试解林西“雹日增多之谜”

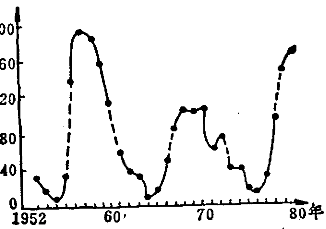
杨 得 宇

（内蒙古昭盟气象科学研究所）

近年来，我国不少地区发现在前阶段人工防雹过程中有雹日增多的现象，因而有“催云可成雹”的说法，认为这是防雹的结果。但又因统计检验不显著而得不到结论。本人曾对林西高炮防雹效果作了统计检验，并提出了“雹日增多之谜”(见《气象》1982年第7期)。那么“雹日增多之谜”应如何解释呢？本文在研究了林西雹日按太阳活动峰谷期的分配之后，作了探讨性的回答。

一、太阳活动的峰、谷期及其划分

据林西县气象站29年的资料，林西县冰雹最早从4月下旬开始，最迟至10月下旬结束。据此绘制自1952至1980年逐年4至10月年平均太阳黑子曲线（如附图)。



附 图

峰期：在一个太阳活动周期中，表现为太阳黑子相对数多的时期，在黑子曲线上表现为峰；在最近两个太阳周期中，由太阳黑子峰值年的前一年和后3年（11年周期）至4年（12年周期）组成（见附图)。

谷期：在一个太阳活动周期中，表现为太阳黑子相对数少的一个时期，在黑子曲线上表现为谷；在最近两个太阳周期中，由谷值年的前3年和后2年计6年组成（见附图)。

按上述原则确定，1952—55年为一个不完整谷期，1961—66年、1973—78年为太阳活动谷期；同样，1956—60年、1967—72年是太阳活动峰期，1979、1980年属于太阳活动峰期。其间包括两个完整的太阳周期和不完整的一个峰期一个谷期。下面找出林西站29年年雹日在太阳活动的峰、谷期分配的统计特征，看能否解释“雹日增多之谜”。

二、林西29年雹日在峰、谷期中的分配

林西气象站29年观测资料给出的逐年4至10月(年)雹日如表1。

1.林西29年年雹日服从正态分布 N(3.5,2.2)

用柯尔莫哥洛夫检验法：

$$D_n = \sup_{-\infty<x<\infty} | F_n^*(x) - F(x) | \approx \max_i | F_n^*(x_i) - F(x_i) |$$

此处 $F(x)$ 是随机变数 x 的理论分布，$F_n^*(x)$ 为 x 的 n 次观测结果小于 x_i 的个数与 n 之比即经验分布，D_n 表示两者的偏离。否定域对给定的 α 有 $\sqrt{n}D_n \geqslant \lambda$。对本例 $F(x)=\phi(u)=\phi(\frac{x_i-\bar{x}}{s})$，$\bar{x}=3.5$，$S=2.2$。列计算表如表2。

表 2

x_i	n_i	$F_n^*(x_i)$	$F(x_i)$	$\lvert F_n^*(x_i)-F(x_i)\rvert$
0.1	2	0.0689	0.0606	0.0083
1.1	6	0.2069	0.1357	0.0712
2.1	10	0.3448	0.2611	0.0837
3.1	17	005862	0.4286	0.1576
4.1	19	0.6552	0.6064	0.0488
5.1	22	0.7586	0.7673	0.0087
6.1	26	0.8966	0.8810	0.0156
7.1	28	0.9655	0.9495	0.0160
8.1	29 = n	1.0000	0.9817	0.0183
max(i)				0.1576

$$D_n \approx \max_i | F_n^*(x_i) - F(x_i) | = 0.1576$$

$$\sqrt{n}D_n = \sqrt{29} \times 0.1576 = 0.8487$$

给定 $\alpha = 1 - Q(\lambda) = 0.05$，查得 $\lambda = 1.36$，$\sqrt{n}D_n < \lambda$，没有理由认为29年年雹日不服从正态分布N(3.5,2.2)。

2.用斯米尔诺夫检验法，验证两个峰期、两个谷期雹日来自同一分布而无显著差异。

如果从两总体中分别抽取大小为 n_1、n_2 的样本，

表 1　林西 1952 至 1980 年年雹日

项＼年	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
年雹日	0	5	4	7	3	5	3	0	3	6	6	4	2	3	1	2	1	6	5	6	1	3	3	7	2	3	8	2	1
峰、谷期	谷期				峰期					谷期						峰期						谷期						峰期	

其对应的经验分布为 $F_{n_1}^*(x)$、$F_{n_2}^*(x)$，我们规定：

$$D_{n_1,n_2}=\underset{-\infty<x<\infty}{SUP}|F_{n_1}^*(x)-F_{n_2}^*(x)|$$

如果两总体分布相同，D_{n_1,n_2} 应比较小，所以，检验两总体分布相同的假设之否定域可写成 $|KD_{n_1,n_2}|\geqslant m(n_1,n_2,\alpha)$。K 是 n_1、n_2 最小公倍数，例如 $n_1=5$，$n_2=6$，$K=30$，它使 $KD_{n_1,n_2}\geqslant m(n_1,n_2,\alpha)$ 只取非负整数。α 为显著水平，$m(n_1,n_2,\alpha)$ 为满足 $P(KD_{n_1,n_2}\geqslant r)\leqslant\alpha$ 中的 r 的最小整数。这里的关键是计算 KD_{n_1,n_2}。

$$KD_{n_1,n_2}=K\max_i\max\{|F_{n_1}(x_i+0)-F_{n_2}(x_i)|,\ |F_{n_1}(x+0)-F_{n_2}(x_i+0)|,|F_{n_1}(x_i)-F_{n_2}(x_i)|,\ |F_{n_1}(x_i)-F_{n_2}(x_i+0)|\}=\max_i L_i。$$

通过峰期的实例说明求 L_i 的方法，见表 3。x_m 表示峰期年雹日样本，x_{mi} 表示峰期序号，$F_n(x+0)$ 表示小于或等于 x 数据个数与 n 之比，$F_n(x)$ 表示小于 x 实验结果数据个数与 n 之比，x_{m_1} 表示把第一峰期（由后向前编）雹日216561按大小顺序编号，1_2 表示年雹日为 1 的出现 2 次。

表 3

$x_{m_1}(n=6)$ / $x_{m_2}(n=5)$	0	1_2	2	3_3	5 / 5	6_2
$x_{m_1}(n=6)$		1_2	2		5	6_2
$x_{m_2}(n=5)$	0			3_3	5	
$KF_{n_1}(x_i+0)$	0	10	15	15	20	30
$KF_{n_1}(x_i)$	0	0	10	15	15	20
$KF_{n_2}(x_i+0)$	6	6	6	24	30	30
$KF_{n_2}(x_i)$	0	6	6	6	24	30
L_i	6	6	9	9	15	10

$KD_{n_1,n_2}=\max\limits_i L_i=15$，$n_1=6$，$n_2=5$，给定 $\alpha=0.01$，查斯米尔诺夫表得 $m(5,6,0.01)=30$，$KD_{n_1,n_2}=15<30=m(6,5,0.01)$，没有理由认为这两组数据有显著差异，而属于同分布。

注意到 $\bar{x}_{m_2}=2.8$，$S_{m_2}=1.8$，而 $\bar{x}_{m_1}=3.5$，$S_{m_1}=2.4$ 的情况，认为可能 $\bar{x}_{m_2}$，$\bar{x}_{m_1}$ 相差很大而不是同一分布了，不妨用正态u检验来检验一下，结果同斯米尔诺夫检验结果是相同的，即没显著差异而属同分布。

同样，对谷期的情况（见表 4）有：

表 4

x_{n_1}	1	2	3	4	5		
x_{n_2}		2	3_3			7	8
$KF_{n_1}(x_i+0)$	1	2	3	4	6	6	6
$KF_{n_1}(x_i)$	0	1	2	3	4	6	6
$KF_{n_2}(x_i+0)$	0	1	4	4	4	5	6
$KF_{n_2}(x_i)$	0	0	1	4	4	4	5
L_i	1	2	2	1	2	1	1

$KD_{n_1,n_2}\approx\max\limits_i L_i=2$，$m(6,6,0.01)=6$，$KD_{n_1,n_2}<m(6,6,0.01)$，即认为在不同太阳周期中谷期同谷期有相同的分布而无显著差异。

3.峰、谷期全雹日是正态分布

对谷期的情况：谷期与谷期的雹日分布无显著差异而来自同一分布；每一谷期均为29年雹日正态分布的一个子样，仍为正态分布；用柯尔莫哥洛夫检验法检验时有：谷期雹日0547664231337238，$n=16$，$\bar{x}_n=4.0$；$S_n=2.3$，列计算表如表 5。

表 5

x_i	n_i	$F_n^*(x_i)$	$F(x_i)$	$\lvert F_n^*(x_i)-F(x_i)\rvert$
1.1	2	0.1250	0.1038	0.0212
2.1	4	0.2500	0.2033	0.0467
3.1	8	0.5000	0.3483	0.1517
4.1	10	0.6250	0.5200	0.1050
5.1	11	0.6875	0.6844	0.0031
6.1	13	0.8125	0.8186	0.0061
7.1	15	0.9375	0.9115	0.0260
8.1	16	1.0000	0.9700	0.0300

$1-Q(\lambda)=0.05$，查得 $\lambda=1.36$，$\sqrt{n}D_n=\sqrt{16}\times0.1517=0.6068<\lambda$。所以，认为谷期雹日是正态分布 $N(4.0,2.3)$，理由是充分的。

同样，对峰期的情况有：峰期与峰期属同分布，来源于 29 年正态分布的子样；用柯尔莫哥洛夫检验时，有在 $1-Q(\lambda)=0.05$ 置信水平上 $\sqrt{n}D_n=\sqrt{13}\times0.1525=0.5498<\lambda$。$n=13$（年），峰期 13年雹日服从正态分布 $N(2.9,2.0)$，理由是充分的。

4.比较峰、谷期两正态分布总体的标准差 σ

设峰、谷期两个正态分布总体为 $N_1(\mu_1,\sigma_1)$ 和 $N_2(\mu_2,\sigma_2)$，我们分别独立地抽取样本 $x_1,x_2\cdots\cdots x_m$ 和 $y_1,y_2,\cdots\cdots y_n$，计算出样本的方差：

$$S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2\quad S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

这时 $(m-1)S_1^2$，$(n-1)S_2^2$ 分别遵从 $m-1$ 和 $n-1$ 个自由度的 x^2 分布，当 $\sigma_1=\sigma_2$ 时，统计量 $F=\frac{S_1^2}{S_2^2}$，遵从自由度 $f_1=m-1$、$f_2=n-1$ 的 F 分布。对本例有：

$S_1^2=2.0^2$，$m=13$，为峰期；$S_2^2=2.3^2$，$n=16$，为谷期。$F=\frac{S_1^2}{S_2^2}=\frac{2.0^2}{2.3^2}=0.76$。

检验假设 $H_0:\sigma_1=\sigma_2$，自由度 $f_1=m-1=12$，$f_2=n-1=16-1=15$。给定显著性水平 $\alpha=0.10$（比 $\alpha=0.01$ 更严格），查F 表得 $F_{\frac{\alpha}{2}}=F_{0.05}=2.48$，$F<F_{\frac{\alpha}{2}}$ 肯定原假设：认为峰谷期雹日正态分布中具有相同的标准差 σ。

5.对峰、谷期雹日比较均值作 t 检验

如果两个正态分布 $N_1(\mu_1,\sigma_1)$，$N_2(\mu_2,\sigma_2)$ 的标准差 $\sigma_1=\sigma_2$，比较 μ_1,μ_2，我们分别从 N_1、N_2 两正态分布中作 n_1、n_2 次观测，令 $\bar{x},\bar{y}$ 表示样本均值，S_1^2,S_2^2 分

别表示样本方差：

$$S_1^2=\frac{1}{m-1}\sum_{i=1}^{m}(x_i-\bar{x})^2,\ S_2^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

我们建立统计量

$$t=\frac{\bar{x}-\bar{y}}{\sqrt{(m-1)S_1^2+(n-1)S_2^2}}\sqrt{\frac{mn(m+n-2)}{m+n}}$$

于是 t 服从 $m+n-2$ 个自由度的t 分布，对于显著性水平 α，检验假设 $H_0:\mu_1=\mu_2$，否定域$|t|>t_\alpha$；检验假设 $H_0:\mu_1\geq\mu_2$，否定域为 $t<-t_{2\alpha}$；检验假设$H_0:\mu_1\leq\mu_2$，否定域为$t>t_{2\alpha}$。

对我们讨论的问题，检验假设$H_0:\mu_m\geq\mu_n$。μ_m为峰期雹日均值，μ_n为谷期雹日均值。

$\bar{x}=\frac{1}{12}\sum_{i=1}^{12}(x_i)$为峰期样本均值，为2.9；

$\bar{y}=\frac{1}{15}\sum_{i=1}^{15}(y_i)$为谷期样本均值，为4.0；

$S_1^2=2.0^2$ 为峰期雹日样本方差，$m=13$；

$S_2^2=2.3^2$ 为谷期雹日样本方差，$n=16$。

计算 $t=\frac{2.9-4.0}{\sqrt{12\times2^2+15\times2.3^2}}\sqrt{\frac{13\times16\times(13+16-2)}{13+16}}$

$=-1.357$，给定 $\alpha=0.10$，查 t 表得 $-t_{2\alpha}=-1.314$，$t<-t_{2\alpha}$，否定原假设 $H_0:\mu_m\geq\mu_n$ 而有 $\mu_m<\mu_n$，即有90%的把握说，在近几个太阳周期中，谷期雹日均值显著大于峰期的雹日均值，是其 1.4 倍。

三、结　论

1.多雹地区林西历史年雹日服从正态分布（3.5, 2.2）。它是由σ相同而均值存在显著差异的两个正态分布组成，一是峰期雹日的正态分布$N_m(2.9,\sigma=2.15)$，二是谷期雹日的正态分布$N_n(4.0,\sigma=2.15)$。其中谷期雹日均值比峰期雹日均值，在置信水平 $\alpha=0.10$上有显著地增多，谷期是峰期的 1.4 倍左右。这一结论至少在现有的29年历史资料中是正确的。

2.年雹日在不同太阳周期的峰期与谷期中具有相同的分布，且基本服从同一正态分布，至少在已经出现的29年历史资料中是这样。

3.十分明显，对前段人工防雹中发现雹日增多的现象，从时段上讲正好是1973—78年左右落入太阳活动的谷期，雹日正是增多的时段。不仅1973—78年防雹阶段雹日有所增多，在同是谷期的1961—66年也同样甚至增得更多，而这一时段并未防雹。这一点至少在林西是这样，在昭盟也是这样。在目前已知的国内外提出防雹后雹日增多的地方，他们防雹的时段也都在太阳活动谷期，如包头麻池、瑞士等。

四、结束语和问题讨论

1.雹日增多，对固定地点林西，我们已经发现不仅 1973—78 年防雹时段如此，而且对应于前一谷期1961—66年雹日，无论同前一个峰期还是同历史均值相比同样有所增多，这就在时间展延上说明了“雹日增多”不是防雹的结果，而是一种年雹日和太阳活动有联系的自然分布结果，即太阳活动谷期多，峰期少。

同样，我们利用农业气象手册中随机给出的昭盟20个站1965—76年雹日资料（见表6），假设1972年前雹日遵从正态分布N（3.2, 4.1），检验 1973—76年是否发生变异。

表 6　昭盟 1965—76 年年雹日数(20个站合计)

年　份	65	66	67	68	69	70	71	72	73	74	75	76
雹　日	34	28	29	26	31	31	38	36	46	28	49	49

1965—72年　$\bar{x}=\mu_0=32$，$\sigma=S=4$；1973—76 年 $\mu=43$；$H_0:\mu\leq\mu_0$　$U=\frac{\bar{x}-\mu_0}{\sigma}\sqrt{4}=\frac{43-32}{4.1}\sqrt{4}=5.3659$，给定$\alpha=0.00000005$，$U_{2\alpha}=5.32672$，$U>U_{2\alpha}$，否定原假设 $\mu\leq\mu_0$ 而有 $\mu>\mu_0$。即在相当高置信水平上认为，远离作业区很少受到影响的气象站的观测资料表明：不是因为防雹而是因为1973—76年处于太阳活动谷期，使此时全盟雹日数极为明显地比以太阳活动峰期为主的1965—72年要高出近 1.4 倍，这与林西的情况是一样的。从这一固定时间，展延空间再次看到，至少在昭盟地区于最近一个太阳周期中，谷期雹日比峰期雹日有明显地增多。这就解开了“雹日增多之谜”，并说明“催云可成雹”是没有根据的，此种防雹作业对雹日似无影响。

2.即使雹日同太阳活动有本质的联系，由于地理纬度、海拔高度、地形地貌的影响，也并非所有地方都表现为雹日谷期多峰期少。就是在对太阳活动反应敏感的北方，也存在有与此相反的情况即峰期多谷期少，如长春雹日就是如此。此外，对同一地点，当观测资料年代足够长，也存在规律从某一周期开始反转的情况。林西29年资料对雹日在峰、谷期分布的研究是十分短的，但对解开“雹日增多之谜”及说明“催云可成雹”没有根据，看来是够了。

3.这一研究不仅在防雹时间系列对比检验和设计中有十分重要的意义，在长期统计预报和寻找预报不准的原因以及改进预报方法上也均有重要意义。特别是将要素进行分层统计之后，规律将更为明显。