

DOI:10.19853/j.zgjsps.1000-4602.2021.01.015

基于主成分分析的松花江流域黑龙江段水质评价

邢 洁^{1,2,3}, 宋男哲¹, 陈祥伟³, 倪红伟², 韩 旭²

(1. 黑龙江省环境科学研究院, 黑龙江 哈尔滨 150056; 2. 黑龙江省科学院 自然与生态研究所, 黑龙江 哈尔滨 150040; 3. 东北林业大学 林学院, 黑龙江 哈尔滨 150040)

摘 要: 对松花江流域黑龙江控制区段 13 个监测断面不同水期的水质指标进行主成分分析,以鉴别主要的污染源,进而提出有针对性的水污染控制对策。结果表明,BOD₅、COD、TP、TN 为主要污染指标,合计贡献率为 49.529%,首要污染源为生活污水与农业废水。第二大污染指标主要为石油类、阴离子表面活性剂、氰化物,主要来源于工业废水。总污染最严重的主要是阿什河口内平水期与龙安桥平水期。龙安桥平水期等对应的河流区段受生活污水及农业废水污染较为严重,苗家平水期等对应的河流区段受工业污染较为严重。

关键词: 松花江; 黑龙江省; 主成分分析; 水质评价

中图分类号: TU992 **文献标识码:** A **文章编号:** 1000-4602(2021)01-0089-06

Water Quality Assessment of Heilongjiang Control Section in Songhua River Basin Based on Principal Component Analysis

XING Jie^{1,2,3}, SONG Nan-zhe¹, CHEN Xiang-wei³, NI Hong-wei², HAN Xu²

(1. Heilongjiang Provincial Research Academy of Environmental Sciences, Harbin 150056, China;
2. Institute of Natural Resources and Ecology, Heilongjiang Academy of Sciences, Harbin 150040, China;
3. School of Forestry, Northeast Forestry University, Harbin 150040, China)

Abstract: Principal component analysis of the water quality indicators of 13 monitoring sections in Heilongjiang control section of the Songhua River Basin in different water periods was carried out to identify the main pollution sources, and then targeted water pollution control countermeasures were proposed. The results showed that BOD₅, COD, TP and TN were the main pollution indicators, with a total contribution rate of 49.529%. The primary pollution sources were domestic sewage and agricultural sewage. The second largest pollution indexes were petroleum, anionic surfactants and cyanide from industrial wastewater. The most serious pollution was mainly in Ash River Estuary and Longan Bridge during normal water period. Longan Bridge during normal water period was seriously polluted by domestic sewage and agricultural sewage. Miaojia during normal water period was seriously polluted by industrial wastewater.

Key words: Songhua River; Heilongjiang Province; principal component analysis; water quality assessment

松花江作为我国七大水系之一,是黑龙江在中国境内的最大支流。松花江流域的水质污染较为严重,在十三五规划期间多数河流的水质仅达到国家地表水Ⅲ、Ⅳ类标准,劣Ⅴ类水体依然存在,在全国地表水水质目标考核中成绩不佳^[1]。流域上游建立的石油、化工基地是主要污染源之一^[2]。同时,农业是黑龙江省的主要产业,农业面源污染也是主要的污染类型,每年有大量化肥和农药随地表径流进入地表水和地下水^[3]。针对松花江流域的水质污染,应确定主要污染源,有针对性地污染进行治理。但是,现有水质监测指标繁多,数据冗杂,并且不同指标间的相关性可能较大,所表达信息重叠,难以根据庞大的监测数据对松花江主要污染物和污染严重断面进行评价。此外,现有研究多对松花江全年进行综合评价,但不同水期的降水量、温度、排污量不同,针对不同水期进行评价,能更明确地制定动态流域治理方案^[4]。

在含有多个指标、变量的问题研究中,由于不同变量之间可能存在一定相关性,造成数据信息重叠。主成分分析(PCA)是一种多元统计分析方法,通过降维的方式来简化影响某一事物的多种变量,力求用较少的具有代表性的变量来反映事物的大部分信息^[5-9]。由于水质评价中监测断面数量较多,涵盖大量监测指标数据和众多变量,使水质分析的复杂性显著增加。因此,运用主成分分析方法可以从众多水质监测指标中选取对水质影响显著的指标,获取关键信息^[10-11]。

笔者基于松花江流域黑龙江区段2018年枯水期、丰水期、平水期的水质监测数据,对不同水期不同监测断面的主要水质指标进行主成分分析,鉴别主要污染源,并分析不同断面及水期的主要污染因子。同时利用水质综合评价,分析污染最严重的监测断面及水期,有针对性地提出松花江流域黑龙江区段水污染控制对策。

1 研究区概况

松花江全长为1 927 km,流经黑龙江省、吉林省和内蒙古自治区,流域总面积达 $55.72 \times 10^4 \text{ km}^2$,其中黑龙江区段为 $26.94 \times 10^4 \text{ km}^2$ ^[12]。黑龙江省松花江流域由嫩江流域、松花江干流流域及其支流组成,流经哈尔滨、齐齐哈尔、鹤岗、双鸭山、大庆、伊春、佳木斯、七台河、牡丹江、黑河、绥化、大兴安岭12个市(地),59个县(市、区)。

2 研究方法

2.1 数据来源及处理

本研究基于松花江流域黑龙江区段2018年枯水期(1月—3月、12月)、丰水期(7月—9月)、平水期(4月—6月、10月—11月)13个监测断面的3 276条水质监测数据。其中,水质监测数据为大顶子山(A)、阿什河口内(B)、呼兰河口内(C)、苗家(D)、牡丹江口内(E)、江桥(F)、龙安桥(G)、佳木斯上(H)、汤旺河口内(I)、梧桐河口内(J)、同江(K)、柴河铁路桥(L)、博霍头(M)等13个监测断面的月度监测值。在综合考虑流域水质各影响因素基础上,选定了12个监测指标作为主要考察对象,即:DO、COD_{Mn}、BOD₅、氨氮、石油类、COD、TN、TP、氟化物、氰化物、阴离子表面活性剂(LAS)、硫化物。

利用SPSS对选取的监测数据进行主成分分析。首先将原始数据进行标准化,得出各水质指标的相关系数。然后对标准化后的数据进行主成分分析,以累计贡献率约为85%为原则,筛选出主成分,计算所得主成分的载荷矩阵。最后根据载荷矩阵分析得出松花江流域黑龙江控制区段主要水质污染指标,并计算水质综合评价得分,对各监测断面及其不同水期水质进行综合评价。基于相同的方法分别对各监测断面的水质污染因子进行主成分分析,得出13个监测断面的主要污染因子。

2.2 数据的质量保证与质量控制

研究所采用的水质监测数据均根据《地表水和污水监测技术规范》(HJ/T 91—2002)所得,即严格按照该规范进行监测断面与监测点位的布设,以及水样的采取与分析。送入实验室的水样按照规定核对采样单、容器编号、包装情况、保存条件和有效期等。每批水样分析时,空白样品对被测项目有响应的均制作实验室空白样。对于水样分析,采用分光光度法校准曲线定量时,检验校准曲线的相关系数和截距是否正常。采用原子吸收分光光度法、气相色谱法等仪器分析方法定量时,校准曲线制作与样品测定同时进行。在精密度控制上,对均匀样品凡能做平行双样的分析项目,分析每批水样时均需做10%的平行双样,样品较少时则每批样品应至少做一份样品的平行双样。测定的平行双样允许差符合规定质控指标的样品,最终结果以双样测试结果的平均值报出;平行双样测试结果超出规定允许偏差时,在样品允许保存期内再加测一次,取相对偏差符

合规定质控指标的两个测定值报出。在准确度控制上,采用标准物质或质控样品作为控制手段,每批样品带一个已知浓度的质控样品。

3 结果与讨论

3.1 松花江流域黑龙江控制区段水质主成分分析

为了消除水质原始数据数量级与量纲对分析的影响,利用 SPSS v26.0 软件对原始水质数据进行标准化,得到如表 1 所示的各水质指标的相关系数,并

筛选与其他指标相关系数极小的指标。由表 1 可知,氨氮与 TN 之间的相关系数为 0.850,COD 与 BOD₅ 的相关系数为 0.709,COD 与 COD_{Mn}的相关系数为 0.697,TN 与 TP 的相关系数为 0.692,表明多个水质指标之间的相关性较强,这些水质指标所反映出的水质信息存在一定程度的重叠,适合利用主成分分析进行降维,利用较少的新变量反映松花江流域黑龙江控制区段水质。

表 1 水质指标相关系数矩阵

Tab. 1 Correlation coefficient matrix of water quality index

项目	DO	COD _{Mn}	BOD ₅	氨氮	石油类	COD	TN	TP	氟化物	氰化物	LAS	硫化物
DO	1.000	-0.314	-0.176	-0.069	0.044	-0.356	-0.008	-0.298	-0.384	0.216	-0.224	0.096
COD _{Mn}	-0.314	1.000	0.239	-0.033	0.059	0.697	0.005	0.290	0.172	-0.189	0.157	-0.064
BOD ₅	-0.176	0.239	1.000	0.505	0.329	0.709	0.600	0.462	0.375	-0.182	0.461	0.115
氨氮	-0.069	-0.033	0.505	1.000	0.406	0.317	0.850	0.649	0.087	-0.216	-0.140	0.013
石油类	0.044	0.059	0.329	0.406	1.000	0.376	0.386	0.133	-0.283	0.264	0.143	-0.191
COD	-0.356	0.697	0.709	0.317	0.376	1.000	0.386	0.465	0.214	-0.184	0.481	-0.111
TN	-0.008	0.005	0.600	0.850	0.386	0.386	1.000	0.692	0.128	-0.204	-0.039	0.056
TP	-0.298	0.290	0.462	0.649	0.133	0.465	0.692	1.000	0.258	-0.486	-0.226	-0.094
氟化物	-0.384	0.172	0.375	0.087	-0.283	0.214	0.128	0.258	1.000	-0.360	0.194	0.488
氰化物	0.216	-0.189	-0.182	-0.216	0.264	-0.184	-0.204	-0.486	-0.360	1.000	0.165	0.207
LAS	-0.224	0.157	0.461	-0.140	0.143	0.481	-0.039	-0.226	0.194	0.165	1.000	-0.062
硫化物	0.096	-0.064	0.115	0.013	-0.191	-0.111	0.056	-0.094	0.488	0.207	-0.062	1.000

对标准化后的数据进行主成分分析,结果见表 2。以累计贡献率约为 85% 为原则,从 12 种成分中筛选出 5 种主成分,分别表示为 F1、F2、F3、F4、F5。累计贡献率达到 84.176%,说明筛选出的 5 种主成分能够反映原始数据 84.176% 的信息。

表 2 水质指标特征值、主成分贡献率及累计方差贡献率

Tab. 2 Eigenvalue, principal component contribution rate and cumulative variance contribution rate of water quality index

成分	特征值	贡献率/%	累计贡献率/%
1	3.906	32.549	32.549
2	2.038	16.980	49.529
3	1.826	15.219	64.748
4	1.481	12.340	77.088
5	0.850	7.088	84.176
6	0.727	6.058	90.234
7	0.394	3.287	93.521
8	0.231	1.929	95.450
9	0.205	1.709	97.159
10	0.112	1.283	98.442
11	0.112	0.935	99.377
12	0.075	0.623	100.000

5 种水质指标主成分的载荷矩阵见表 3。水质

指标主成分载荷矩阵能够体现各水质指标与主成分之间的相关性,载荷矩阵系数的绝对值越大,代表该主成分与该水质指标的相关性越强^[13]。

表 3 水质指标主成分载荷矩阵

Tab. 3 Principal component matrix of water quality index

水质指标	F1	F2	F3	F4	F5
BOD ₅	0.820	0	0.186	0.318	-0.154
COD	0.803	-0.316	0.380	-0.107	0.177
TP	0.788	0.216	-0.358	-0.239	0.100
TN	0.767	0.512	0	0.182	0
氨氮	0.718	0.564	-0.127	0.110	0
氟化物	0.393	-0.538	-0.501	0.406	0
DO	-0.391	0.514	0	0.237	0.196
石油类	0.376	0.427	0.639	0	0
LAS	0.233	-0.525	0.588	0.260	-0.416
氰化物	-0.402	0.131	0.584	0.439	0.237
硫化物	0	-0.156	-0.336	0.814	0.375
COD _{Mn}	0.444	-0.501	0.161	-0.347	0.596

由表 3 可知,对于总方差贡献率为 32.549% 的 F1,BOD₅、COD、TP、TN 占有较高的载荷;对于总方差贡献率为 16.980% 的 F2,氨氮、DO 占有较高的载荷。这反映了松花江流域黑龙江控制区段有机污染

的特征,且该有机污染多来源于生活污水与农业废水^[14]。对于F3,石油类、阴离子表面活性剂、氰化物占有较高的载荷;对于F4,硫化物占有较高载荷;对于F5,COD_{Mn}占有较高载荷,这反映了松花江流域黑龙江控制区段工业污染的特征。

根据表3,可得到松花江流域黑龙江控制区段水质综合评价函数,即主成分表达式(F1的得分记作 F_1 ,其他同此):

$$F_1 = 0.820X_1 + 0.803X_2 + 0.788X_3 + 0.767X_4 + 0.718X_5 + 0.393X_6 - 0.391X_7 + 0.376X_8 +$$

$$0.233X_9 - 0.402X_{10} + 0X_{11} + 0.444X_{12} \quad (1)$$

同上,依次求出 F_2 、 F_3 、 F_4 、 F_5 ,进而得到总得分 F (λ 表示不同主成分对应的特征值):

$$F = (\lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3 + \lambda_4 F_4 + \lambda_5 F_5) / (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5) \quad (2)$$

所求得的松花江流域黑龙江控制区段各监测断面及其不同水期水质综合评价得分结果见图1(字母后数字1、2、3分别代表枯水期、丰水期和平水期),分值越高表示污染越严重^[14]。

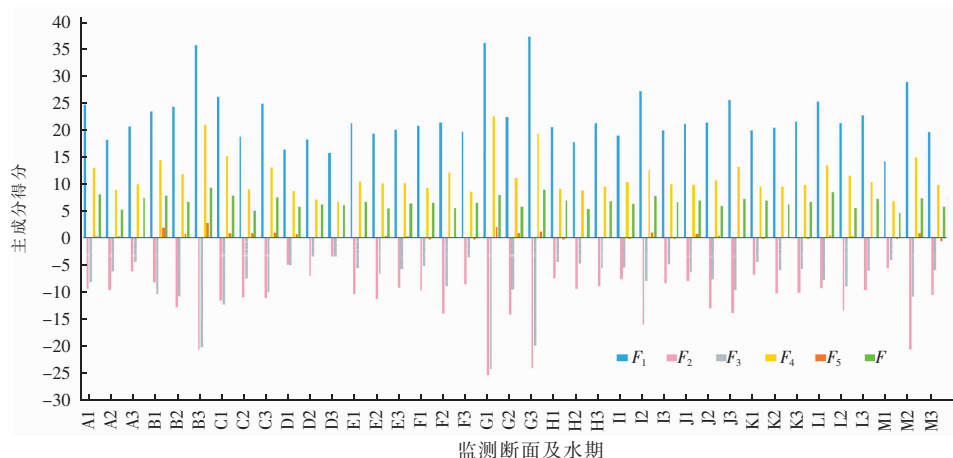


图1 各监测断面水质综合评分

Fig. 1 Comprehensive water quality score of all monitoring sections

由图1可知,从总得分 F 来看,污染最严重的主要是阿什河口内平水期与龙安桥平水期,监测断面对应的河流区段应该被设置为重点治理区段,且在水污染严重的水期更应重点防护。水质最好的是博霍头枯水期、呼兰河口丰水期、大顶子山丰水期、佳木斯上丰水期、牡丹江口内丰水期,现有的治理措施能较好地保护河流水质。

单从代表BOD₅、COD、TP、TN指标的F1主成分来看,污染较严重的主要是龙安桥平水期、龙安桥枯水期、阿什河口内平水期。单从代表氨氮、DO指标的F2主成分来看,污染较严重的主要是苗家平水期和枯水期。说明这几个监测断面对应的河流区段受生活污水及农业废水的污染较为严重。单从代表石油类、阴离子表面活性剂、氰化物指标的F3来看,污染较严重的主要是苗家平水期、苗家丰水期、江桥平水期。对于代表硫化物的主成分F4,污染较严重的主要是龙安桥枯水期、阿什河口内平水期、龙安桥平水期。而对于代表COD_{Mn}的F5,污染较严重的主要

是阿什河口内平水期。说明这几个监测断面对应的河流区段受工业废水污染较为严重。

3.2 各监测断面水质主成分分析

基于相同的方法分别对各监测断面的水质污染因子进行主成分分析,表4所示为13个监测断面的主要污染因子。

从表4可以看出,COD、BOD₅基本都为各河流区段的主要污染指标,这可以解释13个监测断面对应河流区段的主要污染因素为耗氧污染物。而氨氮、总氮也基本都为各河流区段的主要污染指标,这表明氮元素在13个监测断面对应河流区段的含量均比较高。同时,除了大顶子山、博霍头、苗家这三个监测断面以外,其余的监测断面中总磷均为主要污染因子,这说明各河流区段均存在氮、磷含量较高的情况。另外,龙安桥、梧桐河口内、苗家监测断面的氰化物含量超标,呼兰河口内、苗家监测断面的石油类污染物含量超标,这可能是由于工业废水的排放导致的。

表4 各监测断面污染因子的主成分

Tab.4 Principal components of pollution factors in each monitoring section

断面	所属控制单元	污染因子主成分
大顶子山	松花江哈尔滨市 市辖区控制单元	氨氮、COD、TN、DO、BOD ₅
阿什河口内		BOD ₅ 、氨氮、COD、TN、TP
同江	松花江佳木斯市控制单元	COD、氨氮、硫化物、TP、BOD ₅ 、TN
牡丹江口内	牡丹江牡丹江市控制单元	COD、TP、TN、DO、氨氮
柴河铁路桥		COD、TN、TP、氨氮
博霍头	嫩江黑河市控制单元	DO、氨氮、TN
佳木斯上	松花江哈尔滨市辖县控制单元	氨氮、BOD ₅ 、TN、氟化物、DO、TP
汤旺河口内	汤旺河伊春市控制单元	BOD ₅ 、COD、氨氮、TP
江桥	嫩江齐齐哈尔市控制单元	COD、TP、氨氮、氟化物、氰化物、DO、BOD ₅
龙安桥	乌裕尔河齐齐哈尔市控制单元	COD、BOD ₅ 、氨氮、TN、TP、氰化物、硫化物
苗家	拉林河哈尔滨市控制单元	COD、TN、氨氮、DO、BOD ₅ 、石油类、氰化物、LAS
呼兰河口内	呼兰河伊春市—绥化市—哈尔滨市控制单元	DO、氨氮、石油类、TN、TP、BOD ₅ 、COD
梧桐河口内	梧桐河鹤岗市控制单元	BOD ₅ 、氨氮、TN、氰化物、COD、TP

4 结论及对策

① 松花江流域黑龙江控制区段主要污染因子为BOD₅、COD、TP、TN,主要来源为生活污水与农业废水。其中,龙安桥平水期及枯水期、阿什河口内平水期、苗家平水期及枯水期的生活污水及农业面源污染较为严重。从源头上,应控制畜禽养殖等农业污染,转变畜禽养殖方式。同时,改善流域内城镇污水处理设施及工艺,提高生活污水处理能力。

② 污染最严重的主要是阿什河口内平水期、龙安桥平水期,监测断面对应的河流区段应该被设置为重点治理区段,且在水污染严重的水期更应重点防护。

③ 工业废水污染较严重的主要是苗家平水期及丰水期、江桥平水期、龙安桥平水期及枯水期、阿什河口内平水期,应加强对该区域工厂的排污管理,同时改善废水处理设施及工艺,提高污水厂废水处理能力。

参考文献:

- [1] 郎朗,董晓琪,狄静波. SPE-HPLC法测定松花江哈尔滨段水样中11种抗生素[J]. 中国给水排水, 2018,34(20):114-118.
LANG Lang,DONG Xiaoqi,DI Jingbo. Determination of 11 antibiotics in Harbin section of Songhua River by SPE-HPLC[J]. China Water & Wastewater, 2018, 34(20):114-118(in Chinese).
- [2] 魏春风. 松花江干流河流健康评价研究[D]. 长春: 中国科学院东北地理与农业生态研究所,2018.

WEI Chunfeng. Study on the Health Evaluation of the Main Stream of Songhua River [D]. Changchun: Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences,2018(in Chinese).

- [3] 吴玉博,侯保灯,仵峰,等. 基于面源污染的松花江下游水质评价[J]. 人民黄河,2018,40(5):69-72.
WU Yubo,HOU Baodeng,WU Feng,et al. Evaluation of the downstream water quality of Songhua River based on non-point source pollution[J]. Yellow River, 2018, 40(5):69-72(in Chinese).
- [4] 何佳吉,边晓东,张长占,等. 松花江流域融冰期水质分析[J]. 东北水利水电,2020,38(2):45-46,72.
HE Jiagi, BIAN Xiaodong, ZHANG Changzhan, et al. Analysis on water quality during melting ice of Songhua River basin [J]. Water Resources & Hydropower of Northeast, 2020, 38(2): 45-46, 72(in Chinese).
- [5] 高爽. 主成分分析—BP神经网络法在漳河水质评价中的联合应用[D]. 邯郸:河北工程大学,2018.
GAO Shuang. Combined Application of Principal Component Analysis and BP Neural Network in Water Quality Evaluation of Zhanghe [D]. Handan: Hebei University of Engineering, 2018(in Chinese).
- [6] 祁峰,王琳,孙艺珂. KCFs识别法的建立及其在引黄饮用水中的应用[J]. 中国给水排水,2019,35(3):38-43.
QI Feng,WANG Lin,SUN Yike. Establishment of KCFs identification method and its application in drinking water from the Yellow River [J]. China Water & Wastewater, 2019, 35(3): 38-43(in Chinese).
- [7] 陈锋,孟凡生,王业耀,等. 基于主成分分析—多元线

- 性回归的松花江水体中多环芳烃源解析[J]. 中国环境监测, 2016, 32(4): 49–53.
- CHEN Feng, MENG Fansheng, WANG Yeyao, *et al.* The research of polycyclic aromatic hydrocarbons in the river based on the principal component – multivariate linear regression analysis [J]. Environmental Monitoring in China, 2016, 32(4): 49–53 (in Chinese).
- [8] NGUYEN B T, NGOYEN T M T, BACH Q V. Assessment of groundwater quality based on principal component analysis and pollution source-based examination: a case study in Ho Chi Minh City, Vietnam [J]. Environmental Monitoring and Assessment, 2020, 192(6): 395–399.
- [9] HE H, TIAN C, JIN G, *et al.* Principal component analysis and Fisher discriminant analysis of environmental and ecological quality, and the impacts of coal mining in an environmentally sensitive area [J]. Environmental Monitoring and Assessment, 2020, 192(4): 207–213.
- [10] 熊云帆, 朱勤东. 基于主成分—模糊综合评价模型的近岸海域水质综合评价[J]. 水土保持应用技术, 2020, 193(1): 1–4.
- XIONG Yunfan, ZHU Qindong. Comprehensive evaluation of coastal water quality based on principal component – fuzzy comprehensive evaluation model [J]. Technology of Soil and Water Conservation, 2020, 193(1): 1–4 (in Chinese).
- [11] LI Q, ZHANG H, GUO S, *et al.* Groundwater pollution source apportionment using principal component analysis in a multiple land-use area in southwestern China [J]. Environmental Science and Pollution Research, 2020, 27(9): 9000–9011.
- [12] 杨先兴. 松花江哈尔滨段水环境质量时空变化特征分析[D]. 哈尔滨: 哈尔滨工业大学, 2014.
- YANG Xianxing. Research on Temporal and Spatial Variation of Water Environmental Quality in Harbin Section of the Songhua River [D]. Harbin: Harbin Institute of Technology, 2014 (in Chinese).
- [13] 闫媛. 基于主成分分析法的扶风县生态环境质量评价[D]. 杨凌: 西北农林科技大学, 2016.
- YAN Yuan. Evaluation of Ecological Environmental Quality Based on Principal Component Analysis—Take Fufeng County as a Case [D]. Yangling: Northwest A&F University, 2016 (in Chinese).
- [14] 林卉, 李楠, 黄伯当, 等. 基于主成分分析的南流江水质评价[J]. 广东化工, 2020, 47(4): 144–146, 148.
- LIN Hui, LI Nan, HUANG Bodang, *et al.* Water quality evaluation of Nanliu River based on principal component analysis [J]. Guangdong Chemical Industry, 2020, 47(4): 144–146, 148 (in Chinese).

作者简介: 邢洁 (1983 –), 女, 黑龙江哈尔滨人, 博士, 高级工程师, 主要从事流域水环境管理研究。

E-mail: 18686885851@163.com

收稿日期: 2020-07-30

修回日期: 2020-08-20

(编辑: 李德强)

节约每一滴水,

回收每一滴水,

让每一滴水多循环一次