

DOI:10.12171/j.1000-1522.20200173

# 基于贝叶斯模型平均法的森林火灾预测模型构建研究 ——以云南省大理州为例

白海峰<sup>1</sup> 刘晓东<sup>1</sup> 牛树奎<sup>1</sup> 何亚东<sup>2</sup>

(1. 北京林业大学生态与自然保护学院, 北京 100083; 2. 云南大学数学与统计学院, 云南 昆明 650504)

**摘要:**【目的】本文基于贝叶斯模型平均法, 结合二项逻辑斯蒂回归模型, 构建云南省大理州森林火灾发生预测模型, 以期提高林火预测精度, 为研究地区林火管理提供技术支持。【方法】利用 2000—2013 年大理州林火数据及对应的气象数据, 分别运用二项逻辑斯蒂回归模型和贝叶斯模型平均法, 对该地区森林火灾对气象因子的响应进行实证分析。二项逻辑斯蒂回归模型为单一模型, 建模前通过对各解释变量进行多重共线性检验, 剔除有显著共线性的解释变量, 然后通过逐步回归法, 筛选最终变量并进行参数拟合。贝叶斯平均模型为组合模型, 基于贝叶斯模型平均法建模时, 采用奥卡姆窗的方法来适当调整模型空间, 并以 5 个最优模型的后验概率作为权重进行加权建模。将全样本数据随机分成 80% 的训练样本和 20% 的测试样本, 基于训练样本建立模型, 对测试样本进行预测, 通过对比观测值和预测值计算模型的准确率。【结果】通过二项逻辑斯蒂模型拟合, 优度为 0.783, 预测精度为 0.718。通过贝叶斯平均模型拟合, 优度为 0.868, 预测精度为 0.807。2 个模型预测结果对比显示, 在训练集中, 贝叶斯平均模型的预测准确率比二项逻辑斯蒂回归模型高 9.3%; 在测试集中, 贝叶斯平均模型的预测准确率比二项逻辑斯蒂回归模型高 8.9%。【结论】在基于气象因子的大理州林火发生预测模型构建研究中, 贝叶斯平均模型的拟合优度和预测精度均高于二项逻辑斯蒂模型, 表明贝叶斯模型平均法具有一定的现实应用意义, 可用于提高研究地区林火预测精度, 有利于森林火灾的决策管理。

**关键词:** 大理州; 森林火灾; 气象因子; 贝叶斯模型平均法; 逻辑斯蒂回归

**中图分类号:** S762.2 **文献标志码:** A **文章编号:** 1000-1522(2021)05-0044-09

**引文格式:** 白海峰, 刘晓东, 牛树奎, 等. 基于贝叶斯模型平均法的森林火灾预测模型构建研究——以云南省大理州为例[J]. 北京林业大学学报, 2021, 43(5): 44-52. Bai Haifeng, Liu Xiaodong, Niu Shukui, et al. Construction of forest fire prediction model based on Bayesian model averaging method: taking Dali Prefecture, Yunnan Province of southwestern China as an example[J]. Journal of Beijing Forestry University, 2021, 43(5): 44-52.

## Construction of forest fire prediction model based on Bayesian model averaging method: taking Dali Prefecture, Yunnan Province of southwestern China as an example

Bai Haifeng<sup>1</sup> Liu Xiaodong<sup>1</sup> Niu Shukui<sup>1</sup> He Yadong<sup>2</sup>

(1. School of Ecology and Nature Reserves, Beijing Forestry University, Beijing 100083, China;

2. College of Mathematics and Statistics, Yunnan University, Kunming 650504, Yunnan, China)

**Abstract:** [Objective] Based on the Bayesian model averaging method and binomial Logistic regression model, this paper constructs a forest fire prediction model in Dali Prefecture, Yunnan Province of southwestern China, so as to improve the prediction accuracy of forest fire and provide technical support for forest fire management in the study area. [Method] Using the forest fire data and corresponding meteorological data of Dali Prefecture from 2000 to 2013, the binomial Logistic regression model and the Bayesian model averaging method were used to empirically analyze the response of forest fires to

收稿日期: 2020-06-19 修回日期: 2021-01-07

基金项目: 国家自然科学基金项目(31770696)。

第一作者: 白海峰。主要研究方向: 林火生态。Email: haifengbai@sina.com 地址: 100083 北京市海淀区清华东路 35 号北京林业大学生态与自然保护学院。

责任作者: 刘晓东, 博士, 教授。主要研究方向: 林火生态。Email: xd\_liu@bjfu.edu.cn 地址: 同上。

本刊网址: <http://j.bjfu.edu.cn>; <http://journal.bjfu.edu.cn>

meteorological factors in this area. The binomial Logistic regression model is a single model. Before modeling, the explanatory variables with significant collinearity were eliminated by multicollinearity test. Then, the final variables were screened by stepwise regression method and the parameters were fitted. The Bayesian average model is a combined model. When modeling based on the Bayesian model averaging method, the Occam's window method was used to appropriately adjust the model space, and the posterior probabilities of the five optimal models were used as weights for weighted modeling. In this paper, the all sample data were randomly divided into 80% training samples and 20% test samples. A model was built based on the training samples to predict the test samples. The accuracy of the model was calculated by comparing the observations and predictions. [Result] Fitting through the binomial Logistic model, the results showed that: the model fitting goodness was 0.783, and the prediction accuracy was 0.718; through the Bayesian average model fitting, the results showed that: the model fitting goodness was 0.868, and the prediction accuracy was 0.807. The comparison of the prediction results of the two models showed that: in the training set, the prediction accuracy of the Bayesian average model was 9.3% higher than that of the binomial Logistic regression model; and in the test set, the former was 8.9% higher than the latter. [Conclusion] In the prediction model of forest fire occurrence in Dali Prefecture based on meteorological factors, the goodness of fit and prediction accuracy of Bayesian average model were higher than that of binomial Logistic model, indicating that the Bayesian model averaging method had certain practical application significance. It can be used to improve the prediction accuracy of forest fire in the study area, which is beneficial to the decision management of forest fire.

**Key words:** Dali Prefecture; forest fire; meteorological factor; Bayesian model averaging method; Logistic regression

近年来, 全球范围内森林火灾的发生频次不断增大, 不仅威胁着人们的生命财产安全, 同时还造成巨大的资源损失和环境破坏。2017 年为欧洲有史以来遭受森林火灾侵袭最为严重的年份之一, 火灾对西班牙、葡萄牙和意大利等国均造成了灾难性的事件<sup>[1]</sup>, 2019—2020 年火险期澳大利亚的野火持续了 5 个月, 引起了全球性持续关注<sup>[2-3]</sup>。我国也是世界上森林火灾发生较为严重的国家之一, 2000—2015 年发生森林火灾的次数平均为 7 632 次/a, 火场总面积平均为 230 622 hm<sup>2</sup>/a, 受害森林面积平均为 94 864 hm<sup>2</sup>/a, 人员伤亡平均为 111 人/a<sup>[4]</sup>。在全球气候变暖的背景下, 火活动将增加, 火险期延长, 野火发生概率升高<sup>[5-8]</sup>, 准确估计火灾概率对于减少林火的负面影响起着至关重要的作用<sup>[9-11]</sup>。因此, 准确的森林火灾发生预测模型的构建是林火发生预测的重要手段, 对防控森林火灾具有十分重要的意义。

随着对森林火灾发生的认识不断深入, 从早期的线性模型到计数模型, 森林火灾发生与气象因子的关系模型结构日趋复杂, 然而这种模型结构的复杂化不能降低林火发生与气象因子关系模型的不确定性。而贝叶斯模型平均法则是近年来文献报告中处理模型不确定性一个很好的方法, 虽目前尚未将其应用于森林火灾预测, 但在其他方面如医疗、水文、渔业的应用<sup>[12-15]</sup>表明, 贝叶斯模型平均算法具有

较高的预测效果和模型稳定性。云南省地处我国西南, 是我国三大林区之一, 但同时也是我国的林火高发区。由于受地理位置、地形地势、气候以及森林资源分布和人为活动等影响, 云南的森林火灾较为严重<sup>[16]</sup>。因此, 本文基于云南省大理州林火数据和气象数据, 运用 R 统计软件, 分别应用逻辑斯蒂回归模型和贝叶斯模型平均法建立林火发生的单一模型和组合模型, 通过不同模型拟合结果的对比分析, 判断贝叶斯模型平均法在构建区域林火预测模型中的适用性。

## 1 研究区概况

大理州地处云南省中部偏西, 地理位置 98°52' ~ 101°03'E、24°41' ~ 26°42'N, 总面积 2.945 9 万 km<sup>2</sup>, 其中山区面积占总面积的 93.4%。地势西北高, 东南低, 平均海拔 2 090 m, 地貌复杂多样, 土壤类型以紫色土和红壤土为主。气候属于低纬度高原季风气候, 立体气候特点显著, 干湿季节分明, 年温差小、日温差大。年均气温 15.7 °C, 年均降水量 836 mm, 年日照时数 2 072 ~ 2 693 h, 无霜期 225 ~ 345 d。大理州森林面积 173.37 万 hm<sup>2</sup>, 森林覆盖率 61.22%, 森林蓄积量 1.1376 亿 m<sup>3</sup>, 主要优势树种有云南松(*Pinus yunnanensis*)、华山松(*P. armandii*)、铁杉(*Tsuga chinensis*)、冷杉(*Abies fabri*)、马尾杉(*Phlegmariurus*)

*phlegmaria*)、思茅松(*Pinus kesiya*)等。

大理州属于森林火灾高发区,该地区冬春季气温高、降水少、风力大,极易发生森林火灾<sup>[17]</sup>。据统计,2000—2013年间,大理州共计发生森林火灾

552次(图1),过火面积高达11 027 hm<sup>2</sup>,其中有林地面积为7 002 hm<sup>2</sup>。由于大理州森林火灾集中发生于1—6月份(551次),因此本文以该时段为研究对象,选取了该时段内的区域林火数据和气象数据。

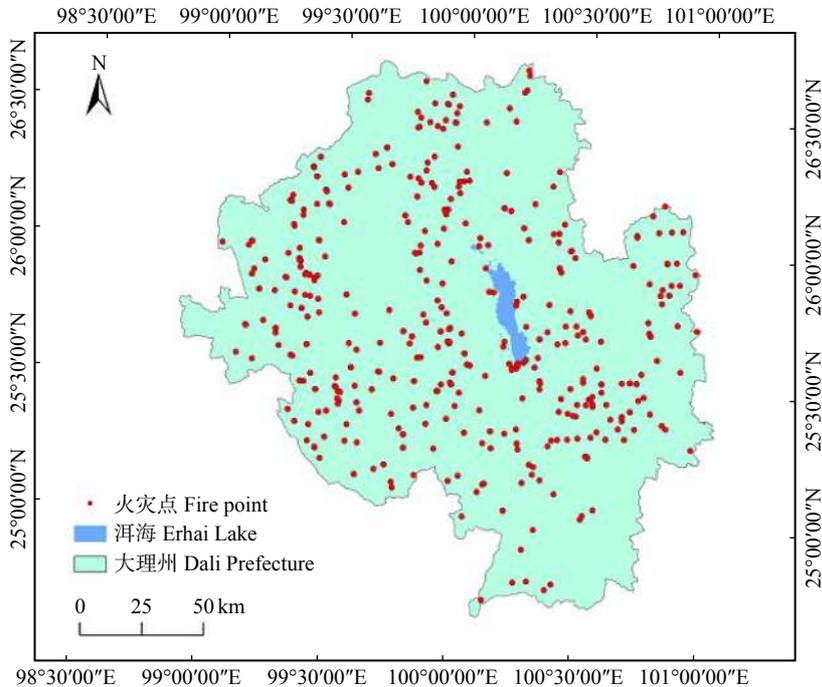


图1 研究区2000—2013年火点分布

Fig. 1 Distribution of fire points in the study area from 2000 to 2013

## 2 数据来源与处理

### 2.1 数据来源

本文森林火灾数据来自云南省大理州森林防火指挥部办公室2000—2013年林火发生情况数据,包括火灾发生时间、起火地点、起火原因、过火面积等。

气象数据来自国家气象科学数据中心基本气象数据,为大理气象站(区站号:56751)2000—2013年1—6月的逐日气象数据。本文通过对气象数据进行预处理,剔除掉缺测数据过多的气象因子,保留11个气象因子:日平均风速(m/s)、日最大风速(m/s)、日照时数(h)、日平均气压(kPa)、日平均气温(℃)、日最高气温(℃)、日最低气温(℃)、日平均水汽压(kPa)、日平均相对湿度(%)、日最小相对湿度(%)、前1日20:00—20:00降雨量(mm)。

另外,考虑到森林火灾的发生不仅与当时的气象条件有关,而且还与前期气象条件密切相关,气象因子对林火发生的影响具有一定滞后性,因此引入加拿大森林火险天气指标系统(fire weather index, FWI)。FWI指标体系以时滞—平衡含水率理论为基础,将气象条件和可燃物含水率有机地联系起来,通过天气条件的变化计算可燃物含水率的变化,进

而确定潜在火险等级,这在一定程度上代表了当前气象条件与前期气象条件的综合效应,因此本文将FWI指标体系看成气象因子,与11个气象因子共同设为自变量,进行森林火灾发生概率模型构建。

### 2.2 数据处理

本文以火灾发生日气象数据为自变量,因变量 $Y=1$ ,同时在构建判别模型时,需要创建一定比例的非火点( $Y=0$ )作为对照。本文按1:1比例选取对照点<sup>[18-19]</sup>,对照非火点创建过程中遵循时间和空间上的完全随机<sup>[20]</sup>。

本文将全样本数据随机分成80%的训练样本和20%的测试样本,基于训练样本建立模型,对测试样本进行预测,通过对比观测值和预测值计算出模型的准确率。本文运用R软件中的glm函数进行二项逻辑斯蒂回归的计算,逐步回归是通过R软件中的step.glm函数实现的,贝叶斯模型平均算法是通过R软件中的BMA、MASS程序包实现的。

## 3 研究方法

### 3.1 林火预测模型选择

逻辑斯蒂回归模型(Logistic regression model, LR)是目前国内外普遍运用的林火预测模型<sup>[18,21-23]</sup>。

将是否发生火灾的概率  $p$  设定为因变量  $Y(Y=1$  为发生林火,  $Y=0$  为未发生林火), 各气象因子设定为自变量  $X$ , 建立 Logistic 回归模型, 模型表达式如下:

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (1)$$

式中:  $p$  为林火发生的概率;  $\beta_i (i=1, 2, \dots, n)$  为模型中影响林火发生的自变量  $X_i (i=1, 2, \dots, n)$  的逻辑斯蒂回归系数,  $n$  为模型自变量的个数。

### 3.2 贝叶斯模型平均法

贝叶斯模型平均法 (Bayesian model averaging, BMA) 是由 Raftery 等<sup>[24]</sup> 提出的一种利用多模型组合进行概率预报的统计方法。BMA 采用贝叶斯公式, 将带有先验信息的未知参数分布与描述对象新信息的似然函数结合, 获得对象的后验分布, 基于后验分布对对象进行统计和推断, 从而不断完善对对象的认识<sup>[14]</sup>。

BMA 是通过估算潜在变量  $X$  的所有可能组合的模型并在需要的组合上构建加权平均来解决问题的。通常情况下, 根据已有的主观信息指定先验分布, 以模型后验概率 (posterior model probabilities, PMP) 为权重对可能的单项模型进行加权平均, 以解释变量的后验包含概率 (posterior inclusion probabilities, PIP) 大小作为选择解释变量的客观标准<sup>[15]</sup>。

在 Logistic 回归模型中, 基于 BMA 方法可得到的每个解释变量系数  $\beta_i$  的后验概率, 即为解释变量的后验包含概率 PIP, 并对参数向量  $\theta(\beta_0, \beta_1, \dots, \beta_n)$  进行相应的统计推断, 可以构建出 BMA-Logistic 回归模型, 具体情况如下。

单个模型的后验概率表达式为:

$$P(M_r|D) = \frac{P(D|M_r)P(M_r)}{\sum_{j=1}^k P(D|M_j)P(M_j)} \quad (2)$$

$$P(D|M_r) = \int P(D|\theta_r, M_r)P(\theta_r|M_r)d\theta_r \quad (3)$$

式中:  $M_r$  代表模型空间  $D$  中的第  $r$  个模型,  $k$  表示模型的个数,  $P(M_r)$  为模型  $M_r$  的先验概率分布,  $P(D|M_r)$  为模型  $M_r$  对应的似然函数积分,  $\theta_r$  表示模型  $M_r$  的参数,  $P(\theta_r|M_r)$  表示模型  $M_r$  所对应的参数先验概率分布,  $P(D|\theta_r, M_r)$  表示模型  $M_r$  所对应的似然函数。

根据贝叶斯公式, 参数  $\theta$  的后验密度分布是模型空间条件下参数  $\theta$  后验密度分布的加权平均值, 权重为模型后验概率  $P(M_r|D)$ 。BMA 还需要指定模型先验概率和参数的先验信息, 即设定  $P(M_r)$ 。一般情况, 可以设定相等的模型先验概率, 即设定均匀分布的模型空间, 而对于参数的先验信息, 可以设定单位信息先验 (unit information prior)。

### 3.3 模型的检验

本文采用 ROC(receiver operating characteristic) 曲线和 AUC(area under curve) 对预测效果进行检验。AUC 作为数值可以直观的评价分类器的好坏, 值越大分类效果越好。一般认为, AUC 值等于 0.5 时相当于一个完全的随机预测, 在 (0.5, 0.7] 区间, 准确性较低; 在 (0.7, 0.8] 区间, 准确性中等; 在 (0.8, 0.9] 区间, 准确性较好; 在 (0.9, 1] 区间, 具有高准确性<sup>[25-26]</sup>。

通过 ROC 曲线分析法可以得到模型的敏感度值和特异性值, 通过约登指数 (Youden index) 公式 (约登指数 = 敏感度值 + 特异性值 - 1) 可以计算分类阈值 (cut-off point), 进而对预测概率进行分类。如果计算得到的预测概率值大于该阈值则认为会发生林火, 小于该阈值则认为不会发生林火<sup>[25,27]</sup>。

## 4 结果与分析

根据大理州 2000—2013 年 1—6 月火点数据, 按 1:1 比例选取对照点, 应用 SPSS 软件对火点和对照点组成的全样本数据进行基本统计 (表 1)。

### 4.1 逐步逻辑斯蒂回归模型分析

#### 4.1.1 多重共线性检验

多重共线性 (multicollinearity) 是指线性回归模型中解释变量之间存在某种密切相关关系。多重共线性是普遍存在的, 通常情况下, 适度的共线性不成问题, 但严重的共线性会导致解释变量的显著性检验失去意义及模型估计产生一定偏差甚至无效。因此在涉及多个解释变量时, 应首先对其进行多重共线性检验。本文运用方差膨胀因子 (variance inflation factor, VIF) 诊断法对解释变量进行多重共线性检验。通常, VIF 值越大, 说明多重共线性就越显著, 一般认为 VIF 大于 10 时, 解释变量之间具有显著的共线性<sup>[28]</sup>。经检验, WIN\_avg、WIN\_max、SSD、PRS\_avg、Tmax、Tmin、RH\_avg、RH\_min、FFMC、ISI 共计 10 个气象因子的 VIF 小于 10 (表 2), 进入模型拟合阶段。

#### 4.1.2 逐步逻辑斯蒂回归模型的拟合

基于训练样本, 运用 R 软件中的 glm 函数对大理州火点和对照点以及 WIN\_avg、WIN\_max、SSD、PRS\_avg、Tmax、Tmin、RH\_avg、RH\_min、FFMC、ISI 共计 10 个气象因子数据进行 Logistic 模型拟合, 通过逐步回归法, 逐步剔除模型中不显著的变量, 得到的最终变量为 WIN\_max、Tmax、Tmin、RH\_avg、RH\_min、FFMC 共计 6 个变量, 选择最终变量进行模型拟合, 来构建 Step\_Logistic (Step\_LR) 模型, 拟合参数如表 3 所示。经逐步回归筛选出的最终变量

表1 模型变量的基本统计描述

Tab. 1 Basic statistical description of model variables

模型变量 Model variable	变量代码 Variable code	最小值 Min. value	最大值 Max. value	均值 Mean	标准差 SD
日平均风速 Daily average wind speed/(m·s <sup>-1</sup> )	WIN_avg	0.80	10.80	3.64	1.38
日最大风速 Daily maximum wind speed/(m·s <sup>-1</sup> )	WIN_max	3.10	20.60	9.21	2.35
日照时数 Sunshine hour/h	SSD	2.20	12.20	9.24	1.80
日平均气压 Daily average pressure/kPa	PRS_avg	79.33	80.75	80.05	0.22
日平均气温 Daily average temperature/°C	Tavg	4.20	24.10	15.88	3.34
日最高气温 Daily maximum temperature/°C	Tmax	12.10	31.00	23.34	3.17
日最低气温 Daily minimum temperature/°C	Tmin	-0.80	18.20	8.59	3.92
日平均水汽压 Daily average water vapor pressure/kPa	VP_avg	0.27	1.68	0.71	0.21
日平均相对湿度 Daily average relative humidity/%	RH_avg	21.00	72.00	41.46	8.27
日最小相对湿度 Daily minimum relative humidity/%	RH_min	6.00	46.00	18.78	5.71
前一日20:00—20:00降雨量 20:00 the day before—20:00 precipitation/mm	Pre	0	3.00	0.03	0.23
细小可燃物湿度码 Fine fuel moisture code	FFMC	79.33	97.56	94.6	1.70
粗腐殖质湿度码 Duff moisture code	DMC	18.18	342.68	113.29	52.55
干旱码 Drought code	DC	61.91	660.71	373.32	98.28
初始蔓延指数 Initial spread index	ISI	1.44	15.63	10.01	1.99
累积指数 Build-up index	BUI	23.92	339.76	128.54	48.63
火险天气指数 Fire weather index	FWI	5.35	49.05	33.47	7.14
火点 Fire point	Fire	0	1	0.50	0.50

注:各模型变量样本数为1 102。Note: sample number of each model variable is 1 102.

表2 变量的多重共线性检验

Tab. 2 Multicollinearity test of variables

变量 Variable	WIN_avg	WIN_max	SSD	PRS_avg	Tmax	Tmin	RH_avg	RH_min	FFMC	ISI
VIF值 VIF value	8.67	1.40	9.13	7.68	6.67	8.12	8.11	3.79	1.95	9.35

表3 Step\_LR 模型参数拟合

Tab. 3 Parameter estimation of Step\_LR model

变量 Variable	估计系数 Estimated coefficient	标准误差 Std error	Z值 Z value	P值 P value
截距 Intercept	-13.006	2.225	-2.923	0.003
WIN_max	0.013	0.003	3.932	0.000
Tmax	0.049	0.004	11.281	0.000
Tmin	-0.030	0.004	-8.448	0.000
RH_avg	-0.129	0.013	-9.903	0.000
RH_min	0.055	0.020	2.733	0.006
FFMC	0.077	0.023	2.091	0.037

与该地区林火发生均有显著相关性,其中 WIN\_max、Tmax、Tmin、RH\_avg、RH\_min 在  $P < 0.01$  水平上极显著相关,FFMC 在  $P < 0.05$  水平上显著相关。

#### 4.1.3 逐步逻辑斯蒂回归模型的检验

应用 ROC 曲线分析法对 Step\_LR 模型的预测能力进行拟合优度检验,并计算林火发生的分类阈

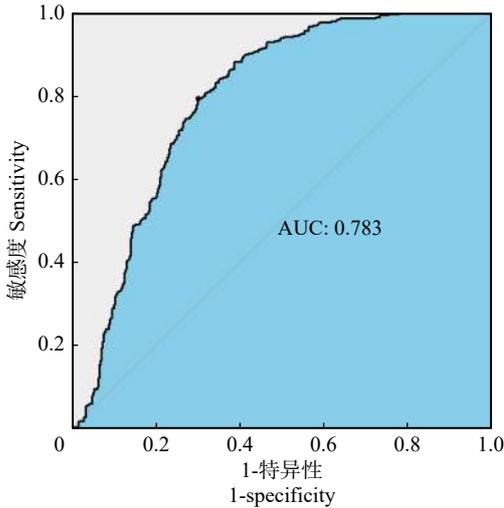
值。图2为预测模型的 ROC 曲线图,ROC 曲线下的面积(AUC 值)为 0.783,显著性水平  $P < 0.001$ ,说明模型的拟合优度为中等水平。根据约登指数公式可得林火发生的分类阈值为 0.498,以该值为分界点,林火发生的预测概率值大于 0.498 视为有林火发生,小于 0.498 则视为无林火发生,进一步计算得到 20% 测试数据集对林火发生的预测概率为 0.718。

## 4.2 贝叶斯平均模型分析

### 4.2.1 BMA 模型的拟合

本文共有 17 个解释变量,利用各解释变量建立林火发生的 Logistic 模型,可能存在的模型个数高达  $2^{17}$  个,贝叶斯模型平均法可遍历模型空间中的每一个模型,根据各模型的后验概率来衡量其对林火发生的相对重要性,通过模型空间调整以确定较优模型。将 Fire(1 和 0)设定为因变量  $Y$ ,各气象因子设定为自变量  $X$ ,建立 BMA 模型,使用 R 程序的“BMA package”进行计算。

本文设定相等的模型先验概率即均匀分布的模



AUC 为曲线下面积。下同。AUC, area under curve. The same below.

图 2 Step\_LR 模型的 ROC 曲线图  
Fig. 2 ROC curve of Step\_LR model

型空间,对于参数的先验信息,设定为单位信息先验;同时,采用了奥卡姆窗(Occam’s window)的方法来适当调整模型空间,即减少一定的模型数量,本文设定当一个模型的后验概率(PMP)小于最佳模型后验概率的 5% 时,则从模型空间中被剔除。程序结果如图 3 所示,奥卡姆窗筛选了 98 个较优模型,每个模型包含了部分变量。

#### 4.2.2 贝叶斯平均模型的构建

本文基于贝叶斯估计的后验概率对 98 个较优模型做了贝叶斯模型平均,程序结果见表 4。

由表 4 可知, Tmax 的后验包含概率为 98.0%, VP\_avg 的后验包含概率为 95.6%, BUI 的后验包含概率为 70.0%, 这 3 个变量的后验包含概率较大,意

式中:  $p$  为林火发生的概率;  $X_1$  为日平均气压(kPa);  $X_2$  为日平均气温(°C);  $X_3$  为日最高气温(°C);  $X_4$  为日最低气温(°C);  $X_5$  为日平均水汽压(kPa);  $X_6$  为日平均相对湿度(%);  $X_7$  为日最小相对湿度(%);  $X_8$  为累积指数。

#### 4.2.3 贝叶斯平均模型的检验

应用 ROC 曲线分析法对 BMA\_LR 模型的预测能力进行拟合优度检验,并计算林火发生的分类阈值。图 4 为预测模型的 ROC 曲线图,且 ROC 曲线下的面积(AUC 值)为 0.868,显著性水平  $P < 0.001$ ,说明模型的拟合优度较高。根据约登指数公式可得林火发生的分类阈值为 0.562,以该值为分界点,林火发生的预测概率值大于 0.562 视为有林火发生,小于 0.562 则视为无林火发生。进一步计算得到 20% 测试数据集对林火发生的预测概率为 0.807,结果显

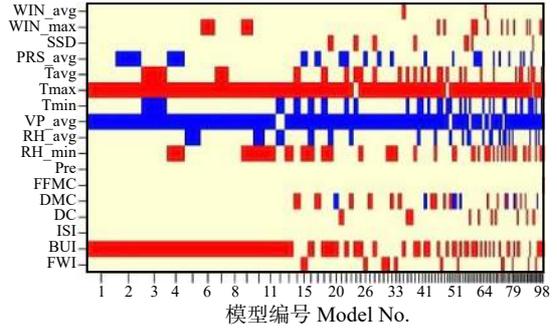


图 3 BMA 模型可视化  
Fig. 3 BMA model visualization

味着这 3 个变量对林火发生有较大影响。其他解释变量的后验包含概率相对较小,说明这些解释变量对林火发生的影响力相对减弱。而从模型的稳定性角度看,最佳模型(model 1)的后验概率仅为 0.062,前 5 个模型的累计后验概率为 0.249,由此可见,模型的不确定性在该数据集中是相当大的。

本文以 5 个最优模型的后验概率作为权重进行加权,来构建 BMA\_Logistic(BMA\_LR)组合模型。加权整合后的平均模型为:

$$P = \frac{1}{1 + e^{-(7.233 - 0.001X_1 + 0.003X_2 + 0.012X_3 - 0.001X_4 - 0.020X_5 - 0.001X_6 + 0.002X_7 + 0.002X_8)}} \quad (4)$$

示模型具有较高的预测能力,可用于大理州林火发生的预测预报。

#### 4.3 不同算法的对比分析

##### 4.3.1 最终指标体系

Step\_LR 模型和 BMA\_LR 模型的最终指标体系对比如表 5 所示。在 BMA\_LR 模型中, Tmax、VP\_avg 和 BUI 3 个变量的后验包含概率最大,这 3 个变量只有 Tmax 进入了 Step\_LR 模型的指标体系。

##### 4.3.2 模型预测准确率

根据模型变量选择结果,分别对 Step\_LR 模型和 BMA\_LR 模型预测准确率进行计算。在训练集中,BMA\_LR 模型的预测准确率比 Step\_LR 模型高 9.3%,在测试集中,BMA\_LR 模型的预测准确率比 Step\_LR 模型高 8.9%,结果显示 BMA\_LR 模型具

表4 基于贝叶斯后验概率的模型平均

Tab. 4 Model average based on Bayesian posterior probability

Variable	p!=0	SD	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	100	45.070	-6.028	82.170	-5.490	86.100	-3.648
WIN_avg	1.3	0.002					
WIN_max	11.5	0.003					
SSD	5.9	0.004					
PRS_avg	23.4	0.005		-0.011		-0.012	
Tavg	26.8	0.021			0.048		
Tmax	98.0	0.015	0.054	0.054	0.028	0.063	0.046
Tmin	26.6	0.011			-0.024		
VP_avg	95.6	0.023	-0.081	-0.082	-0.077	-0.102	-0.064
RH_avg	19.7	0.036					-0.038
RH_min	33.9	0.033				0.061	
Pre	0.0	0					
FFMC	0.3	0.004					
DMC	16.6	0.005					
DC	6.0	0.001					
ISI	0.0	0					
BUI	70.0	0.006	0.008	0.007	0.007	0.007	0.007
FWI	8.3	0.013					
nVar			3	4	5	5	4
BIC			-7 039	-7 038	-7 038	-7 038	-7 037
post prob			0.062	0.057	0.056	0.039	0.035

注: 本表为程序输出表格, 其中Variable表示变量, p!=0为变量回归系数不为零的后验概率, SD为标准差, model 1~model 5为BMA筛选的后验概率最大的5个模型, Intercept为截距项, 从WIN\_avg至FWI为各变量代码, 参考表1, nVar为模型选中的变量数, BIC为贝叶斯信息量, post prob为模型后验概率。Notes: this table is the program output table, where Variables represents the model variables, P!=0 is the posterior probability that the regression coefficient of the variable is not zero; SD is the standard deviation; model 1-model 5 are the 5 models with the largest posterior probability screened by BMA; Intercept is the intercept item, and from Win\_avg to FWI is the variable code, as shown in Tab. 1. nVar is the number of variables selected by the model, BIC is the Bayesian information criterion, and post prob is the posterior probability of the model.

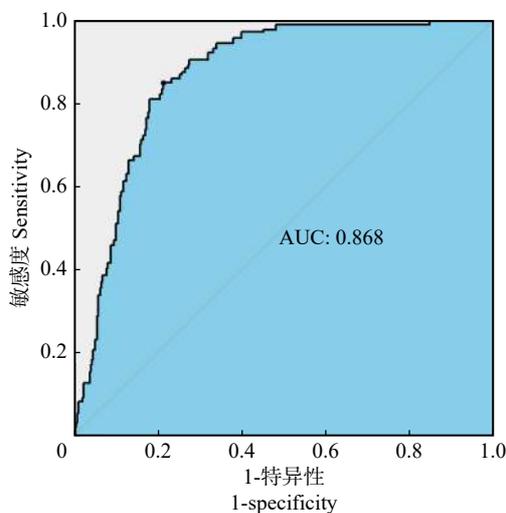


图4 BMA\_LR模型的ROC曲线图  
Fig. 4 ROC curve of BMA\_LR model

有较高的预测能力, 可适用于云南省大理州林火预测。

## 5 结论与讨论

本文分别应用二项逻辑斯蒂回归模型和贝叶斯模型平均法建立了云南省大理州林火发生的单一模型和组合模型。比较模型拟合结果表明, 贝叶斯平均模型的预测准确率比逐步逻辑斯蒂回归模型高8.9%, 说明基于贝叶斯模型平均法的组合模型林火预测效果优于逐步逻辑斯蒂回归模型, 可用于该地区的林火发生预测预报。

本文构建的贝叶斯平均模型包含日平均气压、日平均气温、日最高气温、日最低气温、日平均水汽压、日平均相对湿度、日最小相对湿度、累积指数共计8个气象因子, 表明这8个气象因素是影响该地

表 5 Step\_LR 模型和 BMA\_LR 模型中最终指标体系及预测准确率

Tab. 5 Final indicator system and prediction accuracy in the Step\_LR and BMA\_LR model

模型 Model	模型指标体系 Model index system	预测准确率 Prediction accuracy/%	
		训练集 Training sample (80%)	测试集 Test sample (20%)
Step_LR	WIN_max, Tmax, Tmin, RH_min, RH_avg, FFMC	73.3	71.8
BMA_LR	PRS_avg, Tavg, Tmax, Tmin, VP_avg, RH_avg, RH_min, BUI	82.6	80.7

区林火发生的重要驱动气象因子,特别是日平均水汽压、日最高气温和累积指数后验包含概率均在 70% 以上,表明这 3 个气象因子是影响该地区林火发生的主要驱动气象因子。

在构建逻辑斯蒂回归模型时,首先进行了多重共线性检验,剔除了日平均气温、日平均水汽压、前 1 日 20:00—20:00 降雨量、干旱码、粗腐殖质湿度码、累积指数、火险天气指数共计 7 个气象因子。在这 7 个被剔除的气象因子中,日平均气温、日平均水汽压、累积指数这 3 个气象因子则进入了最终贝叶斯平均模型,可见进行多重共线性检验存在将重要驱动因子提前去除的风险。这是因为进行多重共线性检验考虑的是解释变量之间的相关性,而并未考虑被剔除的变量对被解释变量的影响。因此,提前对解释变量进行多重共线性检验,并根据检验结果对存在多重共线性的变量进行剔除,则有可能剔除了对林火发生有显著影响的气象因子,这就会造成预测准确率的降低。

本文通过分析森林火灾对气象因子的响应讨论了贝叶斯模型平均法对林火发生气象因子的选择及林火预测的适用性,实际上,自然环境中影响林火发生的生态因子和非生态因子很多。已有研究表明,森林火灾发生还与植被类型、林型、地形、人为活动、社会经济<sup>[23,28-31]</sup>等多个因素息息相关,因此若想更加及时有效地预测森林火灾的发生,还需进一步分析上述因素中的重要驱动因子及其对林火发生的影响。今后研究中应纳入以上变量,筛选关键驱动因子,建立组合模型,进一步提高预测效果。

#### 参 考 文 献

- [1] Rigo D D, Giorgio L, Durrant T H, et al. Forest fire danger extremes in Europe under climate change: variability and uncertainty[M]. Luxembourg: Publications Office of the European Union, 2017.
- [2] 田晓瑞,宗学政,舒立福,等. ENSO 事件对中国森林火险天气的影响[J]. 应用生态学报, 2020, 31(5): 65-73.  
Tian X R, Zong X Z, Shu L F, et al. Impacts of ENSO events on forest fire weather of China[J]. Chinese Journal of Applied Ecology, 2020, 31(5): 65-73.
- [3] 白夜,武英达,贾宜松,等. 2019—2020 年澳大利亚气候异常与山火爆发的关系分析及应对策略[J]. 中国应急救援, 2020(2): 23-27.
- Bai Y, Wu Y D, Jia Y S, et al. Link between climate anomaly and Australia bushfires in 2019-2020[J]. China Emergency Rescue, 2020(2): 23-27.
- [4] 赵凤君,舒立福. 森林草原火灾扑救安全学[M]. 北京: 中国林业出版社, 2015.  
Zhao F J, Shu L F. Forest and grassland fire fighting safety[M]. Beijing: China Forestry Publishing House, 2015.
- [5] 岳超,罗彩访,舒立福,等. 全球变化背景下野火研究进展[J]. 生态学报, 2020, 40(2): 385-401.  
Yue C, Luo C F, Shu L F, et al. A review on wildfire studies in the context of global change[J]. Acta Ecologica Sinica, 2020, 40(2): 385-401.
- [6] Marlon J R, Bartlein P J, Gavin D G, et al. Long-term perspective on wildfires in the western USA[J]. Proceedings of the National Academy of Sciences of the United States of America, 2012, 109(9): 3203-3204.
- [7] Westerling A L. Increasing western US forest wildfire activity: sensitivity to changes in the timing of spring[J]. Philosophical Transactions of the Royal Society of London (Series B): Biological Sciences, 2016, 371: 1-10.
- [8] 潘登,郁培义,吴强. 基于气象因子的随机森林算法在湘中丘陵区林火预测中的应用[J]. 西北林学院学报, 2018, 33(3): 175-183.  
Pan D, Yu P Y, Wu Q. Application of random forest algorithm on the forest fire prediction based on meteorological factors in the hilly area, central Hunan Province[J]. Journal of Northwest Forestry University, 2018, 33(3): 175-183.
- [9] North M P, Stephens S L, Collins B M, et al. Reform forest fire management[J]. Science, 2015, 349: 1280-1281.
- [10] Fischer A P, Spies T A, Steelman T A, et al. Wildfire risk as a socioecological pathology[J]. Frontiers in Ecology and the Environment, 2016, 14(5): 276-284.
- [11] Zhang G, Wang M, Liu K. Forest fire susceptibility modeling using a convolutional neural network for Yunnan Province of China[J]. International Journal of Disaster Risk Science, 2019, 10(3): 386-403.
- [12] Murphy T E, Tsang S W, Leo L S, et al. Bayesian model averaging for selection of a risk prediction model for death within thirty days of discharge: the silver-ami study[J]. International Journal of Statistics in Medical Research, 2019, 8: 1-7.
- [13] Huang H, Liang Z, Li B, et al. Combination of multiple data-driven models for long-term monthly runoff predictions based on Bayesian model averaging[J]. Water Resources Management,

- 2019, 33(9): 3321–3338.
- [14] 王倩, 师鹏飞, 宋培兵, 等. 基于贝叶斯模型平均法的洪水集合概率预报[J]. 水电能源科学, 2016(6): 64–66.  
Wang Q, Shi P F, Song P B, et al. Multi-model ensemble flood probability forecasting based on BMA[J]. *Water Resources and Power*, 2016(6): 64–66.
- [15] 张畅, 陈新军. 海洋环境因子对澳洲鲈亲体补充量关系的影响: 基于贝叶斯模型平均法的研究[J]. 海洋学报, 2019, 41(2): 99–106.  
Zhang C, Chen X J. The effect of environmental factors on stock-recruitment relationship of spotted mackerel-based on Bayesian model averaging method[J]. *Haiyang Xuebao*, 2019, 41(2): 99–106.
- [16] 李丽琴. 云南省森林火灾发生与气象因子之间的关系研究[D]. 北京: 北京林业大学, 2010.  
Li L Q. Study on the relationship between forest fires and the meteorological factors in Yunnan[D]. Beijing: Beijing Forestry University, 2010.
- [17] 周明昆, 王永平, 高月忠. 气象因子对云南大理森林火灾的影响[J]. 四川林业科技, 2012, 33(6): 96–99.  
Zhou M K, Wang Y P, Gao Y Z. Effects of meteorological factors on forest fires in Dali, Yunnan[J]. *Journal of Sichuan Forestry Science and Technology*, 2012, 33(6): 96–99.
- [18] Martell D L, Otukol S, Stocks B J. A logistic model for predicting daily people-caused forest fire occurrence in Ontario[J]. *Canadian Journal of Forest Research*, 1987, 17(5): 394–401.
- [19] 苏漳文, 刘爱琴, 郭福涛, 等. 福建林火发生的驱动因子及空间格局分析[J]. 自然灾害学报, 2016, 25(2): 110–119.  
Su Z W, Liu A Q, Guo F T, et al. Driving factors and spatial distribution pattern of forest fire in Fujian Province[J]. *Journal of Natural Disasters*, 2016, 25(2): 110–119.
- [20] 于建龙, 刘乃安. 我国大兴安岭地区森林雷击火发生的火险天气等级研究[J]. 火灾科学, 2010, 19(3): 131–137.  
Yu J L, Liu N A. Lightning-caused wildland fire weather danger rating in Daxing'anling region[J]. *Fire Safety Science*, 2010, 19(3): 131–137.
- [21] Bisquert M, Caselles E, Sánchez J M, et al. Application of artificial neural networks and logistic regression to the prediction of forest fire danger in Galicia using MODIS data[J]. *International Journal of Wildland Fire*, 2012, 21(8): 1025–1029.
- [22] Oliveira S, Oehler F, San-Miguel-Ayanz J, et al. Modeling spatial patterns of fire occurrence in Mediterranean Europe using multiple regression and random forest[J]. *Forest Ecology and Management*, 2012, 275(4): 117–129.
- [23] 陈岱. 基于 Logistic 回归模型的大兴安岭林火预测研究[J]. 林业资源管理, 2019(1): 116–122.  
Chen D. Prediction of forest fire occurrence in Daxing'an Mountains based on logistic regression model[J]. *Forest Resources Management*, 2019(1): 116–122.
- [24] Raftery A E, Gneiting T, Balabdaoui F, et al. Using Bayesian model averaging to calibrate forecast ensembles[J]. *Monthly Weather Review*, 2005, 133(5): 1155–1174.
- [25] 梁慧玲, 林玉蕊, 杨光, 等. 基于气象因子的随机森林算法在塔河地区林火预测中的应用[J]. 林业科学, 2016, 52(1): 89–98.  
Liang H L, Lin Y R, Yang G, et al. Application of random forest algorithm on the forest fire prediction in Tahe Area based on meteorological factors[J]. *Scientia Silvae Sinicae*, 2016, 52(1): 89–98.
- [26] 顾先丽, 吴志伟, 张宇婧, 等. 气候变化背景下江西省林火空间预测[J]. 生态学报, 2020, 40(2): 667–677.  
Gu X L, Wu Z W, Zhang Y J, et al. Prediction research of the forest fire in Jiangxi Province in the background of climate change[J]. *Acta Ecologica Sinica*, 2020, 40(2): 667–677.
- [27] Chang Y, Zhu Z L, Bu R C, et al. Predicting fire occurrence patterns with logistic regression in Heilongjiang Province, China[J]. *Landscape Ecology*, 2013, 28(10): 1989–2004.
- [28] Guo F T, Su Z W, Wang G Y, et al. Understanding fire drivers and relative impacts in different Chinese forest ecosystems[J]. *Science of the Total Environment*, 2017, 605: 411–425.
- [29] Flannigan M D, Krawchuk M A, Groot W J D, et al. Implications of changing climate for global wildland fire[J]. *International Journal of Wildland Fire*, 2009, 18(5): 483–507.
- [30] Loepfe L, Rodrigo A, Lloret F. Two thresholds determine climatic control of forest fire size in Europe and northern Africa[J]. *Regional Environmental Change*, 2014, 14(4): 1395–1404.
- [31] 蔡奇均, 曾爱聪, 苏漳文, 等. 基于 Logistic 回归模型的浙江省林火发生驱动因子分析[J]. 西北农林科技大学学报, 2020, 48(2): 108–115.  
Cai Q J, Zeng A C, Su Z W, et al. Driving factors of forest fire in Zhejiang Province based on logistic regression model[J]. *Journal of Northwest A&F University*, 2020, 48(2): 108–115.

(责任编辑 赵 勃  
责任编委 舒立福)