

低纬高原雹云识别物元可拓模型 及效果检验

李桂华 金少华

(云南省玉溪市气象局, 653100)

提 要

雹云识别是一个多参数的模式识别问题,应用物元可拓思想将单参数雹云识别的不相容性转化为相容性问题。在与成都地区雹云识别的物元可拓模型进行对比后,根据低纬高原地形及气候特点,通过适当增加参数,构造完成低纬高原雹云识别的物元可拓模型,以此进行雹云识别,准确率达 86% 以上。

关键词: 低纬高原 雹云识别 对比 物元 可拓集 关联函数

引 言

目前,利用天气雷达并结合探空资料仍是识别雹云的主要手段^[1],而实际上雹云识别是一个多参数的模式识别问题,用单参数判别雹云结果往往不相容。本文利用蔡文^[2]提出的物元可拓思想,用成都地区雹云识别的物元可拓模型^[3],及其所选定的相关参数在低纬高原构造模型,检验效果很不理想,根据低纬高原地形及气候特点^[4],其地形对雹云的形成影响较大,通过适当增加参数,构造完成低纬高原雹云识别的物元可拓模型,以此进行雹云识别,准确率达到 86% 以上。

1 物元及物元可拓模型

设 $P_0 \subset P$, 对任何待识别对象 $p \in P$, 判别 p 属于 P_0 的程度, 则 p 的 n 个特征 $C_1, C_2, C_3, \dots, C_n$ 及相应量值 $V_1, V_2, V_3, \dots, V_n$ 可用下面 n 维物元表示:

$$R_x = (P, C, V) = \begin{bmatrix} P, & C_1, & V_1 \\ & C_2, & V_2 \\ & C_3, & V_3 \\ & \dots & \dots \\ & C_n, & V_n \end{bmatrix}$$

1.1 经典物元和节域物元

由标准集合 P_0 和它的 n 个特征 C_i 及

其标准量值范围 $V_{0i} = \langle a_{0i}, b_{0i} \rangle$ 组成的物元称为经典物元, 记为:

$$R_0 = (P_0, C, V_0) = \begin{bmatrix} P, & C_1, & \langle a_{01}, b_{01} \rangle \\ & C_2, & \langle a_{02}, b_{02} \rangle \\ & C_3, & \langle a_{03}, b_{03} \rangle \\ & \dots, & \dots \\ & C_n, & \langle a_{0n}, b_{0n} \rangle \end{bmatrix}$$

由集合 P 和它的 n 个特征 C_i 及其将标准量值范围加以拓广了的量值范围 $V_{pi} = \langle a_{pi}, b_{pi} \rangle$ 组成的物元称为节域物元, 记为:

$$R_p = (P, C, V_p) = \begin{bmatrix} P, & C_1, & \langle a_{p1}, b_{p1} \rangle \\ & C_2, & \langle a_{p2}, b_{p2} \rangle \\ & C_3, & \langle a_{p3}, b_{p3} \rangle \\ & \dots, & \dots \\ & C_n, & \langle a_{pn}, b_{pn} \rangle \end{bmatrix}$$

显然有 $V_{0i} \subset V_{pi} (i = 1, 2, 3, \dots, n)$ 。

1.2 计算关联函数值

模式识别是用事物关于某些量值符合要求的程度来表达的,而这种程度通常用由矩定义的关联函数值来刻划。关联函数值由下式计算:

$$K_i(V_i) = \begin{cases} -\rho(V_i, V_{0i})/|V_{0i}| & \text{当 } V_i \in V_{0i} \\ \rho(V_i, V_{0i})/(\rho(V_i, V_{pi}) - (\rho(V_i, V_{0i}))) & \text{当 } V_i \notin V_{0i} \end{cases} \quad (1)$$

其中

$$\begin{aligned} \rho(V_i, V_{0i}) &= |V_i - (a_{0i} + b_{0i})/2| - (b_{0i} - a_{0i})/2 \\ \rho(V_i, V_{pi}) &= |V_i - (a_{pi} + b_{pi})/2| - (b_{pi} - a_{pi})/2 \end{aligned} \quad (2)$$

式(1)中 $|V_{0i}|$, 式(2)中 $|V_i - (a_{0i} + b_{0i})/2|$ 和 $|V_i - (a_{pi} + b_{pi})/2|$ 分别为各参数的模值。

1.3 综合关联度及判别准则

若 $V_k \in V_{pk}$, 则由下式计算 p 与 P_0 的综合关联度:

$$K(p) = \sum \lambda_i K_i(V_i) \quad (i = 1, 2, 3, \dots, n) \quad (3)$$

式中, λ_i 为各特征的权系数, p 是否属于 P_0 的判别准则如下: 当 $K(p) \geq 0$ 时, $p \in P_0$; 当 $-1 \leq K(p) \leq 0$ 时, $p \notin P$, 但 $p \in P_0$; 当 $K(p) \leq -1$ 时, $p \notin P$, $p \notin P_0$ 。

为便于构造物元矩阵, 各参数观测值按下式进行规格化处理:

$$y_i = (x_i - x_{im}) / (x_{iM} - x_{im}) \quad (4)$$

式中, x_{iM} 和 x_{im} 分别为参数 i 的最大值和最小值。

2 雹云识别的物元可拓模型的构造及检验

2.1 资料处理

本文利用地处滇中腹地, 具有典型低纬高原气候特征的玉溪历年 5~10 月(1992~2002 年)天气雷达观测的降雹资料和昆明站高空探测资料进行分析计算, 其中雷达回波资料取自投入实际业务运行的 711B 型天气

雷达, 稳定度指标数值由昆明探空站取得, 在此期间共观测到 254 个雹云和雷雨云样本。由前 210 个作为建模样本, 用其参数值构造经典物元和节域物元, 预留后 44 个样本作为待识别检验样本。用于建模和检验的雹云和雷雨云样本数分布见表 1。

表 1 玉溪地区雷达回波资料样本分布

	雹云样本		雷雨云样本	
	建模样本	待识别样本	建模样本	待识别样本
4、5月	48	14	4	0
6~8月	52	13	31	8
9、10月	5	0	69	9
总计	105	27	104	17

2.2 参数的选取和识别效果的比较

以成都地区雹云识别的物元可拓模型为标准, 用下列 5 个参数作为雹云和雷雨云的判别指标: (1) C_1 : 回波顶高 H_e (km), (2) C_2 : 回波顶温度 T_h ($^{\circ}\text{C}$), (3) C_3 : 负正温区厚度比, (4) C_4 : 等效反射因子 Z_e , (5) C_5 : 稳定度 $\Delta\theta_w$ ($^{\circ}\text{C}$) 构造模型。由玉溪地区的雹云和雷雨云建模样本计算出各参数的均值和方差见表 2。所得判别结果很不理想, 正确率仅为 59.1%。

表 2 各类样本的因子均值和方差

类别	各类样本	C_1		C_2		C_3		C_4		C_5		C_6		C_7	
		均值	方差	均值	方差	均值	方差	均值	方差	均值	方差	均值	方差	均值	方差
雹云	105	0.642	0.165	0.256	0.104	0.522	0.120	0.647	0.125	0.305	0.136	0.765	0.212	0.203	0.104
雷雨云	105	0.402	0.194	0.546	0.236	0.304	0.142	0.302	0.106	0.502	0.164	0.306	0.132	0.102	0.092

在低纬高原强对流天气的发生、发展过程中, 地形抬升和充沛的水汽是产生雹云的主要条件之一。低纬高原地形地貌复杂, 山

岭纵横, 拔海高差悬殊, 气候垂直差异十分显著, 而且山岭旁多有河流穿过, 因此低纬高原强对流性天气明显具有局地性的特点。由于

这一特点,一方面,雹云云体的移动具有突发性、快速性;另一方面,雹云云体的凝结高度变化突然,而且由于地形及海拔高度影响,云体凝结高度明显高于其他地区,因此在考虑以上5个参数的同时增加下列两个因子:(6) C_6 :回波整体移速 $V(\text{km}\cdot\text{h}^{-1})$; (7) C_7 :云体凝结高度 $H_n(\text{km})$ 。用这7个参数来构造低纬高原的雹云识别的物元可拓模型,(所增参数均值和方差见表2),把雹云和雷雨云两类样本的各参数视作服从正态分布的随机变量,其均值为 $\bar{C}_i (i = 1, 2, \dots, n)$, 方差为 σ_i 。根据误差理论,参数取值 C_i 的误差小于 $3\sigma_i$ 的概率为 99.97%, 因此,从理论上讲,当两类样本经典物元的各参数取值范围设定为 $C_i \in [\bar{C}_i - 3\sigma_i, \bar{C}_i + 3\sigma_i]$ 时,判定样本属于某类的误差小于 $3\sigma_i$ 的可靠性为 99.97%, 不过,考虑到参数值 C_i 是按公式(4)规格化为 $[0, 1]$ 之间的值,因此,当 $\bar{C}_i - 3\sigma_i \leq 0$ 时,取区间下限为 0; 当 $\bar{C}_i + 3\sigma_i \geq 1$ 时,取区间上限为 1, 故得雹云和雷雨云的经典物元如下:

$$R_{01} = (P_{01}, C, V_0) = \begin{bmatrix} P_{01}, & C_1, & \langle 0.147, 1.000 \rangle \\ & C_2, & \langle 0.000, 0.568 \rangle \\ & C_3, & \langle 0.162, 0.882 \rangle \\ & C_4, & \langle 0.272, 1.000 \rangle \\ & C_5, & \langle 0.000, 0.713 \rangle \\ & C_6, & \langle 0.129, 1.000 \rangle \\ & C_7, & \langle 0.000, 0.515 \rangle \end{bmatrix}$$

$$R_{02} = (P_{02}, C, V_0) = \begin{bmatrix} P_{02}, & C_1, & \langle 0.000, 0.984 \rangle \\ & C_2, & \langle 0.000, 1.000 \rangle \\ & C_3, & \langle 0.000, 0.730 \rangle \\ & C_4, & \langle 0.000, 0.620 \rangle \\ & C_5, & \langle 0.000, 0.994 \rangle \\ & C_6, & \langle 0.000, 0.702 \rangle \\ & C_7, & \langle 0.000, 0.378 \rangle \end{bmatrix}$$

由于节域物元参数量值范围应是经典物元相应参数量值范围的适当拓广,因此,节域物元可表示为:

$$R_p = (P_{01}, C, V_p) = \begin{bmatrix} P, & C_1, & \langle 0.000, 1.200 \rangle \\ & C_2, & \langle 0.000, 1.200 \rangle \\ & C_3, & \langle 0.000, 1.200 \rangle \\ & C_4, & \langle 0.000, 1.200 \rangle \\ & C_5, & \langle 0.000, 1.200 \rangle \\ & C_6, & \langle 0.000, 1.200 \rangle \\ & C_7, & \langle 0.000, 1.200 \rangle \end{bmatrix}$$

第1个待识别样本的物元可表示为

$$R_{X(1)} = (P, C, V) = \begin{bmatrix} P, & C_1, & 0.452 \\ & C_2, & 0.447 \\ & C_3, & 0.309 \\ & C_4, & 0.304 \\ & C_5, & 0.277 \\ & C_6, & 0.324 \\ & C_7, & 0.407 \end{bmatrix}$$

现由公式(1)~(3),根据玉溪地区44个待识别的雹云和雷雨云样本的7参数的规格化后的数值(见表2),计算样本1分别对雹云集合 N_1 和雷雨云集合 N_2 的关联度 K_1 、 K_2 步骤如下:

首先计算样本1的指标 C_1 对雹云集合 N_1 的关联度 K_{1C_1} ,由式(2)有:

$$\begin{aligned} \rho(V_1, V_{01}) &= |V_1((a_{01} + b_{01})/2) - (b_{01} - a_{01})/2| - \\ &= |0.452 - 0.574| - 0.427 \\ &= -0.305 \end{aligned}$$

$$\begin{aligned} \rho(V_1, V_{p1}) &= |V_1 - (a_{p1} + b_{p1})/2| - \\ &= |0.452 - 0.6| - 0.6 \\ &= -0.452 \end{aligned}$$

$$|V_{01}| = |b_{01} - a_{01}| = 0.853$$

因为 $V_1 \in V_{01}$, 所以

$$k_{1c1} = -\rho(V_1, V_{01})/|V_{01}| = 0.358$$

类似,可以计算出 $k_{1c2} = 0.213, k_{1c3} = 0.204, k_{1c4} = 0.044, k_{1c5} = 0.388, k_{1c6} = 0.224, k_{1c7} = 0.210$ 。若各参数视作等权的,则由公式(3)可得 $K_1 = 0.234$;同理可计算出样本1对雷雨天集合 N_2 的关联度 $K_2 = 0.356$ 。由于 $K_2 > K_1$,因此,判别样本1属于 N_2 。

通过上述计算方法,可得44个待识别样本对雹云集合 N_1 和雷雨天集合 N_2 的关联度 $K_1、K_2$ 及判别结果,所得判别结果正确率达到86.4%。由此可以看出,低纬高原因其不同于其它地区的地形和气候特点,所构造模型也必须与其特点相适应,才能达到较好的应用效果,两种模型的检验效果对比见表3。

表3 两种模型的检验效果对比

模型	建模 样本数	待识别 样本数	识别准 确率/%	漏报 率/%	空报 率/%
5参数	210	44	59.1	22.7	18.2
7参数	210	44	86.4	4.5	9.1

3 讨论

(1)物元可拓的思想是根据事物关于特征的量值来判断事物属于某集合的程度,而关联函数使识别方法更为精细化,因此,物元可拓法为雹云识别开辟了新的途径。

(2)物元可拓法用于模式识别和分类概念清晰,分辨率高,灵活性好。关联函数值的计算简单,用于雹云识别有一定的实用性。

(3)物元可拓模型的构造必须结合当地特点,低纬高原因其特殊的纬度及海拔高度,形成了与其它地区不同的强对流天气生成的特点,结合其特点构造模型取得了很好的效果。

参考文献

- 1 Eccles P J, D Atlas. A new method of hail detection by dual-wavelength radar[C]. Proc. 14th Radar Meteor. Conf. 1970: 1—6.
- 2 蔡文著. 物元模型及其应用. 北京: 科技文献出版社, 1994: 18~106.
- 3 李祚泳, 蒯雷, 邓新民. 雹云识别的物元可拓模型及其效果检验. 高原气象, 2001, 20(2): 197~201.
- 4 秦剑, 琚建华, 解明恩等. 低纬高原天气气候. 北京: 气象出版社, 1997: 93~120.

Recognition Model of Low Latitude Plateau Hail Cloud with Matter Element and Extension Sets and Its Result Verification

Li Guihua Jin Shaohua

(Yuxi Meteorological Bureau, Yunnan Province 653100)

Abstract

Hail cloud recognition is a pattern recognition problem of multi-parameter. The recognition result of hail cloud based on single parameter is always non-compatible. The basic thinking of matter elements and extension set is used for changing the non-compatible problem into compatible problem. The comparison study results show that the accuracy of hail cloud recognition by this method is over 86%.

Key Words: low latitude plateau hail cloud recognition construct matter element extension set dependent function