

利用弹性网对 PLS 佐证分析城市水生态足迹及驱动因素 ——以贵阳市为例

许国钰¹, 任晓冬¹, 杨振华², 肖杰¹, 刘立冰¹, 高洋¹

(1. 贵州师范大学 喀斯特研究院/国家喀斯特石漠化防治工程技术研究中心, 贵州 贵阳 550001; 2. 贵州省山地资源研究所, 贵州 贵阳 550001)

摘要: [目的] 在分析贵阳市水生态足迹时间序列的基础上, 探讨影响贵阳市水生态足迹的驱动因素, 评估其水资源可持续利用程度, 为水资源的合理利用提供决策参考。[方法] 利用水生态足迹理论, 计算出 2002—2016 年贵阳市水生态足迹, 水资源承载力和水资源可持续利用指数。再根据环境压力模型 STIRPAT 建模, 并运用偏最小二乘法 (PLS) 回归对模型进行分析, 再利用弹性网回归对 PLS 进行佐证, 分析出影响城市水生态足迹的几个驱动因素。[结果] ①2002—2016 年贵阳市水生态足迹呈波动式上升趋势, 水生态承载力总体上呈波动下降趋势。水资源可持续利用指数在研究时段内都低于 1, 表明整个城市用水处于不安全状态。②STIRPAT 模型进行最小二乘法 (LS) 回归分析, 结果表明模型中 4 个驱动因素存在多重共线性。③采用 PLS 回归对 STIRPAT 模型进行修正, 消除驱动因素之间的共线性, 但 VIP 只筛选出人均 GDP 的二次项、人口两个有意义的驱动因素。④根据弹性网回归对 STIRPAT 进行分析, 人口, 人均 GDP, 第一二产业占总产业比值, 城市化率都对贵阳市水生态足迹产生影响。第一二产业占总产业比值和城市化率都与贵阳市的水生态足迹间存在环境 EKC 假说。[结论] 贵阳市在发展经济与推动快速城市化的过程中, 必须降低水生态足迹, 才是实现水资源可持续利用的关键。

关键词: 水生态足迹; STIRPAT 模型; 环境 EKC; 偏最小二乘法; 弹性网

文献标识码: B

文章编号: 1000-288X(2018)04-0220-08

中图分类号: TV38

文献参数: 许国钰, 任晓冬, 杨振华, 等. 利用弹性网对 PLS 佐证分析城市水生态足迹及驱动因素[J]. 水土保持通报, 2018, 38(4): 220-227. DOI: 10.13961/j.cnki.stbctb.2018.04.036. Xu Guoyu, Ren Xiaodong, Yang Zhenhua, et al. An analysis on water ecological footprint in Guiyang City and its driving factors based on STIRPAT model using partial least squares and elastic net regression[J]. Bulletin of Soil and Water Conservation, 2018, 38(4): 220-227.

An Analysis on Water Ecological Footprint in Guiyang City and Its Driving Factors Based on STIRPAT Model Using Partial Least Squares and Elastic Net Regression

XU Guoyu¹, REN Xiaodong¹, YANG Zhenhua², XIAO Jie¹, LIU Libing¹, GAO Yang¹

(1. School of Karst Science, Guizhou Normal University, Guiyang, State Key Engineering

Technology Research Center for Karst Rocky Desertification Rehabilitation, Guiyang, Guizhou

550001, China; 2. Institute of Mountain Resources of Guizhou Province, Guiyang, Guizhou 550001, China)

Abstract: [Objective] Based on analyzing the time series of urban water ecological footprint in Guiyang City, this paper made use of the water ecological footprint model to calculate the water ecological footprint and evaluated the sustainable utilization of water resources and to provide some advice for rational utilization of water resources. [Methods] Using water ecological footprint theory, the water ecological footprint, water

收稿日期: 2018-04-09

修回日期: 2018-05-07

资助项目: 贵州省研究生教育创新计划项目“贵州省熊康宁喀斯特环境研究生导师工作室”(黔教研合 GZS 字[2016]04 号); 贵州省十一五科技计划课题“贵州喀斯特石漠化综合防治技术集成与典型模式凝练和推广”(黔科合 SZ(2 008)3077)

第一作者: 许国钰(1986—), 女(汉族), 贵州省贵阳市人, 硕士研究生, 研究方向为喀斯特生态建设与区域经济。E-mail: 249893131@qq.com.

通讯作者: 任晓冬(1965—), 男(汉族), 四川省崇州市人, 教授, 硕士生导师, 主要从事喀斯特生态建设与区域经济方面的研究。E-mail: renxiaodong@hotmail.com.

resources carrying capacity and sustainable utilization index of water resources in Guiyang City from 2002 to 2016 were calculated. Then, according to the environmental pressure model STIRPAT modeling and using partial least squares(PLS) regression and elastic net regression to back up and analyse the driving force factors which affect the water ecological footprint. [Results] ① From 2002 to 2015, as a whole, water ecological footprint in Guiyang City was on the rise. The water ecological carrying capacity was in a large fluctuating range. The water resources sustainable utilization index was lower than 1. The water resource in the whole city was in a insecure state. ② Using LS regression to analyze the STIRPAT model, the results showed that the four driving force factors in the model have multiple collinear. Then, using partial least squares(PLS) to modify the STIRPAT model, the total linearity between the driving force factors was eliminated. But the VIP value only selected two meaningful driver force factors: the quadratic term of per capita GDP and the population. ③ According to the analysis of STIRPAT by elastic net, the results showed that the population, per capita GPD, the first two industries accounted for the ratio of the total industry, and the urbanization rates affected the water ecological footprint of Guiyang City. The water ecological footprint between first two industries accounted for the ratio of the total industry and the urbanization rate both existed the environmental Kuzenets curve. [Conclusion] In the process of developing and promoting rapid urbanization, the water ecological footprint of Guiyang City must be reduced, which is the key to the sustainable development of water resources.

Keywords: water ecological footprint; STIRPAT model; environmental Kuzenets curve; PLS regression; elastic net regression

水资源作为一种战略性资源,关系着国家安全与人民的生产生活。随着经济发展、工业化和城市化不断推进,水资源短缺,水污染,生态恶化等问题,已经成为制约城市发展与生态协调的一个重大问题。开展针对水生态足迹的研究已成为实现水资源可持续开发利用的一个重要环节。生态经济学家 William^[1]提出的生态足迹概念为可持续发展提供了理论依据。Mathis 等^[2-3]对生态足迹计算原理和方法进行进一步的拓展,并广泛应用于不同空间尺度的定量评价体系,并由此衍生出水足迹,水生态足迹等概念。黄林楠等^[4]建立水生态足迹和水资源承载力模型,并对整个水资源账户进行细化。潘真真等^[5]基于水生态足迹并将概念引申为水生态占用,并提出生态系统供给及净化服务功能的水生态占用研究。在此之后,国内学者对水生态足迹的研究,从单一的核算水足迹及其历年变化特征;发展到胡永红等^[6]运用 ARIMA 模型的大样本建模对水生态赤字进行预测。杨振华等^[7-9]分别采用 GRNN 模型、SVR(支持向量回归机)模型对时间短、变异性大的小样本容量对城市水生态足迹进行预测,揭示水生态足迹的时间序列规律。虽然取得了良好的预测精度,但是缺乏对影响水生态足迹驱动因素的分析。本文以地质条件特殊而导致工程性缺水的城市贵阳为例,计算其 2002—2016 年水生态足迹时间序列,引入 STIRPAT 模型对影响其水生态足迹的驱动因素进行分析。虽然传统的线性回归方差最小,但是由于数据拥有大量预测变量,而且存在严重的相关性,导致矩阵病态。之后

采用偏最小二乘法(PLS)及弹性网对 STIRPAT 进行修正,PLS 回归兼具主成分分析和多元回归的优点,克服了变量间存在强烈的交互相关而导致多元共线性^[11]。弹性网是一种是对 PLS 的算法优化,它是对回归系数采用联合惩罚函数的有偏估计。并且在筛选变量方面更加精准。本文在丰富了水生态足迹理论的实践及应用的同时,运用 PLS 及其弹性网对影响城市水生态足迹驱动因素进行分析,且进一步探讨如何减缓贵阳市水生态足迹,对城市水资源的可持续利用提供一定的决策参考。

1 研究区概况

贵州省贵阳市地处云贵高原东斜坡地带,东经 $106^{\circ}71' - 107^{\circ}17'$,北纬 $26^{\circ}11' - 27^{\circ}22'$ 。属于明显的亚热带高原性季风气候,温和湿润。贵阳市地处长江流域的乌江水系和珠江流域的红水河水系,境内虽有 98 条河流。但是由于贵阳市地处我国喀斯特地貌的中心地带,水循环受“二元”结构影响导致地表蓄水能力差^[12]。独特的地质条件导致用水率极低,加上山高坡陡峡谷深,很难建造水库,河水大量流失,属于典型的工程性缺水^[12],对耗水严重的工农业及相关的产业的约束性强。喀斯特地貌地表起伏度大,土层稀薄,多裂隙隙及落水洞发育,易与地下水形成联动污染^[12]。贵阳市已探明矿种有 52 种,主要有煤(9.00×10^8 t)、铁(2.40×10^7 t)、硅、钒土、磷、硫、汞等矿产资源。其中铝土矿的保有储量 4.30×10^8 t,磷矿

4.60×10⁸ t,是全国 3 大磷矿基地之一。2016 年 3 大产业结构比重为 15.61:39.47:44.9,第一、第二产业占国内生产总值的比重过高,农业和大规模工业行业耗水量过大。

2 研究方法

2.1 基本理论

2.1.1 水生态足迹 生态足迹理论是用于比较自然资源消耗量与给定区域内人口与生态承载力的研究,水生态足迹是由生态足迹理论发展而来。水生态足迹是指维持人口对资源消费和废弃物消纳所必需的水域和水资源用地面积^[13]。城市水生态足迹又直指城市人口消费所必需的水域和水资源用地面积,由两部分组成:城市的水量(工业、农业、生活)生态足迹;城市的水污染(氨氮、COD)生态足迹,包括维持城市废弃物消纳所必需的水域和水资源用地面积。但实际在应用层面,城市用水量生态足迹和水质生态足迹两部分叠加即可求出城市水生态足迹^[5,10,14]。其计算模型为:

$$EF_{wf} = WF_{wf} + EF_{wq} + r_w \times U_{ur} / P_w + r_w \times \left(\frac{U_{COD}}{P_{COD}} + \frac{U_{NH_3}}{P_{NH_3}} \right) \quad (1)$$

式中:EF_{wf}——水生态足迹(hm²);EF_{wf}——水量生态足迹(hm²);EF_{wq}——水质生态足迹(hm²);r_w——全球水资源均衡因子;P_w——全球平均水资源生产能力(m³/hm²);U_{ur}——城市用水资源总量 U_{ur} = U_{wri} + U_{wra} + U_{wrc};U_{wri}——工业用水;U_{wra}——农业用水;U_{wrc}——生活用水;U_{COD},U_{NH₃}——COD 和氨氮排放量;P_{COD},P_{NH₃}——水域对污染物 COD 和 NH₃ 的吸纳能力(t/hm²)。

2.1.2 水生态承载力核算 水生态承载力是反映水生态系统的供给能力,是指支持水生态系统在保持健康下维持人类活动的阈值。城市的水生态承载力可通过供需两方面的比较来计算,用于测度城市水生态可持续发展状况,其计算模型为:

$$EC_w = \frac{0.4 \times r_w \cdot \varphi \cdot Q}{P_w} \quad (2)$$

$$\ln EF_i = \eta + \mu \ln P_i + [\theta_1 \ln A_i + \theta_2 (\ln A_i)^2] + [\kappa_1 \ln N_i + \kappa_2 (\ln N_i)^2] + [\omega_1 \ln M_i + \theta_2 (\omega_2 \ln M_i)^2] \ln \epsilon_i \quad (6)$$

再对公式(6)求一阶偏导数,就得出富裕度(即人均 GDP),对水生态足迹的弹性系数为 θ₁ + 2θ₂(lnA_i),结构化(即第一二产业对总产值比重)对水生态足迹的弹性序数为 κ₁ + 2κ₂(lnN_i),城市化对水生态足迹的弹性系 ω 数为 1 + 2ω₂(lnM_i)。

2.3 偏最小二乘法(PLS)

PLS 允许自变量存在多重相关性,并且在样本点个数少于变量个数的条件下进行回归建模^[11];自

式中:EC_w——城市水生态承载力(hm²);φ——区域水资源产量因子;r_w——全球水资源均衡因子;Q——城市水资源总量(m³);P_w——全球平均水资源生产能力(m³/hm²)。

2.1.3 水资源可持续利用指数 水资源可持续利用指数是指水生承载力与水生态足迹的比值,实际上是反应水资源的生态压力指数。水资源可持续利用指数计算公式为:

$$EQ = EC_w / EF_w \quad (3)$$

式中:EQ——水资源可持续利用指数,EQ>1 时说明水资源可持续利用程度高,EQ<1 说明水资源缺口大,水资源可持续利用程度低。

2.2 STIRPAT 模型(环境压力模型)

STIRPAT 模型^[14-15]是通过对人口,财产,技术 3 个影响因素对环境压力之间的关系进行评估。其具体形式为:

$$I_i = \eta P_i^\mu A_i^\theta T_i^\varphi \epsilon_i \quad (4)$$

式中:I_i——环境压力,本文采用贵阳市水生态足迹来表征;P_i——人口总量;A_i(affluence)本文引入的贵阳市人均 GDP;T_i——技术水平;η——模型的系数;μ,θ,φ——公式中几个因素的人文驱动力的指数;ε_i——模型误差项。由于 STIRPAT 模型是一个具有多个自变量的非线性模型,在公式两边取对数处理后:

$$\ln EF_i = \eta + \mu \ln P_i + \theta \ln A_i + \varphi \ln T_i + \ln \epsilon_i \quad (5)$$

式中:EF_i——因变量;P_i,A_i,T_i——自变量;η——常数项;ε_i——误差项,把公式(5)进行多元回归,根据弹性系数的概念,模型中 3 个自变量,如发生 1% 的变化,就会分别引起因变量各自发生 μ%,θ% 和 φ% 的变化。本模型是对人文驱动因素与环境压力的一种非线性关系假设。

本文将公式(5)中技术指标 T_i 分解为结构化指标 N_i(本文指的是第一二产业占总产值比重,因为贵阳市工农业用水比重超过总用水量的 60%)和城市化指标 M_i(城市化率)。即 lnN 和 lnM_i 两项,为了验证各驱动因素与水生态足迹间是否存在倒 U 形的 EKC 曲线^[15-16],将模型调整为公式(6):

变量 X 经标准化处理后的数据矩阵记为 E₀ = (E₀₁, ..., E_{0p})_{n × q},设 p 个自变{x₁, ..., x_p} ,因变量 Y 经过标准化处理后的数据矩阵记为 F₀ = (F₀₁, ..., F_{0q})_{n × q},并设有 q 个因变量{y₁, ..., y_q}。为了研究与 X 和 Y 的统计关系,取 n 个样本点,由此构成了自变量与因变量的数据表 X = (x₁, ..., x_p)_{n × p} 和 Y = (y₁, ..., y_q)_{n × q}。用偏最小二乘法回归分别在 X 与 Y 中提取出 t₁ 作为 x₁, ..., x_p 的线性组合和 u₁ 作为 y₁, ...,

y_q 的线性组合。在第一个成分 t_1 和 u_1 被提取后, 偏最小二乘法回归分别实施 X 对 t_1 的回归以及 Y 对 t_1 的回归。算法将会在方程达到了满意精度后终止。反之, 将利用 X 被 t_1 解释后的残余信息, 以及 Y 被 t_1 解释后的残余信息进行第二轮的成分提取, 直到能达到一个较为满意的精度为止。该方法同时采用自变量投影重要指数 VIP 判断驱动因子的重要性。一般认为, VIP 大于 1.00, 表示该自变量是因变量的重要确定驱动因公式为:

$$VIP_j = \sqrt{\frac{p}{R_d(I; t_1, \dots, t_m)} \sum_{h=1}^m R_d(I; t_h) \cdot \omega_{hj}^2} \quad (7)$$

2.4 岭回归, LASSO, 弹性网

岭回归是一种有偏估计, 其解是在 LS 回归解的基础上, 加了一个正的惩罚参数 λ , 其回归结果中包含了所有的预测变量但是没有进行变量选择, 因此影响了模型的准确性^[17]。LASSO 回归就是解决岭回归中不能进行变量选择的问题, 相对于岭回归中的二次罚 $\sum_{j=1}^p \beta_j^2$, LASSO 中的一次罚 $\lambda \sum_{j=1}^p |\beta_j|$ 收缩程度更小, 能把非 0 的预测变量系数 β_j 向 0 收缩, 因此该模型更加精确^[17]。弹性网的罚函数 $\lambda \sum_{j=1}^p \{ (1-\alpha)\beta_j^2 + \alpha|\beta_j| \}$ 恰好为岭回归函数和 LASSO 罚函数的一个凸线性组合。当 $\alpha=0$ 时, 弹性网为岭回归。当 $\alpha=1$ 时, 弹性网即为 LASSO 回归。它兼顾了岭回归, LASSO 二者的优点能更好选择变量^[19-20]。

3 数据来源及模型参数

本文数据来源于 2001—2017 年《贵阳市统计年鉴》《贵阳市水资源公报》《贵阳市国民经济和社会发展统计公报》。城市化率, 学术界有人口比重指标法, 农村城镇化指标, 现代城市化指标法, 调整序数法, 城镇土地指标法。本文采用的是人口比重指标即城镇人口占总人口的比例。本文关于水生态足迹模型计算参数, 其中计算参数参照黄林楠等^[4]的研究成果, 全球均衡因子在 40 a 内只有微小变化可采用静态值。本文取 2000 年的值作为计算参数, 水质生态足迹中的计算参数根据《地表水环境质量标准 (GB3838-2002)》, 表 1 为公式 (1) — (2) 中计算参数。

表 1 水生态足迹计算参数 t/hm²

r_w	P_{COD}	P_{NH_3}	P_w	φ
5.19	0.062 9	0.003 1	3 140	1.67

4 结果与分析

4.1 贵州省贵阳市水生态足迹时间序列

根据图 1 所示结果, 2002—2016 年贵阳市总的

水量生态足迹呈波动变化。2002—2008 年, 水量生态足迹, 大幅上升后逐步下降, 变化幅度达 $7.75 \times 10^5 \text{ hm}^2$ 。2009—2016 年贵阳市水量生态足迹整体呈缓慢上升趋势, 年均升幅不足 $6.10 \times 10^4 \text{ hm}^2$; 分别从工业、农业、生活用水生态足迹来看, 工业用水生态足迹自 2009 年起大致呈下降趋势, 但仍然是影响贵阳市总的水量生态足迹的最大因素。由于灌溉用水效率提升, 农业用地萎缩, 导致农业用水生态足迹呈逐年下降趋势。同时, 由于十二五以来, 贵阳市推进快速城市化战略, 导致城市人口激增, 生活用水生态足迹从 2002 年的 $3.39 \times 10^4 \text{ hm}^2$ 暴涨至 2016 年的 $7.13 \times 10^5 \text{ hm}^2$, 年均增幅超过 17.175%。

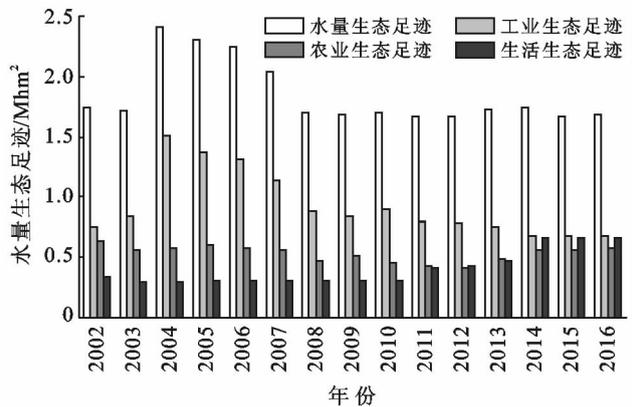


图 1 贵阳市 2002—2016 年水量生态足迹变化特征

由图 2 所示, 2002—2016 年贵阳市水质生态足迹总体呈先下降后又波动式上涨。其原因是贵阳市 2005 年起实施新的排放标准和工业企业改进排污技术。导致水质生态足迹中 COD 与氨氮排放量下降。2011—2016 年, 污水中的 COD 排放量基本稳定, 但是由于城市人口迅猛增长, 生活污水中氨氮排放量增加, 导致氨氮生态足迹上涨, 造成水质生态足迹有强烈反弹上升的趋势。

由图 3 所示, 2002—2016 年贵阳市水生态足迹在 $1.28 \times 10^7 \sim 1.60 \times 10^7 \text{ hm}^2$ 浮动, 呈先上升后微降再上升的趋势。水生态承载力的变化幅度也十分明显, 2006, 2011, 2012, 2013 年这 4 个年份是贵阳市水生态承载力的低点, 主要是由于贵州省出现大旱等极端天气, 导致水资源总量波动, 进而影响水生态承载力波动。

水资源可持续利用指数又是由水生态足迹和水生态承载力共同决定的。2002—2016 年贵阳市水资源利用指数都小于 1.0, 其变化范围 0.315~0.624。表明贵阳市水生态系统常年处在超负荷状态, 水资源持续利用面临很大压力。总体而言, 贵阳市对水资源的消费量远远超过生态系统的承载力范围, 总体处于

不可持续的状态。随着贵阳市经济发展,城市化的快速推进,水生态足迹将进一步增加,贵阳市的可持续发展形势更为严峻。

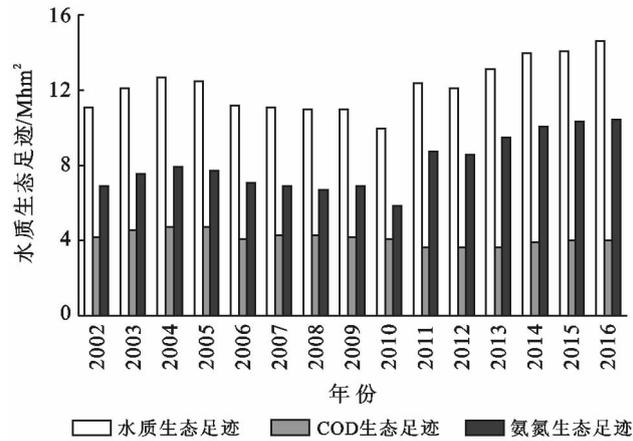


图 2 贵阳市 2002—2015 年水质生态足迹变化特征

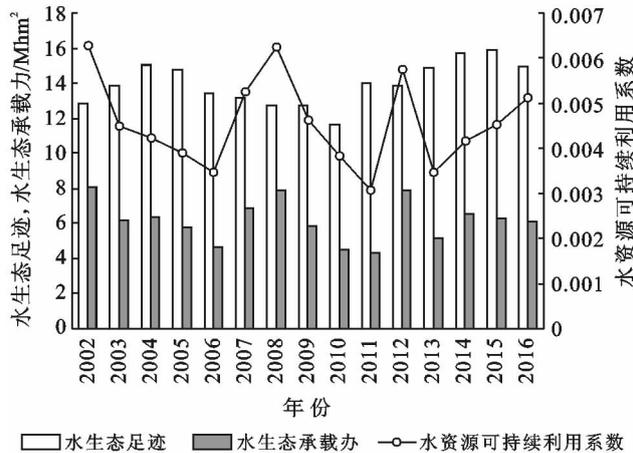


图 3 贵阳市 2002—2016 年水生态足迹、水生态承载力、水资源可持续利用指数变化

4.2 基于 STIRPAT 模型的 LS 回归分析

贵阳市水生态足迹变化的驱动因素主要包括总人口,富裕度,结构化和城市化水平。因此本文选取这 4 个指标,根据公式(6)构建贵阳市水生态足迹变化的 STIRPAT 模型。表 2 模型拟合度介于 24%~66.56%。模型 1 是以 $\ln EF$ 为因变量、 $\ln A$ 和 $(\ln A)^2$ 为自变量的 STIRPAT 方程,其调整 $R^2=0.24$,DW 统计量为 1.378, $\text{sig.}<0.05$, $(\ln A)^2$ 对 $\ln EF$ 存在显著影响。在模型 2 中添加自变量 $\ln P$, $\ln P$ 对 $\ln EF$ 不存在显著性影响。模型 3 添加了自变量 $\ln N$, 在模型 3 中, $\ln A$, $\ln N$, ($\text{sig.}<0.05$)对 $\ln EF$ 解释提升,调整 $R^2>0.24$ 。模型 4 添加自变量 $(\ln A)^2$ ($\text{sig.}<0.01$)和 $(\ln N)^2$ ($\text{sig.}<0.05$), $(\ln A)^2$ 对模型有显著性影响。模型 5,6,7 是在模型 3,4 的基础上考虑 $\ln N$ 和 $(\ln N)^2$, $\ln M$ 和 $\ln M^2$, 但这几个自变量都对 $\ln EF$ 影响不显著。模型 8 是把所有影响 $\ln EF$ 的驱动因素添到回归模型中,调整 $R^2=0.6656$ 。由表 3 得出,对 $\ln EF$ 有影响的自变量主要是 $\ln A$, $(\ln A)^2$, 和 $\ln N$, $(\ln N)^2$ 。在 8 个模型当中 $\ln A$ 的弹性系数在 $-0.546\sim-0.06$ 之间, $(\ln A)^2$ 的弹性系数在 $0.186\sim0.316$ 之间。表明 $\ln A$ 对 $\ln EF$ 存在负影响, $(\ln A)^2$ 系数为正,说明贵阳水生态足迹不存在环境 EKC 假说。 $\ln N$ 和 $(\ln N)^2$ 的弹性系数从分别是 $1.1451\sim6.325$ 和 $-1.128\sim4.086$ 。因为 $(\ln N)^2$ 的弹性系数出现负值,说明第一二产业占总产值比重在某种程度上存在环境 EKC 假说。普遍认为自变量的方差膨胀因子 $VIF>10$ 意味着变量之间具有严重的共线性。尤其是模型 7,8 共线性更为严重,完全不适合用最小二乘法 LS 进行无偏估计。表明用上述方法进行水生态足迹驱动因素分析结果的准确性受到质疑。

表 2 贵阳市水生态足迹变化的 STIRPAT 模型分析

模型	项目							
	1	2	3	4	5	6	7	8
intercept	2.668***	-1.045	3.446***	2.990**	3.665*	3.814*	-36.338	-13.117
$\ln P$		0.635						0.99
$\ln A$	-0.275	-0.281	-0.06	-0.093	-0.047	-0.49	-0.546	-0.535
$(\ln A)^2$	0.186*	0.119	0.217*	0.246**	0.202	0.277	0.316	0.227
$\ln N$			1.451*		2.161			6.325
$(\ln N)^2$				-1.128*	0.565			4.086
$\ln M$						1.475	-108.103	-37.226
$(\ln M)^2$							-74.584	-27.69
F	2.9	2.47	5.08	4.9	3.4	2.02	2.49	4.41
adjr	0.24	0.2684	0.5047	0.4935	0.4444	0.2026	0.3325	0.6656
DW	1.378	1.016	1.984	1.832	2.052	1.512	2.099	2.976
N	13	13	13	13	13	13	13	13

注: * sig.<0.05; ** sig.<0.01; *** sig.<0.001。

表 3 自变量投影重要指数(VIF)值

模型	项目							
	1	2	3	4	5	6	7	8
lnP		7.953 5						16.93
lnA	13.012 6	13.030 9	19.954 7	18.193 8	26.934	64.662 2	65.47	106.95
(lnA) ²	13.012 6	19.297 5	13.462 2	14.730 5	39.079	40.972 7	42.16	65.01
lnN			11.786 2		724.192			1 445.48
(lnN) ²				13.632 1	837.613			1 718.65
lnM							9 659.45	15 272.17

4.3 偏最小二乘法(PLS)分析结果

根据 PLS 回归普遍用于解决自变量的多重共线性问题,对上述引入所有影响水生态足迹的驱动因素(lnP,lnA、(lnA)²,lnN,(lnN)²,lnM)的模型 8 进行修正,其结果由表 4 所示,当 t=1,即提取一个主成分时,调整 R²=0.05,说明提取一个主成分时能达到对因变量 5%的解释;当 t=2 时,调整 R²=0.28,表明提取 2 个主成分能对因变量 28%的解释。同理当 t=5 为最佳成分数,建立基于 PLS 的 STIRPAT 模型,调整 R²=0.714,达到最大值。提取 5 个主成分对因变量产生 71.4%的解释,Press 的最小值达到 5.306 2,此时 lnP 的弹性系数为 0.918,当 lnP 发生 1%的变化时,将会引起 lnEF 发生 0.918 的变化,此时 lnA,(lnA)²,lnN,(lnN)²,lnM,(lnM)² 发生 1%

的变化,分别会引起 lnEF 发生 -0.492,0.282,1.223,-0.003 7,0.778 和 -1.271 的变化。但是 PLS 模型属于未标准化回归方程,其系数大小反映 5 个自变量对因变量的影响,因此需要进一步引入 VIP 值分析各因变量对自变量影响。

VIP 值结果如图 4 所示,lnP,(lnA)²>1 表明总人口,人均 GDP 及二次项是贵阳市水生态足迹变化的重要影响因素,其变化按 VIP 重要性排序为:人均 GDP 的二次项>人口。但是 R² 介于 0.05~0.714 表示各模型拟合效果不够理想,此时主成分并不能解释自变量和因变量之间的关系,基于 PLS 对 STIRPAT 模型建立的水生态足迹与其驱动因素之间的线性回归模型不合理,与文献^[21-22]研究结果不一致。

表 4 基于 PLS 的 STIRPAT 模型

Intercept	lnP	lnA	(lnA) ²	lnN	(lnN) ²	lnM	(lnM) ²	R ²	Press
1.880	0.107	0.011	0.011	-0.024	0.025	-0.054	0.035	0.050	13.374
-0.130	0.340	0.010	0.030	0.100	-0.040	-0.740	0.500	0.280	11.157
-1.982	0.882	0.025	0.074	0.819	-0.479	-0.062	-0.006	0.520	6.757
0.646	0.491	-0.056	0.179	1.079	-0.439	0.047	-0.197	0.552	6.328
-0.667	0.918	-0.492	0.282	1.224	-0.004	0.778	-1.271	0.714	5.306

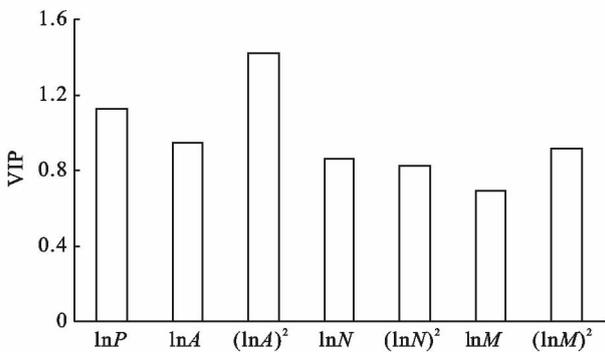


图 4 贵阳市水生态足迹变化的重要影响因素 VIP 值

4.4 基于弹性网对 PLS 的优化

偏最小二乘法适用于观测数少于预测变量数以及预测变量之间存在共线性的情况,但是计算过程需先提取主成分,在一定程度上会同时损失预测变量和

因变量的信息,提取因子的个数越少,损失的信息越多。弹性网(岭回归和 LASSO 回归为其特例)是对回归系数采用联合惩罚函数的有偏估计,既能达到变量选择的目的,又能保证群组效应,结合了岭回归和 LASSO 的优点^[17-19]。本研究纳入 2002—2016 年贵阳市水生态足迹相关数据,利用 STIRPAT 模型探究影响贵阳市水生态足迹 4 个驱动因素,研究水生态足迹和人口,富裕度,结构化,现代化之间的相关性,并探讨是否存在倒 U 形环境库兹涅茨曲线关系。本研究的模型中共纳入 15 条记录,8 个自变量,包含信息相对较少,所以使用弹性网进行数据拟合,所有统计采用 R 软件中 glmnet 包完成。

表 5 结果表明,岭回归的惩罚系数初始值较高,随着 lamda 的增加,系数的变化成波动状,并非逐渐

减小。同时按照倒 U 形结构假设,二次项的系数为正,一次项系数为负,但在岭回归结果中 $\ln N$ 的系数为负,这与模型规律不符。LASSO 和弹性网中在相同的 α 变化情况下,各回归系数的压缩趋势类似,均能逐渐收缩系数,达到变量选择的目的,但 LASSO 回归的回归收缩速率较快。此外,从对应的数据表中可见,LASSO 回归中系数中出现从 0 上升后回到 0 的波动,同时 $\ln A$ 和 $(\ln A)^2$, $\ln N$ 和 $(\ln N)^2$, 以及 $\ln M$ 和 $(\ln M)^2$ 构成 3 个不同群组,但三者的回归系数均在 α 细微的增加情况快速收缩到 0,模型结果无法合理解释。弹性网的结果显示回归系数收缩平滑,并未显示出群组效应,在惩罚系数较大时仍有部分指标系数不为 0。其中人口因素 $\ln P$ 在不同的模型的不同惩罚系数中均持续不为 0,表示其重要作用。最后对在 $\alpha=0.5$ 的条件下进行弹性网

的交叉验证,根据 glmnet 包的建议,选择 lambda 在一个标准误差的情况下为最优条件,选择 $\lambda=0.008$ 是为最优的模型结果。此时 $\ln P, (\ln A)^2, \ln N, (\ln N)^2, \ln M, (\ln M)^2$ 在 STIRPAT 模型当中是有意义的,此时 $\ln P, (\ln A)^2, \ln N, (\ln N)^2, \ln M, (\ln M)^2$ 的弹性系数分别为 0.484, 0.428, -0.097, -0.187, 0.036。由岭回归、LASSO、弹性网 3 种回归的系数变化图可知(图略),当 $\ln P$ 发生 1% 变化时, $\ln EF$ 发生 0.484%, 0.428%, -0.097%, -0.187%, 0.036% 的变化。由于 $\ln N, (\ln N)^2$ 的系数一个出现正值,一个出现负值表明贵阳市的水生态足迹与第一二产业占总产值比重间存在环境 EKC 假说。 $\ln M, (\ln M)^2$ 的系数同样是一个正值,一个负值,这同样表明贵阳市水生态足迹与城市化水平之间存在环境 EKC 曲线。

表 5 统一惩罚系数变化在岭回归、LASSO、弹性网 3 种模型中回归系数的影响

回归类型	Alpha	$\ln P$	$\ln A$	$(\ln A)^2$	$\ln N$	$(\ln N)^2$	$\ln M$	$(\ln M)^2$
岭回归	0.002	0.656 645	-0.065 21	0.116 996	0.803 592	-0.305 45	-0.003 8	-0.158 8
	0.004	0.597 564	-0.019 48	0.085 338	0.639 999	-0.296 88	-0.124 6	-0.003 95
	0.006	0.551 564	-0.003 84	0.069 464	0.542 322	-0.267 09	-0.183 68	0.062 425
	0.008	0.512 635	0.002 951	0.059 551	0.470 860	-0.238 77	-0.218 01	0.100 393
	0.01	0.479 168	0.006 433	0.052 621	0.415 788	-0.214 20	-0.240 69	0.124 772
	0.012	0.450 238	0.008 338	0.047 437	0.371 774	-0.193 15	-0.256 16	0.141 395
	0.014	0.424 864	0.009 455	0.043 395	0.335 641	-0.175 20	-0.267 03	0.153 198
	0.016	0.402 616	0.010 088	0.040 123	0.305 46	-0.159 70	-0.274 48	0.161 622
	0.018	0.382 836	0.010 457	0.037 42	0.279 811	-0.146 30	-0.279 56	0.167 719
	0.02	0.365 1	0.010 668	0.035 146	0.257 717	-0.134 63	-0.282 89	0.172 128
LASSO	0.002	0.570 178	0	0.101 883	1.288 746	0	0	0
	0.004	0.564 474	0	0.069 684	0.992 439	0	0	0
	0.006	0.551 144	0	0.037 375	0.676 097	0	-0.060 26	0
	0.008	0.538 352	0	0.004 993	0.357 359	0	-0.132 04	0
	0.01	0.408 721	0	0	0.162 260	0	-0.196 26	0
	0.012	0.264 691	0	0	0	0	-0.247 35	0
	0.014	0.227 23	0	0	0	0	-0.109 50	0
	0.016	0.193 65	0	0	0	0	0	0
	0.018	0.174 942	0	0	0	0	0	0
	0.02	0.156 233	0	0	0	0	0	0
弹性网	0.002	0.598 378	-0.024 44	0.107 062	0.984 185	-0.184 97	0	0
	0.004	0.562 723	0	0.075 87	0.768 685	-0.183 98	-0.128 37	0
	0.006	0.526 13	0	0.057 229	0.609 699	-0.141 18	-0.185 51	0.004 725
	0.008	0.484 469	0	0.042 797	0.489 235	-0.096 6	-0.187 00	0.035 77
	0.01	0.441 682	0	0.031 278	0.392 470	-0.055 13	-0.190 05	0.055 679
	0.012	0.400 588	0	0.021 624	0.310 608	-0.017 95	-0.191 81	0.068 665
	0.014	0.360 846	0	0.014 413	0.229 036	0	-0.189 90	0.075 285
	0.016	0.326 53	0	0.008 831	0.146 503	0	-0.182 47	0.076 678
	0.018	0.294 563	0	0.004 064	0.075 163	0	-0.172 53	0.075 087
	0.02	0.264 629	0	0	0.013 173	0	-0.162 58	0.070 142

5 结论

(1) 贵阳市的水生态足迹是远大于水生态承载力的,这与该市地处喀斯特发育严重地区有关。其特点是土层稀薄脆弱,岩石渗透严重,地表蓄水能力弱,工程性缺水明显,导致整个市水资源利用效率低。农业水生态足迹受到自然条件的巨大影响。工业主导的水生态足迹存在大量的 COD 以及氨氮排放的,同时由于地表与地下属典型的喀斯特地貌,土层渗透系数较高,地表水与地下水形成联动污染,导致水体污染严重。而且在喀斯特地型状况下,岩体破碎,平缓沉积地形缺乏,虽然整个城市划分 6 个区,但是建成区域异常密集且坡度大,城市人口密集导致供水紧张。由于贵阳市正处于快速城市化和社会经济的高速发展期,虽然在 2002—2008 年生活用水水生态足迹有小幅波动,但是整体处于飞速上涨期。贵阳市水生态足迹的潜在因素表现在人口的增加、产业结构不够合理和城市化水平的过快增长。

(2) 根据 STIRPAT 建模分析贵阳市水生态足迹变化的重要驱动因素,其 LS(最小二乘法)回归结果如下:人均 GDP 及二次项,第一二产业占总产值比重是贵阳市生态足迹变化最重要驱动因素,但是对水生态足迹有重大影响的总人口、城市化指标被排除在外。VIF 值的计算结果表明各个驱动因子之间存在多重共线性。之后采用 PLS(偏最小二乘法)对模型进行修正,当 $t=5$ 时, $R^2=0.714$,得出总人口,人均 GDP 及其二次项是影响贵阳市水生态足迹的总要因素。按 VIP 重要性排序为:人均 GDP 的二次项 > 人口,一个城市的水生态足迹仅仅只有这 2 个影响因素是不符合常理的。于是本文引入弹性网回归对其模型进行佐证,其回归系数采用联合惩罚函数的有偏估计,达到了变量选择目的,又能保证群组效应。 $\lambda=0.008$ 是为最优的模型结果, $\ln P$, $(\ln A)^2$, $\ln N$, $(\ln N)^2$, $\ln M$, $(\ln M)^2$ 在 STIRPAT 模型当中是有意义的。即贵阳市总人口,人均 GDP,第一二产业占总产业比值,城市化率都对贵阳市水生态足迹产生影响。而且第一二产业占总产业比值与贵阳市的水生态足迹间存在环境 EKC 假说,城市化与贵阳市的水生态足迹间也存在环境 EKC 假说。

(3) 由于第一二产业占总产值比重与城市化率和贵阳市水生态足迹存在环境 EKC 假说。虽然受整个论文思路所限,本文没有进一步讨论环境 EKC 假说其临界点到底在哪里。但是这 2 个驱动因素出现了环境 EKC 假说,就表明整个贵阳市在推进快速城镇化的发展过程中需要调整步伐。进一步优化产

业结构,优化水资源配置,按序调控。在优先保证生活用水前提下,合理规划和保障社会经济发展的用水,保证一定的生态用水。把工业、农业的发展规模、区域产业结构定位与水资源的配置结合起来,引导与水环境承载力相协调的发展布局。并推动涉水产业绿色转型与发展,积极发展生态绿色农业,大力降低农业面源污染。工业区应积极发展循环经济,开展清洁生产,中水回用,实行用水梯级循环,不断降低废水污染物排放量,淘汰废水排放量大的落后产能。城市倡导节水生活方式,另在生活污染为主的河流,加大污水收集范围,提高生活污水处理率,减少污水中氨氮排放量。并实施河道垃圾清理、清淤疏浚、生态修复等综合措施,全面降低污染物入河负荷,逐步改善水质。

[参 考 文 献]

- [1] Mathis W, William E R. Perceptual and structural barriers to investing in natural capital: Economics from an ecological footprint perspective[J]. *Ecological Economics*, 1997,20(1):3-24.
- [2] Wackernagel M, Rees W E. Our ecological footprint: Reducing human impact on the earth[J]. *Population & Environment*, 1998,1(3):171-174.
- [3] Hoekstra A Y, Hung P Q. Virtual water trade: A quantification of virtual water flows between nations in relation to international corp trade[J]. *Value of Water Research Report Series*, 2002,11:239-304.
- [4] 黄林楠,张伟新,姜翠玲,等. 水资源生态足迹计算方法[J]. *生态学报*, 2008, 28(3):1279-1286.
- [5] 潘真真,苏维词,王建伟,等. 基于生态系统供给及净化服务功能的贵州省水生态占用研究[J]. *环境科学学报*, 2017,37(7):2786-2796.
- [6] 胡永红,吴志峰,李定强,等. 基于 ARIMA 模型的区域水生态足迹时间序列分析[J]. *生态环境学报*, 2006, 15(1):94-98.
- [7] 曹坤,刘素霞. 上海市水资源生态足迹分析与 SVR 预测[J]. *华东师范大学学报:自然科学版*, 2016,2016(4):139-149.
- [8] 杨振华,苏维词,赵卫权,等. 基于 GRNN 模型的岩溶地区城市水生态足迹分析与预测[J]. *中国岩溶*, 2016, 35(1):36-42.
- [9] 徐建伟. 基于水资源水环境双重约束的产业结构优化方法研究[D]. 北京:中国环境科学研究院,2016.
- [10] 郭晓娜,苏维词,杨振华,等. 城乡统筹背景下重庆市水生态足迹分析及预测[J]. *灌溉排水学报*, 2017,36(2):69-75.
- [11] 王惠文,吴载斌,孟洁. 偏最小二乘回归的线性与非线性方法[M]. 北京:国防工业出版社,2006.

用的技术方法相比:①本研究运用遥感监测加调查单元野外复核的方法,以高分辨率遥感影像为信息源,来确定调查单元的土地利用状况和水土保持措施;水利普查直接采用人工野外调查的方法划分图斑。前者消除了后者的人工误差,在工作精度上得到了提高。②本研究选择了4%的调查单元抽样密度,高于水利普查1%的密度,基本覆盖了铜山区和赣榆区的全部行政区域,更加具有代表性,也提高了插值计算结果的可靠性。③本研究调查单元图斑的上图面积(1 000 m²)要小于水利普查,避免了某一调查单元中小图斑被大图斑合并的状况。今后,该研究方法通过进一步优化因子取值和土壤流失方程参数选取,能为江苏省水土流失动态监测提供高效可行的技术方法。

[参 考 文 献]

- [1] 江苏省水利厅. 关于发布江苏省省级水土流失重点预防区和重点治理区的公告(苏水农[2014]48号)[R]. 江苏南京:江苏省水利厅,2014.
- [2] 李晓松,姬翠翠,曾源,等. 基于遥感和GIS的水土流失动态监测:以河北省赤城县为例[J]. 生态学杂志,2009,28(9):1723-1729.
- [3] 刘耀林,罗志军. 基于GIS的小流域水土流失遥感定量监测研究[J]. 武汉大学学报:信息科学版,2006,31(1):35-38.
- [4] 周斌,杨柏林,洪业汤,等. 基于GIS的岩溶地区水土流失遥感定量监测研究:以贵州省(原)安顺市为例[J]. 矿物学报,2000,20(1):13-21.
- [5] 叶健. 江苏省第一次全国水利普查成果汇编[M]. 江苏南京:河海大学出版社,2015.
- [6] Renard K G. RUSLE Agricultural Manual NR. 703; Predicting Soil Erosion by Water: A Guide to Conservation Planning with the Revised Universal Soil Loss Equation (RUSLE) [M]. USA Springfield: United States Department of Agriculture, 1992.
- [7] McCool D K. Revised slope steepness factor for the universal soil loss equation[J]. Transactions of the American Society of Agricultural Engineers, 1987, 30(5): 1388-1396.
- [8] Wischmeier W H, Smith D D. Agriculture Handbook No. 537; Predicting Rainfall Erosion Losses: A Guide to Conservation Planning with the Universal Soil Loss Equation(USLE)[M]. USA Springfield: United States Department of Agriculture, 1978.
- [9] 符素华,吴敬东,段淑怀,等. 北京密云石匣小流域水土保持措施对土壤侵蚀的影响研究[J]. 水土保持学报,2001,15(2):21-24.
- [10] 张岩,袁建平,刘宝元. 土壤侵蚀预报模型中的植被覆盖与管理因子研究进展[J]. 应用生态学报,2002,13(8):1033-1036.
- [11] 林杰,张金池,顾哲衍,等. 基于叶面积指数的植被覆盖管理措施因子C的遥感定量估算[J]. 林业科学,2013,49(2):86-92.
- [12] 李叶鹏. 对水土保持措施及其减水减沙效益分析[J]. 中国房地产业,2015(S2):114.
- [13] 朱虹,唐婷婷,蔡艳蓉. 土壤侵蚀预报模型中的水土保持措施因子研究进展[J]. 科技展望,2015(21):222-224.
- [14] 章文波,刘宝元. 基于GIS的中国土壤侵蚀预报信息系统[J]. 水土保持学报,2003,17(2):89-92.
- [15] 柯克比. 土壤侵蚀[M]. 北京:水利电力出版社,1987.
- [12] 袁道先,薛禹群,傅家谟,等. 防止我国西南岩溶地区地下河变成“下水道”的对策和建议[R]. 北京:中国科学院院士建议,2007:1-16.
- [13] Ayres R U. Commentary on the utility of the ecological footprint concept[J]. Ecological Economics, 2000, 3(3):347-349.
- [14] 于冰,徐琳瑜. 城市水生态系统可持续发展评价:以大连市为例[J]. 资源科学,2014,36(12):2578-2583.
- [15] Tang Zhong, Xiang Hao. Analysis of major driving forces of ecological footprint based on the STIRPAT model and RR method: A case of Sichuan Province, Southwest China[J]. Journal of Mountain Science, 2011,8(4):611-618.
- [16] 燕华,郭运功,林逢春. 基于STIRPAT模型分析CO₂控制下上海城市发展模式[J]. 地理学报,2010,65(8):983-990.
- [17] 杨万平,袁晓玲. 环境库兹涅茨曲线假说在中国的经验研究[J]. 长江流域资源与环境,2009,18(8):704-710.
- [18] 张哲,梁冯珍. 基于弹性网回归的居民消费价格指数分析[J]. 哈尔滨商业大学学报:自然科学版,2013,29(5):592-597.
- [19] De Mol C, De Vito E, Rosasco L. Elastic-net regularization in learning theory[J]. Journal of Complexity, 2009,25(2):201-230.
- [20] Zou H, Hastie T. Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society, 2005, 67(2):301-320.
- [21] 贾俊松. 河南生态足迹驱动因素的Hi_PLS分析及其发展对策[J]. 生态学报,2011,31(8):2188-2195.
- [22] 肖思思,黄贤金,吴春笃. 江苏省生态足迹时间维度变化及其驱动因素分析:基于PLS方法对STIRPAT模型的修正[J]. 地理与地理信息科学,2012,28(3):76-82.

(上接第227页)