,09	0,36
11	0,77	-0,77	0	5,35	0,06	-4,20	-1,58
12	0,77	-0,77	0	-1,85	6,99	-1,47	-3,67
13	0,77	-0,77	0	6,36	-4,43	1,74	-3,67
14	0,77	-0,77	0	15,82	-12,15	0,0	-3,67
15	0,26	-0,26	0	2,80	3,11	-2,24	-3,67
16	0,26	-0,26	0	9,26	-3,35	-2,24	-3,67
17	0,77	-0,77	0	5,66	0,36	-2,35	-3,67
18	3,27	-3,27	0	12,38	-8,71	-3,67	
19	3,27	-3,27	0	7,00	-3,33	-3,67	
20	3,27	-3,27	0	9,79	-6,12	-3,67	
21	3,27	-3,27	0	7,00	-3,33	-3,67	
22	3,27	-3,27	0	22,71	-18,99	-3,72	
23	3,27	-3,27	0	16,24	-12,52	-3,72	
24	3,27	-3,27	0	22,04	-18,31	-3,72	
25	3,27	-3,27	0	18,82	6,70	-17,82	-7,71
26	-2,50	2,50	0	-13,70	10,46	-7,38	10,62
27	-2,67	2,67	0	-2,79	13,48	-10,69	
28	0,0	2,50	0	1,41	0,0	-1,41	

Точное решение данной задачи требует полного перебора по всем подпространствам, поэтому в алгоритме NMIN используются эвристические приемы последовательного улучшения оценки, которые существенно сокращают схему перебора.

Для реализации этой процедуры признаки упорядочиваются. Процедура упорядочения состоит в следующем. Используя обобщенный портрет ψ для каждого i -го признака ($i = 1, \dots, n$), вычисляется коэффициент K_i :

$$K_i = \sum_{t=1}^T (\Psi_i^t)^2,$$

где Ψ_i^t — значение координаты вектора „обобщенного портрета”, соответствующее t -й градации i -го признака, T — число градаций i -го признака.

Признаки упорядочиваются в соответствии с уменьшением значений коэффициентов K_i . Алгоритм NMIN состоит из трех модификаций. NMIN1 реализует «обратную пошаговую» процедуру, в которой последовательно из общего числа признаков исключается последний по порядку признак. Алгоритм NMIN2 реализует «прямую пошаговую» процедуру, в которой происходит последовательное добавление признаков. Алгоритм NMIN3 является комбинацией алгоритмов NMIN1 и NMIN2.

Таблица 2

Коэффициенты K упорядоченных признаков

K	№ признака	Уровни, гПа	P_3^A	P_3^C	ΔP
0,089	22	150	15	110	— 95
0,083	24	70	50	176	—126
0,078	25	50	100	160	60
0,046	26	30	146	116	30
0,043	23	100	18	155	—131
0,041	14	600	20	27	— 7
0,041	27	20	134	84	50
0,026	4	950	20	21	— 1
0,024	18	250	16	68	— 52
0,015	20	200	17	100	— 83
0,012	16	400	17	42	— 25
0,008	3	975	18	21	— 3
0,008	13	650	20	29	— 9
0,007	19	225	17	73	— 56
0,007	21	175	16	105	— 89
0,007	12	700	20	29	— 9
0,007	5	925	21	21	0
0,005	2	Земля	20	20	0
0,005	17	300	17	42	— 25
0,005	11	750	22	27	— 5
0,004	8	850	21	22	— 1
0,004	15	500	18	26	— 8
0,000	6	900	21	21	0
0,000	7	875	21	20	1
0,000	9	825	21	22	1
0,000	10	800	21	25	4
0,000	28	10	77	46	31

Используя алгоритмы NMIN, был произведен отбор оптимальных признаков для совокупностей вертикальных профилей озона типа А. Исходная выборка включала данные озонзондирования на 4 станциях, расположенных в тропическом поясе (класс 1), и данные озонзондирования на полярной станции Гус–Бэй (класс 2). В табл. 2 приведены упорядоченные коэффициенты K_i , отражающие расстояние между проекциями множеств X и \bar{X} на направление гиперплоскости, а также приведены средние профили парциального давления озона для первого (P_3^A) и второго (P_3^C) классов. Как видно из табл. 2, наиболее значимые коэффициенты K_i были получены на уровнях с максимальной разностью $\Delta P = P_3^A - P_3^C$, которая наблюдается на верхних рассматриваемых уровнях.

В табл. 3 приведены результаты выбора оптимальных признаков по трем алгоритмам: NMIN1, NMIN2, NMIN3. Все три практически обеспечивают одинаковые результаты. Оптимальные уровни в основном сосредоточены на высотах, где наблюдается максимальное содержание озона. Кроме того, в число оптимальных попали также несколько уровней из нижней тропосферы.

Оптимальные уровни

Уровни, гПа	950	925	600	300	250	225	175	150	100	70	50	30	10
NMIN1			+					+	+	+	+		+
NMIN2	+		+		+			+	+	+	+	+	+
NMIN3			+		+			+	+	+	+		+
Кантон		+				+						+	+
Гус-Бэй		+	+	+			+					+	+

Таблица 4

Факторные нагрузки

Давле- ние (гПа)	Станции									
	Гус-Бэй						Кантон			
	Факторы						Факторы			
	i	2	3	4	5	6	1	2	3	4
Земля	0,88	-0,25	0,17	-0,16	0,03	0,06	0,98	-0,25	0,07	-0,17
975	0,94	-0,13	0,09	-0,17	-0,01	0,11	0,94	0,16	-0,08	-0,08
950	0,96	-0,02	0,02	-0,10	-0,07	0,10	0,96	0,16	-0,07	-0,03
925	0,96	0,06	-0,03	0,03	-0,09	0,11	0,97	0,21	-0,10	-0,02
900	0,94	0,01	-0,09	0,08	0,03	0,22	0,96	0,23	-0,12	-0,03
875	0,94	0,09	-0,07	0,08	-0,05	0,24	0,96	0,24	-0,11	0,02
850	0,83	0,21	-0,11	0,06	-0,22	0,29	0,94	0,28	-0,13	0,05
825	0,74	0,00	-0,09	0,06	-0,15	0,46	0,92	0,32	-0,12	0,06
800	0,76	0,11	-0,16	0,02	0,02	0,53	0,90	0,35	-0,13	0,08
750	0,62	0,06	-0,12	0,00	0,04	0,69	0,88	0,36	-0,15	0,09
700	0,57	0,18	-0,10	0,00	0,09	0,73	0,84	0,39	-0,18	0,12
650	0,43	0,18	0,04	0,00	0,09	0,78	0,80	0,48	-0,25	0,12
600	0,36	0,21	0,09	-0,04	0,11	0,84	0,74	0,55	-0,26	0,13
500	0,10	0,23	0,18	0,02	0,09	0,82	0,48	0,60	-0,30	0,12
400	-0,04	0,09	0,76	-0,14	0,13	0,24	0,15	0,41	-0,40	-0,07
300	-0,11	0,07	0,91	-0,10	-0,11	-0,12	0,31	0,81	-0,34	0,00
250	-0,01	-0,04	0,79	0,23	-0,40	-0,01	0,24	0,92	-0,23	-0,09
225	0,04	-0,02	0,55	0,48	-0,51	0,11	0,16	0,94	-0,13	-0,16
200	0,17	-0,16	0,29	0,45	-0,75	-0,03	0,10	0,93	-0,05	-0,17
175	0,08	-0,38	0,05	-0,12	-0,82	-0,16	0,20	0,94	-0,08	-0,11
150	0,08	-0,54	0,07	-0,11	-0,72	-0,14	0,30	0,88	-0,11	-0,01
100	0,09	-0,71	0,06	0,05	-0,51	-0,26	0,38	0,57	0,20	0,05
70	-0,05	0,86	-0,07	-0,09	-0,31	-0,14	-0,25	-0,20	0,79	-0,04
50	0,10	-0,88	-0,05	0,04	-0,23	-0,10	-0,12	-0,01	0,88	-0,15
30	-0,02	0,0	-0,07	0,14	0,00	-0,17	-0,15	-0,16	0,84	0,04
220	-0,19	-0,81	0,00	0,39	-0,01	-0,14	-0,06	-0,26	-0,12	0,97
10	-0,09	-0,25	-0,09	0,93	0,00	0,00	-0,48	0,22	-0,70	-0,65
VAR	8,2	4,4	2,6	1,7	2,8	4,2	11,8	7,4	3,4	1,6
%	35	18	11	7	12	17	49	30	14	6

Алгоритм NMIN осуществляет выбор оптимальных признаков, необходимых для уверенной классификации векторов. Представляет интерес исследовать вопрос: в какой степени выбранные оптимальные признаки могут быть использованы для восстановления вертикальной структуры профилей озона? Последнее очень важно для оптимизации схемы измерений, а также для схемы подготовки априорных данных об озоне. Немаловажное значение имеет и

физическая интерпретация процессов, определяющих содержание озона на оптимальных уровнях. Для исследования этого вопроса использовался метод главных компонент и факторного анализа [4], в частности, процедура FACTOR системы статистического анализа (SAS) [4].

В табл. 4 приведены факторные нагрузки, полученные при обработке данных озонирования на станциях Гус-Бэй и Кантон. Как видно из табл. 4, вертикальный профиль озона можно восстановить, используя 4–6 главных компонент корреляционных матриц озона. Используя максимальные факторные нагрузки, были выбраны оптимальные уровни для решения задачи восстановления профилей озона, которые приведены в табл. 4.

Сравнение данных табл. 4 показывает, что оптимальные уровни, полученные алгоритмом NMIN и процедурой FACTOR, практически совпадают. Это позволяет существенно сократить объем информации об озоне, которую необходимо использовать при подготовке априорных данных для задач дистанционного зондирования атмосферы. Кроме того, выбранные оптимальные уровни могут быть использованы при создании различных схем измерений озона.

1. Вапник В.Н., Глазкова Г.Г., Кошечев В.А., Михальский А.И., Червоненкис А.Я. Алгоритмы и программы восстановления зависимостей. — М.: Наука, 1984. — 816 с.
2. Калайда В.Т., Хамарин В.И. Анализ распределения вероятностей парциального давления озона. — Изв. АН СССР. ФАО, 1986, № 11, с. 1217–1219.
3. Хргиан А.Х. Физика атмосферного озона. — Л.: Гидрометеиздат, 1973. — 292 с.
4. Аффифи А., Эйзен С. Статистический анализ. — М.: Мир, 1982, с. 354–380.

Институт оптики атмосферы
СО АН СССР, г. Томск

Поступила в редакцию
23 сентября 1987 г.

V. T. Kalaida, V. I. Khamarin. **Method of objective determination of polar and tropical ozone profiles.**

The possibilities of the method for constructing an optimal dividing hyperplane («generalized portrait») for classifying vertical profiles of partial ozone pressure are under investigation. It is shown that this method can be used for solving this problem. The optimal characteristics in the problem of vertical ozone-profile classification and restitution are analyzed.