



基于Gompit回归模型的大兴安岭林火预测模型及驱动因子研究

苏漳文, 曾爱聪, 蔡奇均, 胡海清

引用本文:

苏漳文, 曾爱聪, 蔡奇均, 等. 基于Gompit回归模型的大兴安岭林火预测模型及驱动因子研究[J]. 林业工程学报, 2019, 4(4): 135–142.

在线阅读 View online: <https://doi.org/10.13360/j.issn.2096-1359.2019.04.020>

您可能感兴趣的其他文章

Articles you may be interested in

大兴安岭森林火灾NO和SO₂释放量的估算

Estimation of NO and SO₂ emission by forest fires in the Da Hinggan Mountains

林业工程学报. 2015, 29(5): 134–138 <https://doi.org/10.13360/j.issn.1000-8101.2015.05.033>

基于环境卫星数据的森林叶面积指数遥感反演与验证——以大兴安岭加格达奇林区为例

Retrieval and validation of forest leaf area index based on HJ satellite data: taking Jiagadaqi District as a case study

林业工程学报. 2015, 29(4): 126–130 <https://doi.org/10.13360/j.issn.1000-8101.2015.04.031>

基于卫星监测热点的林火分布及发生预报模型

Research on spatial distribution of forest fire based on satellite hotspots data and forecasting model

林业工程学报. 2017, 2(4): 128–133 <https://doi.org/10.13360/j.issn.2096-1359.2017.04.021>

基于气象和时空因子的森林火险等级预测

Prediction of forest fire danger rating based on meteorological and space-time factors

林业工程学报. 2018, 3(3): 102–110 <https://doi.org/10.13360/j.issn.2096-1359.2018.03.017>

广义线性模型在林火发生预报中的应用

Research on generalized linear models applied to forest fire forecast

林业工程学报. 2017, 2(1): 135–142 <https://doi.org/10.13360/j.issn.2096-1359.2017.01.024>

基于 Gompit 回归模型的大兴安岭林火 预测模型及驱动因子研究

苏漳文¹, 曾爱聪², 蔡奇均², 胡海清^{1*}

(1. 东北林业大学林学院, 哈尔滨 150040; 2. 福建农林大学林学院, 福州 350002)

摘要:本研究基于 2000—2016 年林火数据, 选取气象、地形、植被、人为活动等因素作为林火预测变量, 采用 Gompit 回归模型对林火发生的主要驱动因子进行分析, 并建立大兴安岭地区林火发生预测模型。结果表明: 大兴安岭地区林火受气象因素(日累计降水、日平均相对湿度)的影响最大且与林火发生均呈显著负相关; 此外, 大兴安岭林火多发生于缓坡、远离居民区、铁路、公路等人为活动较为频繁地区。模型结果表明: Gompit 回归模型的预测效果较好(准确率 77%), ROC 检验结果表明模型的拟合度较高(效果值为 0.868); 而独立样本的检验显示, 预测准确率为 75.3%, 模型具有较高的适用性。大兴安岭近 17 年的火险等级总体呈南高北低、东高西低的地理分布, 其中高火险和中火险区主要集中在南部、东南部等地, 占整个研究区域的 24.2%; 同时南部和东南部存在大面积低估区, 表明模型对这些地区的预测能力不高。

关键词:大兴安岭; Gompit 回归; 林火预测模型; 驱动因子

中图分类号: S762.2 文献标志码: A 文章编号: 2096-1359(2019)04-0135-08

Study on prediction model and driving factors of forest fire in Da Hinggan Mountains using Gompit regression method

SU Zhangwen¹, ZENG Aicong², CAI Qijun², HU Haiqing^{1*}

(1. College of Forestry, Northeast Forestry University, Harbin 150040, China; 2. College of Forestry,
Fujian Agriculture and Forestry University, Fuzhou 350002, China)

Abstract: In this study, the Gompit regression method was applied to analyze the driving factors of forest fires in Da Hinggan Mountains and establish the forest fire prediction model based on forest fire data from 2001 to 2016 and meteorological, topographic, vegetation, human activity factors that are associated with the occurrence of forest fires in the region. The results of investigation showed that the meteorological factors (daily cumulative precipitation and daily average relative humidity) had the most powerful and important influence on forest fire occurrence. The daily cumulative precipitation and daily average relative humidity were negatively related to the fire occurrence in Da Hinggan Mountains. Multitude of forest fires occurred in the areas with slow slope and far away from residential regions, railways, roads and other human activities. The results of model fitting showed that the Gompit regression method had a good prediction accuracy of 77%, and the area under the ROC curve (AUC) derived from the ROC test was 0.868, indicating a high-level goodness of fit for the Gompit regression method. Same as the modeling result, the result of validation test also indicated a high prediction accuracy (75.3%) and good applicability of the Gompit regression method for fire prediction in Da Hinggan Mountains. In general, the geographical distribution of the fire risk level was high in the South and East parts, low in the North and West parts of Da Hinggan Mountains from 2001 to 2016. The high and medium fire risk level areas were mainly concentrated in the south and southeast of study region, accounting for 24.2% of the entire Da Hinggan Mountains. At the same time, the result of model residual examination revealed that multitude of underestimated areas existed in the southern and southeastern Da Hinggan Mountains, which indicated a weak prediction ability of Gompit regression method employed in these areas.

Keywords: Da Hinggan Mountains; Gompit regression; the prediction model of forest fire; importance of driving factor

林火是影响森林更新和演替的一个重要生态因子, 同时也会对森林生态系统造成破坏^[1]。据

统计,每年约有 5 000~12 000 次林火发生在加拿大,年均过火面积 100~300 万 hm^2 ;我国每年发生林火约 10 000 次,年均过火面积约 82 万 hm^2 。林火的预防预报一直是我国林业管理工作的重点。大兴安岭地区是我国最重要的林业基地,也是林火发生的重灾区^[3]。了解该地区的林火发生特性、建立林火发生预测模型已经成为当地林火管理的关键。

了解林火发生的驱动因子是建立林火预测模型的基础。近年来国内外学者对林火驱动因子进行了广泛的研究,取得了大量的成果^[4-8]。研究表明,气象、植被、地形等环境因素在林火发生和蔓延中起到了重要的作用,而人为活动和社会经济条件也对林火发生具有显著的影响^[6-9]。因此,对林火及其驱动因子的综合分析一直是森林生态、林火预测的研究重点,也是林火管理的重要主题。

Logistic 回归模型一直以来都是国内外林火研究人员广泛使用的预测模型^[4-5,7]。Logistic 回归模型是基于概率的对称累积分布,即数据中因变量 y 的分布是对称的^[10]。然而实际应用中,林火数据常为非对称结构,即未发生林火的情况远大于林火发生。而二项 Gompit 回归模型是基于概率的非对称累积分布(Gompertz 分布),因变量 y 的分布是不对称的^[10-11],当数据中非对称结构明显时,建议使用 Gompit 回归模型^[10]。目前,该模型被广泛应用于生物学、医学、农业科学、信息科学、社会科学等多个学科领域^[11],但在林火研究领域还鲜有报道。本研究运用 Gompit 回归模型建立林火发生概率模型,判断预测变量的相对重要性,并根据预测结果及模型分类阈值划分大兴安岭地区的火险等级。研究结果将为了了解大兴安岭地区林火发生特性及主要影响因子提供重要的信息,并有助于优化当地的林火管理及防火资源的合理配置。

1 研究区概况

研究区域位于中国东北大兴安岭(50°10'~53°33'N,121°12'~127°00'E),是全球北方针叶林生物群落的最南端,总面积达 $8.46 \times 10^6 \text{ hm}^2$ 。该地区属寒温带,年均气温 -2~4 °C,年均降水量 460 mm^[5]。大兴安岭森林覆盖率高达 74.1%,拥有大面积的天然林区,是我国最重要的林业基地之一。主要树种包括兴安落叶松(*Larix gmelinii* (Rupr.) Kuzen.)、樟子松(*Pinus sylvestris* L. var. *mongolica* Litv.)、白桦(*Betula platyphylla* Suk.)、山杨(*Populus davidiana*)和蒙古栎(*Quercus mongolica* Fisch. ex

Ledeb)。干燥的气候条件和丰富的可燃物使该地区面临极高的火灾风险,是中国年平均过火面积最大的区域。据统计,2004—2014 年期间,大兴安岭共发生森林火灾 240 起,过火面积高达 385 725 hm^2 ,其中雷击火占 75.5%(数据来源于中国林业统计年鉴)^[5]。

2 数据来源与处理

研究中使用的因变量为二分类变量:火点数据($Y=1$)和随机点数据($Y=0$);自变量包括:地形、植被、气象、人为基础设施、社会经济等因素。

2.1 火点提取和生成随机点

本研究使用的卫星火点数据来源于每日火产品 MOD14A1(空间分辨率为 1 km)。MOD14A1 是 MODIS 火产品中的一种,MODIS 是搭载在 terra 和 aqua 卫星上的一个重要传感器,近年来在林火的多数研究中得到了广泛的应用^[12]。但是由于卫星无法分辨火灾的类型,因此该产品也包括了一些城市和建设用地火、农田火等非森林火。基于此,本研究结合 1 km 分辨率的土地利用类型图,剔除落在城市和建设用地、农田的火点。此外,从时间的角度,将火点数据限制在防火期内,提取防火期内的火点及其信息。大兴安岭有两个防火期:春季防火期为 3 月 15 日到 7 月 15 日,秋季防火期为 9 月 15 日到 11 月 15 日。数据显示 2000—2016 年大兴安岭共 6 444 个火点(图 1a)。

本研究使用 Gompit 回归模型解释大兴安岭林火发生及其驱动因子,并建立林火预测模型。由于模型是二项式模型,因此,将已知火点赋值为 1,同时使用 Arcgis10.2 在研究区域内创建随机点赋值为 0^[4-5,13]。为了避免创建的随机点与已知火点重合或临近,设置任意两个随机点的最小允许距离为 1 km;为了保证充分的随机,研究中创建了 4 倍于火点的随机点并重复 10 次,同样剔除落在城市等建设用地和农田的随机点,再从中分别取火点的 1,2,3,4 倍作为模型因变量的随机点(点个数分别为 6 444,12 888,19 332,25 776)。

2.2 气象因素

气象因素是较为重要的林火驱动因子,在早期的研究中就被用于预报森林火险。如日相对湿度低于 55%时,随湿度的降低,林火越容易发生^[14]。日累积降水直接影响可燃物含水率,一般情况下降水量在 2~5 mm 时,可燃物的燃烧性就会降低;此外,降水也会使空气中的相对湿度升高^[8];温度与林火发生的关系也十分密切,通过影响可燃物自身

温度进而影响森林火险的变化^[7];风是决定林火发生和林火蔓延的重要因素,有研究发现,风能不断补充氧气,增加助燃条件,加速燃烧过程^[15]。本研究中日气象数据来源于中国气象数据网(<http://data.cma.cn/>)^[4],从中选取4个气象参数参与模型拟合,包括:日平均相对湿度、日累计降水、日最高气温、日平均风速。先运用 ArcGIS 计算火点和随机点到最近气象站的距离;再通过 Excel 提取与火点和随机

点的信息(坐标和时间)相对应的气象数据。

2.3 地形因素

本研究基于地理空间数据云提供的 25 m 高分辨率数字高程模型(DEM)(图 1c),运用 Arcgis10.2 软件对 DEM 进行坡度计算(图 1d),并从 DEM 和坡度栅格数据中提取与火点和随机点相应的高程和坡度信息^[4-5,13]。

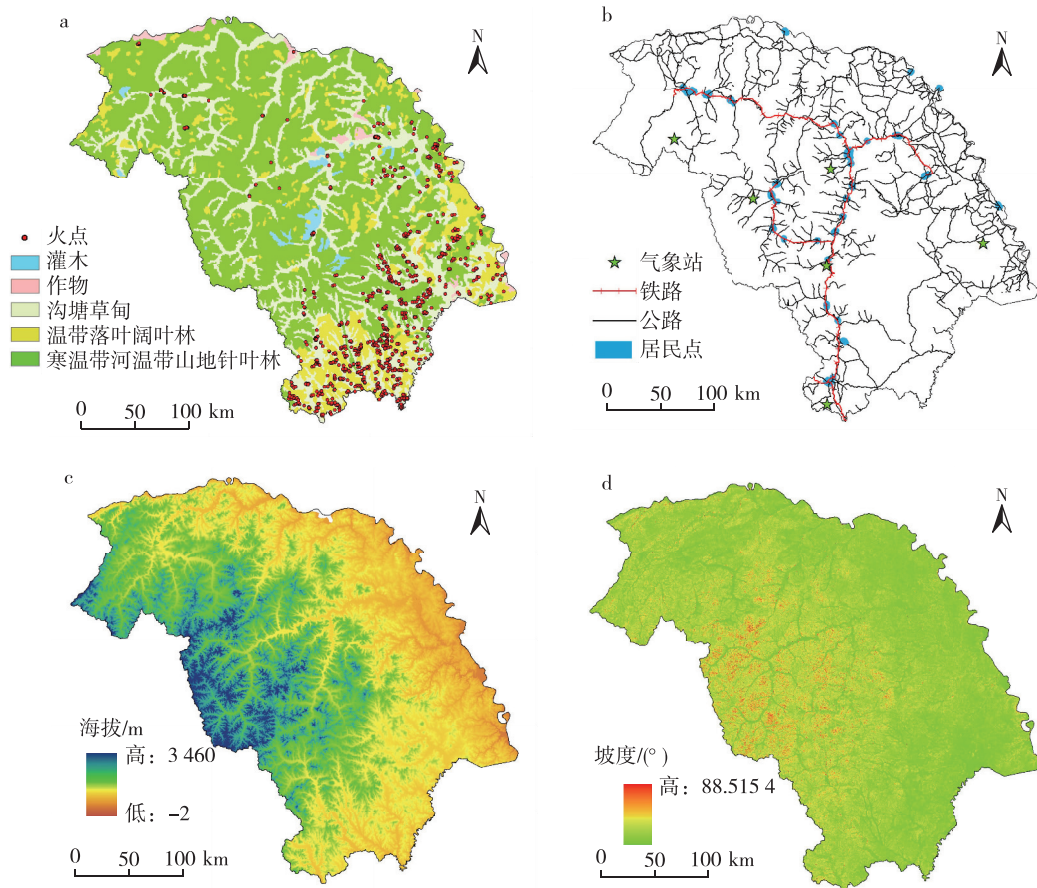


图1 研究地区火点分布和植被类型(a)、基础设施及气象站点(b)、海拔(c)、坡度(d)
Fig. 1 Fire point and vegetation type (a), infrastructure and meteorological stations (b), elevation (c), slope (d) throughout the study area

2.4 植被因素

植被覆盖度(fractional vegetation cover, FVC)反映了地表植被状况,通常指一定区域范围内植被的叶、茎、枝在地面垂直投影面积的百分比。植被覆盖度可以有效地表示地表活或死可燃物的总量。由于 GIS 和 RS 技术的推广,遥感测量植被覆盖度被广泛地使用,利用植被指数近似估算植被覆盖度已成为较为实用的方法,常用的植被指数为归一化植被指数(normalized difference vegetation index, NDVI)^[16]。该数据来源于地理空间数据云(<http://www.gscloud.cn/>)提供的 MODIS 植被归一化指数(NDVI),其空间分辨率为 500 m^[4-5]。计算

公式如下:

$$F_v = (I - I_{soil}) / (I_{veg} - I_{soil}) \quad (1)$$

式中: F_v 为植被覆盖度 FVC; I 为归一化植被指数 NDVI; I_{soil} 为完全是裸土或无植被覆盖像元的 NDVI 值; I_{veg} 为完全被植被所覆盖像元的 NDVI 值。

2.5 人为基础设施

基础设施数据为 1:25 万基础设施矢量图,来源于国家基础地理信息中心^[4-5]。使用 Arcgis10.2 计算火点和随机点到所有基础设施的最短距离,即点到铁路、公路、居民区等基础设施的最短距离(图 1b)。

2.6 社会经济因素

社会经济数据主要是人口密度和人均 GDP,来源于中国科学院资源环境科学数据中心(<http://www.resdc.cn/DOI>)2010年1 km空间分辨率的网格

数据^[4-5]。根据国家统计年鉴计算2000—2016年每年人口和GDP增长率,使用Arcgis10.2中的栅格计算器工具计算2000—2016年的人口和GDP栅格数据。上述数据来源及具体描述见表1。

表1 参与模型拟合的自变量概况

Table 1 Overview of the independent variables involved in model fitting

变量类型	变量名称	代码	尺度	数据来源
气象(防火期)	日平均风速	DW	0.1 m/s	中国气象数据网(http://data.cma.cn/) ^[4]
	日累计降水量	DP	0.1 mm	
	日最高气温	DMT	0.1 °C	
	日平均相对湿度	DR	%	
地形	高程	ELEV	1 m	地理空间数据云 ^[4-5,13]
	坡度	SLOPE	1°	
植被	植被覆盖度	FVC	%	地理空间数据云 ^[4-5,13]
基础设施	到铁路的距离	DRAIL	1 m	国家基础地理信息中心 ^[4-5]
	到公路的距离	DROAD	1 m	
	到居民点的距离	DRESID	1 m	
社会经济因素	人均GDP	GDP	元/km ²	资源环境数据云平台 ^[4-5,17]
	人口密度	POP	人/km ²	

3 研究方法

3.1 模型方法

Gompit 回归模型(又称 complementary log-log 模型)用于描述二分类变量和一个或多个解释变量之间的关系。Gompit 回归服从累积分布函数(Gompertz 分布),其累积概率函数为:

$$P(Y=1) = F(z) = F(x\beta) = 1 - \frac{e^{-e^z}}{1 - e^{-e^z}} \quad (2)$$

式中: $P(Y=1)$ 表示对于给定的 x 相应个体选择1的概率; $F(z)$ 是极值分布函数; Y 通过极值分布函数被转换为概率。

$$\ln[-\ln(1-P)] = z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \quad (3)$$

式中: P 为林火发生的概率; m 为解释变量的个数; $\beta_1, \beta_2, \dots, \beta_m$ 为各个解释变量的相关系数; x_1, x_2, \dots, x_m 为影响林火发生的解释变量; z 与若干个预测变量 x 呈线性关系^[10-11]。

将原始数据随机划分为80%作为建模数据集,20%作为独立检验数据集。为了减少训练样本的分布情况对实验结果的影响,本研究中将建模数据集随机分成75%作为训练样本,剩下的25%作为测试样本,并重复10次随机划分。从而得到10组不同的样本集和一组建模数据集及其对应的独立验证数据集^[18]。

3.2 预测变量选择

3.2.1 多重共线性检验

在线性回归模型中,解释变量间通常存在高度

的相关性,这种关系会给模型的估计值带来一定的误差,这种关系称为多重共线性。本研究使用多重共线性检验中的方差膨胀因子(variance inflation factor, VIF)对模型的解释变量进行检验,排除有显著共线性的解释变量。一般情况下,当某个解释变量的VIF显著大于10时,说明该变量与其他解释变量之间具有显著的共线性,需要剔除,国内外已有学者将此方法运用到林火的研究中^[4-5,18]。

3.2.2 逐步回归

逐步回归是一种线性回归模型自变量选择方法,该方法已在模型研究领域得到大量的应用^[2-5,7]。本研究应用逐步回归方法对10个训练样本拟合的中间模型进行变量筛选($\alpha < 0.05$),然后选择在10个中间模型中出现7次及以上的变量进入最后建模数据的拟合计算^[5]。

3.3 模型预测准确度

ROC曲线检验方法是二分类模型效果评价的常用方法。该方法是以曲线下面积(area under the curve, AUC)的值(取值范围0~1)来判断模型预测效果。近年来该方法被广泛地运用于判断二分类模型的拟合水平^[2-3,5,19]。其中包括了一些林火概率预测模型的研究^[4-5]。

在分类模型中,尤其是二分类模型,其分类阈值是评判模型预测概率的关键。近年来多数学者运用ROC(receiver operating characteristic)中的“约登指数(敏感性+特异性-1)”判断“Cut-off”值(分类阈值),从而对预测值进行分类。该方法也已被

用于林火概率预测领域,如果模型的预测概率大于该最佳界值,则认为有林火发生,反之,则认为没有林火发生^[20]。

3.4 预测变量相对重要性

在林火驱动因子的相关研究中,研究者感兴趣于评价选定模型中各因素对结果变量的影响大小,即每个驱动因子对林火发生的解释能力。通常,通过估计标准回归系数或优势比来评价各自变量的相对重要性^[21]。本研究采用“标准化系数”法,计算各变量对火灾发生的相对重要性。标准化系数是调整后的系数,即对不同预测变量进行伸缩变换的,目的是使不同度量之间的特征具有可比性,即系数的大小可以直接进行比较或排序。当模型中其他预测变量不变时,其中一个预测变量标准化系数的绝对值越大,该变量变化1个标准差时,预测

结果的概率变化越大^[22]。标准化的 Gompit 回归系数可以通过 SAS 中 PROC LOGISTIC 程序的 STB 选项计算得到。

4 结果与分析

4.1 模型的预测变量

根据多重共线性检验结果,所有解释变量之间不存在显著的共线性,均用于模型的逐步拟合。对每种比例的10个训练样本进行 Gompit 回归拟合。在4种比例数据的拟合中,对变量的选择略有不同。其中“坡度”、“到公路距离”、“到铁路距离”、“到居民点距离”、“人口密度”、“GDP”、“植被覆盖度”、“日累积降水”、“日最高气温”、“日平均相对湿度”在4种比例的结果中均显著出现,对于不同数据比例而言,上述10个变量稳定出现(表2)。

表2 Gompit 回归模型的变量选择
Table 2 Variables selected for Gompit regression models

变量	显著样本个数(P_{min}/P_{max})			
	1:1	1:2	1:3	1:4
到铁路的距离 DRAIL	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
到公路的距离 DROAD	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
到居民点的距离 DRESID	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
海拔 ELEV	-	-	9(<0.000 1/0.015 8)	9(<0.000 1/0.045 9)
坡度 SLOPE	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
植被覆盖度 FVC	8(<0.000 1/0.009 4)	8(<0.000 1/0.029 6)	10(<0.000 1/0.005 5)	8(<0.000 1/<0.000 1)
人口密度 POP	10(<0.000 1/0.009 6)	10(<0.000 1/0.004 4)	10(<0.000 1/0.000 2)	10(<0.000 1/0.000 1)
人均 GDP	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
日平均风速 DW	10(<0.000 1/0.018 5)	-	-	-
日累积降水 DP	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
日最高气温 DMT	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)
日平均相对湿度 DR	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)	10(<0.000 1/<0.000 1)

从4种数据比例的拟合结果中可以看出,训练样本的预测准确率随着比例的扩大而增大,而临界值则随之减小。ROC 检验结果表明,模型的拟合效果(AUC)基本一致(0.864~0.873),说明每个中间模型都能较好地对数据进行描述(表3)。

表3 Gompit 回归模型4种数据比例
拟合结果的预测准确率

Table 3 Prediction accuracy of four proportions
by Gompit regression model

数据比例(火点: (训练样本非火点) 取值范围)	临界值	ROC 曲线下面积(训练样本取值范围)	预测准确率/%	
			训练样本(取值范围)	测试样本(取值范围)
1:1	0.447~0.567	0.865~0.873	77.9~78.7	77.3~79.9
1:2	0.301~0.384	0.864~0.871	79.8~80.7	76.2~79.9
1:3	0.237~0.306	0.864~0.871	83.1~83.8	78.0~81.8
1:4	0.171~0.208	0.866~0.869	85.4~86.1	77.2~81.8

4.2 建模结果

根据模型分析结果,基于1:4的数据拟合结果建立林火发生概率模型:

$$P = 1 - e^{-[-1.14 + 0.01x_1 + 0.04x_2 + 0.01x_3 - 0.15x_4 + 0.62x_5 + 0.0002x_6 + 0.001x_7 - 0.28x_8 + 0.02x_9 - 0.03x_{10}]}$$

式中: P 为林火发生概率; x_1 为到铁路距离,km; x_2 为到公路距离,km; x_3 为到居民点距离,km; x_4 为坡度; x_5 为植被覆盖度; x_6 为人口密度; x_7 为人均GDP,元/km²; x_8 为日累积降水量,mm; x_9 为日最高气温,℃; x_{10} 为日平均相对湿度,%。

拟合结果显示,在0.5的默认分类标准下,模型整体的预测准确度为85.9%,其中,“火点”的预测准确度为42.7%，“非火点”的预测准确度为96.1%。ROC 检验结果显示,模型的AUC为0.867,模型的拟合效果较好。此外,根据ROC 检验结果计算得出模型的阈值为0.17,以0.17为分类标准,

总体预测准确率有所下降(77%)。其中,计算得到模型对“火点”的预测准确率大幅提高,为80.2%;而“非火点”的预测概率有所降低,为76.2%。模型对独立样本的检验结果显示,根据临界值0.17,独立检验样本预测准确率为75.3%,并且“火点”和“非火点”的预测准确率与建模结果相近(82.9%和73.4%)(表4)。

表4 建模样本和独立检验样本 Gompit 回归模型的准确预测结果

Table 4 The result of modeling and independent test samples by Gompit regression model

样本	临界值	预测准确率/%		
		火点	非火点	总预测准确率
建模样本(80%)	0.50	42.7	96.1	85.9
	0.17	80.2	76.2	77.0
独立检验样本(20%)	0.17	82.9	73.4	75.3

4.3 林火驱动因子分析

基于 Gompit 回归分析得到最终模型,并且得到林火发生驱动因子的相关重要性排序。气象因素和地形对于林火的影响能力显著高于人为因素。

根据最终模型拟合的结果及林火驱动因子的相对重要性排序表明,气象因素对大兴安岭地区林火发生概率影响最大。这一结果与其他北方森林的林火研究一致^[2]。在 Gompit 回归模型结果中,“日累计降水”和“日平均相对湿度”为最重要的气象因素,并且与林火发生均呈显著负相关;而“日最高气温”的影响都较低。在其他的研究中暂未发现日最高气温对林火的影响,而本研究的结果中日最高气温的影响较大且与林火发生呈显著正相关(表5、图2)。

虽然地形因子中只有坡度与林火发生呈显著负相关,但其对林火发生的影响仅次于日累积降水。从图1d和图3a亦可看出中等以上火险区多分布在低坡度。这与目前大部分研究结果相同。

人为因素对林火的影响程度不及气象和地形

因素,但对林火发生也具有显著的影响。值得注意的是所有人为因素与林火发生均呈显著正相关。

表5 Gompit 回归模型的参数估计

Table 5 Parameter estimation of Gompit regression model

参数	估计系数	标准误差	卡方	显著性
常量	-1.138	0.079 8	203.571 2	<0.000 1
到铁路距离	0.008 95	0.000 724	153.061 8	<0.000 1
到公路距离	0.041 9	0.002 15	381.003 4	<0.000 1
到居民点距离	0.012 1	0.000 972	154.231 1	<0.000 1
坡度	-0.150 5	0.005 17	846.698 5	<0.000 1
植被覆盖度	0.621 3	0.101 4	37.507 9	<0.000 1
人口密度	0.000 259	0.000 034	57.760 2	<0.000 1
GDP	0.001 13	0.000 075	229.526 9	<0.000 1
日累计降水	-0.284 4	0.023 3	148.387 0	<0.000 1
日最高气温	0.021 4	0.001 97	117.938 9	<0.000 1
日平均相对湿度	-0.029	0.001 12	665.449 0	<0.000 1

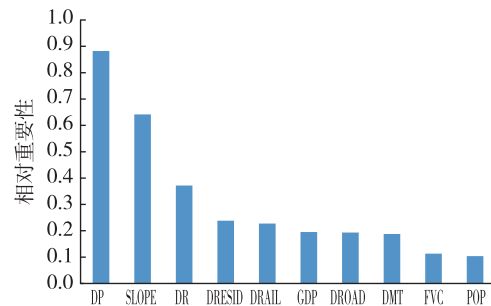


图2 基于 Gompit 回归模型的林火驱动因子相对重要性
Fig. 2 The relative importance of fire drivers based on Gompit regression model

4.4 林火发生概率

基于上述林火概率预测模型,对大兴安岭地区林火空间概率分布进行插值分析(图3a)。由图3b可知,林火发生概率地域分异明显。值得注意的是,大兴安岭南部加格达奇、松岭大部分地区发生林火的概率远高于其他地区,并且全区面积半数以上的火险等级较高,这与前人的研究结果基本一致^[5,19];呼玛林业局全区林火发生概率均很高,且全区几乎无低火险区;另外,韩家园南部、西林吉西部也有中火险区,这些中火险和高险区分别占整个

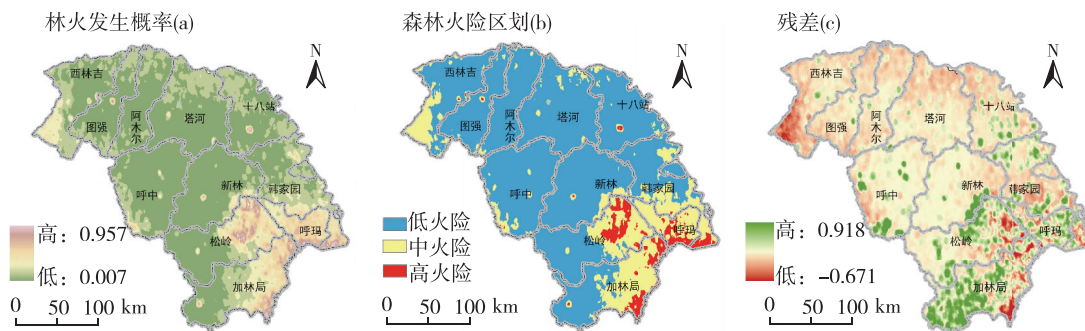


图3 研究区域林火发生概率(a)、森林火险区划(b)和残差(c)
Fig. 3 The probability of forest fire (a), fire danger class (b) and residuals (c) in the study area

大兴安岭的 19.9%和 4.3%(表 6)。根据残差分析结果,模型低估区域较为集中,主要分布在加格达奇、呼玛和松岭 3 个林业局,韩家园、新林和呼中等林业局也出现小部分林火低估区,这些低估区约占整个大象案例区域面积(约 19 139.2 km²)的 22.92%,这些地区森林火灾隐患依然较大(图 3c)。

表 6 火险等级分类

Table 6 Classification of fire risk rank

林火发生概率	火险等级	面积/km ²	面积百分比/%
0.007 ~ 0.17	低火险区	63 218.6	75.7
0.18 ~ 0.50	中火险区	16 618.9	19.9
0.51 ~ 0.641 4	高火险区	3 591.0	4.3

5 结 论

应用二项 Gompit 回归模型对大兴安岭地区林火及其驱动因子的关系进行探索性研究。拟合结果表明:模型的应用条件基本符合大兴安岭地区林火发生的数据模式,并且预测能力较强,在解释大兴安岭地区的林火发生概率上具有较高的可信度。

基于标准化系数,对大兴安岭的林火发生驱动因子的影响进行分析。结果表明,气象因素对 2000—2016 年大兴安岭地区林火发生影响最大。其中,日累积降水和日平均相对湿度的重要性程度最高。坡度对林火的发生也有很重要的作用;人为活动因素对于近 17 年来该地区林火发生的影响程度远不及气象因素,“到居民点的距离”、“到铁路的距离”、“到公路的距离”、“人口密度”和“人均 GDP”对林火发生也起到了显著影响,但其相对重要性较低,这可能是由于近年来,该地区人口增长较慢、人口向外迁徙所致,并且经济发展相对较慢。

基于预测模型得到的林火发生概率和火险等级图显示,大兴安岭东部、南部大部分地区处于中火险和高火险区域,特别是呼玛林业局、松岭林业局、加格达奇林业局三地。此外,模型对于加格达奇、松岭、呼玛、韩家园等 4 个林业局及松岭和新林交界处等地区的预测能力较差,这些地区同样也是林火高发区,同时模型低估区也较为集中。说明这些地区未来火险形势严峻,需加强林火管理尤其是那些处于高火险的地区。生成的火灾概率类图和火险区划图有助于当地森林管理人员进行林火管理工作,为了节省有限的财力物力和人力资源,重要的是优先考虑各个林火高发区及模型低估区域。

6 讨 论

当前对于大兴安岭地区的林火模型研究中,二

分类模型主要以 Logistic 回归模型为主,还包括随机森林、地理加权 Logistic 回归模型^[2-3]。本研究在前期研究的基础上扩大研究时间,结果表明,与前期的研究结果相比,模型的预测能力基本相当;以二分类数据形式对大兴安岭地区的林火进行研究,结果可信度较高,表明此类模型在大兴安岭地区具有较高的适用性^[2]。然而模型对于驱动因子的分析结果却有所差异:前期研究在使用官方记录火点数据(1980—2005 年)的情况下表明人为因素对林火的影响大于气象因素^[23];而本研究使用的是卫星火点数据(2000—2016 年),结果表明气象因素的主导作用大于人为因素。造成这种差异可能有以下几个原因:1)拟合数据不同,官方数据和卫星火点数据之间有差异;2)研究区域近年来人口增长较慢、人口向外迁徙。再下一步的研究中可以在校验官方数据和卫星数据后,取同一时间内两部分的数据进行模型拟合分析。

尽管气象因素的影响很大,但人为因素的影响也不能忽视^[17,23-25]。本研究结果也表明:“到居民点的距离”、“到铁路的距离”、“人均 GDP”对林火发生都具有显著的影响。研究结果与 Chang 等^[13]、Wu 等^[23]的研究一致。然而,笔者发现远离居民点的地区林火发生概率却越高。Penman 等^[24]发现远离居民区的林火发生越多,因为相对偏远地区的可燃物通常都有较高的连续性,并且植被的完整性相对较高;而城市及边缘地区的可燃物水平连续性差,无法形成连续性较高的可燃物床层。社会经济因素(人口密度和人均 GDP)的重要性程度相对低于气象因素及其他人为因素,但在模型中表现出与林火发生呈显著正相关。Prestemon 等^[25]研究表明低收入和经济欠发达地区容易导致人为纵火。Aldersley 等^[26]的研究结果表明 GDP 对研究区域内的过火面积呈负相关,这与本研究结果略有不同。在大兴安岭地区,森林生产活动在地区经济中占有重要地位,可能会增加林火发生概率。大量研究结果表明人口密度与火灾的发生呈正相关^[17,25],这与本研究结果一致。

本研究考虑的地形因子较少,主要是因为大兴安岭地区地貌多为低山丘陵,分布较为规律。其他同类的研究提出,平面曲率通过影响地表径流的汇聚和分散,进而改变地表可燃物的湿度^[27]。此外,本研究中使用日最高气温对林火的影响显著且重要性也较高,然而与林火的关系却呈负相关,在同地区或同一纬度带的林火研究中暂未发现这个现象^[7]。可燃物是林火发生的重要因素^[14],而在本

研究中未使用这类变量的原因在于与可燃物相关的数据获取过程难度较大,如可燃物含水率、可燃物载量等指标,很难大尺度范围内进行量化。这些参数的缺乏,一定程度上影响了变量选择的结果,在未来的研究中需要进一步分析和解决。

参考文献(References):

- [1] MCKENZIE D, SHANKAR U, KEANE R E, et al. Smoke consequences of new wildfire regimes driven by climate change [J]. *Earth's Future*, 2014, 2(2): 35-59. DOI:10.1002/2013ef000180.
- [2] GUO F T, INNES J L, WANG G Y, et al. Historic distribution and driving factors of human-caused fires in the Chinese boreal forest between 1972 and 2005 [J]. *Journal of Plant Ecology*, 2015, 8(5): 480-490. DOI:10.1093/jpe/rtu041.
- [3] GUO F T, SELVALAKSHMI S, LIN F F, et al. Geospatial information on geographical and human factors improved anthropogenic fire occurrence modeling in the Chinese boreal forest [J]. *Canadian Journal of Forest Research*, 2016, 46(4): 582-594. DOI:10.1139/cjfr-2015-0373.
- [4] GUO F T, WANG G Y, SU Z W, et al. What drives forest fire in Fujian, China? Evidence from logistic regression and Random Forests [J]. *International Journal of Wildland Fire*, 2016, 25(5): 505. DOI:10.1071/wf15121.
- [5] 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, 605/606: 411-425. DOI:10.1016/j.scitotenv.2017.06.219.
- [6] 石晶晶. 浙江省林火发生格局及预测模型研究 [D]. 杭州: 浙江农林大学, 2014.
- SHI J J. Study on occurring space and forecasting model of forest fires in Zhejiang province [D]. Hangzhou: Zhejiang A&F University, 2014.
- [7] 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: 117-129. DOI:10.1016/j.foreco.2012.03.003.
- [8] PRADHAN B, DINI HAIRI BIN SULIMAN M, ARSHAD BIN AWANG M. Forest fire susceptibility and risk mapping using remote sensing and geographical information systems (GIS) [J]. *Disaster Prevention and Management: an International Journal*, 2007, 16(3): 344-352. DOI:10.1108/09653560710758297.
- [9] VILAR L, WOOLFORD D G, MARTELL D L, et al. A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain [J]. *International Journal of Wildland Fire*, 2010, 19(3): 325. DOI:10.1071/wf09030.
- [10] GURIŞ S, ÇAĞLAYAN E, UN T. Estimating of probability of home-ownership in rural and urban areas: Logit, probit and Gompit model [J]. *European Journal of Social Sciences*, 2011, 21(3): 405-411.
- [11] MINAB M, OIA M, ASL M G. Analysis of the effect of corporate governance on the financial health of companies Using Logit, Probit and Gompit models [C] // The 12th Iranian National Conference on Accounting, Iran, 2014.
- [12] JUSTICE C O, GIGLIO L, KORONTZI S, et al. The MODIS fire products [J]. *Remote Sensing of Environment*, 2002, 83(1/2): 244-262. DOI:10.1016/s0034-4257(02)00076-7.
- [13] 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. DOI:10.1007/s10980-013-9935-4.
- [14] 胡海清. 林火生态与管理 [M]. 北京: 中国林业出版社, 2005.
- HU H Q. Forest fire ecology and management [M]. Beijing: China Forestry Publishing House, 2005.
- [15] CHANDLER C, CHENEY P, TRABAUD L, et al. Fire in forestry. Volume I. Forest fire behavior and effects [M]. Florida: Krieger Publishing Company, 1991.
- [16] GUTMAN G, IGNATOV A. The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models [J]. *International Journal of Remote Sensing*, 1998, 19(8): 1533-1543. DOI:10.1080/014311698215333.
- [17] STURTEVANT B R, CLELAND D T. Human and biophysical factors influencing modern fire disturbance in northern Wisconsin [J]. *International Journal of Wildland Fire*, 2007, 16(4): 398. DOI:10.1071/wf06023.
- [18] SU Z W, HU H Q, WANG G Y, et al. Using GIS and Random Forests to identify fire drivers in a forest city, Yichun, China [J]. *Geomatics, Natural Hazards and Risk*, 2018, 9(1): 1207-1229. DOI:10.1080/19475705.2018.1505667.
- [19] 邓欧, 李亦秋, 冯仲科, 等. 基于空间 Logistic 的黑龙省林火风险模型与火险区划 [J]. *农业工程学报*, 2012, 28(8): 200-205. DOI:10.3969/j.issn.1002-6819.2012.08.031.
- DENG O, LI Y Q, FENG Z K, et al. Model and zoning of forest fire risk in Heilongjiang province based on spatial Logistic [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28(8): 200-205.
- [20] LITTELL J S, MCKENZIE D, PETERSON D L, et al. Climate and wildfire area burned in western US ecoprovinces, 1916-2003 [J]. *Ecological Applications*, 2009, 19(4): 1003-1021. DOI:10.1890/07-1183.1.
- [21] 张波, 代鲁燕, 黄启凤, 等. Logistic 回归模型中自变量相对重要性的优势分析 [J]. *浙江预防医学*, 2012, 24(8): 13-15, 29. DOI:10.3969/j.issn.1007-0931.2012.08.005.
- ZHANG B, DAI L Y, HUANG Q F, et al. Dominance analysis for the relative importance of independent variable in logistic regression model [J]. *Zhejiang Journal of Preventive Medicine*, 2012, 24(8): 13-15, 29.
- [22] MENARD S. Six approaches to calculating standardized logistic regression coefficients [J]. *The American Statistician*, 2004, 58(3): 218-223. DOI:10.1198/000313004x946.
- [23] WU Z W, HE H S, YANG J, et al. Relative effects of climatic and local factors on fire occurrence in boreal forest landscapes of northeastern China [J]. *Science of the Total Environment*, 2014, 493: 472-480. DOI:10.1016/j.scitotenv.2014.06.011.
- [24] PENMAN T D, BRADSTOCK R A, PRICE O. Modelling the determinants of ignition in the Sydney Basin, Australia: Implications for future management [J]. *International Journal of Wildland Fire*, 2013, 22(4): 469-478.
- [25] PRESTEMON J P, BUTRY D T. Time to burn: modeling wildland arson as an autoregressive crime function [J]. *American Journal of Agricultural Economics*, 2005, 87(3): 756-770. DOI:10.1111/j.1467-8276.2005.00760.x.
- [26] ALDERSLEY A, MURRAY S J, CORNELL S E. Global and regional analysis of climate and human drivers of wildfire [J]. *Science of the Total Environment*, 2011, 409(18): 3472-3481. DOI:10.1016/j.scitotenv.2011.05.032.
- [27] POURGHASEMI H R. GIS-based forest fire susceptibility mapping in Iran: a comparison between evidential belief function and binary logistic regression models [J]. *Scandinavian Journal of Forest Research*, 2016, 31(1): 80-98. DOI:10.1080/02827581.2015.1052750.