

DOI:10.13476/j.cnki.nsbdkj.2019.0116

陈帅,于鲁冀,万红友,等.基于耦合模型的流域污染负荷分配实例[J].南水北调与水利科技,2019,17(5):108-114. CHEN S, YU L J, WAN H Y, et al. Case study of watershed pollution load distribution based on coupling model[J]. South-to-North Water Transfers and Water Science & Technology, 2019, 17(5): 108-114. (in Chinese)

基于耦合模型的流域污染负荷分配实例

陈 帅¹, 于鲁冀¹, 万红友¹, 王燕鹏², 赵雪霞¹

(1. 郑州大学 水利与环境学院, 郑州 450001; 2. 郑州大学 环境政策规划评价研究中心, 郑州 450003)

摘要:流域污染负荷分配是实施流域精细化管理的重要措施。在指标选取过程中,为了定量描述评价指标对流域污染负荷分配的重要性,引入随机森林算法对各指标作重要性评价,依据重要性程度筛选涵盖全部影响因素的评价指标,并进一步确定各指标权重。将评价指标体系与环境基尼系数模型相结合,构建污染负荷分配耦合模型,并以清溪河流域(许昌段)为例,制定基于控制单元的COD负荷分配方案。结果表明,满足重要性评分要求的指标分别是居民人均收入、单位GDP污水产生量、工业企业水消费量、地表水资源量、控制单元面积,评分分别为13.97、9.65、8.77、5.21、1.09。依据耦合模型制定的2020年清溪河流域(许昌段)各控制单元COD负荷分配方案中,控制单元QYH-3的削减量和削减率均最大,分别为617.14 t/a和30.44%。分配方案基本符合流域实际,结果较为科学合理。

关键词:评价指标体系;随机森林;环境基尼系数;污染负荷分配

中图分类号:X522 文献标志码:A 开放科学(资源服务)标识码(OSID):



Case study of watershed pollution load distribution based on coupling model

CHEN Shuai¹, YU Luji¹, WAN Hongyou¹, WANG Yanpeng², ZHAO Xuexia¹

(1. School of Water Conservancy and Environment, Zhengzhou University, Zhengzhou 450001, China;

2. Research Center for Environmental Policy Planning & Assessment, Zhengzhou University, Zhengzhou 450003, China)

Abstract: Watershed pollution load distribution is an important measure to implement watershed fine management. In order to quantitatively describe the significance of assessment indicators on the distribution of pollution load in river basin, the random forest algorithm was adopted to evaluate the importance of the selected indicators. The evaluation indices, which can reflect all the influence factors, were selected based on their importance, then, the weight coefficient of each indicator was further determined based on its score. The coupling model of pollution load distribution is constructed by combining the evaluation index system with the environmental gini coefficient model. Additionally, allocation plan of COD load was formulated based on control unit in the target Qingyi river basin (Xuchang Section). We found that the indicators, which fulfill the requirements of importance score, include per capita income, sewage production per unit of GDP, water consumption of industrial and enterprises, the amount of surface water resources and area of control-unit, and the scores are 13.97, 9.65, 8.77, 5.21 and 1.09, respectively. The allocation plan of COD load in the river basin on 2020 was further formulated based on the coupling model, and the pollutant reduction and reduction rate of the control-unit QYH-3 are the largest, which are 617.14 t/a and 30.44%, respectively. The allocation plan is reasonable for reflecting the actual situation of the basin.

Key words: evaluation index system; random forest; environmental Gini coefficient; pollution load distribution

收稿日期:2019-03-18 修回日期:2019-05-05 网络出版时间:2019-06-10

网络出版地址: <http://kns.cnki.net/kcms/detail/13.1334.TV.20190606.1614.004.html>

基金项目:国家水体污染控制与治理科技重大专项项目(2015ZX07204-002-005)

作者简介:陈 帅(1995—),男,河南兰考人,主要从事水污染控制理论与技术研究。E-mail: cshuai@95.com

通信作者:于鲁冀(1962—),男,河南开封人,教授,主要从事环境经济政策,水生态净化与修复等方面研究。E-mail: yuluji@126.com

流域污染负荷分配是我国实施水污染总量控制的重要举措^[1-3],关于污染负荷分配的研究,传统的方法包括层次分析法^[4]和等比例削减法^[5]等,此类方法简单却缺乏对公平和效率的考虑;后有学者提出单目标构建模型的方法,如按贡献率^[6]、费用最小和水质改善最优分配^[7-8]等,这些方法是将各排污单元的排放量进行单目标优化,难以适应复杂的流域现状^[9]。为了克服这一问题,有学者提出模糊数学法^[10]、多目标决策法^[11]、非支配多种群蚁群算法^[12]和基于模糊目标的遗传算法^[13]等,用来解决污染负荷多目标优化问题,在兼顾流域现状的同时,也存在指标选择和权重设计的主观性较强,模型设计过于复杂导致实用性较差的问题^[14]。基于基尼系数法^[15-17]的污染负荷分配方法操作性强,能满足负荷分配对公平性的要求,然而在实际研究中对于评价指标的选取,往往参考以往的经验^[18],定性地给出模型的评价指标,忽视了各指标对流域负荷分配的重要性评价,使得模型难以兼顾流域水环境现状。

本文综合考虑影响污染负荷分配的社会、经济、环境、技术发展 4 个因素,共选取清漯河流域(许昌段)14 个备选指标,同时引入随机森林算法,构建

基于环境基尼系数法的污染负荷分配耦合模型,对各指标近 5 年的数据进行回归分析,依据随机森林算法定量评价各指标的重要性,进而将 4 个影响因素中得分相对较高的指标构建本流域污染负荷评价指标体系。以 2014 年许昌市统计数据为基础,制定清漯河流域(许昌段)控制单元的污染负荷分配方案。

1 材料与方法

1.1 研究区域概况

清漯河是沙颍河流域的一条重要水系,发源于新郑市沟草园,于鄢陵县的陶城闸下游汇入沙颍河,全长 149 km,流域面积 2 362 km²,清漯河(许昌段)面积为 1 585 km²,属于小尺度流域,流经的行政区主要包括禹州市、长葛市、建安区及魏都区,主要汇入支流有石梁河、灞陵河和小泥河等。清漯河流域天然径流匮乏,受人工干扰程度较强,工业化和城镇化带来的工业污染与农业面源污染叠加,具有典型的混合型流域特征^[19]。清漯河流域(许昌段)2014 年底总人口 344.91 万人,国民生产总值为 1 411.16 亿元,2014 年排入流域的 COD 负荷为 1.125 万 t,导致流域 COD 负荷超标严重。各控制单元基本信息见图 1。



图 1 清漯河流域(许昌段)基本概况

Fig. 1 Location and basic information of the Qingyi river basin (Xuchang Section)

1.2 环境基尼系数法原理

基尼系数是经济学家用来衡量一个国家或地区居民收入分配现状的指标^[20]。基尼系数可以反映居民收入分配的不公平程度,已经成为表征国家或地区居民收入分配差距的一项重要指标^[21-23]。在经济学中,受制于社会经济发 展的局限性,基尼系数低于 0.2 的可能性很小;在环境领域中,由于所选指标

存在资源分配公平的情况,因此环境基尼系数可以趋近于 0;本研究中,为了突出分配方案的公平性将环境基尼系数的合理范围设置为 0~0.2^[16]。

1.3 环境基尼系数的计算

基尼系数的计算方法很多,这里采用梯形面积法计算^[24]。将流域控制单元作为计算环境基尼系数的目标,并依据各控制单元对应的污染负荷按照

升序排列,计算各控制单元的评价指标累积百分比作为洛伦兹曲线的横坐标,对应的污染负荷累积百分比作为纵坐标,绘制洛伦兹曲线,并计算出环境基尼系数。

$$G_j = 1 - \sum_{i=1}^n (X_{j(i)} - X_{j(i-1)}) \times (Y_{j(i)} - Y_{j(i-1)}) \quad (1)$$

式中: j 为各评价指标编号; i 为控制单元编号; G_j 为指标 j 的环境基尼系数; $X_{j(i)}$ 为 i 个控制单元 j 指标的累积百分比(%); $Y_{j(i)}$ 为第 i 个控制单元的污染负荷累积百分比(%);当 $i=0$ 时, $(X_{j(i)}, Y_{j(i-1)})$ 为 $(0, 0)$ 。

1.4 随机森林算法模型

随机森林算法^[25]是一种基于组合决策树的回归与分类模型,其用 bagging 的方法从原始数据中随机挑选样本组合^[26],产生不同的训练集,对于每一个新的训练集,利用随机特征选取的方法生成决策树,进行不剪枝式生长^[27]。它运行速度快,并且在部分数据缺失的情况下仍有良好的拟合性^[28]。

随机森林算法在给数据分类的同时,可以给出各指标在分类或回归分析中的重要性排序^[29]。在一次分类过程中,未被选取的数据被称为袋外数据(out of bag),将袋外数据作为测试样本可以对每一颗决策树生成误差估计,将随机森林中的每棵决策树的误差估计取平均值,可得到泛化误差估计,以评估各个变量在决策中的重要程度,用平均精度减少量(Mean Decrease Accuracy)表示^[30]。利用袋外数据计算指标重要性程度实现过程如下。

(1)对于每棵决策树,利用袋外数据进行误差估计并记录下来。其对每棵树的误差表示为: $\text{vote}_1, \text{vote}_2, \dots, \text{vote}_b$;

(2)随机变换指标数据,形成新的袋外数据,再利用袋外数据对每棵决策树的结果进行验证,其每个指标的误差是: $\text{vote}_{11}, \text{vote}_{12}, \dots, \text{vote}_{1b}$ 。

(3)对于指标来说,计算其重要性就是变换后的误差与原来相比差的均值,即

$$V(X_j) = \sum_{j=1}^k (\text{vote}_{ij} - \text{vote}_j) \times b^{-1} \quad (2)$$

式中:指标 X_j 的改变导致 $V(X_j)$ 值变大说明该指标越重要; vote 表示决策树对袋外数据的误差; b 表示样本个数; k 表示指标个数。

随机森林在得到各指标的评分后,根据各指标的评分,按降序排列,在每个影响因素中保留得分大于 0 的指标,构建的评价指标体系为保证体系的全面性,应涵盖全部影响因素。

为了表征各评价指标对污染负荷分配的重要程度,基于重要性评价计算各指标的权重系数为

$$w_j = V(X_j) \times \left(\sum_{j=1}^k V(X_j) \right)^{-1} \quad (3)$$

式中: w_j 表示第 j 个指标的权重系数。

1.5 污染负荷优化分配耦合模型

在流域污染负荷分配方案制定的过程中,指标的选取是首先要面临的问题,而构建涵盖流域基本概况的评价指标体系是决定方案合理性的关键。然而在实际操作中,这类指标往往数目众多,这就对统计信息匮乏型的流域带来了很大的困扰。针对这一现实情况,基于环境基尼系数模型,耦合随机森林模型,通过算法完成指标的重要性评价,构建低维度的评价指标体系,制定符合流域现状的污染负荷分配方案。

以各评价指标环境基尼系数总和最小为目标函数,各控制单元的污染负荷分配量为决策变量,在污染负荷削减总量、各指标现状环境基尼系数和各控制单元削减比例为约束条件下,建立数学模型,并在 MATLAB 中编程求最优解,从而确定各控制单元的优化分配方案,主要计算公式如下。

目标函数

$$\min F = \sum_{i=1}^5 w_i \times G_i \quad (4)$$

各指标现状环境基尼系数约束

$$G_i \leq G_{oi} \quad (5)$$

污染负荷削减约束

$$\sum_{j=1}^9 E_j = R \times \sum_{j=1}^9 E_{oj} \quad (6)$$

各控制单元的削减比例约束

$$\min_R \leq e_j = \frac{E_j - E_{oj}}{E_{oj}} \leq \max_R \quad (7)$$

式中: w_i 为各评价指标的权重; G_{oi} 为初始环境基尼系数值; G_i 为污染负荷优化分配后 j 指标对应环境基尼系数值; e_j 为污染负荷优化分配后第 j 单元的负荷削减率; E_j 为污染负荷优化分配后第 j 单元的污染排放量; E_{oj} 为第 j 单元的现状排放量; R 为流域污染负荷总体削减率; \min_R 和 \max_R 分别为各单元污染负荷削减率上下限。

2 结果与讨论

2.1 指标筛选

指标筛选是运用环境基尼系数法进行流域污染负荷分配的关键步骤^[23]。影响污染负荷分配的因素主要涉及社会、经济、资源、科技发展 4 个方面:(1)社会因素,主要从控制单元人口数、人口自然增长率、非农业人口比重、登记失业率、控制单元面积等 5 项指标来衡量控制区的社会现状水平,体现了

人类在生存繁衍过程中污染负荷排放的差异;(2)经济因素,经济发展对污染物排放量的差异主要体现在国民收入方面,选取国内生产总值、规模以上企业数量、居民人均收入、第三产业比率 4 个指标来衡量各控制单元的经济发展水平;(3)资源环境因素,选取地表水资源量、废水排放总量、废水处理总量 3 个指标来衡量环境因素;(4)科技发展因素,科技进步对污染物排放有着直接的影响,选取生活污水处理率、单位 GDP 污水产生量、环境治理投资占 GDP 比重等 3 个指标来表征。将不同影响因素包含的 14 个指标编号并作为备选指标,具体信息见表 1。

表 1 污染负荷分配备选指标

Tab. 1 Alternative indicators for pollution load distribution

影响因素	指标	符号
社会因素	控制单元人口数	a_1
	人口自然增长率	a_2
	城镇化率	a_3
	控制单元面积	a_4
经济因素	国内生产总值	b_1
	规模以上企业数量	b_2
	居民人均收入	b_3
	第三产业比率	b_4
资源环境因素	废水排放总量	c_1
	污水处理总量	c_2
	地表水资源量	c_3
科技发展因素	工业企业水消费量	d_1
	单位 GDP 污水产生量	d_2
	环境投资占 GDP 比重	d_3

2.2 流域负荷分配指标筛选与重要性分析

清渭河流域污染负荷指标的数据来自于 2009 年至 2014 年许昌市统计年鉴。14 个指标 9 个控制单元近 5 年来的指标数据构成机器学习的初始样本集,随机森林算法在开源软件 R 中运行 random Forest 程序包,得出各指标与污染负荷分配的重要性排序,选取重要性得分大于 0 并且能涵盖全部影响因素的指标,作为环境基尼系数污染负荷分配的评价指标。各影响因素的指标重要性排序见图 2。

由图 2 全体指标重要性排序可知,在影响污染负荷分配的 4 个因素中得分大于 0 的指标分别是 b_3 (居民人均收入)、 c_3 (地表水资源量)、 d_1 (工业企业水消费量)、 d_2 (单位 GDP 污水产生量)和 a_4 (控制单元面积),其中得分最高的是 b_3 ,最低是 a_1 。科技发展因素中有指标 d_1 和 d_2 入选。将选取的 5 个指标作为评价指标再次引入随机森林中得出重要性并评价,并依据公式(2)计算各指标的权重系数,见表 2。

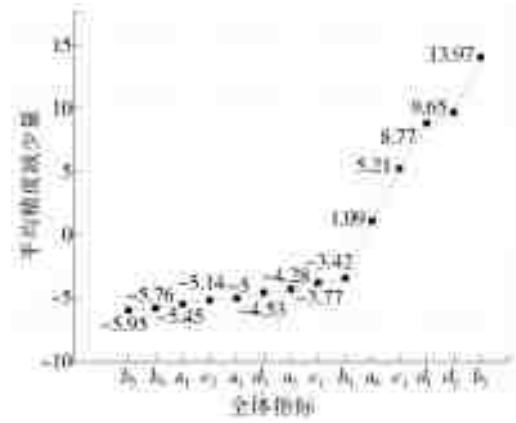


图 2 全体指标重要性排序

Fig. 2 Ordination graph of the importance of all the indicators

表 2 评价指标重要性大小与权重

Tab. 2 The importance and weight coefficient of the assessment indicators

指标	a_4	c_3	d_1	d_2	b_3
重要性评分	0.13	5.43	6.11	12.89	12.98
权重	0.003	0.145	0.163	0.343	0.346

在 5 个评价指标中关于科技发展因素的指标有 2 个,总体权重达到 50.6%,经济因素的指标权重为 34.6%,位居第二。在 4 个影响因素中科技发展因素的指标权重远高于另外 3 个影响因素,这与以往的分配方案强调社会因素和经济因素对污染负荷分配的影响不同,随着经济发展以及社会在科技领域的资金投入,科技发展对于流域污染负荷分配的影响也越来越大,这也符合自“十二五”以来许昌市建设生态文明城市的规划。

2.3 公平性分析

环境基尼系数可以量化流域污染物分布的公平性现状。首先,根据公式(1)计算基于评价指标和各控制单元对应的 COD 负荷现状环境基尼系数,结果见表 3。

表 3 现状环境基尼系数

Tab. 3 Current status of environment gini coefficient

评价指标	现状基尼系数
控制单元面积	0.36
居民人均收入	0.34
地表水资源量	0.40
工业企业水消费量	0.48
单位 GDP 污水产生量	0.55

在清渭河流域(许昌段)现状 COD 负荷分配中,各评价指标的环境基尼系数均大于 0.3,处于比较不公平性的状态,说明流域的 COD 负荷分布与流域环境现状存在较大的冲突。其中指标 GDP

污水产生量指标的现状环境基尼系数达到了 0.55,其次是工业企业水消费量为 0.48,评价结果表明各控制单元的 COD 负荷现状分布格局不公平性与当地的科技发展水平和经济发展状况有很大的相关性,这一公平性分布现状也与指标的重要性评价一致。

2.4 流域 COD 负荷优化分配

根据许昌市的“十三五”节能减排规划,到 2020 年 COD 负荷削减率为 18.4%,为保证分配方案的

可实施性,各控制单元的削减率应为[1%,40%]。在保证计算过程中各评价指标按照 COD 负荷降序排列的前提下,根据污染负荷分配耦合模型中式(4)至(7),以环境基尼系数之和最小为目标函数,在 MATLAB 软件中进行模型编程,求解最优分配削减率和污染物分配量。计算出不同控制单元 COD 的削减率和分配排放量,结果见表 4。由分配方案中 COD 负荷分配排放量,根据公式(1)计算优化后的环境基尼系数,结果见表 5。

表 4 各控制单元 COD 负荷分配方案

Tab. 4 The allocation plan of COD load in each control unit

控制单元	现状排放量/(t·a ⁻¹)	削减量/(t·a ⁻¹)	削减率/%	分配排放量/(t·a ⁻¹)	
长葛市	QYH-1	2 175.27	515.10	24.51	1 586.64
	SLH-2	1 063.26	72.09	6.56	1 027.11
	小计	3 238.53	587.19	18.34	2 613.75
魏都区	QYH-3	2 027.39	617.14	30.44	1 410.25
	QNH-1	1 338.23	226.43	15.85	1 202.27
	小计	3 365.62	843.57	24.41	2 612.52
建安区	QYH-2	418.35	14.22	3.93	347.57
	QYH-4	126.23	0.03	0.02	117.67
	XNH-1	1 463.17	148.66	10.52	1 265.06
	XHH-1	1 749.39	355.13	21.01	1 335.13
	小计	3 757.14	518.03	14.46	3 065.43
禹州市	SLH-1	888.03	120.24	11.93	887.85
	小计	888.03	120.24	11.93	887.85
总计	11 249.32	2 069.03	18.39	9 179.55	

表 5 优化后环境基尼系数

Tab. 5 Optimized environment gini coefficient

评价指标	优化后基尼系数
控制单元面积	0.10
居民人均收入	0.11
地表水资源量	0.20
工业企业水消费量	0.37
单位 GDP 污水产生量	0.30

由表 5 可知,经过优化后环境基尼系数均有明显的降低,并且不存在环境基尼系数大于 0.4 这样的极端不公平情况。从表 4 中 COD 削减分配方案可见,流域各控制单元的削减率存在较大的波动,并且最终的削减分配方案并不是现状排放量越多削减量就越大。例如,位于长葛市的控制单元 SLH-2 的现状排放量为 1 063.26 t/a,高于禹州市 SLH-1 的 888.03 t/a 的排放量,而 SLH-2 的分配的削减量为 72.09 t/a,小于 SLH-1。这是因为控制单元 SLH-1 在 2014 年的人均 GDP 为 38 376 元/人远高于 SLH-2 的人均 GDP,为了不给 SLH-2 地区的经济带来过

重的负担,削减量不宜过大。而在衡量科技发展因素的指标中,SLH-1 的单位 GDP 污水产生量为 2 761 t/万元远大于 SLH-2,可见 SLH-1 地区的人均 GDP 虽然较高,但经济结构并不合理,分配较高的削减量,也是为了倒逼企业转型,达到改善经济结构的目的。

3 结 论

(1)本文在环境基尼系数模型的基础上引入随机森林模型,对颍河流域(许昌段)14 个指标的 5 年数据进行处理,依据其特有的变量重要评价,定量评价并构建污染负荷分配评价指标体系,实现了评价指标的定量化选取。借助指标体系的重要性评分,计算出各指标的权重,并且由于随机森林算法对于缺失数据有极好容忍度,避免了人为判断中由于信息缺失导致的偏差。

(2)清颍河流域(许昌段)2014 年的现状环境基尼系数表明,科技发展因素指标的环境基尼系数均超过 0.4,最高达到了 0.55,存在较大的不公平性,

需要分配调整。在影响流域污染负荷分配的4个因素中,科技发展因素和经济因素的指标权重之和达到了0.852,而以往的分配方案更多关注经济和社会因素,可见流域环境现状之间的差异对污染负荷分配有很大的影响。

(3)通过构建流域污染负荷分配耦合模型,实现清溪河流域(许昌段)2020年的COD负荷分配,结果表明:位于魏都区的控制单元总体削减率最高,达到了24.41%,控制单元QYH-3的削减率最大,达到30.44%,削减量达到了617.14 t/a,削减结果基本符合许昌市经济发展现状。

参考文献(References):

- [1] ZHANG Y, WANG X Y, ZHANG Z M, et al. Multi-level waste load allocation system for Xi'an-Xianyang Section, Weihe River [J]. *Procedia Environmental Sciences*, 2012, 13: 943-953. DOI: 10.1016/j.proenv.2012.01.088.
- [2] 赵永宏, 邓祥征, 吴锋, 等. 乌梁素海流域氮磷减排与区域经济发展的均衡分析[J]. *环境科学研究*, 2011, 24(1): 110-117. (ZHAO Y H, DENG X Z, WU F, et al. Equilibrium analysis of nitrogen and phosphorus emission reduction and regional economic development in wuliangsuhai basin[J]. *Research of Environmental Sciences*, 2011, 24(1): 110-117. (in Chinese)) DOI: 10.13198/j.res.2011.01.112. zhaoyh. 017.
- [3] 秦迪岚, 黄哲, 罗岳平, 等. 洞庭湖区污染控制区划与控制对策[J]. *环境科学研究*, 2011, 24(7): 748-755. (QIN D L, HUANG Z, LUO Y P, et al. Pollution control regionalization and control countermeasures in dongting lake area [J]. *Research of Environmental Sciences*, 2011, 24(7): 748-755. (in Chinese)) DOI: 10.13198/j.res.2011.07.40. qindl. 009.
- [4] 李如忠, 钱家忠, 汪家权. 水污染物允许排放总量分配方法研究[J]. *水利学报*, 2003, 34(5): 112-115. (LI R Z, QIAN J Z, WANG Jia-quan. Study on the total allowable discharge of water pollutants[J]. *Journal of Hydraulic Engineering*, 2003, 34(5): 112-115. (in Chinese))
- [5] 孟伟, 张远, 郑丙辉. 水环境质量基准、标准与流域水污染物总量控制策略[J]. *环境科学研究*, 2006, 19(3): 1-6. (MENG W, ZHANG Y, ZHENG B H. Water environment quality standards and total water pollutant control strategies [J]. *Research of Environmental Sciences*, 2006, 19(3): 1-6. (in Chinese)) DOI: 10.13198/j.res.2006.03.7. mengw. 001.
- [6] JOSHI V, MODAK P. Heuristic algorithms for waste load allocation in a river basin [J]. *Water Pollution Research & Control Brighton*, 1988, 21(8-9): 1057-1064.
- [7] 郝信东. 基于信息熵的水污染物总量分配与控制策略研究[D]. 天津: 天津大学, 2010. (HAO X D. Research on total water pollutant allocation and control strategy based on information entropy [D]. Tianjin: Tianjin university, 2010. (in Chinese))
- [8] 郝韶楠, 李叙勇, 江燕, 等. 水污染负荷优化分配模型研究进展[J]. *环境科学与技术*, 2016, 39(1): 1-6. (HAO S N, LI X Y, JIANG Y, et al. Research progress on optimal allocation model of water pollution load [J]. *Environmental Science & Technology*, 2016, 39(1): 1-6. (in Chinese)) DOI: 10.3969/j.issn.1003-6504.2016.01.001.
- [9] 单保庆, 王超, 李叙勇, 等. 基于水质目标管理的河流治理方案制定方法及其案例研究[J]. *环境科学学报*, 2015, 35(8): 2314-2323. (SHAN B Q, WANG C, LI X Y, et al. A case study of river management scheme formulation based on water quality objective management [J]. *Acta Scientiae Circumstantiae*, 2015, 35(8): 2314-2323. (in Chinese)) DOI: 10.13671/j.hjkxxb.2015.0029.
- [10] NIKOO M R, KERACHIAN R, KARIMI A, et al. Optimal water and waste-load allocations in rivers using a fuzzy transformation technique: A case study. [J]. *Environmental Monitoring & Assessment*, 2013, 185(3): 2483-2502. DOI: 10.1007/s10661-012-2726-6.
- [11] 李如忠, 舒坤. 基于多目标决策的水污染负荷分配方法[J]. *环境科学学报*, 2011, 31(12): 2814-2821. (LI R Z, SHU K. Water pollution load distribution method based on multi-objective decision [J]. *Acta Scientiae Circumstantiae*, 2011, 31(12): 2814-2821. (in Chinese)) DOI: 10.13671/j.hjkxxb.2011.12.031.
- [12] MOSTAFAVI S A, AFSHAR A. Waste load allocation using non-dominated archiving multi-colony ant algorithm [J]. *Procedia Computer Science*, 2011, 3(none): 64-69.
- [13] SAADATPOUR M, AFSHAR A. Waste load allocation modeling with fuzzy goals; simulation-optimization approach [J]. *Water Resources Management*, 2007, 21(7): 1207-1224. DOI: 10.1007/s11269-006-9077-4.
- [14] LIANG S, JIA H, YANG C, et al. A Pollutant load hierarchical allocation method integrated in an environmental capacity management system for Zhushan Bay, Taihu Lake [J]. *Science of the Total Environment*, 2015, 533: 223-237. DOI: 10.1016/j.scitotenv.2015.06.116.
- [15] 李如忠, 舒琨. 1526 基于基尼系数的水污染负荷分配模糊优化决策模型[J]. *环境科学学报*, 2010, 30(7): 1518-1526. (LI R Z, SHU K. A fuzzy optimization decision model for water pollution load distribution

- based on gini coefficient [J]. *Acta Scientiae Circumstantiae*, 2010, 30 (7): 1518-1526. (in Chinese) DOI:10.13671/j. hjkxxb. 2010. 07. 022.
- [16] 秦迪岚, 韦安磊, 卢少勇. 基于环境基尼系数的洞庭湖区水污染总量分配[J]. *环境科学研究*, 2013, 26(1): 8-15. (QIN D L, WEI A L, LU S Y. Total water pollution distribution in dongting lake area based on environmental gini coefficient [J]. *Research of Environmental Sciences*, 2013, 26(1): 8-15. (in Chinese)) DOI:10.13198/j. res. 2013. 01. 12. qindl. 006.
- [17] 杨芳. 基于基尼系数法的九龙江流域水污染物排放总量分配研究[J]. *环境科学与管理*, 2012, 37(5): 30-35. (YANG F. Research on total emission allocation of water pollutants in jiulong river basin based on gini coefficient[J]. *Environmental Science and Management*, 2012, 37(5): 30-35. (in Chinese))
- [18] 李泽琪, 张玥, 王晓燕, 等. 基于不同层级排污单元的水污染负荷分配方法[J]. *资源科学*, 2018, 40(7): 1429-1437. (LI Z Q, ZHANG Y, WANG X Y, et al. Water pollution load distribution method based on different levels of pollutant discharge units [J]. *Resources science*, 2018, 40 (7): 1429-1437. (in Chinese)) DOI:10.18402/resci. 2018. 07. 11.
- [19] 赵雪霞, 于鲁冀, 王燕鹏. 清溪河流域(许昌段)水生态环境功能分区指标体系构建[J]. *水利水电技术*, 2018, 49(9): 165-172. (YU L J ZHAO X X, WANG Y P. Qingyi river basin (Xuchang) water environment functional zoning index system to build [J]. *Water Resources and Hydropower Engineering*, 2018, 49 (9): 165-172. (in Chinese)) DOI:10.13928/j. cnki. wrahe. 2018. 09. 022.
- [20] 田平, 方晓波, 王飞儿, 等. 基于环境基尼系数最小化模型的水污染物总量分配优化——以张家港平原水网区为例[J]. *中国环境科学*, 2014, 34(3): 801-809. (TIAN P, FANG X B, WANG F E, et al. Optimization of total water pollutant distribution based on environmental gini coefficient minimization model—a case study of water network area in zhangjiagang plain [J]. *China Environmental Science*, 2014, 34 (3): 801-809. (in Chinese))
- [21] BOSI S, SEEGMULLER T. Optimal cycles and social inequality: what do we learn from the gini index[J]. *Research in Economics*, 2006, 60(1): 35-46. DOI: 10.1016/j. rie. 2005. 12. 003.
- [22] 王媛, 牛志广, 王伟. 基尼系数法在水污染物总量区域分配中的应用[J]. *中国人口·资源与环境*, 2008, 18(3): 177-180. (WANG Y, NIU Z G, WANG W. Application of gini coefficient method in regional distribution of total water pollutants [J]. *China Population Resources and Environment*, 2008, 18(3): 177-180).
- [23] 赵骞. 河流和海洋污染物总量分配研究述评[J]. *中国人口·资源与环境*, 2014, 24(3): 82-86. (ZHAO Q. Review of studies on total pollutant distribution in rivers and oceans [J]. *China Population Resources and Environment*, 2014, 24(3): 82-86. (in Chinese)) DOI: 10.12062 /cpre. 20170307.
- [24] 何帮强, 洪兴建. 基尼系数计算与分解方法研究综述[J]. *统计与决策*, 2016(14): 13-17. (HE B Q, HONG X J. Research review on gini coefficient calculation and decomposition methods [J]. *Statistics & Decision*, 2016(14): 13-17. (in Chinese)) DOI: 10.13546/j. cnki. tjyc. 2016. 14. 003.
- [25] CUTLER A, CUTLER D R, STEVENS J R. Random forests [J]. *Machine Learning*, 2004, 45(1): 157-176. DOI:10.1007/978-1-4419-9326-7_5.
- [26] BIFET A, HOLMES G, PFAHRINGER B. Leveraging bagging for evolving data streams [C]. *European Conference on Machine Learning & Knowledge Discovery in Databases*, 2010. DOI: 10.1007/978-3-642-15880-3_15.
- [27] 冯盼峰, 温永仙. 基于随机森林算法的两阶段变量选择研究[J]. *系统科学与数学*, 2018, 38(1): 119-130. (FENG P F, WEN Y X. Two-stage variable selection based on random forest algorithm [J]. *Journal of Systems Science and Mathematical Sciences*, 2018, 38 (1): 119-130. (in Chinese))
- [28] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. *统计与信息论坛*, 2011, 26(3): 32-38. (FANG K N, WU J B, ZHU J P, et al. Research review of random forest method [J]. *Statistics & Information Forum*, 2011, 26(3): 32-38. (in Chinese))
- [29] RAMÓN DÍAZ-URIARTE, SARA ALVAREZ DE ANDRÉS. Gene selection and classification of microarray data using random forest [J]. *Bmc Bioinformatics*, 2006, 7(1): 3-0. DOI:10.1186/1471-2105-7-3.
- [30] GREGORUTTI B, MICHEL B, SAINT-PIERRE P. Correlation and variable importance in random forests [J]. *Statistics and Computing*, 2017, 27(3): 659-678. DOI:10.1007/s11222-016-9646-1.