

# 朝阳地区降水概率预报模型构建

郭佰汇<sup>1</sup>, 吴丹<sup>2</sup>, 范兰艳<sup>3</sup>, 冯雪菲<sup>1</sup>, 王梦琳<sup>1</sup>

(1. 辽宁省朝阳市气象局, 辽宁 朝阳 122000; 2. 辽宁省朝阳市龙城区气象局, 辽宁 朝阳 122005; 3. 辽宁省朝阳县气象局, 辽宁 朝阳 122000)

**摘要:** 以朝阳地区2005—2013年5—9月降水量记录为基础资料, 通过经验法和Spearman、Pearson、Kendall's Tau-b相关系数3种相关性检验法, 选出与预报对象相关性好的气象要素作为预报因子。用Logistic回归方法进行有无降水的概率预报, 建立了朝阳地区降水概率预报模型。该预报方程具有较高的历史拟合率, 为90.7%。将2014年5月1日至10月31日个例作为样本对方程进行检验, 当规定降水概率 $\geq 40\%$ 为有降水时, 晴雨预报准确率最高, 为86.26%, 高于朝阳地区本地晴雨预报准确率3.29个百分点, 表明该方法对朝阳地区降水具有很好的预报效果。

**关键词:** 降水概率预报; Logistic回归; 相关系数法

**中图分类号:** P456.8

**文献标识码:** A

**文章编号:** 1001-1463(2015)07-0031-03

**doi:** 10.3969/j.issn.1001-1463.2015.07.011

降水是大系统相互作用的结果, 它具有小尺度易变性等特点, 在时空分布上具有随机性, 因此发布绝对准确的降水预报是不可能的。以往的定性预报用简单的“有”或“无”来描述降水, 人为的增大了预报误差。概率预报则以百分率形式对降水出现的可能性大小作出判断, 较真实地反映了降水的不确定性, 使预报更科学, 更客观, 更具参考价值。

Logistic回归模型在建立公式时较简单, 在理论、数学模型及实用上却都具有很强的生命力。该方法首先是Cox D. R. 提出, 后经Day N. E. 和Korriage D. F. 发展, 又由Anderson J. A. 改进<sup>[1-3]</sup>。由于大气是一个高度非线性的混合系统, 而模式的初始场只是大气真实状态的近似, 数值模式所描述的大气过程也是非真实的大气过程, 所以单一的确定性预报水平的提高已经变得越来越困难, 概率预报成为天气预报发展的必然趋势。Logistic回归模型适用于大量的观测因变量是二分类变量<sup>[4]</sup>, 符合降水发生和不发生的特性, 并且

它是非线性的, 符合大气是非线性的系统的本质。

## 1 资料与方法

### 1.1 资料选取

选用朝阳地区国家基准站朝阳县气象观测站(站号54324)2005—2013年5—9月常规地面气象观测站观测历史数据文件(A文件)实况气象资料作为建立预报方程的基础资料, 选用2014年5—10月朝阳县站的A文件资料作为预报方程检验的资料, 选用朝阳市2014年5—10月晴雨预报资料作为方程检验结果的对比资料。

### 1.2 方法

**1.2.1 相关系数** Pearson相关系数用来判定两个数据集合是否在一条线上, 它用来衡量定距变量间的线性关系。当两个变量都是正态连续变量, 而且两者之间呈线性关系时, 表现这两个变量之间的相关程度用Pearson相关系数。

Spearman相关系数利用两变量的秩次大小作线性相关分析, 对原始变量的分布不作要求, 属

收稿日期: 2015-04-23

基金项目: 辽宁省朝阳市项目《基于logistic模型的降水概率预报》

作者简介: 郭佰汇(1989—), 女, 辽宁朝阳人, 助理工程师, 主要从事天气预报预警研究工作。联系电话: (0)13591876815。

E-mail: guobaihui@sina.com

- 定[J]. 果树学报, 2006, 23(2): 196-199.
- [23] 刘威生, 张加延, 唐士勇, 等. 李属种质资源的抗寒性鉴定[J]. 北方果树, 1999, 1(2): 6-8.
- [24] 全月澳, 周厚基. 果树营养诊断法[M]. 北京: 农业出版社, 1982: 250.

- [25] 张宪政, 谭桂茹, 黄元极, 等. 植物生理学实验技术[M]. 沈阳: 辽宁科技出版社, 1989: 333-337.
- [26] 章文才. 果树研究法[M]. 北京: 农业出版社, 1979: 450.

(本文责编: 陈伟)

于非参数统计方法,适用范围要广些。对于服从 Pearson 相关系数的数据亦可计算 Spearman 相关系数,但统计效能要低一些。

Kendall's Tau-b 相关系数用于反映分类变量相关性的指标,适用于两个分类变量均为有序分类的情况。

1.2.2 logistic 回归模型 利用 SAS9.2 软件中的 Logistic 回归模型,建立朝阳地区降水概率预报模型。SAS 总设  $Y=1$  表示某事物不发生;  $Y=0$  表示某事物发生;  $1-E(Y)=p$  表示某事物发生的概率;  $X_1 \cdots X_m$  是预报因子;  $\ln(p/(1-p))=b_0+b_1X_1+\cdots+b_mX_m+\varepsilon=b_0+\beta'X+\varepsilon$  是 Logistic 回归模型。Logistic 回归模型的建模问题,就是从观测数据求出  $b_0, b_1 \cdots b_m$ 。通常用极大似然法估计  $b_0, b_1 \cdots b_m$ , 即由

$$l = \prod_{i=1}^k \frac{\exp(b_0 + \beta'X_i)}{1 + \exp(b_0 + \beta'X_i)} \prod_{j=k+1}^n \frac{1}{1 + \exp(b_0 + \beta'X_j)} = \max$$

求解,其中  $n$  是观测次数;  $k$  是事件发生次数;  $X_i$  是事件发生时预报因子的观测值,  $X_j$  是事件不发生时预报因子的观测值<sup>[5-8]</sup>。

1.2.3 因子筛选 因子的选择是影响预报模型准确性的一个重要因素。先对因子进行粗选,以将日常进行降水预报的经验为依据,选择降水预报时会考虑到的各气象要素,包括降水、气压、10

min 风向风速、相对湿度、总云量、低云量作为入选因子;这些入选物理量的平均值、最大值、最小值、变化值都作为建立预报模型的一个因子,其中考虑引入 24 h 变压(P24)、平均气压(Pagv)、最高气压(Pmax)、最低气压(Pmin)、平均相对湿度(Uagv)、最大相对湿度(Umax)、最小相对湿度(Umin)、平均总云量(Zagv)、最大总云量(Zmax)、最小总云量(Zmin)、平均低云量(Dagv)、最大低云量(Dmax)、最小低云量(Dmin)、10 min 平均风速(FV10agv)、2 时次 10 min 风向(F2)、8 时次 10 min 风向(F8)、14 时次 10 min 风向(F14)、20 时次 10 min 风向(F20),将这些要素作为待选的预报因子,降水量(R)作为预报对象。

完成因子粗选后进行因子的精选,通过计算粗选因子与预报对象 R 之间的线性相关系数,挑选相关系数的绝对值较大的各物理量为入选因子,入选因子要通过信度为 0.05 的相关检验,才能作为精选因子,最后建立预报模型。

在进行相关性检验时,选用 Spearman 相关, Pearson 相关和 Kendall's Tau-b 相关 3 种相关性检验方法对粗选因子进行筛选,根据以上 3 种相关性检验方法,计算得降水量 R 与各要素之间的相关性如表 1。

表 1 R 与各要素间相关性检验

气象要素	Kendall相关		Pearson相关		Spearman相关	
	相关系数	统计量	相关系数	统计量	相关系数	统计量
P24	0.015 63	0.479 6	0.011 06	0.681 7	0.019 06	0.479 8
Pagv	-0.155 05	<0.000 1	-0.193 40	<0.000 1	-0.189 36	<0.000 1
Pmax	-0.137 85	<0.000 1	-0.169 00	<0.000 1	-0.168 36	<0.000 1
Pmin	-0.144 87	<0.000 1	-0.182 16	<0.000 1	-0.176 94	<0.000 1
Uagv	0.365 88	<0.000 1	0.428 13	<0.000 1	0.444 28	<0.000 1
Umax	0.265 89	<0.000 1	0.281 69	<0.000 1	0.314 19	<0.000 1
Umin	0.338 61	<0.000 1	0.436 95	<0.000 1	0.411 55	<0.000 1
ZYagv	0.289 84	<0.000 1	0.339 31	<0.000 1	0.340 33	<0.000 1
ZYmax	0.208 55	<0.000 1	0.282 73	<0.000 1	0.236 23	<0.000 1
ZYmin	0.261 19	<0.000 1	0.316 95	<0.000 1	0.298 24	<0.000 1
DYagv	0.515 78	<0.000 1	0.610 53	<0.000 1	0.594 68	<0.000 1
DYmax	0.388 85	<0.000 1	0.442 62	<0.000 1	0.428 34	<0.000 1
DYmin	0.297 90	<0.000 1	0.277 33	<0.000 1	0.301 66	<0.000 1
FV10agv	-0.047 62	0.033 0	-0.051 43	0.056 4	-0.057 47	0.033 0
F2	0.003 44	0.876 4	-0.034 71	0.198 1	0.004 19	0.876 5
F8	-0.029 55	0.181 0	-0.050 82	0.059 5	-0.036 08	0.181 1
F14	-0.042 62	0.053 5	-0.036 51	0.175 9	-0.052 07	0.053 5
F20	0.025 04	0.257 3	0.001 50	0.955 5	0.030 54	0.257 4
平均值		0.104 489		0.118 17		0.104 516 67

对各检验方法得出的检验统计量进行分析, Pearson 相关的检验统计量的平均值明显大于 Spearman 相关和 Kendall's Tau-b 相关, Spearman 相关和 Kendall's Tau-b 相关的检验统计量的平均值基本相同, 表明这两种相关性检验中各要素对 R 的总体相关性更优。以 Spearman 相关和 Kendall's Tau-b 相关检验的结果作为因子精选的参考, 发现这 2 种检验方法筛选出的因子是相同的, 最终选取了 13 个物理量 (Pagv、Pmax、Pmin、Uagv、Umax、Umin、Zagv、Zmax、Zmin、Dagv、Dmax、Dmin、FV10agv) 作为建立预报模型的因子。

## 2 结果与分析

利用 2005—2013 年 5—9 月共 1 377 d 的历史个例样本, 由所筛选出的 13 个气象要素因子, 通过 Logistic 回归方法建立朝阳地区降水概率预报方程如下。

$$\ln(p/(1-p)) = 20.9265 - 0.055 P_{agv} + 0.0492 P_{max} + 0.00251 P_{min} + 0.00763 U_{agv} + 0.0666 U_{max} + 0.000528 U_{min} + 0.0235 Z_{agv} - 0.0477 Z_{max} - 0.00111 Z_{min} + 0.3498 D_{agv} + 0.291 D_{max} - 0.0002 D_{min} - 0.0119 FV10agv$$

经检验, 该预报方程具有较高的历史拟合率, 为 90.7%, 采用似然比检验、SCORE 检验、Wald 检验,  $Pr > \chi^2_{sq}$  且都小于 0.0001, 说明预报因子对预报对象的影响是显著的, 所建立的降水概率预报方程是有意义的。将 2014 年 5 月 1 日至 10 月 31 日 8:00~8:00 时段共 182 d 的个例作为样本进行试报, 182 d 的天气个例中, 48 d 出现降水 (其中 5 d 为微量降水), 134 d 无降水。分别以 10%~50% 作为划分是否预报出现降水的概率, 对降水预报结果如表 2。

表 2 晴雨预报准确率 %

划分概率	晴雨准确率
10	80.22
20	84.07
30	84.62
40	86.26
50	85.71

从结果可以看出, 当划分概率为 40% 时, 晴雨预报准确最高, 为 86.26%。对比朝阳市本地预

报结果, 在 2014 年 5—10 月期间朝阳地区 8:00~8:00 时时段的晴雨预报准确率为 82.97%, 这比本文预报方法的准确率低了 3.29 个百分点, 可见用 logistic 回归模型建立的降水预报模型对于朝阳地区的预报效果是非常好的。

## 3 小结与讨论

将降水作为预报对象, 以 A 文件中部分资料作为预报因子, 并用经验法和相关性检验法对预报因子进行粗选和精选, 通过 logistic 回归方法, 建立了朝阳地区降水概率预报模型。利用 2005—2013 年 5—9 月资料建立的方程具有较高的历史拟合率, 为 90.7%。将 2014 年 5 月 1 日至 10 月 31 日个例作为样本对方程进行检验, 当规定降水概率大于等于 40% 为有降水时, 晴雨预报准确率最高, 为 86.26%, 高于朝阳地区本地晴雨预报 3.29 个百分点, 表明该方法对朝阳地区降水具有很好的预报效果。当然, 该预报方法还有一定的不足, 在建立方程以及检验方程时, 只应用了降雨资料, 未考虑降雪以及秋冬季、冬春季交替时常出现的雨夹雪等情况, 这些需要在以后的研究中不断补充。

## 参考文献:

- [1] COX D R. Some procedure associated with the logistic qualitative response cure[C] //Neyman Ed, F N David Research Rapers in Statistics: Festschrift for J. New York: Wiley, 1966.
- [2] BOCCHIERI J R. Use of the logimodel to transform predictors for precipitation type forecasting [C]//Amer: Preprint 6th Conf. on Probability and Statistics in Atmos Sci Amer: Meteor Soc, 1979.
- [3] 纪玲玲, 王昌雨, 张志华. Logistic 回归及其在概率降水预报中的应用[J]. 解放军理工大学学报, 2003, 4(5): 92-94.
- [4] 汪海波, 罗莉, 吴为, 等. SAS 统计分析与应用[M]. 北京: 人民邮电出版社, 2013.
- [5] 吕纯谦, 陈杰伦. Logistic 及其在气象上的应用[J]. 南京气象学院学报, 1982, 5(1): 112-123.
- [6] 万夫敬, 袁慧玲, 宋金杰, 等. 南京地区降水预报研究. 南京大学学报, 2012, 48(4): 513-525.
- [7] 汤浩. 新疆降水概率预报技术研究. 新疆气象, 2003, 26(1): 5-7.
- [8] 黄永新. 南宁市降水概率预报方法研究. 广西气象, 1997, 18(1): 49-52.