

主成分分析在气候区划中的应用

缪启龙

李兆元

窦永哲*

(陕西省气象科学研究所)

(山西省气象局)

提 要

对 n 个站、 p 个气候要素的气候场,可由主成分分析用前 m 个主成分表征气候场的气候特征;并由因子载荷量判断各主成分表示的气候意义。对陕西省 100 个站、17 个气候要素进行主成分分析,结果前三个主成分的累积贡献率达 88.8%。并由前三个主成分的分布将陕西省划分为 5 个气候亚带、11 个气候区、26 个气候小区的三级气候区划。

气候区划能为工农业生产布局、国民经济发展规划提供有价值的气候依据。做好气候区划的关键在于由众多的气候要素中找出气候区域间的分异规律。长期以来,气候学家们为此提出各种各样的综合表征气候特征的指标(包括各种指数、系数),并以此进行各种类型的气候分类、气候区划。但指标的选择、设计都是人为确定,常因人而异,且指标设计中不能使用很多气候要素,只能选用主要的 1、2 个。因而,指标法进行气候区划常有较大的人为影响。近年来,主成分分析广泛应用于气象科学,为综合表征气候特征提供了一种数值分类的方法。本文试图以陕西省为例,探讨由主成分分析进行大尺度范围内气候要素场的气候特征分析,并由此进行气候区划的可能性。

一、气候要素场的主成分分析

主成分分析是一种多元分析法,能将相互间有相关特性的多个变量的气候要素场用几个综合变量来拟合^[1,2]。这几个综合变量应尽可能反映原来气候场的信息,且是相互独立的。

设气候场中有 n 个站, p 个气候要素,即

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{pmatrix}$$

1986年7月24日收到,10月27日收到修改稿

*我馆1986届毕业

可由

$$Y = V X$$

来拟合。式中

$$V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1p} \\ v_{21} & v_{22} & \cdots & v_{2p} \\ \vdots & \vdots & & \vdots \\ v_{p1} & v_{p2} & \cdots & v_{pp} \end{pmatrix}, \quad Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{p1} & y_{p2} & \cdots & y_{pn} \end{pmatrix}$$

式中 V 为由气候要素场的相关矩阵 R 求得的特征向量; Y 为主成分, 若某一个主成分的分方差贡献越大, 则该主成分提供的气候场的信息就越多。

通常认为, 当前 m 个主成分的累积贡献率 $G(m) \geq 80\%$ 时^[3], 就表示它们反映了整个气候场的大部分信息, 就可由前 m 个主成分来表征气候场的主要特征。

各主成分所代表的气候意义, 可由因子分析, 求各气候要素在各主成分上的载荷量, 即主成分与 p 个气候要素的相关矩阵。它反映了每个气候要素与各主成分的相关程度, 载荷率越大, 相关越紧密。某一主成分中某一气候要素的载荷率越大, 则该主成分越明显地反映该气候要素所具有的气候意义^[1]。由此, 气候场能用 m 个主成分来反映, 而 1、2、…、 m 个主成分又分别与不同的气候要素有较高的相关, 即不同的主成分, 反映的气候特征的主要方面是不同的。这样, 主成分分析是由气候要素的相关矩阵 R 的特征值大小决定主成分, 由各主成分中气候因子载荷率的大小, 决定主成分表示的气候意义。第 i 个主成分在某地越大, 表明该地第 i 个主成分所反映的气候意义越显著。于是, 表示气候特征不同方面的 m 个主成分在各地的大小, 就可以表明各地气候特征上的差异, 从而可以进行气候分类和区划。

在因子载荷阵中, 有时气候因子的载荷率差异不甚显著, 使得各个主成分所表示的气候特征不很清晰。故使用方差极大正交旋转, 使原先的载荷阵结构简化, 载荷率明显分离, 各主成分所表示的气候特征更明显、更突出^[1]。旋转后载荷阵仍是因子载荷阵, 每个主成分的分方差贡献略有变化, 而总方差贡献不变。

二、气候资料

本文资料取自陕西省气象局30年整编资料, 除少数站点外, 绝大部分资料年代都在20—30年间。

在气候要素的选取上, 我们注意了选择与生产密切相联, 又表示气候主要特征的一些要素。笔者认为气候区划中地理因子具有重要意义, 但在各种指标、指数中都无法体现, 人们常在区划、分类划图时参考, 显然带有人为因素。为此, 我们试将地理因子数值化直接参与主成分分析。考虑到陕西省南北长, 东西窄, 且气候要素的等值线多呈东西走向, 所以以纬度表示南北水平差异, 省略经度的东西差异。又因陕西省境内地形复杂、差异甚大, 故用海拔高度表示垂直方向上的差异。植被、地貌既是气候的影响因子, 又是气候的产物。为了引进植被因子, 本文将不同植被进行等级区分(表1), 它仅考虑了植被的高、低、疏、密和类型等自身的特征^[4], 是否合理还有待进一步研究、探

表1 植被等级

等级	0	1	2	3	4	5	6	7	8	9	10
植被	戈壁漠	荒漠	草原	灌木	农作物	稀疏林	针阔混交林	落叶阔叶林	常绿针叶林	常绿阔叶林	热带雨林

讨,但这种等级区分能反映植被之间基本特征的差异。

于是,由陕西省100个站点的气候要素、地理因子共17种气候因子,构成 17×100 的气候要素场。

X_1 : 4—6月降水量, X_2 : 7—9月降水量, X_3 : 5—10月平均气温, X_4 : 12—3月平均气温, X_5 : 年降水量, X_6 : 年平均气温, X_7 : 1月平均最低气温, X_8 : 7月平均最高气温, X_9 : 4—6月平均最小相对湿度, X_{10} : 7—9月平均最小相对湿度, X_{11} : 最大积雪深度, X_{12} : 积雪期, X_{13} : 霜期, X_{14} : 大风日数, X_{15} : 海拔高度, X_{16} : 站点纬度, X_{17} : 站点植被。

三、主成分分析的结果分析

表2 方差极大正交旋转后的因子载荷阵

变 量	主 成 分		
	1	2	3
X_1	0.955	0.057	0.043
X_2	0.865	-0.266	-0.090
X_3	0.086	0.924	-0.267
X_4	0.728	0.620	-0.160
X_5	0.974	-0.014	0.026
X_6	0.423	0.859	-0.262
X_7	0.737	0.647	-0.054
X_8	-0.211	0.819	-0.449
X_9	0.845	0.377	-0.152
X_{10}	0.832	0.363	-0.158
X_{11}	-0.028	-0.357	0.878
X_{12}	-0.241	-0.585	0.690
X_{13}	-0.602	-0.716	0.105
X_{14}	-0.114	-0.196	0.832
X_{15}	0.013	-0.883	0.363
X_{16}	-0.854	-0.421	-0.009
X_{17}	0.827	-0.007	-0.170
方差贡献	7.165	5.391	2.539
方差贡献率 累计(%)	42.2	73.9	88.8

表2是经过方差极大正交旋转后的因子载荷阵。表中第一主成分的方差贡献为7.165,方差贡献率为42.2%;第二主成分方差贡献为5.391,方差贡献率为31.7%;第三主成分的方差贡献为2.539,方差贡献率为14.9%。前三个主成分累积贡献率为88.8%,几乎反映了气候场的绝大部分信息。因此,可用前三个主成分进行陕西省气候场的特征分析。

第一主成分中载荷率大的因子是年降水量、4—6月、7—9月降水量及平均最小相对湿度、植被等,载荷率都在0.8以上。即第一主成分主要反映了各地水分状况的分异规律,尤其是年降水量和4—6月降水量,载荷率在0.95以上。因此,第一主成分大的地方降水多、潮湿、植被好,尤其是年降水量和4—6月降水多。反之,降水少、干燥、植被差,尤其是年降水量和4—6月降水量少。第一主成分中站点纬度的载荷量有较大的负载荷,达-0.854,表示第一主

成分的分布具有明显的纬度影响。即第一主成分大的地方纬度低、降水多、潮湿；第一主成分小的地方纬度高、降水少、干燥。

第二主成分中载荷率大的因子是年平均气温、7月平均最高气温、载荷率均达0.80以上。平均霜期和海拔高度有较大的负载荷，分别为-0.716和-0.883。因此，第二主成分主要反映了陕西省各地热量状况的差异。即第二主成分大的地方年平均温度高、夏季温度高、霜期短、海拔高度低；第二主成分小的地方年平均温度低、夏季温度低、霜期长、海拔高度高。海拔高度有较大的负载荷，表明第二主成分的分布与地形的高低有较好的一致性。

第三主成分载荷率大的因子是最大积雪深度和大风日数，载荷率分别为0.878和0.832。积雪期也有较大载荷为0.690。因此，第三主成分反映陕西省各地积雪和风速的分异。即第三主成分大的地方积雪深、积雪期长、大风日数多，反之亦然。

上述分析表明各主成分，特别是第一、二主成分具有明显的气候意义。由此可以看到，在陕西省的气候分析、区划、资源利用中，气候的区域分异最明显的是水分状况，它起着主要的、决定性的作用。其次是热量。这在广大西北干旱、半干旱地区可能都应如此，这与东部湿润地区热量是主要因素有明显的不同。在西北各省区目前已进行的气候区划和农业气候区划中这一点多数还未引起注意。

图1是第一主成分的分布图。由图可见，陕西省第一主成分由北向南逐渐增加，呈带状分布，反映陕西省的水分状况是由北向南逐渐变好。最小值在长城一线以北，这里是毛乌素沙地、白于山山地，是陕西最为干燥的地区，年降水量小于400毫米，降水日数只有70天左右，且集中于夏季，夏季最小相对湿度仅有10%，干燥少雨，自然植被差，为草原和半荒漠。最大值在陕南山区，年降水量为900—1200毫米，降水日数达140—160天，潮湿，夏季最小相对湿度也达34%，自然植被好，为常绿阔叶林、落叶阔叶林混交林为主的亚热带森林。

图2是第二主成分分布状况，较第一主成分分布复杂，虽然基本趋势为由北向南增加，但分布规律大致与地形高低相一致。低值位于长城以北及白于山一带和秦岭高山区。这两个地区是全省热量条件最差的地区，年平均气温在8.0℃左右，霜期长达200天，海拔高度较高，北面白于山达2000米，秦岭更在2000米以上，主峰太白山达3767米。两个高值区位于渭河平原和安康河谷地区，是全省热量条件最好的两个区域。年平均气温前者为13℃左右，后者为15℃左右，

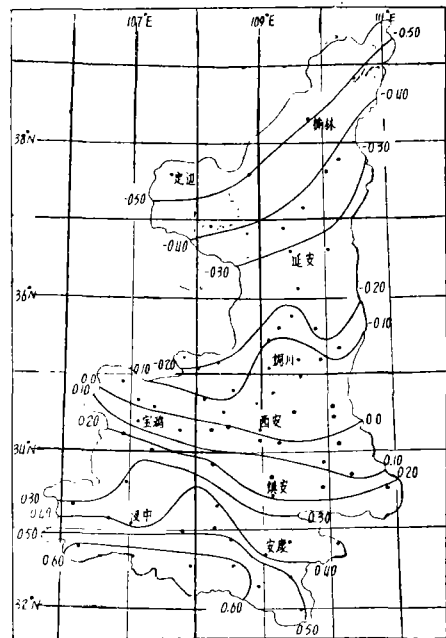


图1 陕西省第一主成分分布

夏季气温高,是全省两个高温中心。且都位于河谷地带,拔海高度低。该图在渭河平原和秦岭之间第二主成分的等值线十分密集,这可能由于秦岭北坡是一大断层,北坡陡峭,水平间温度的垂直变化大所致。

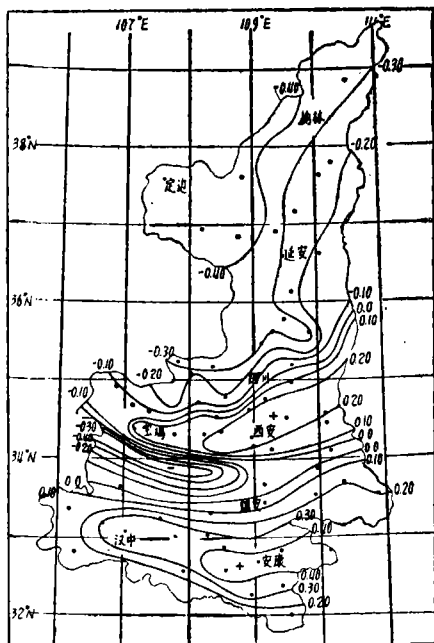


图2 陕西省第二主成分分布

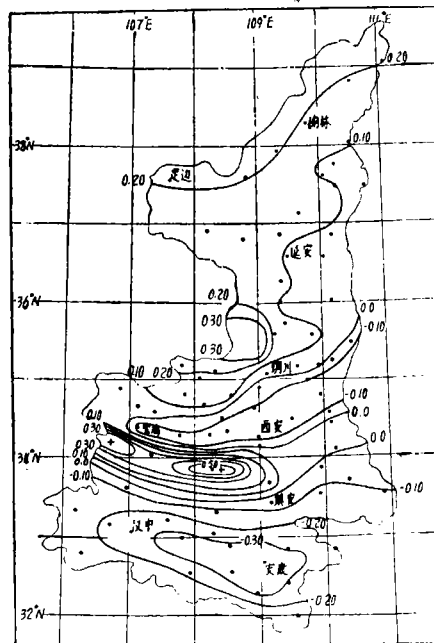


图3 陕西省第三主成分分布

第三主成分分布(图3)也较复杂。高值区位于秦岭高山区、桥山山地及长城以北。秦岭高山区是陕西主要的积雪区,积雪期长达180天,而长城以北是陕西主要的风沙区,平均风速大,大风日数达20天以上。而渭河平原和陕南汉江河谷为低值区,积雪少,风速也小。

上述分析可看到,三个主成分都有零值线在秦岭的北坡通过,此线正好与秦岭成为陕西省南北气候分野相一致。该线以北为干燥、温凉气候,以南为湿润、温暖气候。主成分的分布也表明秦岭是南北气候的分界。还可看到各主成分的分布具有明显的地带性,与地理因子的分布相一致,表明地理因子在分析气候特征时是重要的因子,也说明本文将地理因子数值化参与主成分分析是可行的,有意义的。

四、陕西省气候区划

如果把前三个主成分分别看作三维空间的三个坐标轴,那末某站点的三个主成分大小就决定了该站点在三维空间中的位置点。则可由下式求出三维空间中该点距原点的距离D

$$D = y_1^2 + y_2^2 + y_3^2$$

式中 y_1 、 y_2 、 y_3 分别为该点的第1、第2、第3主成分的值。

同一象限中, 距离相近的点集聚在一起, 即各主成分大小相近。那末由上述分析可知, 它们的水、热等气候状况也相近, 可以判断它们属于气候特征相近的一个气候区。如果点之间的距离较远, 表明点之间的气候特征差异也较大, 则分别属于不同的气候区^[5-7]。由此, 可以根据点间距离的大小, 将各地气候进行分划。若将所有点间距离分为四等分, 则原则上8个象限可分为32个气候群体(气候小区)。实际上有些气候群中没有站点, 因此陕西省分为26个气候群体——气候小区。

若将气候小区 i 与 j 之间的距离 D_{ij} 较近者视为一个气候区, 则可由下式将26个气候小区合并为11个气候区(如表3)。

$$D_{ij}^2 = \sum_{k=1}^m (\bar{y}_{ki} - \bar{y}_{kj})^2 \quad (*)$$

式中 $\bar{y}_{ki}$ ($k=1, 2, 3$; $i=1, 2, \dots, 26$)为第 i 气候小区第 k 主成分的平均值, $\bar{y}_{kj}$ ($k=1, 2, 3$; $j=1, 2, \dots, 26$)为第 j 气候小区第 k 主成分的平均值。

表3 各气候区的主成分平均值及各区间的距离

主成分 距离	A	B	C	D	E	F	G	H	I	J	K
第一主成分	-0.427	-0.262	-0.157	-0.074	-0.008	0.048	0.364	0.128	0.171	0.388	0.582
第二主成分	-0.307	-0.189	-0.306	-0.045	0.182	-0.617	0.060	0.008	0.194	0.368	0.279
第三主成分	0.162	0.072	0.222	0.034	-0.111	0.665	-0.117	0.057	-0.108	-0.266	-0.221
A		0.223	0.227	0.458	0.699	0.758	0.916	0.647	0.826	1.142	1.228
B			0.217	0.240	0.485	0.794	0.700	0.437	0.605	0.920	1.009
C				0.332	0.609	0.579	0.721	0.455	0.683	0.995	1.041
D					0.277	0.860	0.475	0.210	0.615	0.747	0.846
E						1.115	0.392	0.277	0.179	0.464	0.608
F							1.082	0.876	1.127	1.397	1.405
G								0.298	0.235	0.343	0.326
H									0.252	0.549	0.597
I										0.321	0.435
J											0.218
K											

若再用(*)式将11个气候区合并为气候亚带, 则可将陕西省分为5个气候亚带(如表4)。

因此可将陕西分为5个气候亚带、11个气候区、26个气候小区的三级气候区划(表5、图4)。这与陕西气候特点、植被、地形等相吻合,与已完成的《陕西省气候区划》*分

表4 各气候带主成分平均值及带间距离

主成分 距 离		I	II	III	IV	V
平均值	第一主成分	-0.427	-0.147	-0.008	0.179	0.456
	第二主成分	-0.307	-0.154	0.182	-0.108	0.337
	第三主成分	0.162	0.095	-0.111	0.140	0.250
	I		0.326	0.699	0.638	1.096
	II			0.418	0.332	0.793
	III				0.427	0.608
	IV					0.536
	V					

为4个气候亚带、13个气候地区、20个气候小区的三级区划以及《陕西综合农业气候区划》**分为4个农业气候带、9个农业气候区、20个农业气候小区的三级区划基本上一致,但也有不同。《陕西省气候区划》将秦岭和渭河平原划为一个气候亚带

(即关中和秦岭山地暖温带南部、温和气候亚带III)。而本文将这一区域分为两个气候亚带,即渭河平原半干旱温暖气候亚带(III)和秦岭山地半湿润温凉气候亚带(IV)。无论在气候状况还是在地形、自然植被上两地都有明显的差异,我们划分为两个气候亚带表

表5 陕西省气候区划

气候亚带		气候区		气候小区	
I	陕北干旱寒冷气候亚带	A	陕北干旱寒冷气候区	1	长城沿线、定边、靖边地区
				2	白于山山地
				3	府谷、神木、榆林、横山、米脂、子洲、绥德
				4	志丹、子长、安塞
II	渭北高原半干旱温凉气候亚带	B	延安半干旱温凉气候区	5	吴堡、延川、延长
				6	清涧、延安、宜川、富县、甘泉
		C	黄土高原顶部半干旱温凉气候区	7	洛川、黄龙、黄陵
				8	宜君
				9	洵邑、长武、麟游
		D	黄土高原南部关山山地半干旱温凉气候区	10	淳化、铜川、永寿、跃县、白水、合阳、彬县
				11	陇县、千阳、凤翔

* 陕西省气候区划, 陕西省气象局, 1964

** 陕西农业气候区划, 陕西省气象局, 1985

续表 5

气候亚带	气候区	气候小区
Ⅲ 渭河平原半 干旱温暖气 候亚带	E 渭河平原半干旱温暖气 候区	12 韩城、蒲城、大力、华阳、潼关
		13 西安、长安、临潼、渭南、兴坪、 卢县
		14 礼泉、乾县、高陵、三原、咸阳
		15 宝鸡、眉县、扶风、武功、岐山
		16 兰田
Ⅳ 秦岭山地 半湿润温和 气候亚带	F 秦岭高山湿润寒冷气候 区	17 秦岭、太白、双庙
		18 华山
	G 秦岭山地半湿润温和气 候区	19 略阳、留坝、佛坪、宁陕
	H 洛南、镇安半湿润温和 气候区	20 洛南、商县、柞水、镇安
	I 山阳半湿润温和气候区	21 丹凤、山阳、商南
Ⅴ 陕南湿润温 暖气候亚带	J 汉江河谷湿润温暖气 候区	22 勉县、洋县、汉中、南郑
		23 西乡、石泉、汉阴、安康、白河
		24 平利、紫阳、岚皋
	K 陕南山地湿润温和气候 区	25 宁强、镇巴
		26 镇坪

明在水、热两方面的不同是合理的。《陕西综合农业气候区划》虽然将秦岭山地与渭河平原分为两个气候带,但又将秦岭山地与陕北划为一个冬季寒冷农业气候带(Ⅱ),将渭河平原与秦岭南坡划为一个冬季温和农业气候带(Ⅲ)。这是不甚妥当的。虽然秦岭、陕北冬季均为寒冷,渭河平原、秦岭南坡冬季均为温和。但它们各自之间夏季温度、降水、湿度均有较大的差异。从农业角度讲,降水、湿度及夏季温度的差异给农业生产的影响较之冬季的冷、暖更突出、更重要。且二者之间的农业气候条件差异显著,自然植被明显不同,将两处划为一个农业气候带,易使人产生误解和错觉。本文区划使上述两个区划中的矛盾得到较好的解决,将为陕西省进行更合理的气候和农业气候区划提供有益的参考。本文的区划与《陕西省气候区划》另一主要不同在于本文区划将属于后者的陕北、渭北暖温带北部寒温气候亚带(Ⅱ)的神木、米脂、子洲、子长、志丹等无定河中游、延水上游地区划为陕北干燥寒冷气候亚带(Ⅰ)。实际上,上述地区与长城一带虽然在热量上积温有一定差异(长城一线为2800℃,上述地区为3000℃),但其它方面如年平均气温均在8.0℃左右,日平均气温 $\geq 0^{\circ}\text{C}$ 的天数都在250天左右,无霜期也均在150天左右。故以热量指

标分为两个气候亚带也不甚妥当,况且水分条件差异更小,都属于干燥地区。所以本文

均划为干燥寒冷气候亚带是合理的,这样划分综合地反映了气候上的相似性与差异性。

由上可见,由主成分分析而做的陕西气候区划是可行的,能综合各种气候要素组成的气候场内存在的主要气候特征的相似与差异。较大程度上减少了区划中的人为因素。这对于诸如国土区划、土壤区划、环境区划……等一系列自然区划和经济区划是有意义的,使之向数值化、定量化、客观化发展,提供了一种可以试验推广的数值分类方法。

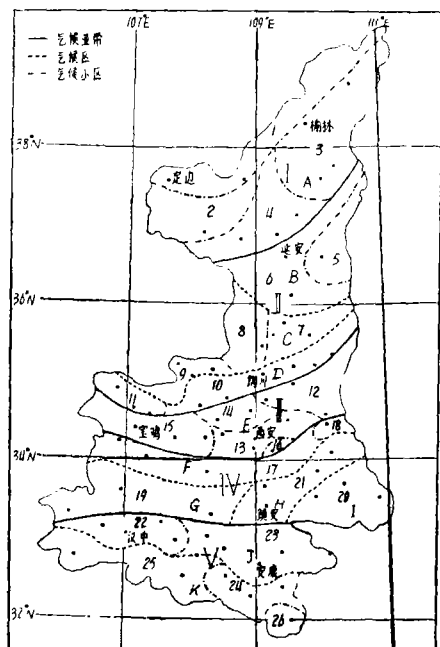


图4 陕西省气候区划

五、小 结

1. 陕西省气候要素场可由前三个主成分来表示。第一主成分主要反映陕西水分状况;第二主成分主要反映热量状况;第三主成分主要反映冬季积雪和大风日数的状况。前三个主成分累积贡献率达88.8%。

2. 地理因子数值化参与统计分析,使得气候要素场的特征分析与地理因子的分布更趋一致。这一方面说明地理因子数值化是可行的,另一方面也说明地理因子数值化在气候分析中是重要的、必需的。

3. 由主成分分析进行的陕西气候的三级区划与已进行的陕西气候三级区划基本上一致;与陕西省气候特征、地形、地貌、植被相吻合。且本区划的结果更合理、更与实际相符。

4. 本文选用的气候要素是否合适还有待于研究,但改变其中部分要素不会使前 m 个主成分表示的气候意义及主成分分布状况有较大的改变,因而气候区划的结果也不会有重大变化。

参 考 文 献

- [1] 王学仁,地质数据的多变量统计分析,269—283,科学出版社,1982。

- [2] 张尧庭、方开泰, 多元统计分析引论, 329—339, 科学出版社, 1982。
[3] Kendall, M., 中国科学院计算中心概率统计组译, 多元分析, 16—76, 科学出版社, 1983。
[4] 中华人民共和国地图集, 地图出版社, 1982。
[5] 小岛忠三郎, 由主成分分析进行的东北地方的详细气候区划, 农业气象, 29卷, 3期, 1973。
[6] Ayode, J.D., On the use of multivariate techniques in climatic classification and regionalization, Arch. Met. Geoph. Biocl. Ser. B, 24(1977), 257—267。
[7] Richman, M.B., Obliquely rotated principal components, an improved meteorological map typing technique, J. Appl. Met., Vol. 20(1981), 1145—1159。

ANALYSIS OF MAJOR CLIMATIC FACTORS AND ITS APPLICATION TO CLIMATIC DEMARCATIION

Miao Qilong Li Zhaoyuan* Dou Yongzhe**

ABSTRACT

For a climatic field having p climatic elements in a region embedding n weather stations, the climate may be characterized by the first m major factors in the analysis with the climatic implication of each major factor discriminated by the factor loading amount. Examination of the major factors for 17 climatic elements at 100 weather stations of Shaanxi Province shows that the accumulative contribution rate of the first three major factors comes to 88.8 percent. Based on the distribution of these three major factors, the Province can be delineated into a 3-level climatic demarcation consisting of five climatic sub-zones, eleven climatic regions and twenty-six sub-regions.

*Affiliated with the Meteorological Research Institute, Shaanxi Province

**Affiliated with the Shanxi Provincial Meteorological Bureau