

文章编号：1006-4354 (2007) 04-0003-04

基于多因子非线性叠套模型的雷电潜势预报

高菊霞¹, 庞亚峰², 武麦凤³(1. 陕西省防雷中心, 西安 710016; 2. 西安市气象局, 西安 710016;
3. 渭南市气象局, 陕西渭南 714000)

摘要：利用统计学原理和数学分析理论, 建立以雷电为预报对象的多因子非线性叠套模型, 试图从理论和实践两方面, 避免统计学的线性假设和模式输出值物理意义不明确的统计预报问题。特别是针对雷电这种小概率事件的预报问题, 通过统计分析建模, 并对 24 h 雷电潜势预报结果检验, 该模型预报效果较好, 能够业务使用。

关键词：雷电; 模型; 潜势预报**中图分类号：**P457.9**文献标识码：**A

多因子预报常用的统计模型有回归、判别、聚类, 最常用的是回归和判别。这些模型在理论上比较成熟, 但实践中存在两个问题: (1) 线性假定, 即假定预报对象与各因子均是线性相关, 假定与实际差距较大。(2) 模型给出的是各因子综合值, 只是一个相对数, 有时物理意义不明确。多因子非线性叠套模型试图从理论和实践上对以上问题予以改进。模型无需线性假定, 而是假定各因子与预报对象均是非线性相关, 线性关系只是这种关系的一种特例; 模型所给出的各因子综合值是当日预报对象将要发生的概率。

1 资料

从 1967—1992 年 5—9 月随机抽取 15 个月, 共计 459 个雷暴日, 利用 500 hPa、700 hPa 和 850

hPa 逐日历史天气图和全省各站逐日雷暴天气现象资料计算 $30^{\circ}\sim 40^{\circ}\text{N}$, $104^{\circ}\sim 117^{\circ}\text{E}$ 的 850 hPa 的 θ_{e} 场, 并标明 $30^{\circ}\sim 40^{\circ}\text{N}$, $105^{\circ}\sim 110^{\circ}\text{E}$ 范围内各站反映层结、能量、湿度及上升运动的物理量场、风场, 选取多个预报因子, 用于建立雷电潜势预报模型。

2 模型建立的原理

如何利用因子值算出预报对象发生的概率^[1]? 若 1 个因子, 根据统计方法可以得到; 2 个因子, 根据 2 个因子的因子值, 做雷电有无的二元点聚图, 在图上分析等概率线; 2 个或 3 个以上, 利用点聚图叠套的概率计算公式。设雷天气为事件 A, 其发生概率(先验概率)为 $P(A)$, 不发生概率为 $P(\bar{A})$, 显然, $P(A) + P(\bar{A}) =$

收稿日期：2007-03-08

作者简介：高菊霞 (1972-), 女, 陕西眉县人, 学士, 工程师, 从事雷电天气预报及相关研究。

年开始在汛期作为常规业务每天制作发布, 2005 年向省防汛办发送面雨量 60 余次, 在电视上播报 20 余次, 特别是在 2005 年国庆期间渭河流域发生的洪涝灾害中起到了重要作用; 2006 年向省防汛办发送 10 余次流域面雨量预报, 取得一定的服务效果, 尤其对陕西防汛决策部门安排防汛工作提供了科学依据。随着中尺度模式的改进和完善, 面雨量的预报效果和服务水平会进一步提高。

参考文献:

- [1] 张志秀, 王承伟. 利用 MM5V3 中尺度预报产品对松花江流域面雨量的客观预报 [J]. 黑龙江气象, 2005, 3: 16-17.
- [2] 赵榆飞, 杜继稳. 基于 MICAPS 平台的四维天气诊断分析系统[G]//杜继稳. 青藏高原东北侧突发性暴雨分析研究与应用. 北京: 气象出版社, 2005, 251-255.

1。 n 个预报因子或 n 时预报雷电的点聚图, 以 B_1 , $B_2 \dots B_n$ 表示因子落入相应点聚图划定区域的条件, 在上述条件下, 求天气 A 发生的概率(后验概率)。

根据概率乘法定理: $P(A \cdot B) = P(A|B) \cdot P(B)$ 有 $P(A|B) = P(A \cdot B) / P(B)$ 。式中 $P(A \cdot B)$ 表示天气 A 与因子落入划定区域 B 时发生的概率, $P(B)$ 为因子落入划定区 B 的概率。 $P(A|B)$ 表示在 B 条件下出现天气 A 的概率。

因为 $P(A \cdot B) + P(\bar{A} \cdot B) = P(B)$ 代入上式则

$$P(A|B) = 1 / \{1 + [1/P(A) - 1] P(B|\bar{A}) / P(B|A)\},$$

若使用 n 张点聚图或 n 个因子, 同理可导出

$$P(A|B_1 B_2 \dots B_n) = \frac{1}{1 + \left[\frac{1}{P(A)} - 1 \right] \frac{P(B_1 B_2 \dots B_n | \bar{A})}{P(B_1 B_2 \dots B_n | A)}},$$

式中 $P(A|B_1 B_2 \dots B_n)$ 就是第 1 个因子落入 B_1 区, 第 2 个因子落入 B_2 区……第 n 个因子落入 B_n 区时, 天气 A 发生的概率。如果几张点聚图或因子相互独立, 则有

$$P(A|B_1 B_2 \dots B_n) = \frac{1}{1 + \left[\frac{1}{P(A)} - 1 \right] \frac{P(B_1 | \bar{A})}{P(B_1 | A)} \cdot \frac{P(B_2 | \bar{A})}{P(B_2 | A)} \cdots \frac{P(B_n | \bar{A})}{P(B_n | A)}},$$

令 $C_i = P(B_i | \bar{A}) / P(B_i | A)$, 则有

$$P(A|B_1 B_2 \dots B_n) = 1 / [1 + (1/P(A) - 1) \prod C_i], \quad i=1, 2, \dots, n \quad (1)$$

式中 $P(A)$ 为预报对象的先验概率, 即 A 的气候概率, 可简记为 P_0 。 $P(A|B_1 B_2 \dots B_n)$ 是 n 个因子的特定组合(因子 x_1 出现在 B_1 区, x_2 出现在 B_2 区…… x_n 出现在 B_n 区)下 A 的后验概率, 即条件概率, 也就是预报日当天 A 将要发生的概率, 简记为 P , 这样上式可写为

$$P = 1 / [1 + (1/P_0 - 1) \prod C_i]. \quad i=1, 2, \dots, n \quad (2)$$

3 雷电潜势预报模型建立

3.1 模型建立的方法

根据公式(2)可以看出, 雷电潜势预报模型

的建立实际上是根据不同的预报落区和时段建立各因子的 C 函数, 任取一因子 x 做关于有无雷电的一元点聚图, 然后根据点的分布分成 m 个区域。计算每个区域的 C 值: $C = P(B_i | \bar{A}) / P(B_i | A)$, $P(B_i | \bar{A})$ 为第 i 区对无雷电(\bar{A})的概括率, $P(B_i | A)$ 为第 i 区对有雷电(A)的概括率。这样 m 个区就有 m 个值。利用非线性回归把 m 个离散的值变成一个连续的 C 函数: $C = f(x)$ 。若有 n 个因子就可以找到 n 个 C 函数: $C_i = f(x_i)$ 。在各因子相互独立的假设下, 某地某日发生雷电的概率为公式(2)中的 P , 式中 P_0 为该地该季雷电的气候概率。 C 函数的性质:(1) C 的值域 $(0, +\infty)$, 无量纲。(2) 当 $C=1$ 时, 它代表因子对预报对象的发生与否不表态, 对应的因子值称为该因子的“中性值”, 中性值类似于因子的临界值。

3.2 潜势预报模型实例

以 7—8 月陕南的雷电潜势预报为例。陕南雷电预报因子: ①汉中和安康的湿位涡 MPV 。②汉中和安康的梯度和 $\sum T$ 。③汉中和安康预报区上游的最大涡度 G 。以第③个因子涡度 G 为例, 说明建立 C 函数的步骤和方法:

第一步, 用历史数据, 以涡度 G 为因子做陕南有无雷电的一元点聚图。

第二步, 根据点子的分布情况分区。此例分成 5 个区域, 分区统计数据如表 1。

表 1 涡度因子在点聚图中的分区统计表

G	$G > 6$	$4 \sim 6$	$2 \sim 4$	$0 \sim 2$	$G < 0$	合计
无雷电	5	20	51	41	17	134
雷电	25	51	84	24	4	192
C	0.2866	0.5619	0.8699	2.098	6.09	/

根据分区确定因子 G 每组的中值, 边界组中值的确定参照临近组的组距, 并令 $X = G + 5$ (表 2)。据表 2 计算出陕南 7—8 月雷电发生的气候概率: $P_0 = 192 / (192 + 134) = 0.589$ 。

表 2 因子的分区参数值

G 的中值	7	5	3	1	-1
$X = G + 5$	12	10	8	6	4

第三步, 计算各区的 C 值, $C=P(B_i|\bar{A})/P(B_i|A)$, $P(B_i|\bar{A})$ 为第 i 区对 \bar{A} 无雷电的概率, $P(B_i|A)$ 为第 i 区对有雷电的概率。如: 第一区 ($G > 6$), $P(B_1|\bar{A})=5/134$, $P(B_1|A)=25/192$, $C_1=0.2866$ 。第二区 (G 值 $4 \sim 6$), $P(B_2|\bar{A})=20/134$, $P(B_2|A)=51/192$, $C_2=0.5619$ 。余类推, 5 个区有 5 个 C 值。

第四步, 用非线性回归建立 C 与因子 G 的函数关系。

中国科学院数学研究所《回归分析》中讲到化非线性函数为线性函数的回归问题时, 给出 7 种曲线的线性变换公式, 即先把非线性函数变为线性函数, 然后进行线性回归:

① $y=\alpha+b\log x$, 若令 $x'=\log x$, 则 $y=\alpha+bx'$ 。

② $y=de^{bx}$, 若令 $y'=\ln y$, $\alpha=\ln d$, 则 $y'=a+bx$ 。

③ $y=de^{b/x}$, 若令 $y'=\ln y$, $x'=1/x$, $\alpha=\ln d$, 则 $y'=\alpha+bx'$ 。

④ $y=dx^b$, 若令 $y'=\log y$, $x'=\log x$, $\alpha=\log d$, 则 $y'=\alpha+bx'$ 。

⑤ $y=1/(\alpha+be^{-x})$, 若令 $y'=1/y$, $x'=e^{-x}$, 则 $y'=\alpha+bx'$ 。

⑥ $y=\alpha+bx$ 。

⑦ $y=x/(\alpha x+b)$, 若令 $y'=1/y$, $x'=1/x$, 则有 $y'=\alpha+bx'$ 。

令 $x=G+5$, (线性变换时, $x \geq 0$), $y=C$, 将两组数据非线性回归, 算出 7 种曲线的线性变换后的回归结果 (表 3)。

表 3 7 种曲线的线性变换回归结果

曲线	参数 b	参数 a	相关系数 r
①	-11.887	12.35	0.933 5
②	-0.371 5	3.008 5	0.987 7
③	17.579 2	-2.432 6	0.983 8
④	-2.736 9	2.438 5	0.998 1
⑤	-100.08	1.836 0	0.603 0
⑥	-0.657 1	7.238 4	0.866 3
⑦	-16.401 7	3.790 1	0.828 2

选择相关系数 (r) 较大者试算, 看外延会不会出现不合理的结果, 如 $C \leq 0$; 中值与经验是否

一致, 因中值相当于因子的临界值。它是否合适对预报结果影响较大。计算结果选择第④种曲线: $y=dx^b$, 已算得 $b=-2.736 9$, 而 $\alpha=\log d$, 所以 $d=10^\alpha=10^{2.4385}=274.47$, 又因 $x=G+5$, $y=C$, 有 $C=274.47(G+5)^{-2.7369}$ 。

同理可以得到因子湿位涡 (MPV) 的 C 函数。若 $MPV > 0$ $C=1.741+E^{1.125}$, $E=MPV$; 若 $MPV \leq 0$ $C=1.8422e^{[-0.07774(-E+1)]}$, $E=MPV$ 。

因子梯度和 ($\sum T$) 的 C 函数: $C=2.712 8e^{(-0.07666 \times T)}$, $T=\sum T$ 。

因子编序为: $x_1=MPV$, $x_2=\sum T$, $x_3=G$, 对应的 C 函数也顺次为: C_1 、 C_2 、 C_3 。

利用此模型做雷电潜势预报, 如某日 08 时, 汉中 $MPV=-11.0$, $\sum T=22.3$, $G=2.5$ 。则 $C_1=1.8422e^{[-0.07774(11+1)]}=0.7248$, $C_2=2.7128e^{(-0.07666 \times 22.3)}=0.4910$, $C_3=274.47(2.5+5)^{-2.7369}=1.10545$, $P_0=0.589$

将 C_1 、 C_2 、 C_3 、 P_0 值代入公式 (2) 计算, 当日下午到晚上汉中出现雷暴的概率 $P=0.785$, 实况是当日下午汉中地区出现强雷电。

4 试报

试验所用的预报因子取自 2006 年 5—8 月的常规天气观测资料。主要因子有 850 hPa θ_{se} 、湿位涡 MPV , 梯度和 $\sum T$, 500 hPa、700 hPa 和 850 hPa 的涡度 G , 500 hPa、700 hPa 和 850 hPa 的散度 D 等。雷电实况取闪电定位监测结果。雷电潜势预报分为陕北北部、陕北南部、关中东部、关中西部、陕南东部和陕南西部共 6 片。

评分采用概率天气预报中最常用的 Brier 评分法^[2], 评分公式:

$$B=1/N \sum (P_i - O_i)^2, \quad (i=1, 2, \dots, N)$$

N 为预报次数, P_i 为预报概率, O_i 为实况 (天气事件出现时取 1, 不出现时取 0)。 B 得分在 0 和 1 间, 预报完全正确得分为 0, 完全不正确为 1。评分结果见表 4, 全省 Brier 评分在 0.140 至 0.653 之

自动气象观测站 2 个问题的处理方法

1 21时自动站正点资料不能自动上传的处理

现象 在地面气象测报业务系统软件升级为 V3.0.9 和 V3.0.10 后,发现网络显示正常,主辅通道指示灯为绿色,21时自动站正点资料却不能自动上传,可手工上传,或重新启动电脑后,只要传输时间在本站参数设置范围内,能自动上传,但 22 时及其后时次有时能自动上传,有时又不能自动上传。

处理方法 在 20:30 日数据文件上传后,重新启动电脑,21 时自动站正点资料可按时自动上传,且 22 时及以后时次也都正常上传。若日数据文件在主通道上传了,可以在 20:30 前重新启动电脑,21 时及其后时次自动站正点资料也能按时自动上传;若日数据文件没有上传前重新启动电脑,21 时自动站正点资料还是不能自动上传。西乡站按上级文件规定设置主通道的启动时间为正点后 03 分,最大延时为 5 分,辅通道启动时间为正点后 30 分(文件规定为 31 分,软件中最大为 30 分),最大延时为 5 分,所以大部分时间,日数据文件都是在辅通道上传的,故值班员每天在

20:30 后,且日数据文件上传后(30 min)重新启动电脑,21 时及以后时次自动站正点资料都能够按时自动上传。

2 非汛期 A 文件中降水要素方式位的处理

现象 陕西省雨量传感器的启用时间为每年的 4 月 1 日—10 月 31 日或 5 月 1 日—10 月 31 日,其余时段停用雨量传感器。停用期间,在业务软件中,需将台站参数的自记降水选为“无”。在 A 文件维护时,由于台站参数的自记降水在此状态下的默认状态为“有,人工”,这时,如将 A 文件维护中台站参数的自记降水选为无,存盘后,重新打开 A 文件维护,台站参数的自记降水仍为“有,人工”。造成一些台站上报的 A 文件降水的方式位有时正确,有时不正确。

处理方法 在 A 文件维护和预审完毕上传 A 文件前,将 A 文件维护中的台站参数的自记降水重新选为“无”,存盘,然后不要打开 A 文件维护,即刻上传。

(李爱华)

间,评分结果良好,7—8 月的预报结果比 5—6 月好,在 0.140~0.430 之间。雷电过程漏报次数较少,空报次数相对较多一些。试报结果表明,该模型对于雷电潜势预报是可行的,其预报效果较好。

表 4 2006 年 5—8 月雷电发生概率预报 Brier 评分结果

区域	陕北	陕北	关中	关中	陕南	陕南
	北部	南部	东部	西部	西部	东部
5 月	0.327	0.363	0.324	0.271	0.346	0.480
6 月	0.588	0.653	0.297	0.244	0.516	0.422
7 月	0.258	0.140	0.382	0.409	0.192	0.190
8 月	0.225	0.283	0.308	0.430	0.261	0.250

5 小结与讨论

5.1 对备选因子,通过逐步回归或逐步判别对因子进行初选,保证各因子相互独立。进入模型的因子数决定于因子质量,质量越高,所需数目越少;也决定于预报对象的气候概率 P_0 , P_0 越小,

所需因子数越多。

5.2 用非线性回归建立各因子的 C 函数有一定的经验性,表现在做点聚图后的分区和对曲线的选择上。

5.3 多因子非线性模型较有效的避免了统计学中的线性假设和模式的输出值物理意义不明确的问题,其试报结果表明该模型运用于雷电潜势预报是可行的,效果良好。

在此对李左老师的指导表示真挚的感谢!

参考文献:

- [1] 朱伯承.统计天气预报 [M].上海:上海科学技术出版社,1981: 364—367.
- [2] 邵明轩,刘凤辉,程维中,等.用非线性多因子动态组合方法作降水概率预报 [J].气象科技,2003,31 (4): 206—210.