

副热带高压强度变化的模糊聚类诊断预测*

洪梅¹⁾²⁾ 张韧¹⁾²⁾³⁾ 吴国雄¹⁾ 何金海³⁾ 余丹丹²⁾

¹⁾(中国科学院大气物理研究所 LASG,北京 100029)

²⁾(解放军理工大学气象学院海洋与空间环境系,南京 211101)

³⁾(南京信息工程大学 KLME 暨大气科学博士后流动站,南京 210044)

摘要

采用模糊 C 均值聚类(FCM)、遗传算法(GA)和模糊减法聚类(FSC)交叉融合、优势互补思想进行副热带高压影响因子的综合聚类分析和副热带高压指数的诊断预测。在统计分析的基础上,通过选择若干与副热带高压指数关系密切的影响因子构成了高维特征空间,进行了综合聚类分析,实现了副热带高压指数的聚类判别和诊断预测。该文提出的综合聚类方法既可克服 FCM/GA 算法全局/局部寻优的不足,又可客观确定聚类数目。试验结果表明,该方法具有良好的分类效果,判别结果与实况基本相符。

关键词:副热带高压面积指数;季风系统;模糊聚类;遗传算法

引言

西太平洋副热带高压是连接热带和中高纬大气环流的重要纽带,其强度和位置变化直接影响制约热带和中高纬大气环流演变和东亚地区的天气气候。作为东亚夏季风系统的重要成员,西太平洋副热带高压与东亚夏季风系统成员之间存在着相互作用、互为反馈的相关性。西太平洋副热带高压异常与东亚夏季风活动异常经常相伴随,并导致长江流域出现洪涝和干旱灾害。弄清副热带高压与东亚夏季风系统成员关联的天气事实和变化规律,对预测副热带高压活动具有十分重要的意义。黄荣辉等^[1]的研究强调了热带西太平洋对流对副热带高压活动的影响;喻世华等^[2]研究指出,副热带高压在东亚副热带地区的活动与东亚大陆季风雨带和南海季风槽雨带关系密切,强对流降水凝结潜热的热力强迫作用对副热带高压有明显的反馈作用;张韧^[3-4]基于变分原理和不稳定理论从系统能量角度论证了上述观点,提出了东亚雨带和季风槽雨带及环流分布影响副热带高压稳定性的动力机理和能量判据,随后用小波包分解重构方法讨论了印度洋和南海地区的夏

季风活动对西太平洋副热带高压的影响,揭示了两 者之间的一些相关特性^[5-6]。

上述研究多是针对副热带高压与夏季风系统中某些重要因子进行分析讨论,由于副热带高压活动受多种因子的共同影响制约,彼此共处于非线性系统之中。因此,讨论副热带高压与有限的季风系统因子之间的相互作用或将单个因子讨论结果的累加起来替代多因子的作用无疑是有缺陷或是不完整的。

针对上述问题,本文通过统计分析,选择了多个显著的季风影响因子,将其构成高维特征空间映射点集,随后引入遗传算法、模糊 C 均值聚类和模糊减法聚类等方法及其优势互补的研究思想,通过对季风影响因子的特征空间聚类和映射落区判别,实现了副热带高压强度的聚类判别和诊断预测。

1 研究资料与影响因子

所用研究资料为 NCEP/NCAR 提供的 $2.5^{\circ} \times 2.5^{\circ}$ 网格 10 年平均逐候再分析资料,包括 1958—1967 年、1968—1977 年、1978—1987 年以及 1988—1997 年 4 个时间段 10 年平均的逐候再分析资料,每个时段序列长度为 73 候。参照中央气象台副热带

* 国家自然科学基金项目(40375019)、江苏省气象灾害重点实验室开放课题(KLME0507)及中国博士后科学基金项目(2004036012)共同资助。

2005-04-27 收到,2005-10-08 收到再改稿。

高压面积指数的定义^[7]计算了上述 4 个时段之和的时间序列(共 $4 \times 73 = 292$ 候)的副热带高压面积指数,并选择上述相同时段序列中若干重要区域夏季风系统要素格点平均值作为候选因子。

通过将若干影响因子与副热带高压面积指数作时滞相关分析,在关注相关性显著的前提下,基于预报目的考虑,选择超前副热带高压面积指数 1 候的 11 个因子作为初始影响因子:① 马斯克林冷高压强度指数(A): $15^{\circ} \sim 25^{\circ} \text{S}, 40^{\circ} \sim 60^{\circ} \text{E}$ 区域范围内的海平面气压格点平均值;② 澳大利亚冷高压强度指数(B): $15^{\circ} \sim 25^{\circ} \text{S}, 120^{\circ} \sim 140^{\circ} \text{E}$ 区域范围内的海平面气压格点平均值;③ 中南半岛感热通量(CI): $10^{\circ} \sim 20^{\circ} \text{S}, 95^{\circ} \sim 110^{\circ} \text{E}$ 区域范围内的感热通量格点平均值;④ 索马里低空急流指标(D): $5^{\circ} \text{S} \sim 5^{\circ} \text{N}, 40^{\circ} \sim 50^{\circ} \text{E}$ 区域范围内 850 hPa 经向风格点平均值;⑤ 南海低空急流指标

(EI): $5^{\circ} \text{S} \sim 5^{\circ} \text{N}, 100^{\circ} \sim 110^{\circ} \text{E}$ 区域范围内的 850 hPa 经向风格点平均值;⑥ 印度季风潜热通量(FLH): $10^{\circ} \sim 20^{\circ} \text{N}, 70^{\circ} \sim 85^{\circ} \text{E}$ 区域范围内的潜热通量;⑦ 印度季风 OLR 指标(FULW): $10^{\circ} \sim 20^{\circ} \text{N}, 70^{\circ} \sim 85^{\circ} \text{E}$ 区域范围内的外逸长波辐射 OLR 格点平均值;⑧ 江淮梅雨对流降水率(M): $25^{\circ} \sim 35^{\circ} \text{N}, 110^{\circ} \sim 125^{\circ} \text{E}$ 区域范围内对流降水率(MCP);⑨ 青藏高压活动指数(XZ):200 hPa 位势高度($25^{\circ} \sim 30^{\circ} \text{N}, 95^{\circ} \sim 105^{\circ} \text{E}$) ~ ($25^{\circ} \sim 30^{\circ} \text{N}, 75^{\circ} \sim 95^{\circ} \text{E}$) 范围格点平均值;⑩ 孟加拉湾纬向风环流指数(JIU): $0^{\circ} \sim 20^{\circ} \text{N}, 80^{\circ} \sim 100^{\circ} \text{E}$ 区域范围内的 u_{850} 与 u_{200} 差值格点平均值;⑪ 孟加拉湾经向风环流指数(JIV): $0^{\circ} \sim 20^{\circ} \text{N}, 80^{\circ} \sim 100^{\circ} \text{E}$ 区域范围内的 v_{850} 与 v_{200} 格点平均值。

上述影响因子与副热带高压面积指数的时滞相关分析结果如表 1 所示。

表 1 各影响因子与副热带高压面积指数的时滞相关分析

影响因子	相关系数	最大相关系数
马斯克林冷高压强度指数	-1 ~ 0 候正相关显著(>0.65)	0 候时达到 0.694
澳大利亚冷高压强度指数	-4 ~ -2 候正相关显著(>0.5)	-3 候时达到 0.559
中南半岛感热通量	-2 ~ 0 候负相关显著(>-0.5)	-1 候时达到 -0.518
索马里低空急流指标	-1 ~ 0 候正相关显著(>0.6)	0 候时达到 0.720
南海低空急流指标	-2 ~ 0 候正相关显著(>0.6)	-1 候时达到 0.671
印度季风潜热通量	-1 ~ 0 候正相关显著(>0.6)	0 候时达到 0.644
印度季风 OLR 指标	0 ~ 1 候负相关显著(>-0.6)	0 候时达到 -0.684
江淮梅雨对流降水率	-1 ~ 0 候正相关显著(>0.65)	0 候时达到 0.677
青藏高压活动指数	-1 ~ 1 候负相关显著(>-0.69)	0 候时达到 -0.764
孟加拉湾纬向季风环流指数	-1 ~ 1 候正相关显著(>0.8)	0 候时达到 0.942
孟加拉湾经向季风环流指数	-1 ~ 1 候正相关显著(>0.8)	0 候时达到 0.915

注:相关系数均通过 0.01 置信度检验。

考虑到 20 世纪 70 年代末—80 年代初大气环流与副热带高压系统存在较明显的年代际突变^[5],为便于分析对比,我们将 1958—1997 年资料分为 1958—1977 年前 20 年和 1978—1997 后 20 年两段来分别进行处理,将 1958—1977 年前 20 年时段的副热带高压面积指数划分成以下 5 类:异常偏弱 A1(面积指数 0~40)、偏弱 A2(面积指数 41~80)、正常 A3(面积指数 81~120)、偏强 A4(面积指数 121~160)、异常偏强 A5(面积指数 161~200)。1978 年以后的副热带高压面积指数有增强的变化趋势^[5],故 1978—1997 年后 20 年时段的副热带高压面积指数参照以上标准每段数值作适当放大后予以划分。

上述 11 个影响因子与副热带高压面积指数存在较好的时滞相关性,但能否有效诊断和区分不同

强度的副热带高压面积指数尚需要作进一步的统计分析,为此针对不同强度的副热带高压面积指数,分别制作上述 11 个影响因子的统计线箱图,以选择能够较好区分和判别不同副热带高压面积指数情况的季风影响因子。

2 影响因子的聚类选择

分别进行 1958—1977 年期间前述 11 个初始季风影响因子与 5 类副热带高压面积指数的统计箱线图分析(图 1,只选择其中 4 幅典型图,其余略)。从资料可知样本总数是 145,图中盒子中间的线为样本中位数,分别表示 5 类副热带高压面积指数的统计中值,盒子的上底和下底间为内四分位间距,盒子

上下的两条线分别为样本的 25% 和 75% 分位数,表示中值附近 50% 样本的分布域;虚线贯穿盒子上下,表明样本其余部分(野值例外);“+”表示野值。

从图 1a 可以看出,对马斯克林冷高压强度指数来说,A1, A2, A3 三类不同强度的副热带高压面积指数与 A4, A5 两类不同强度的副热带高压面积指数基本没有交叉,可以完全区分开;A1, A2, A3 三类不同强度的副热带高压面积指数之间虽有部分重合,但中位数相对独立,故亦有一定程度的可分性,A4, A5 情况亦类似,故马斯克林冷高压强度指数可以较好区分 5 类不同强度的副热带高压面积指数,选其为聚类分析因子。图 1b 所示的青藏高压活动指数有类似情况,亦能较好区分 5 类不同强度的副热带高压面积指数(尤其是 A2, A3, A4 三类),可选其为聚类分析因子。图 1c 的江淮梅雨对流降水率(MCP)在 A1, A2, A3 三类不同强度的副热带高压面积指数之间存在明显重合现象,很难区分,不适合

作聚类分析因子,故予剔除。同理,孟加拉湾经向季风环流(JIV)对应的 5 类不同强度的副热带高压面积指数之间亦存在较明显的重合现象,不适宜作聚类分析因子,亦予以剔除。

基于上述类似的分析判别原则,通过统计箱线图分析,最后选取得到 7 个较好的季风影响因子:马斯克林冷高压强度指数、澳大利亚冷高压强度指数、中南半岛感热通量、索马里低空急流指标、印度季风 OLR 指标、青藏高压活动指数、孟加拉湾纬向季风环流指数。

对 1978—1997 年后 20 年的影响因子,通过相同的方法分析,亦可找到能够较好区分 5 类副热带高压强度的 7 个季风因子:马斯克林冷高压强度指数、澳大利亚冷高压强度指数、中南半岛感热通量、南海低空急流指标、印度季风 OLR 指标、孟加拉湾纬向季风环流指数、孟加拉湾经向季风环流指数。

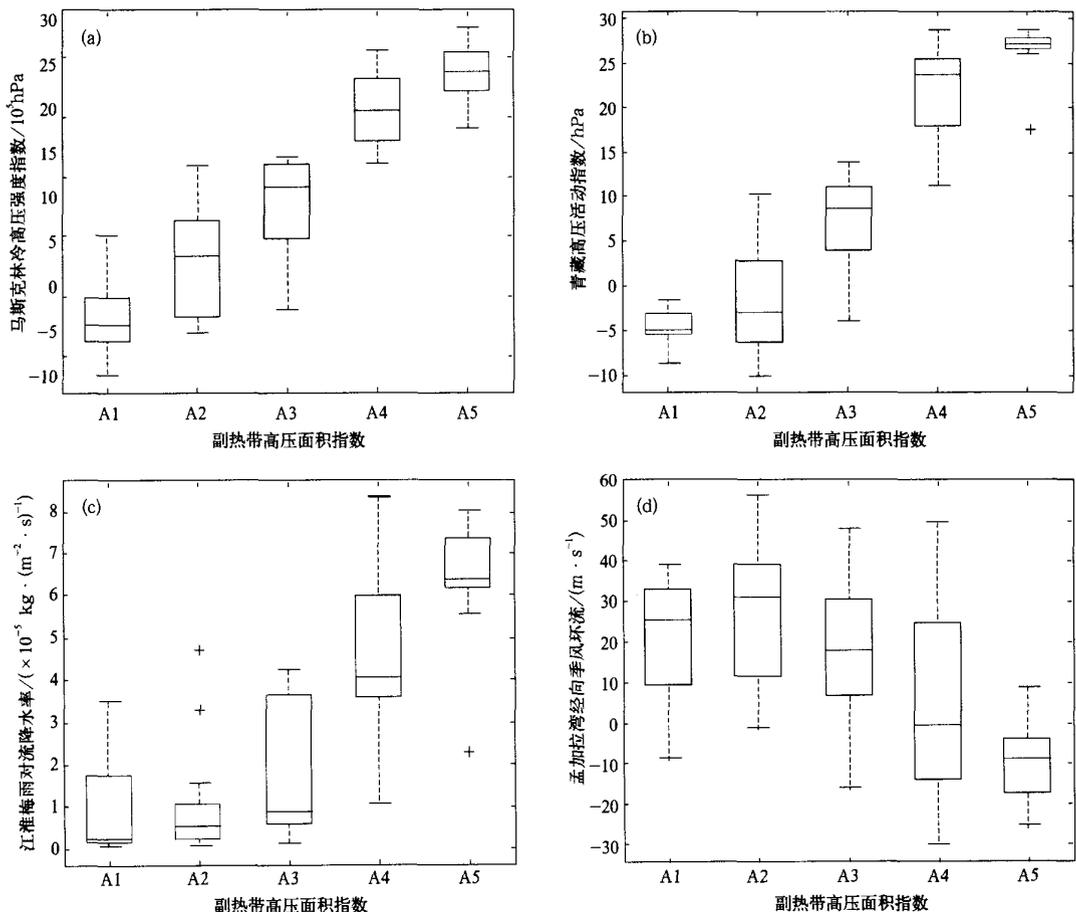


图 1 4 个季风影响因子的箱线图

3 方法的基本思想

用上述 7 个季风影响因子构成 7 维特征空间, 分别将 5 类副热带高压强度对应的季风影响因子的 7 维样本序列点集进行特征空间投影, 然后进行聚类分析, 以划分出各类副热带高压强度对应的高维季风影响因子点集在特征空间中的投影映射域, 进而实现对副热带高压强度的划分。针对不同聚类方法存在的优缺点, 本研究基于综合优化方法思路进行聚类分析。

模糊 C 均值聚类(Fuzzy C-Means, FCM)是一种局部搜索和目标逼近能力强、应用广泛的聚类算法。由于各类强度的副热带高压面积指数之间并无严格的界限, 加上观测资料本身不可避免的一些误差, 因此 FCM 算法及其隶属度概念, 适宜于刻画和描述副热带高压指数分类的上述模糊特征。但 FCM 方法存在两个固有缺点: 迭代收敛过程中的误差局部小和聚类结果对初值的敏感性。遗传算法(Genetic Algorithm, GA)是一种基于生物自然选择与自然遗传机理的全局优化算法, 具有全局搜索优势, 但其局部优化能力有限。因此, FCM 与 GA 算法存在很好的优势互补性。但是, 无论 FCM 算法还是 GA 算法, 都需要预先人为给出聚类数目, 这既缺乏充分的客观性, 也不利于副热带高压指数强度的自动分类实现。为此, 可以借助于模糊减法聚类(Fuzzy Subtractive Clustering, FSC), 该方法将每个数据点均作为可能的聚类中心, 根据该数据点周围的样本数据点密度来估算该点作为聚类中心的可能性, 第一个聚类中心选出后, 继续采用类似方法选择下一个聚类中心, 直到所有剩余数据点作为聚类中心的可能性低于某一阈值。减法聚类的上述思想使其适宜于实时副热带高压强度聚类数的客观估算。但由于 FSC 只是单次判别算法, 用它来搜索和逼近各聚类中心的最佳位置, 效果并不好。

综合 FCM, GA 和 FSC 方法各自的优势和缺点, 提出如下技术途径和解决方案: 首先用 FSC 方法客观估算出各类强度的副热带高压面积指数样本在特征空间中的聚类数, 之后利用遗传算法全局搜索优势进行聚类分析, 确定出具有较好全局结构特性的聚类结果, 随后利用 FCM 局部寻优特性对 GA 聚类结果进行局部调整优化, 最后得出副热带高压面积指数综合优化聚类结果和各类副热带高压强度

对应的季风影响因子在特征空间中的映射区域, 以此作为副热带高压强度分类的判据。模糊 C 均值聚类、遗传算法和模糊减法聚类均是成熟可靠的计算方法, 其算法原理可参考相关的文献专著。

4 综合聚类分析

以 1958—1977 年 20 年 5 类副热带高压面积指数及其对应的 7 维季风影响因子的样本点集进行综合聚类分析。由于副热带高压活动的复杂性与观测资料误差, 各类副热带高压面积指数对应的影响因子的高维样本点在特征聚类空间中的投影分布表现得较为散乱, 影响因子样本点集若未经处理即汇集映射于同一特征空间, 则各类投影点交叉重叠, 变得杂乱无章, 根本无法进行聚类分析(图略)。

为此, 基于第 3 章所述的方法思路, 首先对各类副热带高压面积指数样本作独立聚类分析, 在此基础上将各类副热带高压指数独立的聚类结果进行集成并作必要的分割处理, 最后得到 7 维特征空间中各类副热带高压面积指数分别对应的特征隶属区域(每类副热带高压强度可以有多个独立的特征隶属区域)并以影响因子在特征空间中投影点的落区位置属性作为副热带高压强度划分的依据。

第一阶段: 先对各类副热带高压面积指数的影响因子样本在 7 维特征空间中的投影点集单独进行聚类, 找出其空间分布特征, 确定各类副热带高压指数的影响因子样本在特征空间中的聚类中心和隶属范围, 并通过设定适当临界隶属度, 将远离聚类中心的疏散点作为噪声去除。

① 先对 7 维特征空间中的影响因子样本数据进行模糊减法聚类, 客观确定出聚类中心数目。经减法聚类确定的每类副热带高压强度对应的影响因子的聚类数均为 2 类, 即每类副热带高压强度对应的影响因子在 7 维特征空间中均有两个聚类中心和隶属映射区域。

② 引入遗传算法(GA)分别对每类副热带高压强度的影响因子样本进行聚类中心搜索, 确定出各聚类中心对应的隶属范围, 通过设定适当的临界隶属度, 滤除边远零散点。遗传算法的聚类结果从全局意义上较好找出和确定了每类副热带高压强度的聚类中心和隶属范围, 但聚类结果的局部效果仍有待改进完善。

③ 采用模糊 C 均值聚类方法(FCM)对 GA 的

聚类结果进行局部调整,即用 GA 确定的聚类中心替代常规 FCM 算法中随机产生的初始聚类中心,再作进一步的 FCM 聚类优化。

经上述步骤处理后,可得到 5 类副热带高压强度各自对应的高维特征空间中的影响因子的聚类中心和隶属域分布。显然,这些独立的特征空间中的聚类中心和隶属区域尚不能用于实况副热带高压面积指数的判别划分,必须将它们放入同一特征空间中集成,并作必要界定和分割处理方可用作聚类分析判据。

第二阶段:将上述 5 类副热带高压强度对应的 5 个独立特征空间以及空间中的聚类中心和隶属域

合并于同一特征空间,旨在最终界定和分割出每类副热带高压强度在同一高维特征空间中各自的隶属区域范围。为便于直观显示,每类副热带高压强度对应的聚类中心和隶属域分别被分解为若干三维图像表示。

经上述步骤综合处理后的结果表明,高维特征空间中,5 类副热带高压强度的聚类中心位置彼此相互分离,各类副热带高压强度的隶属域基本处于相对独立的位置(图 2 ~ 图 5),表明所选的影响因子能够较好地甄别和区分 5 类副热带高压强度,所建立的特征空间聚类结果具有较好的代表性和可分性。

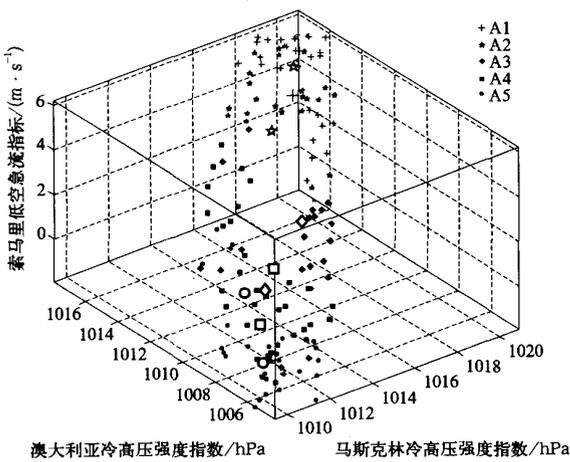


图 2 马斯克林冷高压强度指数-澳大利亚冷高压强度指数-索马里低空急流指标三维特征空间聚类图

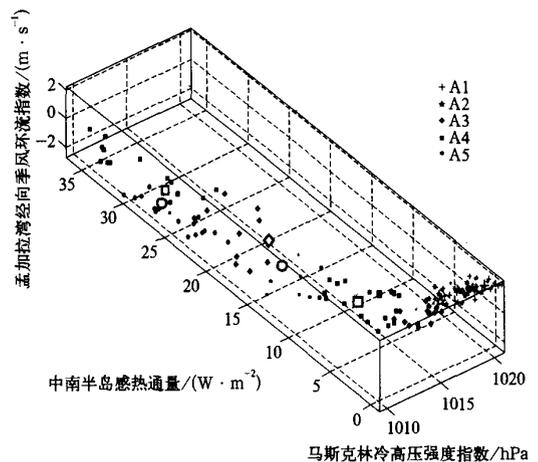


图 3 马斯克林冷高压强度指数-中南半岛感热通量-孟加拉湾经向季风环流指数三维特征空间聚类图

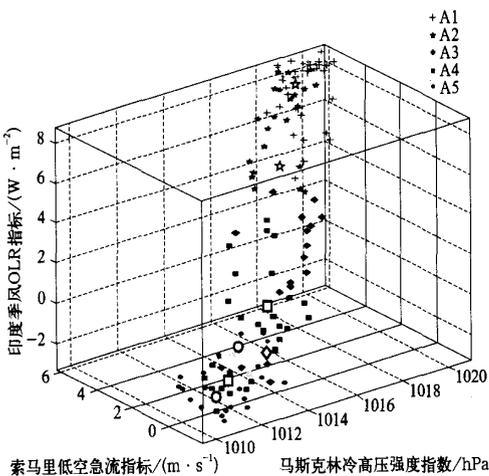


图 4 马斯克林冷高压强度指数-索马里低空急流指标-印度季风 OLR 指标三维特征空间聚类图

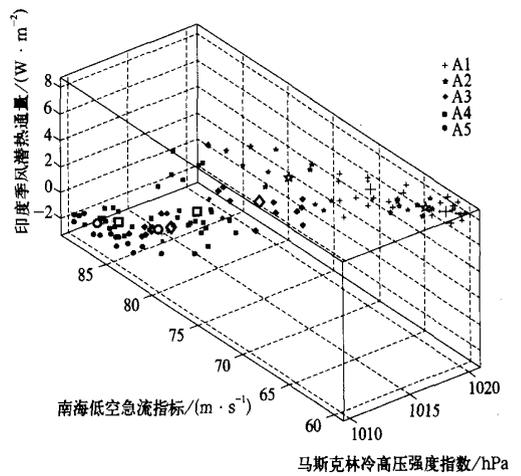


图 5 马斯克林冷高压强度指数-南海低空急流指标-印度季风潜热通量三维特征空间聚类图

对 7 个影响因子组合可以得到 13 幅类似图 2 ~ 5 的三维聚类图(其余 9 幅略),经同样步骤处理的 1978—1997 年后 20 年资料可以得到 12 幅类似的聚类图(图略)。经以上两个阶段处理后所得的 7 维特征空间的聚类中心和隶属区域即可作为实际副热带高压面积指数强度划分的判据。

5 应用试验

为检测和验证所建特征空间聚类模型的副热带高压分类预报效果,将 1958—1977 年和 1978—1997 年两个时段每类副热带高压强度的影响因子样本分别带回相应的特征空间分类模型进行判别,以检验模型的分类效果,由于每类副热带高压强度对应的季风影响因子均超前副热带高压 1 候,因此判别结果有 1 候的预报意义。具体判别步骤为:选择副热带高压最弱的 31 候(分别对应的是第 1—3, 7—8, 10, 12—16, 20—22, 24, 66, 72—84, 95—96 候;也就是对应于 1958—1967 年 1 月 1—15 日, 1958—1967 年 1 月 31 日—2 月 9 日, 1958—1967 年 2 月 16—19 日, 1958—1967 年 2 月 24 日—3 月

21 日, 1958—1967 年 4 月 5—19 日, 1958—1967 年 4 月 26—30 日, 1958—1967 年 11 月 22—26 日, 1958 年 12 月 22 日—1977 年 2 月 24 日, 1968—1977 年 12 月 16—25 日)和副热带高压最强的 13 候(分别对应的是第 38—39, 44—45, 107—108, 111, 116, 118—121, 127 候;也就是对应于 1958—1967 年 7 月 4—13 日, 1958—1967 年 8 月 1—10 日, 1968—1977 年 5 月 14—23 日, 1968—1977 年 6 月 4—8 日, 1968—1977 年 6 月 29—7 月 3 日, 1968—1977 年 7 月 8—27 日, 1968—1977 年 8 月 21—25 日)7 个影响因子的时间序列,构成 7 维特征点序列并映射投影到特征聚类空间,通过判断该特征映射点在特征聚类空间中落区位置的属性,即可判定相应的副热带高压强度类别,进而实现对副热带高压面积指数的诊断预测。

表 2 是 1958—1977 年样本所建的副热带高压聚类模型的判别效果。其中 5 种强度类型的副热带高压面积指数判别正确率均大于 60%,特别是最弱和最强两种类型的副热带高压判别准确率更是高于 80%。上述结果表明,该时段模型的副热带高压判别预报效果是准确可行的。

表 2 1958—1977 年时段模型的判别拟合效果比较

实际情况	判别结果(样本数)					判定的正确率/ %
	最弱	较弱	正常	较强	最强	
最弱	25	5	1	0	0	80.6
较弱	4	25	1	3	0	75.8
正常	1	3	16	5	0	64
较强	0	0	5	31	7	72.1
最强	0	0	0	2	11	84.6

1978—1997 年样本所建聚类判别模型的副热带高压判别预报准确率均大于 70%(表 3),其中最弱类型副热带高压判别效果达到 100%,总体判别预报效果优于 1958—1977 年聚类判别模型的副热带高压分类效果。

为进一步验证所建聚类模型的副热带高压分类效果,将 1978—1997 年样本序列代入 1958—1977

年样本所建的分类型模型进行预报判别。结果表明,模型对最强和最弱两种极端情况的副热带高压面积指数的判别预报效果依然很好,属性判别基本正确,只有个别最弱/最强的样本被误判为较弱/较强(图略)。试验判别结果如表 4,平均判别正确率大于 70%,但略低于同期分类模型的判别准确率(表 3)。

表 3 1978—1997 年时段模型的判别拟合效果比较

实际情况	判别结果(样本数)					判定的正确率/ %
	最弱	较弱	正常	较强	最强	
最弱	9	0	0	0	0	100
较弱	5	34	5	0	0	77.3
正常	0	4	21	1	1	77.7
较强	0	0	4	33	4	80.5
最强	0	0	0	3	21	87.5

表 4 1978—1997 年样本代入 1958—1977 年时段模型所得判别拟合效果比较

实际情况	判别结果(样本数)					判定正确率/ %
	最弱	较弱	正常	较强	最强	
最弱	7	2	0	0	0	77.8
较弱	6	33	5	0	0	75
正常	0	4	20	2	1	74.1
较强	0	0	4	32	5	78
最强	0	0	0	3	21	87.5

将 1958—1977 年的样本代入 1978—1997 年样本所建的判别模型中进行判别试验,结果见表 5。判别结果尽管略低于同期模型的拟合效果(表 2),

但平均判别准确率大于 60%,仍然有较好的适用性和有效性。

表 5 1958—1977 年样本代入 1978—1997 年时段模型所得的判别拟合效果比较

实际情况	判别结果(样本数)					判定正确率/ %
	最弱	较弱	正常	较强	最强	
最弱	26	2	3	0	0	83.9
较弱	6	22	5	0	0	66.7
正常	1	4	15	4	1	60
较强	0	0	5	31	7	72.1
最强	0	0	0	3	10	76.9

6 小 结

通过对副热带高压面积指数与亚洲夏季风系统诸重要区域要素因子进行相关分析,提取出 11 个候选的副热带高压影响因子,随后对候选的 11 因子再次进行统计箱线图分析,确定出能较好划分副热带高压强度的 7 个季风影响因子。用这 7 个影响因子构成 7 维特征空间,然后对 5 类不同强度的副热带高压面积指数所对应的影响因子点集序列进行特征空间映射投影,再通过聚类分析方法划分出不同强度类型的副热带高压所对应的影响因子在特征空间的聚类中心和隶属区域,进而实现了对副热带高压面积指数强度类型的判别划分。

针对传统聚类方法存在的缺点,本文采用模糊 C 均值聚类(FCM)遗传算法(GA)和模糊减法聚类(FSC)优势互补的思想方法途径,得到高维特征空间中各类副热带高压面积指数强度的优化聚类中心和隶属区域判据。实际应用时,通过计算和判断 7 个季风影响因子映射点在特征聚类空间中的落区位置类属,即可确定其所对应的副热带高压面积指数强度类别,进而实现副热带高压面积指数强度的自动分类。试验结果表明,本文提出的副热带高压分类判别预报思想和算法模型能较为客观准确地分析判别副热带高压面积指数的强度类型,判别结果与

实际情况基本相符,对分析和诊断、预测副热带高压活动有参考应用意义。此外,本文方法途径还可方便地应用于建立副热带高压脊点和副热带高压脊线等其他副热带高压指数的分类判别模型,具有较好的普适性和实用性。

另外,需要指出的是,由于本文主要侧重于分析方法研究,因此所选择的 10 年平均的逐候资料难以有效描述和表现副热带高压的季内和年际变化,这也是下一步将改进、完善之处。

参 考 文 献

- [1] 黄荣辉,李维京. 夏季热带西太平洋上空的热源异常对东亚上空副热带高压的影响及其物理机制. 大气科学,1989,13(特刊):107-116.
- [2] 喻世华,杨维武. 副热带季风环流圈特征及其在东亚夏季环流中的作用. 应用气象学报,1991,2(3):242-247.
- [3] 张韧,史汉生,喻世华. 西太平洋副热带高压非线性稳定性问题研究. 大气科学,1995,19(6):687-700.
- [4] 张韧,史汉生,沙文钰. 夏季东亚副热带反气旋进退的非线性机理讨论. 应用数学和力学,1999,20(4):418-426.
- [5] 张韧,余志豪,蒋全荣. 南海夏季风活动与季内副高形态和西伸. 热带气象学报,2003,19(2):113-121.
- [6] 张韧,何金海,董兆俊. 南亚夏季风影响西太副高活动的小波包能量诊断. 热带气象学报,2004,20(2):113-121.
- [7] 中央气象台长期预报组. 长期天气预报技术经验总结(附录). 北京:气象出版社,1976:5-6.

Fuzzy Clustering Diagnosing and Forecasting on Subtropical High Intensity Index

Hong Mei¹⁾²⁾ Zhang Ren¹⁾²⁾³⁾ Wu Guoxiong¹⁾ He Jinhai³⁾ Yu Dandan²⁾

¹⁾ (LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029)

²⁾ (Institute of Meteorology, PLA University of Science and Technology, Nanjing 211101)

³⁾ (KLME, Nanjing University of Information Sci & Tech, Nanjing 210044)

Abstract

In order to analyze and diagnose the correlative characters and phase relation between subtropical high and the east-Asia summer monsoon circulation and thermal factors as well as southern hemispheric pressure field, an idea of synthetically combining FCM, GA with FSC to make clustering analyze and diagnosis on subtropical high and its influence factors is presented. Based on observational and statistical facts, some initial subtropical high influence factors in east-Asia monsoon system are chosen to form a high dimensional feature spaces. Firstly, the FSC technique is used to objectively estimate the clustering number of samples in the feature space, then the GA technique is used to make global optimization search and the FCM technique is used to make further local optimization searching and adjusting on the GA global optimization results. Finally, the synthetical optimized clustering results of subtropical high indexes and its influence factors in the feature mapping space are taken as the classification criterion for subtropical high intensity index classification. By the clustering analyze and diagnosis of using above synthetical techniques, some important portentous characteristic facts of subtropical high changing are revealed and discovered, and a corresponding clustering diagnosis and prediction are carried out.

On the basis of above researches, some examples of subtropical high index and its monsoon influence factors are analyzed, diagnosed and predicted by using the synthetical clustering technique, the experiments results show that the classification correctness and forecast precision of 5 grades subtropical high indexes are superior to 60%, especially, the correctness and precision in the case of the weakest and the strongest subtropical high indexes are better than 80%, and similar better subtropical high classification results are also gotten by different examples clustering experiments, the validity and reliability of the synthetical clustering technique are well supported and approved by above experiments.

The idea and technique presented not only can overcome the shortcoming of GA/FCM short of sufficient local/global optimization, but also can objectively determine the number of clustering centers. In the actual application, by calculating and judging the mapping location of monsoon influence factors in their high-dimensional feature spaces, the intensity grade of subtropical high index are accordingly determined, and that the subtropical high index diagnosis and forecast are also carried out.

Key words: subtropical high index; monsoon system; fuzzy clustering; genetic algorithm