

基于逻辑回归模型的雷电潜势预报方法

梁巧倩¹, 蒙伟光², 林良勋¹, 万齐林²

(1. 广州中心气象台, 广东广州 510080; 2. 广州热带海洋气象研究所, 广东广州 510080)

摘 要: 利用闪电定位资料、广州地面观测资料和清远探空观测资料, 分析了广州的雷电日分布特点, 提取了相关的环境场物理量, 应用逻辑回归模型技术, 针对广州地区 3~6 月的雷电活动, 研发了适用于该地区日间和夜间雷电活动的预报方程, 为雷电潜势预报提供了一种比较客观的预报方法。检验结果表明, 该方法的预测成功率超过 76%, 日间段预报效果总体优于夜间段, 具有较大的业务参考价值。

关键词: 应用气象学; 雷电; 逻辑回归模型; 潜势预报

中图分类号: P49

文献标识码: B

文章编号: 1007-6190(2011)04-0044-04

随着社会经济建设的快速发展, 人民生活水平的日益提高, 由雷电灾害带来的经济损失越来越大, 由雷击造成的人身伤亡事故也呈上升趋势。如何更加有效地防范由雷电带来的灾害, 在对雷电活动总体特征进行分析的基础上, 进一步开展雷电活动预报方法的研究很有必要。

雷电的预警大多是在潜势预报的基础上, 进行短时临近的监测、跟踪和发布短时临近警报^[1]。基于天气分型和预报指标的雷电潜势预报方法, 是一种比较常用的主观定性预报方法, 一定程度上提高了雷电活动的预报准确率^[2-3]。随着雷电监测技术的发展, 闪电定位资料已经在气象预报和研究方面得到应用, 基于这些闪电定位观测资料和各种统计分析方法的雷电活动客观预报方法也得到了发展^[4-5]。这些方法可具体分析雷电的活动特点, 并通过对多年资料的统计分析, 建立相应的预报方程, 推进雷电预报的客观性。

广东省闪电密度的分布特征表明, 闪电密度最大值位于清远至广州一带, 平均每年出现闪电的次数达到 10~20 次/km², 广州是广东省闪电发生频繁的地区之一^[6]。但目前业务上运行的雷电潜势预报方法还比较少。针对广州地区雷电潜势预报, 借助统计回归方法和最新观测资料研发客观有效的预报方法是必要的。

因此, 本文基于闪电定位资料和环境场物理量, 应用二分变量逻辑回归模型, 研发用于广州地区 3~6 月雷电活动的潜势预报方法, 以便在业务上尝试雷电客观定量化预报。

1 资料说明和有关定义

所选用资料包括: (1) 广东省电力部门提供的闪电定位资料; (2) 源于广州市五山气象台的广州雷暴日资料; (3) 位于广州市北部约 60 km 的清远站探空观测资料。

所选用的 3~6 月份资料样本长度为 2001~2005 年, 共 5 年。其中 2001~2004 年的资料用于预报方程的建立, 2005 年的资料用于预报效果检验。

广州地区的范围是指以广州城区为中心, 半径 50

km 的范围, 如图 1 所示, 该范围主要包括了广州市区、番禺、花都、增城和从化的部分地区以及佛山的部分地区。



圆圈是以广州为中心半径 50 km 的范围; 粗黑点为清远探空站位置
图 1 广州地区地图

定义广州的日间雷电为: 当日 08:00~20:00 (或夜间雷电为: 当日 20:00~次日 08:00), 广州地区出现云闪电纪录 ≥ 50 次, 则确定该日广州地区日间 (或夜间) 出现了雷电。

2 广州地区雷电的主要特点

广州地区的电闪数逐月变化呈准“双峰”型^[7] (如图 2 所示), 初雷一般出现在 3 月份, 高峰期在汛期 6~8 月, 次峰在 4 月前后。4 月前后是广东前汛期, 北方冷空气的影响仍然存在, 而海洋暖湿气流开始加强, 两者交汇易产生激烈天气, 是电闪急剧增加的时期。正电闪数的逐月变化主要呈“单峰”分布, 峰值在 4~5 月, 与总电闪的次峰值正好对应。可见时间的变化上, 广州地区的雷电活动在 3~6 月份具有较一致的阶段变化。逐月的以

及平均的电闪总数空间分布表明,广州地区的雷电活动具有空间一致性,年平均最频繁的雷电活动出现在广州的北侧。从影响系统来看^[7],春末夏初,西风槽类天气型是广州地区雷电频发的主导天气型,直到6月份,随着副热带高压的逐渐加强西伸北跳,副热带高压影响类的天气型才逐渐明显。由此可见,3~6月广州地区的雷电活动具有较一致的天气型。因此,探讨广州地区雷电预报技术,利用3~6月份来建立相应的预报方程是适宜的。

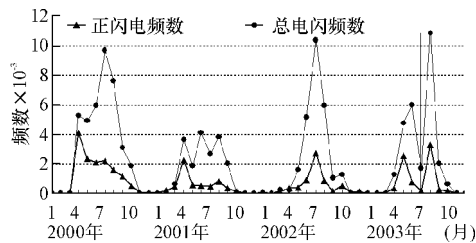


图2 2000~2003年逐月的总闪电和正闪电曲线

3 技术方法基本原理^[8-10]

回归分析是一种比较客观有效的定量统计分析技术,它通过分析因变量和自变量的关系,建立数学模型。传统的回归分析技术一般要求因变量是连续变量,对于因变量是离散的情形(分类变量或二分变量),传统的回归分析技术就不再适用,但逻辑回归模型可以适用。逻辑回归模型没有关于自变量分布的假设条件,从而避开了分析中面临的难以满足的前提假设,运用逻辑回归模型对问题进行分析方便可行,适用范围更广。Everitt^[11]的研究表明,对于相同的预报因子变量和数据,采用逻辑回归模型建立的预报方程比用线性回归方法建立的预报方程,可使预报技能提高48%左右。因此,基于闪电定位资料和环境场物理量,可以应用二分变量逻辑回归模型技术建立关于广州地区有无雷电发生的概率预报方程。

逻辑回归模型通过非线性变换将自变量的线性组合变换为因变量取值的概率刻度,从而可以用来建立分类规则。一般地,贝努利事件的发生与否可以用二分变量来表示。考虑一个有限的独立观察数据集 (y_i, x^i) , $i=1, 2, \dots, n$, 其中 $x^i = [x_1^i, \dots, x_s^i]^T$ 是 s 维的自变量,是贝努利变量,其分布依赖于 y_i 的取值。若事件属于第1类($y=1$, 设定为雷电发生)的概率 p ,属于第0类($y=0$, 设定为无雷电发生)的概率为 $1-p$,则概率 p 的逻辑变换 $\text{logit}(p)$ 为:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = a + b^T x \tag{1}$$

其中 a 为截距 $b = [b_1, \dots, b_s]^T$ 为自变量 x 的系数,自变量 x 可以是连续的、二分的、离散的或者是混合形式的。由式(1)可见,因变量是事件发生概率的自然对数,因此一般不需要像线性回归那样,假定因变量与自变量之间的线性关系以及自变量为服从独立正态分布的变量。根据式(1)可知,事件发生的概率 $p(x)$ 是自变量的非线性函数,形式如下:

$$p(x) = \frac{1}{1 + e^{-\text{logit}(p(x))}} = \frac{1}{1 + e^{-(a + b^T x)}} \tag{2}$$

其中 $p \in [0, 1]$ 。若 $a + b^T x \rightarrow -\infty$ 时 $p \rightarrow 0$;若 $a + b^T x \rightarrow \infty$ 时 $p \rightarrow 1$ 。通常,由于逻辑回归模型是非线性的,可以应用极大似然法估计回归系数 a 和 b 。当事件(如雷电)发生的概率 $p(x) > 0.5$ 时预测事件将会发生, < 0.5 时预测事件将不会发生。

应用SPSS统计分析软件,实现二分变量逻辑回归模型的建立。在建立回归方程时,遵循一个原则:精简实用。具体地说,既要尽可能地提高拟合的精确度,又要尽可能地使模型简单^[12]。

4 预报方程的发展和检验

4.1 雷电日和预报量的确定

根据第1章关于雷电发生的具体定义,对2001~2005年3~6月份的闪电定位资料进行统计,得到雷电日的统计结果如表1所示。在2001~2005年3~6月份期间,除掉记录不完整(主要是探空观测资料)的个例,日间的样本数为534 d,其中有雷电的天数为168 d;夜间的样本数为529 d,其中有雷电的天数为76 d。可见日间雷电发生的概率可达到31.46%,而夜间发生的概率仅为14.37%。剔除日间和夜间重复的个例,符合定义的雷电日为148 d,与广州观象台记录到的该时段年平均雷暴日数135 d相当接近。根据逻辑回归模型对因变量的要求,确定预报方程中雷电是否发生为因变量,当预报时段(日间或夜间)有雷电发生时,预报量被置为“1”,无雷电发生时预报量被置为“0”。

表1 2001~2005年3~6月份不同类型雷电现象的统计

雷电类型	资料样本天数	无雷电天数	有雷电天数	雷电发生概率/%
日间型	534	366	168	31.46
夜间型	529	453	76	14.37
总计	1 063	819	244	22.95

4.2 待选预报因子的选择

雷电现象往往伴随着强风暴所带来的强对流天气而出现,因此有利于强风暴天气出现的因子对于雷电的活动也是重要的。强对流天气发生的层结特征分析表明,珠江三角洲地区强对流天气的大气层结有“上干下湿”型和湿层深厚型^[13-14]。与温带地区相比,“中层干冷”的条件在这里并不是必要的。因此强对流天气的发生发展与温度、湿度的平流表现出密切的关系,高层冷平流与低层暖平流的叠置成为珠江三角洲地区强对流天气环境场的最基本特征之一。蒙伟光等^[15]通过对有无雷电活动的对流天气过程的比较分析发现,广州地区雷电易出现在具有更大垂直风切变、更高相对风暴螺旋度以及有更多对流抑制能量的环境场中,与对流系统的组织性和强度有关系。国外的一些研究则表明,风暴环境场0~3 km的风切变(廓线)、云中冻结高度层以下对流有效位能(CAPE)大小以及边界层混合比的大小等因素可影响到风暴的结构和演变,从而影响到云中电荷的分布结构

和雷电的活动^[16]。由此可见,环境场物理量表现出来的特征,对雷电的活动有一定的指示意义。

因此,基于清远站(距离广州最近的探空站)探空观测资料,提取和计算了 20 个分别表征气层冷暖平流、环境风切变、对流程度等相关物理量作为预报方程的待选因子如表 2 所示。其中日间样本的相关物理量用 08:00 的探空资料算出,夜间样本的用 20:00 探空资料算出。

另外,根据第 2 章的分析表明,3~6 月广州地区引发雷电的天气系统和环流形势较为相似,因此在预报因子的选择上不再考虑天气分型。

表 2 预报方程的待选因子

因子名(量符号)	因子名(量符号)
地面温度(t)	地面气压(p)
地面露点温度(t_d)	混合层比湿(q_m)
抬升凝结高度(lcl)	抬升指数(li)
k 指数(k)	总温度指数(t_t)
莎瓦特指数(si)	对流有效位能($cape$)
对流抑制位能(cin)	对流温度(t_c)
风暴相对螺旋度($sreh$)	风暴相对速度(stv)
威胁指数(sw)	500 hPa 温度(t_5)
0~6 km 垂直切变(s_6)	0~3 km 垂直切变(s_3)
850 hPa 温度递减率(r_8)	700 hPa 温度递减率(r_7)

4.3 预报方程建立

1) 日间雷电预报方程的发展。

参与预报方程发展的日间样本共 415 例,其中无雷电发生 296 例,有雷电发生 119 例。各待选因子和因变量进入逐步二分变量逻辑回归分析。最后有 3 个因子 q_m 、 li 和 k 进入预报方程,得到二分变量逻辑回归预报方程的系数 $\logit(p)$ 为:

$$\logit(p) = 0.501 \times q_m + 0.159 \times li + 0.141 \times k - 13.159$$

系数通过了 $\alpha = 0.05$ 的显著性检验。把 $\logit(p)$ 代入式(2)的概率函数就可以得到雷电发生的概率值。利用预报方程反算 415 例样本,拟合率达 78.3%。

2) 夜间雷电预报方程的发展。

夜间样本共 407 例,其中无雷电 355 例,有雷电 52 例。同日间的分析步骤进行逐步二分变量逻辑回归分析,有两个因子 q_m 和 k 进入回归方程,得到二分逻辑回归方程的系数 $\logit(p)$ 为:

$$\logit(p) = 0.209 \times q_m + 0.227 \times k - 12.904$$

系数通过显著性($\alpha = 0.05$)检验。 $\logit(p)$ 代入式(2)的概率方程反算 407 例,拟合率为 87.7%。

3) 讨论。

由日间和夜间预报方程的入选因子可以发现,在判别雷电发生的指示意义上,混合层比湿和 k 指数优于其它因子。雷电的发生通常和对流天气联系在一起。对流天气的基本条件包括:水汽条件、不稳定层结条件和抬升力条件^[17]。此前已有的研究表明, k 指数被作为预报雷暴活动的首选指数^[1,18-20],其描述的是雷电发生时大气

层结的不稳定特性。广东省天气预报技术手册中对强对流天气的总结指出:珠江三角洲强对流天气区附近的大气层结可划分为“上干下湿”型和湿层深厚型两类^[14],两类层结类型,其低层都是非常潮湿的。混合层比湿描述的也正是低层的湿度情况,和雷电的发生存在着正相关,相关系数在表 2 所列的因子中是较大的。湿度的加大,也有利于位势不稳定层结的建立。因此,日间和夜间预报方程中的入选因子,对雷电发生的描述是有物理意义的。

雷电的活动是一个复杂的物理过程,尽管已有的研究结果表明雷电的活动与诸如风切变、环境相对螺旋度等一些因子有密切关系^[15-16],但这些因子作为待选因子并没有被选入预报方程中。没被选入方程中并不代表其不能反映雷电活动的物理过程,只是在拟合的过程中,入选因子具有更高的敏感性和使构成的方程最优化。从方程的逐步回归计算方法来说,入选因子的组合对方差贡献是最大的,非入选因子的加入不会再次改善方程的方差贡献度。

4.4 预报检验

分别日间和夜间,利用上述分析得到的二分变量逻辑回归方程对 2005 年 3~6 月进行试报。2005 年日间样本共有 119 例,夜间 122 例。试报结果的评分如表 3 所示。

表 3 2005 年 3~6 月试报评分结果 %

预报类型	命中率	空报率	成功率
日间	85.6	7.8	79.8
夜间	79.5	5.1	76.2

由表 3 可见,分别日间和夜间的二分变量逻辑回归预报方程对雷电是否发生的预测都有 76% 以上的预报成功率,其中日间的成功率和命中率都比夜间的略高,同时空报率也偏高。总体的预报效果日间的优于夜间的,表明了基于环境场物理量的统计回归分析对于雷电的预报来说是可行的。

在预报方程的发展中选用了清远站探空观测资料来描述广州地区的环境场特征,这种因变量和自变量间是同步的关系,在业务实际应用中预报能力受到一定的限制,但数值预报模式探空输出从一定程度上可以改善其预报能力。因此,在将以上方法转化为业务应用的过程中,选择对广州地区环境场描述较好的模式,利用模式环境物理量代入预报方程,将可得到更长预报时效的潜势预报。另外,一个友好的界面平台,使所需要的数据能方便的输入和快速输出,也将大大增加预报方法的日常业务可用性。

5 结论

基于闪电定位资料和环境场物理量资料,应用二分逻辑回归模型技术,分别发展了用于预报广州地区 3~6 月日间和夜间雷电活动的预报方程,为雷电预报提供了一种比较客观的统计预报方法。

(下转第 51 页)

参考文献:

- [1] 蒋维梅, 孙鉴宁, 曹文俊, 等. 空气污染气象学教程[M]. 2 版. 北京: 气象出版社, 2007: 261–263.
- [2] 洪伟, 古国榜, 杨立君. 大气扩散高斯模式的灵敏度系统分析[J]. 环境工程, 2003, 21(2): 62–64.
- [3] 吴艳标, 刘蔚, 冯小冰. 大气扩散模型在气象应急保障中的应用[J]. 广东气象, 2009, 31(2): 4–6.
- [4] 王自发, 谢付莹. 嵌套网格空气质量预报模式系统的发展与应用[J]. 大气科学, 2006, 30(5): 778–790.
- [5] 罗淦, 王自发. 全球环境大气输送模式(GEATM)的建立及其验证[J]. 大气科学, 2006, 30(3): 504–518.
- [6] 魏晓林, 贺佳佳, 王安宇. 热带气旋影响下高臭氧污染的调控机理[J]. 广东气象, 2010, 32(3): 32–34.
- [7] 吴兑, 毕雪岩, 邓雪娇, 等. 珠江三角洲大气灰霾导致能见度下降问题研究[J]. 气象学报, 2006, 64(4): 510–515.
- [8] 吴兑. 霾与雾的区别和灰霾天气预警建议[J]. 广东气象, 2004, 24(4): 1–4.
- [9] 吴兑. 沿海工业城市灰霾天气增多与海盐气溶胶粒子的关系[J]. 广东气象, 2009, 31(2): 1–3.
- [10] GRELL G A, DUDHIA J, STAUFFER D R. 1995: A description of the fifth generation Penn State/NCAR Mesoscale Model (MM5). NCAR Tech. Note NCAR/TN–398_STR, 138 pp. [Available from NCAR Publications Office, P O Box 3000, Boulder, CO80307–3000.
- (上接第46页)
- 尽管所用资料的样本时间不长, 但两个预报方程对于雷电的活动均表现出较高的预报能力, 表明基于环境场物理量的统计回归分析对于雷电的预报来说是可行的。
- 但是雷电活动是一种很局地的现象, 如何更加精细地考虑结合数值预报产品和天气分型的指标、雷电发生的气候概率、前期的雷电活动情况、不同地域的特征、分区、分强度等级, 完善全年不同季节预报等, 都将是今后雷电预报发展值得考虑的问题。
- 参考文献:
- [1] 许小峰, 郭虎, 廖晓农, 等. 国外雷电监测和预报研究[M]. 北京: 气象出版社, 2003: 271–272.
- [2] 陈立详, 刘运策. 广东地区强对流天气的统计特征和分类特征[J]. 热带气象学报, 1989, 5(2): 170–179.
- [3] Watson A L, Holle R L, Lopez R E. Cloud-to-ground lightning and upper-air patterns during burst and breaks in the southwest monsoon[J]. Mon Wea Rev, 1994, 122: 1726–1739.
- [4] Burrows W R, Price C, Wilson L J. Warm season lightning probability prediction for Canada and the northern United States[J]. Wea Forecasting, 2005, 20: 971–988.
- [5] Kathryn K Hughes. Probabilistic lightning forecast guidance for aviation[J]. 11th Conference on Aviation, Range, and Aerospace, October 2004, Hyannis, MA.
- [6] 易燕明, 杨兆礼, 万齐林. 广州市闪电密度特征分析[J]. 资源科学, 2006, 28(1): 151–156.
- [7] 林良助, 梁巧倩, 温晶, 等. 广州地区雷电产生的时空分布特点及天气学分析[J]. 气象科技, 2006, 4(6): 679–683.
- [8] 李品芳, 周兆明, 陈帅. 基于多元回归分析的财务危机预警研究[J]. 同济大学学报: 自然科学版, 2006, 34(5): 701–706.
- [9] 徐晓燕, 王昱, 张斌. 一种集成 logistic 回归与支持向量机的判别分析规则[J]. 系统工程理论与实践, 2007(4): 41–46.
- [10] 黄嘉佑. 气象统计分析与预报方法[M]. 北京: 气象出版社, 2004.
- [11] Everitt J A. An improved thunderstorm forecast index for Cape Canaveral[J]. Florida M S Thesis, 1999, AFIT/GM/ENP/99M–06, Department of Engineering Physics, Air Force Institute of Technology: 98.
- [12] 林杰斌, 林川雄, 刘明德, 等. SPSS12 统计建模与应用实务[M]. 北京: 中国铁道出版社, 2006.
- [13] 罗会邦. 珠江三角洲局地强风暴[M]. 广州: 中山大学出版社, 1994.
- [14] 广东省气象局编写组. 广东省天气预报技术手册[M]. 北京: 气象出版社, 2006.
- [15] 蒙伟光, 易燕明, 杨兆礼, 等. 广州地区雷暴过程云-地闪特征及其环境条件[J]. 应用气象学报, 2008, 19(5): 611–619.
- [16] Gilmore M S, Wicker L J. Influences of the Local Environment on Supercell Cloud-to-Ground Lightning, Radar Characteristics, and Severe Weather on 2 June 1995[J]. Mon Wea Rev, 2002, 130: 2349–2372.
- [17] 朱乾根, 林锦瑞, 寿绍文, 等. 天气学原理和方法[M]. 北京: 气象出版社, 2007.
- [18] Reap R M. New 6-h thunderstorm probability forecasts for the west. NWS Tech[J]. Proc Bull. NO. 362, NOAA, U. S. Department of Commerce, 1986: 8.
- [19] Reap R M, Foster D S. Automated 12–36 h probability forecasts of thunderstorms and severe local storms[J]. J Appl Meteor, 1979, 18: 1304–1315.
- [20] Reap R M. Climatological characteristics and objective prediction of thunderstorms over Alaska[J]. Wea and Forecasting, 1991, 3: 309–319.