

利用天气雷达回波资料识别雹云的 统 计 学 方 法

戴铁丕 王振会 赵玲玲

(镇江地区气象局)

提 要

本文利用江苏省天气雷达回波资料建立多元和逐步回归方程。并对所用几个多元回归和逐步回归方程来识别雹云的准确性进行了比较。还利用同一省不同地区的1985年新资料和上海、宁波两市1979—1984年资料对这些方程的运用范围进行了检验。结果表明方程(6)可以在江苏本省各雷达站试验使用,其效果较理想,但在邻近省、市使用时要慎重。

冰雹是我国主要的灾害性天气之一,但由于它属于中、小尺度天气系统,空间尺度小,时间尺度短,发展、演变迅速,因此用常规的天气图方法难以发现和跟踪。新近随着中尺度天气学和雷达探测技术的发展,对中尺度天气系统的发生、发展、演变和强对流天气雷达回波结构模式以及它们的相互作用有了一定认识,为利用雷达短时预报强对流天气提供了有利条件。另外,据雷达探测理论,用双波长雷达、偏振雷达、脉冲多卜勒雷达等方法可以确定云内粒子的大小、形状和相态,从而判别有无冰雹存在,效果比较好^[1,2]。70年代开始,由于计算机技术迅速发展,可将天气雷达、卫星和地面中尺度站网等现代化探测手段所取得的大量资料组合输入计算机,通过实时处理提高预报准确率。但目前我国由于探测和通讯技术的限制,对强对流天气的监测还不能实时,这就失去了识别和预报冰雹云的意义。

为了能够在我国现有设备和技术条件下,发挥711、713雷达作用,更有效和充分的利用雷达探测到的历史资料来客观和半定量地识别雹云,又考虑到迄今为止还未能用一个明确的定量指标在雷达屏幕上把雹云和雷雨云有效地区别开来,我们研究并采用了统计预报方法,找出雹云与一些雷达回波定量参数、气象因子之间的统计关系,得到了一些有意义的结果。

一、资 料

为了建立识别冰雹等强对流天气的多元和逐步回归方程, 选取江苏省1979—1984年4—6月发生强对流天气的101个样本, 其中47个样本为冰雹等强对流天气, 54个样本为一般雷雨大风天气。建立回归方程所需的雷达回波参数资料主要取自南京、徐州、盐城的713雷达, 少数样本由江苏省其它雷达站提供。气象要素资料通常由该三地07时探空资料经过简单运算得到。没有探空资料的雷达站, 其高空气象资料由上风方向最靠近雷达站的探空站提供。

用于检验回归方程有效性的资料有两部分, 一部分取自1985年江苏省扬州、南通两地, 其中雷达参数资料由该两地雷达站提供, 气象资料分别由南京和上海探空站提供, 共取样本15个。另一部分28个样本取自1979—1984年宁波和上海两市雷达站和探空站, 其中宁波的探空资料由杭州探空站提供。所有取资料的雷达均经过标定。不同型号、波长的雷达所测得的回波参数差异经过统计检验, 证明可以混合统计。

在回归计算中, 我们把强对流天气按收集到的受灾程度、雹块平均和最大尺度、雷达回波参数资料强弱综合分为三类, 分别以0、1、2表示雷雨、一般冰雹和强冰雹(或龙卷)等强对流天气。

二、判别因子的选取

为了获得稳定而有效的判别方程, 在建立多元回归和逐步回归方程以前, 必须依据强对流天气发生、发展的物理因子, 天气条件和雷达气象学理论综合考虑对判别因子的筛选。由于雷达回波参数能较好地反映对流发展的程度, 又能直接从雷达探测中得到, 且经过多年实际证明, 这些参数对判别、预测冰雹有一定效果, 因此我们选取的9个判别因子中大部分为雷达参数。

1. 回波顶高度 H 。对流云回波顶高度一般指的是雷达灵敏度处在最大时所探测到的回波顶高, 它是指示雹云存在的较好的指标。回波顶愈高, 降雹概率愈大。而且, 回波顶

高还能反映接近云的顶部上升气流的强弱, 它是否穿过第一对流层顶是判别雹云存在的一个极有用的指标。考虑到回波顶高有月际变化, 不便于在回归方程中使用, 所以各月回波高度必须以某一个月(如5月)为标准进行月际订正。表1即为由150个样本得到的订正值, 而本文使用的回波顶高度等于实测值加上月订正值。

表 1

雷 达 型	月 份 订 正 值		
	4	5	6
713	2.18	0	-0.061
711	0.59	0	-0.97

2. 强回波顶高度 H_s 。经衰减后得到的 H_s 比 H 具有更好的指示雹云的意义, 这是因为雷达探测证实在雹云中上部上升气流最强。当这支最强上升气流所达高度愈高, 强对流天气愈激烈, 被托住的冰雹也愈大。据散射理论, 当干冰球直径从0.5厘米到3.0厘米

时,对3厘米雷达波长,后向散射截面几乎增加4个数量级(从 8×10^{-3} 厘米²增加到27.07厘米²)。另外,雹云中零度层高度有一定的区间(在江淮流域,这个高度约为4千米左右)。因此, H_z 愈高,冰雹在其中生长的含水量累积区的垂直厚度愈厚,有利于冰雹增长。综上所述,用3厘米波长雷达监视雹云内6—10千米高度强反射率核的数量、增长和演变可判断雹块增长。另外据文献[3],用强回波顶高识别雹云时, Z 值标准定得愈高,该值保守性愈好。在本文中,强回波顶高是指衰减30分贝后回波顶的高度。

回波顶高度与强回波顶高度的差异,还能反映回波垂直梯度的大小,而该值也是指示雹云存在极好的指标。

3.回波强度 Z 和 Z_e 。雹云回波的反射率因子 Z 或等效反射率因子 Z_e 的大小反映回波的强弱,也是判别雹云存在的一个重要指标。地面降雹概率随着强雷暴内 Z_e 极大值增大

而增加。据 Z 和 Z_e 值定义式 $Z = \sum_{i=1}^n N_i D_i^6$, $Z_e = \frac{\lambda^4}{\pi^5} \left| \frac{m^2 - 1}{m^2 + 2} \right|^{-2} \sum \sigma_M$,可见前式仅

与粒子直径 D 及数密度 N 有关,而后式除与 D 、 N 有关外,还与雷达波长 λ 、复折射指数 m (反映相态)及米散射雷达截面 σ_M 有关,因此雷达波长不同,用 Z_e 值判别雹云指标也不同。冰粒子愈大,进入米散射区以后 σ_M 值迅速增加,这也就是为什么 Z_e 值能指示雹块存在的物理原因。

4.高空风切变 高空风的垂直切变是区别强对流天气中冰雹和暴雨的一个很有用的指标。切变小,降水效率高,容易形成暴雨;而切变太大,云体倾斜大,超级单体雹云不能准稳定地维持,对形成冰雹不利。经过统计分析这个值在 0.2×10^{-3} 秒⁻¹到 6.0×10^{-3} 秒⁻¹之间容易形成冰雹天气。

除选取上述4个判别因子外,还选了其它5个判别因子,详见表2。

表 2

名 称	判 别 因 子	代 号
H_e	回波顶高度 (千米)	x_1
Z	回波强度 (dBz)	x_2
h	回波负温区厚度 (千米)	x_3
$\Delta\theta_{se}$	稳定度 (500—850百帕)	x_4
H_s	强回波顶高度 (千米)	x_5
T_{sH}	强回波顶高度处温度 (°C)	x_6
$\partial v / \partial z$	高空风切变 (1/1000秒 ⁻¹)	x_7
$H_0^{\circ C}$	零度层高度 (千米)	x_8
T_{eH}	回波顶高处温度 (°C)	x_9

三、判别方程的建立

利用前面介绍的资料, 根据多元回归和逐步回归方法, 建立判别因子和判别对象的定量关系式作为判别冰雹的工具。

1. 选取全部 9 个判别因子

全部使用在表 2 中所取的 9 个判别因子, 得容量为 18 的样本, 其中南京站占大多数, 样本平均值和标准差如表 3 所示。

表 3

因 子	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	y
平均值	14.06	53.92	10.17	-9.79	6.04	-13.57	2.73	3.93	-58.48	1.06
标准差	2.25	2.70	2.02	8.77	1.52	9.46	1.31	0.59	6.76	0.80

(1) 多元回归分析 对强对流的判断作多元回归分析, 得

$$y = 0.778 - 0.632x_1 + 0.138x_2 + 0.193x_3 - 0.022x_4 - 0.654x_5 - 0.091x_6 - 0.571x_7 - 0.244x_8 - 0.048x_9 \quad (1)$$

代表该方程显著性的 F 为 1.45, 复相关系数 $R = 0.79$, 剩余标准差 $y_N = 0.72$ 。据回归方程显著性检验方法, 取显著性水平 $\alpha = 0.05$, 得 $F_{\alpha}(9, 8) = 3.39$ 。可见 $F < F_{\alpha}$, 方程回归效果不显著。

(2) 逐步回归分析 在逐步回归分析中, 按预先选定的方差贡献显著标准 F_{α}' 对每个因子进行 F 检验, 以决定引进重要的因子。据 F_{α}' 的选取方法, 在本问题中, 若取 $\alpha = 0.05$, 则逐步回归开始时, 应取 $F_{\alpha}'(1, 16) = 4.49$ 。但实际发现, 由于各自变量与因变量的单项相关较差, 方差贡献最大的变量 x_7 的 F 值也只有 2.00, $F_{\alpha}' = 4.49 > 2.00$ 即没有任何自变量与因变量线性相关显著。为了利用现有资料进行研究, 本文适当地降低了显著性标准, 取 $F_{\alpha}' = 2.00$, 得

$$y = 1.502 - 0.315x_1 + 0.104x_2 - 0.587x_7 \quad (2)$$

以及 $F = 4.14$, $R = 0.69$, $y_N = 0.64$ 。由于 $F_{\alpha}(3, 14) = 3.24 < 4.14$, 故回归方程显著。

为了衡量判别方程技术水平的优劣, 引用下面几个指标, 它们的定义为: 拟合率 = 预报成功次数/总预报次数, 冰雹空报率 = 冰雹误报次数/实际无雹次数, 冰雹漏报率 = 冰雹漏报次数/实际有雹次数, Hedike 评分^[4] $S = (R - E)/(T - E)$ 。其中, R 是预报成

功次数, T 是总预报次数, $E = \sum_{i=1}^N R_i C_i / T$ 表示盲目预报可望成功的次数, R_i 是预报第

i 级天气出现的次数, C_i 是第 i 级天气实际出现的次数。一般确认 S 值比拟合率更能反映预报技术水平。

为了检验方程(2)的有效性, 现按回报拟合率最大的原则, 求得降雹判别指标如下: $y \leq 0.55$ 为无雹, $0.55 < y \leq 1.13$ 为弱冰雹, $y > 1.13$ 为强雹。实况与预报比较见表4。由表可计算得到拟合率=66.7%, 冰雹空报率=20%, 冰雹漏报率=7.7%, $E=6$ $S=0.5$ 。

2. 部分判别因子

由于考虑的判别因子愈多, 资料的样本愈少, 所建立的判别方程在实际使用中也不方便。另外, 据主因子筛选法, 只要选出一个最好的主因子, 然后再选出一些搭配因子即可。因子太多, 可能重复, 不能取长补短。为此又做了下面分析:

(1) 判别因子筛选 由于 x_3 、 x_9 和 x_1 相关较好, x_5 由 x_6 相关也很好, 且雷达并不能直接测出温度参数。故从备选因子中去掉 x_3 、 x_6 和 x_9 。样本容量增大到21个, 由多元回归分析得

$$y = 3.510 - 0.228x_1 + 0.056x_2 - 0.010x_4 - 0.052x_5 - 0.431x_7 - 0.221x_8 \quad (3)$$

$F=2.75$, $R=0.74$, $y_N=0.65$ 。由于 $F_{0.05}(6, 14)=2.85 > 2.75$, 为此该多元回归方程显著性仍不理想。故再用逐步回归法, 取 $F_{\alpha}'=4.41$ (因为 $F_{0.05}(1, 18)=4.41$), 得

$$y = 5.211 - 0.215x_1 - 0.407x_7 \quad (4)$$

$F=7.85$, $R=0.68$, $y_N=0.62$ 。由于 $F_{0.05}(2, 18)=3.55 < 7.85$, 因而方程(4)回归效果显著。其余结果一起列入表5中。

(2) 高空风切变因子 x_7 的方差贡献最显著, 但其数值很难得到, 故在方程(3)中再删除 x_7 , 样本容量增加到51。多元回归与逐步回归方程分别为

$$y = 0.1230 - 0.0148x_1 + 0.0364x_2 - 0.0147x_4 + 0.0620x_5 - 0.3869x_8 \quad (5)$$

$$y = 0.2173 + 0.0324x_2 - 0.0140x_4 + 0.0553x_5 - 0.3955x_8 \quad (6)$$

其计算结果全部列入表5。

(3) 为使雷达一开机就可进行预报、判别, 在备选判别因子中只取 x_1 、 x_2 、 x_5 和 x_8 , 样本容量为56。多元回归与逐步回归方程分别为

$$y = -0.230 - 0.0005x_1 + 0.0416x_2 + 0.0147x_5 - 0.3255x_8 \quad (7)$$

$$y = -0.2410 + 0.0429x_2 - 0.3204x_8 \quad (8)$$

计算结果也全部列入表5。表中拟合率和S评分值是对三级判别而言的, 它比分两级预报时低。由表4可见, 由于实际上3次弱雹被回报为强雹, 1次强雹被回报为弱雹, 故使拟合率和S值大大降低。若分两级预报, 拟合率增加到89%。这种情况在检验回归方程效果的表6中也存在。对流天气从强到弱是连续的, 分级时存在人为的不确定因素。

表 4

		实 况			
		强	弱	无	合计
回 报	强	5	3	0	8
	弱	1	3	1	5
	无	0	1	4	5
	合计	6	7	5	18

表 5 多元回归和逐步回归分析结果

备 因 个 数	选 子 个 数	方 程 序 号	样 本 数	$F'_{0.05}$	F	$F_{0.05}$	R	y_N	判别指标* y' y''	拟 合 率 %	S	冰雹 空报率 (%)	冰雹 漏报率 (%)
9	9	1	18	0.00	1.45	3.39 不显著	0.79	0.72					
	3	2	18	2.00	4.14	3.24 显著	0.69	0.64	0.55, 1.13	66.7	0.50	20.0	7.7
6	6	3	21	0.00	2.75	2.85 不显著	0.74	0.65					
	2	4	21	4.41	7.85	3.55 显著	0.68	0.62	0.80, 1.08	76.2	0.64	14.3	0.0
5	5	5	51	0.00	5.28	2.43 显著	0.61	0.61	0.60, 0.80	68.6	0.44	22.9	6.3
	4	6	51	0.74	6.70	2.57 显著	0.61	0.61	0.64, 0.81	68.6	0.43	22.9	12.5
4	4	7	56	0.00	6.49	2.56 显著	0.58	0.60	0.62, 0.83	71.4	0.47	16.2	15.8
	2	8	56	4.02	13.43	3.17 显著	0.58	0.59	0.61, 0.92	73.2	0.49	18.9	21.1

因而 y 的回归值在判别指标临界值附近, 就易出现级的错误, 不过有经验的预报员, 是可能避免这种差错的。

四、检 验

方程(1)一(8)是用1979—1984年江苏一些雷达站历史资料得到的, 表5得到的结果只是历史回报。现在利用同一省、不同地区1985年新资料和上海、宁波两市1979—1984年资料对该方程适用范围进行检验。为此我们选择表5中由逐步回归得到的方程(2)、(4)、(6)、(8)进行分析, 检验结果一起列入表(6)。

1. 上海、宁波两地资料检验

取1981年5月10日宁波雷达站资料 $x_2=60$, $x_4=-13$, $x_5=6$, $x_8=4.4$, 将这些值代入方程(6), 计算得 $y_6=0.94$, 由降雹判别指标 y'' 大于0.81(见表5)就有强雹出现, 这与实况 $Y=2$ 有强雹一致。类似有这样资料的样本数13个, 把它们一一代入方程(6), 可计算得预报和实况如表7。由表7可计算得拟合率=46.2%, 空报率=50%, 漏报率=14%, $E=3.5$, $S=0.27$ 。可见进行三级判别不理想, 若改用冰雹出现与否两级预报, 则拟合率可提高到70%。

2. 江苏省扬州、南通两地1985年资料检验

取1985年5月4日扬州雷达站资料 $x_2=38$, $x_4=-5.1$, $x_5=5.8$, $x_8=4.6$ 代入方程(6)可计算得预报值 $y_6=0.023$, 由该方程的降雹判别指标(见表5)可知此预报值与实

表6 逐步回归方程检验结果

方程序号	检验地区	样本数	拟合率 %	S	空报率 %	漏报率 %
(2)	上宁 海波	24	62.5	0.44	25.0	6.3
(4)	上宁 海波	25	36.0	0.11	50.0	17.7
	扬南 州通	15	93.3	0.86	0.0	0.0
(6)	上宁 海波	13	46.2	0.27	50.0	14.3
(8)	扬南 州通	15	86.7	0.67	0.0	40.0
	上宁 海波	28	50.0	0.24	20.0	38.9

测值 $y_6=0$ (无雹)相符。类似有这样资料的样本数有15个,其预报、实况如表8,可发现15次天气过程中有5次出现降雹,全部报对,没有空、漏报现象。但有一次弱雹报为强雹,因此拟合率=93.3%, $S=0.86$ 。同样也对方程(8)作了检验,其结果统一列于表6中。可见利用方程(6)判别江苏各地冰雹出现与否效果很好。而利

用方程(8),判别冰雹是否出现,也有一定效果,但漏报现象较严重。因为建立该两个方程的资料主要由南京、盐城、徐州资料回归而得,而用来检验方程的资料却是扬州和南通的,因此方程(6)可在江苏省各雷达站业务中试验使用。

表7

		实 况			
		强	弱	无	合计
预报(回报)	强	2	3	3	8
	弱	0	1	0	1
	无	0	1	3	4
	合计	2	5	6	13

表8

		实 况			
		强	弱	无	合计
预 报	强	3	1	0	4
	弱	0	1	0	1
	无	0	0	10	10
	合计	3	2	10	15

五、结论与存在的问题

1. 在九个判别因子中,除个别因子以外,大部分因子与判别对象之间单项相关较差,利用多元回归特别是逐步回归以后提高了判别冰雹准确率,这正是我们所期望的。

2. 据表5可知逐步回归方程优于多元回归方程。选取最佳的因子组合,建立判别方程,使计算方便,而又使显著性提高。

3. 判别因子选取很重要,在选取物理意义明确、相关好且稳定、各因子能取长补短的同时,还要注意样本容量大,因子数目要适当。为了提高回归效果和显著性,还须对强对流天气形成机制及其雷达回波参数作更深入的研究。

4. 用方程(6)判别江苏各地冰雹出现与否效果较好,可以在业务中试验试用。但判

别上海、宁波两地冰雹, 除方程(2)以外, 三级预报效果均不理想, 但进行两级预报仍有一定效果。这可能与各地雷达性能、取样标准、下垫面性质等因素有关。

5. 建立和检验回归方程的样本数均不够多, 这就会降低方程的代表性。当样本数增加时, 效果如何还要在试验应用中去检验。

6. 多元回归效果不佳, 这与样本数太少而判别因子太多有关。一般认为判别因子数目最好不超过样本容量的1/10。当然判别因子太少也会导致信息量不足, 使预报能力降低。

在完成本文过程中, 顾炳刚、邵玲玲、朱家栋、钱来林、朱竟成、唐询昌同志提供了部分资料, 丁裕国、魏文遂两同志曾对本文提出过宝贵意见, 特此致谢。

参 考 文 献

- [1] Eccles, P.J. and Atlas, D., A new method of hail detection by dual-wavelength radar, Proc. 14th Radar Meteor. conf., 1—6, 1970.
- [2] Barge, B.L., Radar signal depolarization by hydrometeors in Alberta hailstorms, Proc. 15th Radar Meteor. conf., 1—6, 1972.
- [3] 葛润生, 雷达测量雹云回波顶高的误差讨论, 雷达气象文集, 99—105, 气象出版社, 1981.
- [4] 王宗皓、李麦村等, 天气预报中的概率统计方法, 科学出版社, 1974.

A STATISTIC APPROACH TO HAIL CLOUD DISCRIMINATION WITH WEATHER RADAR DATA

Dai Tiepi Wang Zhenhui Zhao Lingling

ABSTRACT

Multiple and stepwise equations are established with weather radar data in Jiangsu province. Comparison is made between the equations used for the accuracy of distinguishing hail clouds from others. Test of these equations with newly-collected data (1985) at other locations in Jiangsu and data (1979-1984) in cities of its adjacent provinces shows that these equations can be used at radar stations in Jiangsu with rather good results. However, care should be taken when they are used for hail cloud discrimination in the adjacent provinces.