

基于主成分非线性回归的湖南省生态公益林需求量预测模型

彭佳红^{a,b}, 邹冬生^b, 杨友^b

(湖南农业大学 a 信息科学技术学院; b 生态学研究所, 湖南 长沙 410128)

摘要: 针对2000—2010年湖南的人口、地区生产总值、财政总收入与支出、年降水量等15个指标和湖南省公益林面积数据, 首先用主成分分析方法确定影响公益林面积变化的2个主成分, 即经济与人口发展水平(包含年生产总值GDP、年末总人口、非农业人口、财政收入、财政总支出、第一产业产值、第二产业产值、第三产业产值、工业产值、建筑业产值、人均纯收入、固定资产投资、城市投入、农村投入)和自然环境水平(年降水总量), 然后基于多元线性回归模型, 运用2个主成分得分进行湖南省公益林需求总量的定量预测, 再通过非线性回归中的三次曲线回归模型和2个主成分得分对湖南省公益林需求总量进行修正的定量预测。结果表明, 建立的模型在一定程度上能够解释社会经济与公益林系统相互作用的内在联系, 并对公益林需求量预测进行滚动修正, 辅助公益林需求量的控制和预测。

关键词: 公益林需求量; 主成分分析; 回归模型; 湖南

中图分类号: S718.55⁺7; F224.9

文献标志码: A

文章编号: 1007-1032(2015)06-0691-04

Demand forecasting model of Hunan ecological public welfare forest based on main ingredient nonlinear regression

Peng Jiahong^{a,b}, Zou Dongsheng^b, Yang You^b

(a.College of Information Science and Technology; b.Institute of Ecology, Hunan Agricultural University, Changsha 410128, China)

Abstract: Aiming at fifteen indices of the population the gross regional production, the fiscal revenue and expenditure, agriculture, industry, service, the annual precipitation of Hunan province from the years of 2000 to 2010, as well as the historical data on the area of Hunan's public welfare forest, this paper firstly decided the economy & population development level and the natural environment level as the two main ingredients of the variation of public welfare forest with the method of main ingredient analysis, then both the quantitative forecast of the total demand of Hunan's public welfare forest through analyzing the scores of above two main ingredients based on the multiple linear regression model and its rectified quantitative forecast based on the cubic curve nonlinear regression model were carried out. The results showed the effectiveness of this model, which could carry out rolling rectifying of the demand of public welfare forest and provide scientific reasons for the control and forecast of the demand of public welfare forest.

Keywords: demand of public welfare forest; main ingredient analysis; regression model; Hunan

生态公益林 (non-commercial forest, 简称NCF) 主要是指提供公益性、社会性产品或服务的森林、林木、林地^[1]。生态公益林能够改善、优化生态环境, 提供对社会公众至关重要的服务。生态公益林

系统服务功能的维持对经济的可持续发展会产生依赖。生态公益林系统不同于自然的森林生态系统, 它明显受到人和人造环境的影响。剧烈的人类活动和明显的环境改变都将影响生态公益林系统

的服务功能。生态公益林需求总量主要受经济发展水平、人口等因素的影响。一般而言,随着社会大众收入水平的提高,社会公众对扩大生态公益林需求量的愿望就越强烈。目前,社会公众对生态公益林的规模需求愿望与政府对生态公益林的有效供给之间存在一定矛盾。生态公益林需求总量预测是一项重要的决策支持技术,具有不确定性、条件性、时间性及多方案性。生态公益林需求量的变化受到社会经济、自然环境因素制约,并且这种制约关系难以定量描述。科学、准确预测生态公益林需求量对生态公益林事业的健康发展,乃至对整个国民经济的发展都有着十分重要的意义。针对生态公益林需求量的变化影响因素的复杂性和不确定性,即多变量和随机性特征,笔者将定量预测方法中的主成分分析与回归分析技术应用于生态公益林需求量定量预测和修正的定量预测,旨在为辅

助生态公益林规模数量控制和预测提供理论依据。

1 材料与方法

1.1 数据来源

本研究选择年生产总值GDP(X_1 , 亿元)、年末总人口(X_2 , 万人)、非农业人口(X_3 , 万人)、年财政总收入(X_4 , 亿元)、年财政总支出(X_5 , 亿元)、第一产业产值(X_6 , 亿元)、第二产业产值(X_7 , 亿元)、第三产业产值(X_8 , 亿元)、年工业总产值(X_9 , 亿元)、建筑业产值(X_{10} , 亿元)、年人均纯收入(X_{11} , 元)、年固定资产投资(X_{12} , 亿元)、城市投入(X_{13} , 亿元)、农村投入(X_{14} , 亿元)、年降水总量(X_{15} , 亿 m^3)等15个指标,构成湖南省生态公益林面积(Y , hm^2)变化驱动指标体系。

原始数据来源于湖南统计年鉴2011^[2]和2000—2010年湖南省森林资源统计年报^[3],见表1。

表1 湖南省2000—2010年生态公益林面积与驱动指标数据

Table 1 Area and the driving data of EPWF during 2000–2010 years in Hunan province

年份	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}
2000	965 100	3 551.5	6 562.1	1 952.2	177.00	347.8	784.9	1 293.2	1 473.4	1 094.8	198.4	3 393.6	1 066.3	763.9	302.4	3 126.0
2001	1 056 600	3 831.9	6 595.9	2 031.5	205.40	431.7	825.7	1 412.8	1 593.4	1 180.4	232.4	3 680.2	1 210.6	878.4	332.3	2 870.0
2002	1 914 700	4 151.5	6 628.5	2 121.1	231.20	533.0	847.3	1 523.5	1 780.8	1 265.7	257.8	3 857.5	1 355.9	996.2	359.7	4 154.0
2003	2 083 182	4 659.9	6 662.8	2 232.0	268.70	573.8	886.5	1 777.7	1 995.8	1 484.9	292.8	4 255.2	1 557.0	1 201.9	355.0	2 595.0
2004	2 902 112	5 641.9	6 697.7	2 377.7	320.60	719.5	1 022.5	2 190.5	2 428.9	1 824.1	366.4	4 889.5	1 981.3	1 659.3	291.0	3 164.0
2005	3 341 945	6 596.1	6 732.1	2 490.9	395.30	873.4	1 100.7	2 612.6	2 882.9	2 195.3	417.2	5 488.2	2 563.9	2 203.9	360.0	2 751.0
2006	3 591 491	7 688.7	6 768.1	2 619.9	477.93	1 064.5	1 272.2	3 187.1	3 229.4	2 707.6	479.4	6 144.2	3 242.4	2 785.3	457.08	3 064.5
2007	4 065 581	9 439.6	6 805.7	2 752.9	606.55	1 357.0	1 626.5	3 977.7	3 835.4	3 397.7	580.0	7 297.8	4 294.4	3 749.1	545.38	2 716.7
2008	3 805 798	11 555.0	6 845.2	2 885.3	722.71	1 765.2	1 892.4	5 028.9	4 633.7	4 310.1	718.8	8 436.3	5 649.7	4 995.6	654.08	2 595.0
2009	3 806 082	13 059.7	6 900.2	2 980.9	847.62	2 210.4	1 969.7	5 687.2	5 402.8	4 819.4	867.8	9 305.2	7 695.3	6 880.0	815.26	2 654.5
2010	3 948 744	16 037.9	7 089.5	3 069.8	1 081.70	2 702.5	2 325.5	7 343.2	6 369.3	6 305.1	1 038.0	10 360.0	9 821.1	8 775.5	1 045.55	3 279.0

1.2 方法

运用主成分分析与非线性回归分析方法,采用SPSS 18.0软件的相关分析模块对数据进行处理^[4-5]。

2 结果与分析

2.1 驱动影响因子分析结果

对所选各驱动因子与公益林面积进行线性相关性分析,得出的相关系数见表2。

从表2可以看出,除年降水总量外,其余的所有因子与公益林面积变化呈高度正相关,为影响显著的因子,影响程度由强到弱的因子依次为非农业

表2 湖南省2000—2010年生态公益林面积与驱动因子的相关系数

Table 2 The correlation analysis of the 2000–2010 area of Hunan's ecological public welfare forests and its driving factors

驱动因子	相关系数	驱动因子	相关系数
年生产总值	0.817**	年工业总产值	0.800**
年末总人口	0.835**	建筑业产值	0.813**
非农业人口	0.938**	人均纯收入	0.853**
年财政总收入	0.813**	年固定资产投资	0.748**
年财政总支出	0.788**	年城市投入	0.754**
第一产业产值	0.818**	年农村投入	0.672*
第二产业产值	0.802**	年降水总量	-0.302
第三产业产值	0.833**		

***示在0.01水平差异显著; **示在0.05水平差异显著。

人口、人均纯收入、年末总人口、第三产业产值、第一产业产值、年生产总值、年财政总收入、建筑业产值、第二产业产值、年工业总产值、年财政总支出、年城市投入、年固定资产投资、年农村投入。这一结果表明，湖南省生态公益林面积变化主要受人类社会经济发展、社会环境改变和人口增加的影响，而降水等自然因素不是约束条件，这可能与湖南省地处亚热带季风湿润气候区有关。

2.2 主成分分析结果

2.2.1 各驱动因子总方差分解

对15个驱动影响因子进行主成分分析的各驱动因子总方差见表3。

表3 各驱动影响因子总方差

主成分序号	各成分特征值	方差贡献率/%	累计方差贡献率/%
1	13.825	92.166	92.166
2	0.981	6.542	98.708
3	0.132	0.881	99.589
4	0.036	0.240	99.829
5	0.015	0.102	99.931
6	0.008	0.055	99.986
7	0.001	0.008	99.994
8	0.001	0.005	99.999
9	0.000	0.001	100.000
10	1.25E-005	8.37E-005	100.000
11	2.39E-016	1.59E-015	100.000
12	1.33E-016	8.85E-016	100.000
13	-1.32E-016	-8.80E-016	100.000
14	-2.33E-016	-1.56E-015	100.000
15	-4.17E-016	-2.78E-015	100.000

从表3可知，前2个主成分的特征值大于或者接近于1，且其累积贡献率为98.708%，说明前2个主成分可以表达全部因子对公益林面积变化的驱动影响力。

2.2.2 各驱动因子对应的 2 个主成分载荷系数

主成分因子载荷表示主成分和因子之间的相关系数，其中各驱动因子的载荷系数表明其在相应主成分中的重要性。各驱动因子对应的2个主成分载荷系数见表4。

表 4 各驱动因子对应的 2 个主成分载荷系数

Table 4 Two main ingredients load corresponding to driving factors

驱动因子	第1主成分载荷系数(C ₁)	第2主成分载荷系数(C ₂)
年生产总值	1.000	0.005
年末总人口	0.982	0.063
非农业人口	0.960	-0.114
年财政总收入	0.995	-0.047
年财政总支出	0.994	-0.019
第一产业产值	0.999	0.020
第二产业产值	0.999	0.021
第三产业产值	0.999	0.013
年工业总产值	0.999	-0.006
建筑业产值	1.000	0.019
年人均纯收入	0.971	0.127
固定资产投资	0.998	0.039
年城市投入	0.993	0.061
年农村投入	0.994	0.055
年降水总量	-0.243	0.967

表4表明，第1主成分主要分布于除年降水总量外的所有因子，反映的是经济、社会环境和人口因子对公益林面积变化的影响，可称之为经济与人口发展水平；第2主成分主要集中在年降水总量上，可称之为自然环境水平。

各年份公益林面积和主成分对应得分值见表5。

表5 各年份公益林面积和主成分对应得分值

Table 5 The area of non-commercial forrest in each year and corresponding principal component scores

年份	公益林面积/hm ²	主成分 1 对应得分值	主成分 2 对应得分值
2000	965 100.0	21 890.231 31	3 811.531 01
2001	1 056 600.0	23 419.196 57	3 611.216 83
2002	1 914 700.0	24 562.567 85	4 885.198 26
2003	2 083 182.5	27 216.352 74	3 446.082 45
2004	2 902 112.0	31 254.075 45	4 119.464 62
2005	3 341 945.6	35 825.284 55	3 858.922 93
2006	3 591 491.6	40 926.493 64	4 315.875 21
2007	4 065 581.6	49 099.220 66	4 244.428 19
2008	3 805 798.0	58 898.885 00	4 425.909 62
2009	3 806 082.4	68 179.275 37	4 808.279 55
2010	3 948 744.3	81 886.468 58	5 802.541 36

2.3 回归分析结果

2.3.1 线性回归分析

使用表5数据作线性回归分析，得线性回归方程：

$$Y = 2\ 013\ 432.359 + 55.543 \times C_1 - 346.331 \times C_2。$$

此回归方程的复相关系数为0.814，复相关系数平方(多重判定系数)为0.663，修正多重判定系数为0.579，回归方程的估计标准误差为754 771.742 03，可以进行拟合优度检验。但参考修正多重判定系

数,其拟合优度一般,因此,生态公益林需求量变化与社会经济、人口,自然环境之间并不是线性关系,可能存在某种曲线相关关系,故采用曲线回归模型即非线性回归模型进行修正预测。

2.3.2 非线性回归分析

先通过绘制散点图考察图形的数据分布特点,即曲线关系,确定应采用的拟合曲线回归模型;再用变量代换法将曲线转换为直线,按线性模型求解参数,最后再变换为曲线模型。同时考虑到自然环境水平对生态公益林需求量影响不大,故非线性回

归分析仅考虑社会经济与人口因素,即上述的经济与人口发展水平,记为 X 。本问题通过散点图观测,可以用如下三次曲线回归模型拟合:

$$Y=b_0+b_1X+b_2X^2+b_3X^3。$$

模型的多重判定系数为0.983,修正多重判定系数为0.975,表明拟合优度好。

F检验的方差分析结果见表6。F检验统计量的观测值为133.523,对应的概率 P 值为0.000,结果表明方程拟合效果好。

表6 非线性回归方程F检验的方差分析结果

模型变异来源	离差平方和	自由度	均方	F检验观测值	概率 P 值
回归	13 304 910 955 392.180	3	4 434 970 318 464.062	133.523	0.000
残差	232 504 593 282.433	7	33 214 941 897.491		
总计	13 537 415 548 674.610	10			

回归系数的显著性检验结果见表7。模型回归系数的标准误差为11.175, P 值为0.000,说明在回归方程中 X 变量的存在具有统计学意义。

表7 非线性回归分析回归系数的显著性检验结果

Table 7 The significant checking of the regression coefficients in nonlinear regression

模型	偏回归系数	偏回归系数标准误差	t检验观测值	概率 P 值
X	646.920	11.175	8.042	0.000
X^2	-0.011	-18.844	-6.400	0.000
X^3	5.89E-008	8.617	5.362	0.001
常数	-8 620 993.831	11.175	-7.479	0.000

由此得出非线性回归方程的湖南省生态公益林需求量预测模型为:

$$Y=-8\ 620\ 994+646.92X-0.011X^2+5.89\times 10^{-8}X^3。$$

式中: Y 为湖南省生态公益林需求量; X 为社会经济与人口影响因素。

从预测模型可以看出,2000—2007年湖南生态公益林面积快速增长;2008—2010年表现为略有下降;2010年后生态公益林规模需求数量呈现相对平稳状态,这一趋势与同期GDP年增长率增长趋势同步,表明湖南省生态公益林系统的存在与发展,明显受到人类经济活动和社会环境改变的影响。当然,从实际出发,生态公益林需求量预测是复杂和

多因素的,且呈现不确定和非线性关系,难以全面定量描述,如土地利用的有限性、经济增长的有限性以及社会公众对生态公益林服务功能的支付意愿变化等因素,都会影响生态公益林需求量的预测。此外,即使面积相同,因林种结构布局以及其功能状态的不同,其服务功能也存在差异,也会影响生态公益林需求量的预测,但模型在一定程度上能够解释社会经济与生态公益林系统相互作用的内在联系,不失为一种有益尝试。

参考文献:

- [1] 国家林业局植树造林司.生态公益林建设导则[M].北京:中国标准出版社,2001.
- [2] 湖南省统计局.湖南统计年鉴2011[M].北京:中国统计出版社,2011.
- [3] 湖南省森林资源检测中心.湖南省森林资源统计年报[R].2000—2010.
- [4] 卢纹岱.SPSS for Windows 统计分析[M].第3版.北京:电子工业出版社,2007:294—365.
- [5] 张文彤,董伟.SPSS 统计分析高级教程[M].北京:高等教育出版社,2006:91—117.

责任编辑:苏爱华

英文编辑:梁和