

春播期亚欧环流形势的客观分型及其 在物理-统计预报中的应用

朱盛明

(江苏省气象研究所)

张守智 夏定中

(南京大学数学系)

提 要

本文对1972—1978年七年春播期(3月27日—5月10日)500mb亚欧环流形势,应用变 K -均值聚类法进行客观分型,共得九型。对它们和春播期四种主要中期天气过程——连阴雨和低温连阴雨、过程性天气、连晴,用对应分析方法数量化。以环流形势与同期的北半球500mb波谱分析中选出的物理量作为因子,对天气过程类型作出了逐步回归和AID聚类的统计推断。

对1979、80年春播期作了预报检验,效果尚好。

一、引 言

用环流形势划分天气阶段,为天气预报提供背景条件的办法,已为预报员们所熟知。但是主观定性方法分型结果,往往因人而异,不便应用。近年来,客观分型方法已有发展,它是用定量计算目标函数的方法来确定图形之间的亲疏关系,进行分门别类的。这样做的优点是客观性、可重复性和一致性,便于推广应用。

客观分型方法中目标函数的形式很多,但大致可分为两种类型:一是取为相似系数的形式^[1,2]

$$r_{ij} = \frac{\sum_{m=1}^N \Delta h_{im} \Delta h_{jm}}{\left[\sum_{m=1}^N \Delta h_{im}^2 \sum_{m=1}^N \Delta h_{jm}^2 \right]^{1/2}} \quad (1)$$

$m=1, N$, N 是网格点总数, r_{ij} 表示 i 日和 j 日两天环流形势的相似系数, Δh_{im} 和 Δh_{jm} 分别代表 m 点高度对 i 日图面平均高度和 j 日图面平均高度的偏差值。(1)式直接与两张图面对应点的偏差符号有关,图形槽脊位置相似较好,但不能充分反映数值接近程度。另一种类型是取为距离形式

$$D_{ij} = \left\{ \sum_{m=1}^N W_m [h_{im} - h_{jm}]^p \right\}^{1/p} \quad (2)$$

D_{ij} 表示 i 日与 j 日两天环流形势的距离, h_{im} 和 h_{jm} 分别代表 i 日和 j 日图上 m 点

的高度, W_m 为权函数, p 为实数. 特别地, $W_m = 1$, $p = 2$, 就是欧几里德距离. 有时为简化计算常采用绝对值距离^[3]的形式

$$D_{ij} = \sum_{m=1}^N |h_{im} - h_{jm}| \quad (3)$$

(2)、(3)两式直接与两张图上对应点高度差有关, 表示高度场总的数值相似较好, 但对槽脊分布的反映不足.

客观分型中的另一问题是目前所用的聚类方法大多采用单链算法 (Single-linkage Algorithm), 把两张最靠近的图归成一类, 然后把两张次靠近的图归成一类, 如此类推. 由于只考虑单个样本间的亲疏远近, 只要与已归入类内的任一样本充分接近就会被划入类内. 这样会使同一类内天气图象并不集中, 它不符合天气形势分型的要求.

针对以上问题, 本文采用相似指数^[4,5], 并模拟一个天气图分型的变 K -均值聚类方法, 对 1972—78 年春播期 (3月27日—5月10日) 500mb 亚欧环流形势进行客观分型, 用于春播期中期预报, 取得一定效果.

二、方 法

1. 相似指数

为了考虑高度场总的数值相似, 也考虑到图形的槽脊位置相似, 定义相似指数为:

$$I_{ij} = r_{ij} \left(1 - \frac{E_{ij}}{n\sigma} \right) \quad (4)$$

式中

$$E_{ij} = \sqrt{\frac{1}{N} \sum_{m=1}^N [h_{im} - h_{jm}]^2} \quad (5)$$

是两天图面的欧氏距离,

$$\sigma = \frac{1}{2} (\sigma_i + \sigma_j) \quad (6)$$

为两天环流图面均方差的平均值, 其中

$$\sigma_i = \sqrt{\frac{1}{N-1} \sum_{m=1}^N \Delta h_{im}^2} \quad \sigma_j = \sqrt{\frac{1}{N-1} \sum_{m=1}^N \Delta h_{jm}^2}$$

n 是正实数, 分母上放 $n\sigma$ 的作用, 除了消除 E 的量纲外, 主要是起控制它的作用, n 取得太小, $\left(1 - \frac{E_{ij}}{n\sigma} \right)$ 会出现负值; 当 r 也是负号时, 就会出现假象, 但若 n 取得过大则 E 值基本不起作用. 曾对亚欧 500mb 高度场作试验^[4], 分别取 $n = 2, 3, 4, 5$, 对同一张图用 I_{ij} 求出相似图形, 结果得出, 当 $n = 2, 3$ 时相似形势的顺序号与单用 E_{ij} 求得的顺序号相同, 而 $n = 5$ 时则又与由 r_{ij} 求得的相近, 而当 $n = 4$ 时求得的相似形势顺序与前面两种情况不同, 比较起来能较好兼顾槽脊位置和高度场数值两方面的相似性, 所以取 $n = 4$ 为宜.

2. 变 K -均值聚类法

K -均值聚类法^[6]是逐步聚类法的一种。 K 就是类数,均值就是同类样本的平均值——平均图。三参数的 Mac Queen 方法^[7],四参数的 Wishert 方法^[7],以及美国斯坦福大学研究所创立,并已有效地用于卫星云图分析的 ISODATA 法^[7]都是有名的逐步聚类方法。基本思路都是在聚类过程中,根据某种最优准则,通过参数来控制 K 的增减和迭代过程以求得优良分类。我们设计的适用于天气图分类的变 K -均值聚类法,步骤如下:

(1) 确定四个参数:归类阈值 T ,并类阈值 φ ,容许前后两次聚类后变类的样本数上限 $TOL1$,0 类样本数上限 $TOL2$ (0 类是指不易群集的少数样本)。

(2) 由经验给出种子样本序号,从参加聚类的天气图中调出,作为初始类均值。

(3) 对所有天气图逐日求出与当步类均值的最近距离 D_M ,若 $D_M < T$ 则归入最近类 LM 中;若 $D_M > T$,且总分类数 $K < 20$ 时,如果上一次归类时属单样本类,则归入 0 类;否则, $K = K + 1$ 成立新类。若 $K \geq 20$ 时,亦归入 0 类。归类后,和上次聚类结果作比较,若改变属类则 $NT = NT + 1$,累计变类的样本数。

(4) 计算类均值,若两类间均值距离 $< \varphi$,则并类并重新计算类均值。

(5) 把所有仅含一个样本的类并入 0 类中,重新整理留下的类编号。

(6) 若 $NT < TOL1$ 则认为聚类已收敛,下接第(7)步;否则转第三步。

(7) 若 0 类样本总数 $> TOL2$,则改变 T, φ 转第二步,否则算法停止,印出结果。

3. 零阶对应分析^[8]与数量化方法

设有 $m \times n$ 阶样本矩阵 A (表 1),其中元素 $a_{ij} > 0$,若

$$\rho X = R^{-1}AY \quad \rho Y = C^{-1}A^T X \quad (7)$$

则称 (ρ, X, Y) 是 A 的零阶对应分析 $C_0(A)$ 的解。简称对应分析。(7)式中 X 为 $m \times 1$ 维随机向量, Y 为 $n \times 1$ 维随机向量, $R = \text{diag} \cdot$

$(a_{i.})$, $C = \text{diag}(a_{.j})$, $a_{i.} = \sum_{j=1}^n a_{ij}$, $a_{.j} = \sum_{i=1}^m a_{ij}$, ρ 称为随机向量 X 和 Y 关于矩阵 A 的相关系数。

事实上,由(7)式有

$$\rho(R^{\frac{1}{2}}X) = (R^{-\frac{1}{2}}AC^{-\frac{1}{2}})(C^{\frac{1}{2}}Y) \quad (8)$$

$$\rho(C^{\frac{1}{2}}Y) = (R^{-\frac{1}{2}}AC^{-\frac{1}{2}})^T(R^{\frac{1}{2}}X) \quad (9)$$

(8)(9)两式互相代入便得

$$\rho^2(R^{\frac{1}{2}}X) = (R^{-\frac{1}{2}}AC^{-\frac{1}{2}})(R^{-\frac{1}{2}}AC^{-\frac{1}{2}})^T(R^{\frac{1}{2}}X) \quad (10)$$

$$\rho^2(C^{\frac{1}{2}}Y) = (R^{-\frac{1}{2}}AC^{-\frac{1}{2}})^T(R^{-\frac{1}{2}}AC^{-\frac{1}{2}})(C^{\frac{1}{2}}Y) \quad (11)$$

作变换令 $Z = R^{-\frac{1}{2}}AC^{-\frac{1}{2}}$,由(10)(11)显见,若 ρ 是 ZZ^T 的非零特征根,则同时也是 Z^TZ 的非零特征根。由代数学知, ZZ^T 和 Z^TZ 的秩是相等的,所以它们的非零特征根个

表 1 样本矩阵 A

天气过程 环流型	a_{11}	a_{12}	$\dots$	a_{1n}	$a_{1.}$
1	a_{11}	a_{12}	$\dots$	a_{1n}	$a_{1.}$
2	a_{21}	a_{22}	$\dots$	a_{2n}	$a_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	a_{m1}	a_{m2}	$\dots$	a_{mn}	$a_{m.}$
	$a_{.1}$	$a_{.2}$	$\dots$	$a_{.n}$	

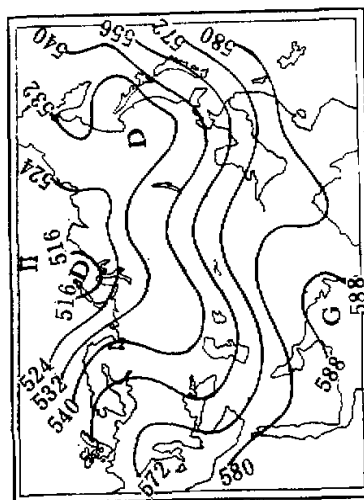
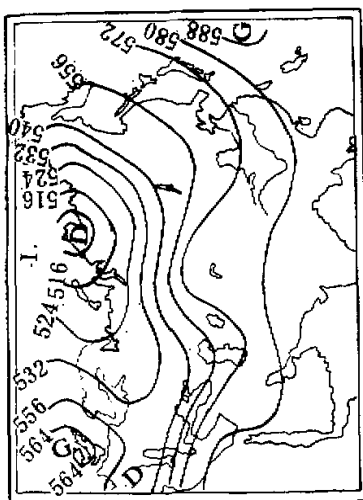
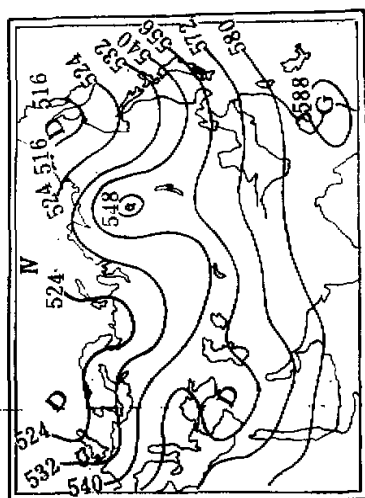
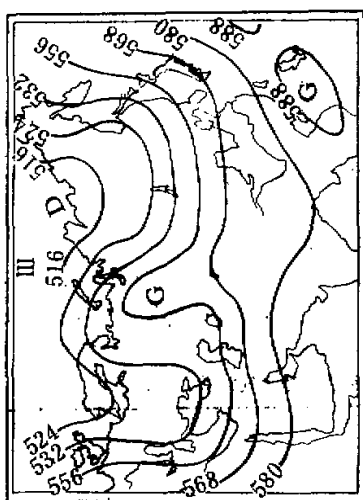
数相等,特征根值相同。并且,若 $R^{\frac{1}{2}}X$ 是 Q 型因子分析的特征向量,则 $C^{\frac{1}{2}}Y$ 就是 R 型因子分析的特征向量。所以,对应分析兼有 Q 型和 R 型因子分析的优点。求得 Y 由(7)式可推算 X ,反之由 X 也可推算 Y 。又因 ZZ^T 与 Z^TZ 具有相同的非零特征根,所以在空间 R^m 中第 K 个因子与空间 R^n 中的第 K 个因子有相同的方差,就可以把天气过程和环流型放在同一数轴上来分析它们的关系。所以,用对应分析求得特征向量 X 、 Y 具有标度 (Scaling) 功能,我们利用这一性质对天气过程、环流分型同时数量化。

三、客观分型及数量化结果

本工作是以寻找春播期中期天气过程的环流背景为主要目的,对 1972—1978 年七年春播期 (3 月 27 日—5 月 10 日) 500mb 亚欧环流形势,应用变 K -均值聚类法进行客观分型。取归类阈值 $T = 0.20$, 并类阈值 $v = 0.10$, 容许两次聚类的变类样本数 $TOL1 = 15$, 0 类样本数上限 $TOL2 = 15$, 并由预报经验选出 15 个种子样本。聚类收敛以后,共

表 2 环流客观分型和江苏省大型天气关系

型	环流型	天 气 型	平均持续天数
I	欧洲槽—新疆暖脊东移型	72—(3) 72—(5) 74—(4) 76—(6) 76—(8) B→A B→A B→A C→B B 77—(4) 78—(2) 78—(6) A A B→A	7.4
II	欧洲脊—新疆脊东移型	74—(1) 76—(9) 78—(3) 78—(5) A B A B	6.5
III	乌东脊西倾—弱副高型	72—(1) 73—(2) 76—(1) 75—(1) 77—(3) B C→A C→B A B	6.5
IV	贝(阻)高—弱副高型	73—(8) 76—(2) B B	5.0
V	乌山 50°N 以南弱脊—弱副高型	73—(4) 73—(6) 74—(6) 75—(2) 75—(5) C→B C B A→B B 76—(3) 76—(7) 77—(8) 78—(4) 78—(7) B C B B B	7.2
VI	欧洲平直气流—弱副高型	72—(6) 73—(1) 73—(7) 74—(2) 74—(5) C A→C B A→B A 75—(7) 76—(5) 77—(2) B C A→C	8.2
VII	北亚主弱平直气流—弱副高型	73—(5) 74—(3) 77—(1) 77—(7) B B→C B→A D	7.1
VIII	乌西(阻)高平直气流—17~20°N 副高型	72—(4) 73—(3) 75—(4) 75—(6) 76—(4) B C D D→B D 77—(6) D	7.3
IX	乌东(阻)高平直气流—17~20°N 副高型	72—(2) 75—(3) 77—(5) B→D D C	7.3



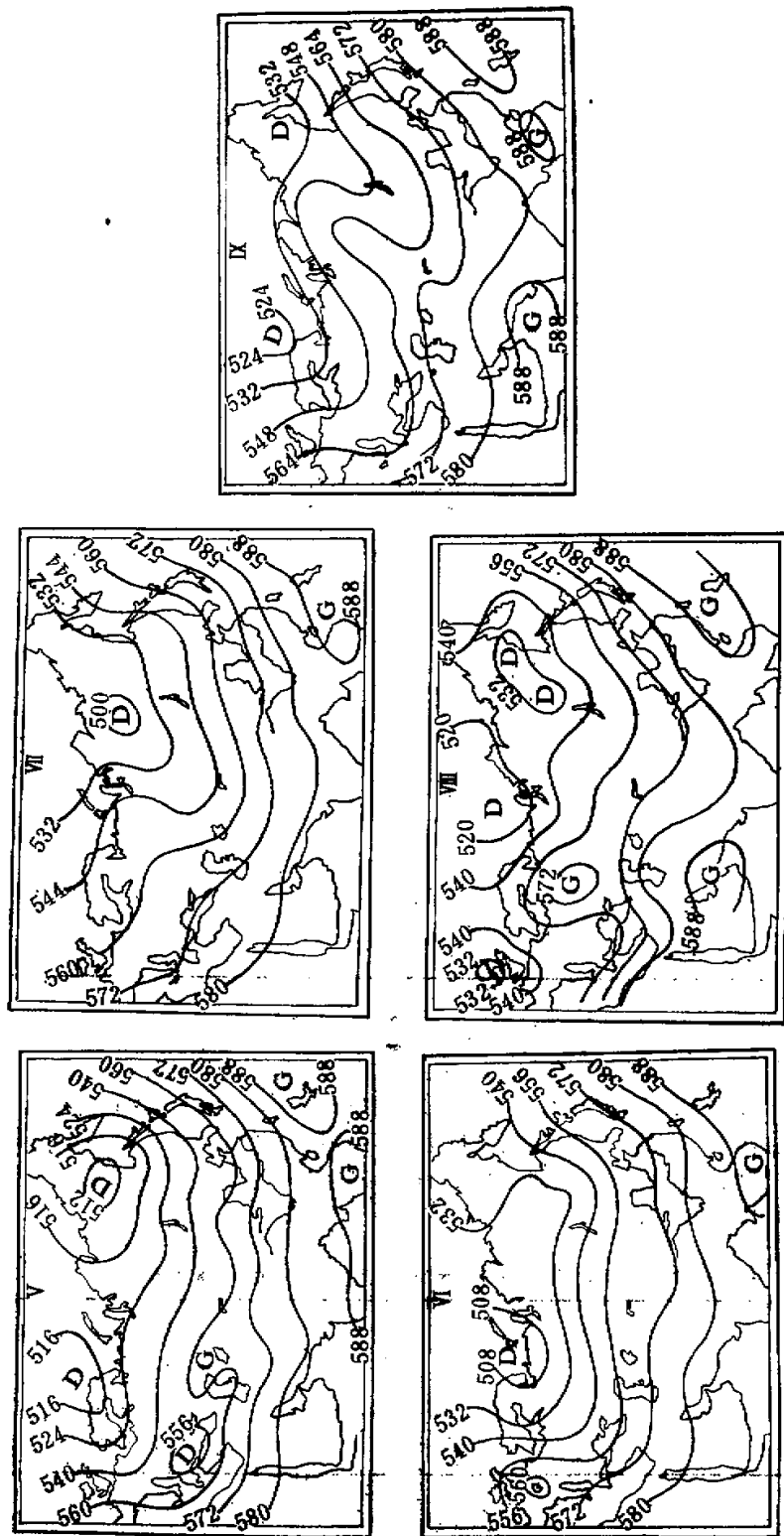


图1 春、夏、秋、冬 500mb 亚欧环流形势 客观分型平均图

得九型(图 1)。

各型与中期天气过程的关系如表 2 所示,表中 A 表示连晴, B 表示过程性天气, C 表示连阴雨, D 表示低温连阴雨,符号上面的数字表示年份,括号内数字表示该年内分型时段顺序号,并给出了每型的平均持续天数。

由图 1 和表 2 可见,由于目标函数采用了相似指数, K-均值方法在聚类过程中又是以样本和类中心的相似程度为依据,所以同类样本集中,平均图特征才比较清楚,异类的区别也明显。这九个类型不仅和预报员经验一致,而且与江苏省大型天气现象有较好的关系,前三型以连晴为主, IV、V、VI 型以过程性天气为主,最后三型中以连阴雨天气为主,其中 VIII、IX 型尤多低温连阴雨天气。

表 3 环流形势与未来四天后天气过程列联表

过程 环流	连 晴	过 程	连 阴 雨	低温连阴雨	Σ
I	4	9	4	2	19
II	13	24	16	1	54
III	10	11	9	1	31
IV	0	7	3	0	10
V	3	3	12	4	22
VI	4	22	20	7	53
VII	12	43	8	3	66
VIII	0	5	2	15	22
IX	0	20	11	7	38
Σ	46	144	85	40	315

把环流形势和未来四天、六天后的天气过程(分为连阴雨、低温连阴雨、过程性天气、连晴四种)制作两方式列联表,如表 3。对它们进行对应分析,求得最大非零特征根值为 0.56217 和 0.50756。可以证明^[6],零阶对应分析等价于 Fisher 的两方式列联表分,特征根也就是典型相关系数。这也从统计学角度证明了客观分型的结果与天气过程类型的关系是较密切的。我们取最大特征根对应的特征向量作为表征环流形势和天气过程类型关系的数量化值,如表 4。

表 4 表征环流形势和未来四天和六天后天气过程关系的数量化值

	连晴	过程	连阴雨	低温连阴雨	I	II	III	IV	V	VI	VII	VIII	IX
四天后	-97	-33	-10	253	-71	-23	-74	-52	-57	17	44	51	323
六天后	-96	-26	27	251	-65	-24	-55	4	-58	8	11	37	339

四、应用举例

以对应分析所得到的我省春播期四种天气类型的数量化值为因变量，应用文献 [9] 的结果，取与中期天气过程变化关系较密切的波长所对应的物理量三天滑动平均值，以及

表 5 用于预报未来四天后天气类型的初选因子表

x_1	65N3 波的角动量输送	x_{12}	65N3 波的纬向动能
x_2	55N4 波的角动量输送	x_{13}	55N3 波的纬向动能
x_3	55N5 波的角动量输送	x_{14}	45N5 波的经向动能
x_4	45N5 波的角动量输送	x_{15}	35N6 波的经向动能
x_5	45N6 波的角动量输送	x_{16}	65N3 波的经向动能
x_6	35N7 波的角动量输送	x_{17}	65、55、45N 三纬圈 6 波的平均动能
x_7	65N3 波的角动量输送的散度	x_{18}	65、55、45N 三纬圈 5 波的平均动能
x_8	55N4 波的角动量输送的散度	x_{19}	65、55、45N 三纬圈 3 波的平均动能
x_9	45N5 波的角动量输送的散度	x_{20}	65、55、45N 三纬圈 4 波的平均动能
x_{10}	35N7 波的角动量输送的散度	x_{21}	500MB 亚欧环流型
x_{11}	35N8 波的角动量输送的散度		

表 6 逐步回归和 AID 聚类的拟合情况

方法 拟合指标 预报时效	逐步回归					AID				
	F 值	信度	复相关系数	标准差		样本数	分类数	类内偏离平方和		Q_2/Q_1
				原始	剩余			原始 Q_1	分类后 Q_2	
四天	3.0	<0.1	0.8169	99.26	58.10	245	11	2404391.8	264483.1	0.11
六天	2.7	0.1	0.8049	98.30	59.19	245	11	2357624.4	212186.2	0.09

表 7 逐步回归和 AID 聚类法的预报效果

准确率 项目		方法	逐步回归	AID
总次数评定	四天		53/70=76%	55/70=79%
	六天		50/70=70%	51/70=72%
转折天气(1)	四天		迟一天	迟一天
	六天		同时	迟二天
转折天气(2)	四天		迟一天	迟一天
	六天		迟二天	迟一天

转折天气 (1): 1979 年 4 月 22—23 日过程性天气转低温连阴雨

转折天气 (2): 1980 年 5 月 2—3 日过程性天气转连阴雨

九种环流型的数量化值为因子。经相关系数初选(取信度 $\alpha = 0.1$), 以预报未来四天后天气类型为例, 初选因子如表 5 所示。

应用逐步回归和 AID^[7] 聚类方法, 分别求出预报方程及 AID 聚类表。仍以未来四天和六天的预报为例, 历史资料的拟合情况如表 6。

1979、80 年春播期共试报 70 次, 二种方法的检验如表 7 所示, 一般准确率尚好, AID 方法较好。如果转折性天气评定允许误差 1 天, 则逐步回归和 AID 准确率都是 3/4。AID 方法在日常业务预报长期使用中, 效果均较优, 其原因是分类方法可以更合理地处理非线性相关关系的缘故。

五、小 结

1. 以相似指数为目标函数, 用变 K-均值法聚类是行之有效的客观分型方法, 所得结果同类样本集中, 类平均图特征清楚, 异类差别显著, 分型结果能表征中期天气过程的环流形势背景。这种聚类方法也可用于从信息化的数据库中及时、准确地寻找相似样本等预报业务工作。

2. 对应分析方法兼有 Q 型和 R 型因子分析的优点; 用此法对类型变量进行数量化, 比较合理。

3. 引用造成大型环流发展和阻尼的物理量因子, 感到物理意义较清楚, 使用效果也不错。

4. 统计与动力相结合是天气预报发展的趋势, 本文只是把中期天气过程的动力分析结果用于统计预报。今后应努力向改造动力方程, 引进随机项, 统计系数等方面发展。

参 考 文 献

- [1] Peagle, J. N., Synoptic climatology of 500-mb winter flow types, Jour. of Appl. Meteor., March 1974.
- [2] 游景炎, 数理统计在暴雨预报中的应用, 应用数学学报, 1976 年第 2 期。
- [3] 无锡气象台、苏州地区气象台、中央气象局研究所概率统计组, 500 毫巴天气形势的客观分型, 全国气象站天气预报会议论文集, 科学出版社, 1980。
- [4] 江苏省气象台、南京大学数学系, 应用聚类分析方法作中期天气预报的尝试, 全国多元分析学术讨论会材料, 1977。
- [5] 野本真一、桑保夫, “類似示数”による型の分類と予想天気図の検証, Jour. of the meteor. Soc. of Japan, Dec. 1971.
- [6] Hartigan, J. A., Clustering Algorithm, John Wiley & Sons, 1975.
- [7] Anderberg, M. R., Cluster analysis for application, Academic press, 1973.
- [8] Hill, M. O., Correspondence analysis: A neglected multivariate method, Appl. statist. Vol. 3, No. 3, 1974.
- [9] 朱盛明, 春播期北半球 500mb 角动量输送和能量变化的波谱分析, 全国中长期预报会议材料选编(中期部份), 1979。

OBJECTIVE CLASSIFICATION OF ASIA-EUROPE CIRCULATION DURING SPRING SOWING PERIOD AND ITS APPLICATION IN PHYSICAL-STATISTICAL WEATHER FORECASTING

Zhu Sheng-ming

Zhang Shou-zhi Xia Ding-zhong

(*Institute of Meteorology, Jiangsu Province*) (*Department of Mathematics, Nanjing University*)

Abstract

In this paper, an objective classification of Asia-Europe circulation pattern at 500 mb during the spring sowing periods (Mar. 27 to May 10) in 1972—1978 has been done by K-means clustering algorithm, and nine circulation patterns have been obtained. In addition, there are four major medium range weather patterns in the spring: the persistent rainy weather, the persistent rainy weather with lower temperature, the weather with alternative processes of rainy and sunny days and the persistent sunny days. The corresponding analysis has been made for these weather patterns. The weather patterns are considered as dependent variables, while the circulation patterns and the selected factors in spectral analysis of the angular momentum transfer and kinetic energy variability at 500 mb are considered as independent variables. The statistical inference is made by using the stepwise regression analysis and automatic interactive detection (AID) clustering algorithm.

The method was used in operational forecasting during the spring sowing period in 1979 and 1980, and the accuracy was rather satisfactory.