

# 基于改进线性回归方法的江苏省 综合用水效率研究

倪剑华<sup>1</sup>, 史安基<sup>2</sup>, 王怀群<sup>2</sup>

(1. 丹阳市抗旱排涝处, 江苏 镇江 212300; 2. 丹阳市瑞禹水利工程有限公司, 江苏 镇江 212300)

**摘要:**结合江苏省农业、工业和生活用水现状,构建以万元GDP用水量为研究目标的江苏省综合用水效率指标体系,在对数学模型理解的基础上,采用多种方法进行江苏省用水效率影响因素的辨识。通过理论分析以及专家咨询确定了21个影响因素指标,收集江苏省1997—2010年各影响因素指标数据,引入基于主成分分析的逐步回归方程进行分析,得到江苏省万元GDP用水量与其影响因素之间的响应方程式。通过回归方程可以看出,影响江苏省万元GDP用水量最重要的4个指标分别为人均灌溉面积、工业废水达标排放率、教育支出占财政支出比重、第三产业产值占GDP比重。

**关键词:**用水效率; 万元GDP用水量; 主成分分析; 逐步回归

中图分类号:TV21 文献标识码:B 文章编号:1007-7839(2020)01-0044-06

## Study on the comprehensive water use efficiency in Jiangsu Province based on the improved linear regression method

NI Jianhua<sup>1</sup>, SHI Anji<sup>2</sup>, WANG Huaiqun<sup>2</sup>

(1. Drought Relief and Drainage Engineering Office of Danyang, zhenjiang 212300, Jiangsu;

2. Danyang City Ruiyu Water Conservancy Engineering Co., Ltd., zhenjiang 212300, Jiangsu)

**Abstract:** Combined with the current situation of agricultural, industrial and domestic water use in Jiangsu Province, a comprehensive water efficiency index system of Jiangsu Province with water consumption per 10000 GDP as the research target was constructed. Based on the understanding of mathematical models, various methods were used to analyze the factors affecting water efficiency in Jiangsu identification. Through theoretical analysis and expert consultation, 21 influencing factor indicators were identified. Data of each influencing factor index in Jiangsu Province from 1997 to 2010 were collected, and a stepwise regression equation based on principal component analysis was introduced for analysis, so that the response equations between the water consumption per 10000 GDP in Jiangsu Province and its influencing factors were obtained. As a result, the major four factors have been got: irrigation area per capita, industrial waste water emission on standard, proportion education cost accounting for the financial expenditure, proportion the output value of tertiary industry accounting for the GDP.

**Key words:** water use efficiency; water consumption per 10000 GDP; principal component analysis; stepwise regression

收稿日期:2019-06-20

基金项目:国家基金重大研究计划项目(91647209)

作者简介:倪剑华(1970—),男,工程师,主要从事农水工作。

## 1 概述

我国人均淡水资源占有量约 2100 m<sup>3</sup>, 仅为世界平均水平的 28%, 目前全国城市中有约三分之二缺水, 约四分之一严重缺水, 水资源短缺已成为制约社会经济持续发展的重要因素之一<sup>[1]</sup>。同时, 我国水资源利用方式粗放、用水效率不高, 万元 GDP 用水量约为世界平均水平的 2.5 倍<sup>[2]</sup>。水资源从需求型开发转向以促进社会可持续发展为目的的清洁发展, “节水、降耗、减污、增效” 势在必行。近年来, 江苏省万元 GDP 用水量不断下降, 2012 年比全国平均水平低 77 m<sup>3</sup>。但江苏省水资源在特殊时段、部分地区仍存在结构性矛盾<sup>[3]</sup>, 水生态环境保护压力仍然较大, 需进一步实施最严格的保护制度, 加快转变经济发展方式。省政府批复的《江苏省水资源综合规划》提出, 到 2020 年, 省用水总量力争控制在 590 亿 m<sup>3</sup> 以内; 全省万元 GDP 用水量降低到 90 m<sup>3</sup> 以下, 比 2010 年降低 51% 以上<sup>[4]</sup>。本文就江苏省的现状, 以万元 GDP 用水量作为效率评价指标, 研究水资源的利用效率问题, 找出影响这一效率指标的关键因子并进行分析, 这一研究对于深刻认识江苏水资源问题, 制定切合实际的水资源战略及政策具有重要的指导意义。

## 2 万元 GDP 用水量影响因素指标

本文指标体系遵循指标全面性、客观性、可获得性的建立原则<sup>[5-6]</sup>, 采用理论分析和专家咨询相结合的方法, 对影响江苏省万元 GDP 用水量的重要因素或投资主体收益、群众比较关注的因素等, 进行分析、比较、综合, 在初步拟定的基础上, 征询专家意见, 并进行调整得到以下指标体系。

江苏省万元 GDP 用水量影响因素指标体系分为人口因素、资源因素、经济因素、技术因素、管理因素等 5 个方面。其具体内容见表 1。

## 3 模型构建

为了量化研究江苏省万元 GDP 用水量, 探索影响因素与目标变量之间的作用方式, 本文建立基于主成分分析的多元线性回归方程, 通过研究万元 GDP 用水量与主要影响因素之间的响应关系式, 为未来江苏省万元 GDP 用水量的调控提供参考。具体步骤如下:

(1) 主成分分析法降维。由于变量较多且变量间有一定的相关性, 直接建立回归方程时, 回归的变量很少, 虽然达到了剔除变量的目的, 但也造成了一部分数据信息的缺失, 不利于之后调控研究; 另一方面, 运用计算机进行逐步回归时, 计算机只是根据统计规律进行选择, 忽视了逻辑思考, 可能使模型失去意义。因此, 在建立回归方程之前, 先采用主成分分析对因素做降维处理, 将原来众多具有一定相关性的变量重新组合成新的少数几个互相无关的综合变量来代替原来变量, 保留原有信息。

(2) 逐步回归法建立方程。多元回归模型是研究变量之间作用关系的较为成熟的模型。但在应用回归分析处理实际问题时, 首先要解决的问题是自变量的选择和回归函数的选择。而当自变量的个数不多时, 从所有可能的回归模型中寻找最优回归方程是可行的。但由于自变量个数较多, 求出所有的回归方程是很不容易的。因此, 在建立模型时采用较为简便实用的快速选择最优方程的方法——逐步回归法。在逐步回归的过程中, 将由 21 个影响因素所新组成的主成分因子作为自变量, 万元 GDP 用水量作为因变量。

(3) 主因子的换算。通过逐步回归得到的方程, 自变量为主因子(21 个指标信息的“载体”), 可视化程度不高, 不方便直接研究, 需要根据 21 个指标在主因子上的投影系数, 分解主因子, 获得 21 个指标的回归系数, 得到新的回归方程。

## 4 模型实践

### 4.1 数据标准化

根据 1997—2010 年江苏省统计年鉴和水资源公报、中国统计年鉴、中国环境统计年鉴、中国水利统计年鉴、城市供水统计年鉴等, 得到江苏省 1997—2010 年万元 GDP 用水量及其影响因素数据, 并用式(1)对数据进行标准化处理, 以消除各变量的度量单位差异影响。

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{var}(x_j)}} \quad i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (1)$$

其中,  $\bar{x}_j$  和  $\sqrt{\text{var}(x_j)}$  分别是第  $j$  个变量的平均值和标准差。利用 SPSS 进行数据的标准化处理,

表 1 江苏省万元 GDP 影响因素结构表

因变量	主要类型	影响因素
江苏省万元 GDP 用水量(年用水总量除以年万元 GDP 增加值的数值)	人口因素	总人口(TP)、城市化率(UZ)
	资源因素	地表水资源(SW)、地下水资源(GW)、降水量(PT)、人均水资源量(WP)、人均用水量(PC)、年均气温(AT)、干旱指数(DI)、人均灌溉面积(IC)、人均耕地面积(PA)
	经济因素	第一产业产值占 GDP 比重(PG)、第二产业产值占 GDP 比重(SG)、第三产业产值占 GDP 比重(TG)、工业产值占 GDP 比重(IG)、高耗水行业产值占工业总产值比重(HA)
	技术因素	科学技术支出占财政支出比重(SF)、工业废水达标排放率(ID)、教育支出占财政支出比重(EF)、高新技术产业产值(TV)
	管理因素	水价(TF)

得到相应的标准化数据。

#### 4.2 主成分分析

以最大保留原始信息为原则(提取原则根据直接设定的主成分数目),对标准化之后的数据运用 SPSS 做主成分分析,降维处理。表 2 是全部解释方差表。初始特征根一栏给出了按顺序排列的主成分得分的方差(合计),在数值上等于相关系数矩阵的各个特征根  $\lambda$ , 因此得以直接根据特征根计算每一个主成分的方差百分比。

根据各主因子的得分系数矩阵(表 3),将影响因素的数据代入,得到 13 个主因子的数值矩阵。因子得分信息表达式:

$$f_1 = 0.094x_1 + 0.091x_2 + 0.013x_3 - 0.046x_4 + 0.001x_5 - 0.002x_6 + 0.066x_7 + 0.001x_8 - 0.035x_9 - 0.093x_{10} - 0.051x_{11} - 0.091x_{12} + 0.063x_{13} + 0.074x_{14} + 0.068x_{15} - 0.085x_{16} + 0.075x_{17} + 0.068x_{18} - 0.066x_{19} + 0.091x_{20} + 0.077x_{21} \quad (2)$$

表 2 解释的总方差

成份	初始特征值			提取平方和载入		
	合计方差	方差的百分比	累积百分比	合计方差	方差的百分比	累积百分比
1	14.541	63.222	63.222	14.541	63.222	63.222
2	4.055	17.630	80.852	4.055	17.630	80.852
3	1.934	8.407	89.259	1.934	8.407	89.259
4	1.078	4.686	93.945	1.078	4.686	93.945
5	0.526	2.288	96.233	0.526	2.288	96.233
6	0.372	1.616	97.850	0.372	1.616	97.850
7	0.271	1.178	99.028	0.271	1.178	99.028
8	0.085	0.369	99.397	0.085	0.369	99.397
9	0.069	0.302	99.698	0.069	0.302	99.698
10	0.036	0.157	99.856	0.036	0.157	99.856
11	0.015	0.065	99.920	0.015	0.065	99.920
12	0.013	0.055	99.975	0.013	0.055	99.975
13	0.006	0.025	100.000	0.006	0.025	100.000

表 3 主成分得分系数矩阵

	1	2	3	4	5	6	7	8	9	10	11	12	13
TP	0.094	0.011	0.01	-0.042	0.013	-0.011	0.3	-0.264	-0.264	0.025	-0.041	0.875	-4.374
UZ	0.091	0.012	0.06	-0.099	-0.02	-0.078	0.361	-0.688	0.067	-0.037	4.49	0.246	-11.864
SW	0.013	0.255	0.019	0.05	0.07	0.197	-0.053	0.298	1.489	-0.234	-1.941	-1.474	-10.437
GW	-0.046	0.203	-0.093	0.056	-0.179	0.094	0.673	-0.993	-1.912	1.97	-0.436	1.014	2.202
PT	0.001	0.26	-0.006	0.032	0.041	-0.002	-0.16	0.27	-0.799	-1.46	3.739	1.426	7.89
WP	-0.002	0.259	0.003	0.058	0.056	0.16	-0.131	0.302	1.231	0.072	-0.474	-0.344	0.338
PC	0.066	-0.129	-0.043	0.213	-0.027	0.767	0.26	0.149	0.147	-0.881	0.689	0.623	-0.283
AT	0.001	-0.014	0.004	0.376	0.737	-0.589	0.34	0.257	-0.394	0.074	-0.43	0.371	0.243
DI	-0.035	0.006	0.156	-0.252	0.558	0.864	-0.55	-0.249	-1.124	0.167	-0.257	0.019	-0.623
IC	-0.093	-0.024	-0.016	-0.012	0.015	-0.138	0.083	0.546	-0.068	0.433	6.374	-4.367	-3.407
PA	-0.051	-0.011	0.249	-0.171	0.011	0.003	0.569	1.149	0.929	1.481	1.596	2.417	5.385
PG	-0.091	-0.023	-0.076	0.026	-0.015	0.041	-0.239	0.211	0.195	0.706	-0.459	2.222	-2.502
SG	0.063	0.003	0.193	0.25	-0.106	0.132	-0.064	-0.243	-0.146	0.594	0.836	1.138	-1.647
TG	0.074	0.033	-0.084	-0.301	0.128	-0.195	0.433	-0.066	-0.111	-1.796	-0.288	-3.541	7.621
IG	0.068	0.001	0.156	0.275	-0.097	0.296	0.019	-0.332	0.197	0.431	1.274	-4.705	9.614
HA	-0.085	-0.003	0.072	0.114	-0.178	0.266	1.319	0.795	-0.614	-2.232	-0.948	2.114	-1.531
SF	0.075	0	-0.174	-0.05	0.065	0.274	0.843	1.412	-0.62	1.722	-1.293	-4.317	-2.385
ID	0.068	0.022	0.19	-0.158	0.096	-0.293	1.013	-0.387	-0.051	0.198	-0.851	2.167	-1.743
EF	-0.066	-0.039	-0.194	0.007	0.242	0.35	-0.066	-0.039	-0.194	0.007	0.242	0.35	2.887
TV	0.091	0.006	-0.081	-0.054	0.007	0.018	0.091	0.006	-0.081	-0.054	0.007	0.018	10.411
TF	0.077	0.005	-0.182	0.024	0.067	0.134	0.077	0.005	-0.182	0.024	0.067	0.134	-3.26

式中:未知数  $x_1, x_2, \dots, x_{21}$  依次分别代表总人口(TP -  $x_1$ )、城市化率(UZ -  $x_2$ )、地表水资源(SW -  $x_3$ )、地下水资源(GW -  $x_4$ )、降水量(PT -  $x_5$ )、人均水资源量(WP -  $x_6$ )、人均用水量(PC -  $x_7$ )、年均气温(AT -  $x_8$ )、干旱指数(DI -  $x_9$ )、人均灌溉面积(IC -  $x_{10}$ )、人均耕地面积(PA -  $x_{11}$ )、第一产业产值占 GDP 比重(PG -  $x_{12}$ )、第二产业产值占 GDP 比重(SG -  $x_{13}$ )、第三产业占 GDP 比重(TG -  $x_{14}$ )、工业产值占 GDP 比重(IG -  $x_{15}$ )、高耗水行业产值占工业总产值比重(HA -  $x_{16}$ )、科学技术支出占财政支出比重(SF -  $x_{17}$ )、工业废水达标排放率(ID -  $x_{18}$ )、

教育支出占财政支出比重(EF -  $x_{19}$ )、高新技术产业产值(TV -  $x_{20}$ )、水价(TF -  $x_{21}$ )21 个影响因素指标,其他 12 个主成分的组成方式与  $f_1$  一致,在此不一一列出,最终得到表 4 的数据。

### 4.3 多元线性逐步回归分析

将 13 个主因子作为自变量,江苏省万元 GDP 用水量作为因变量,应用 SPSS 做多元线性逐步回归分析,结果见表 5。

其中  $R^2 = 0.999$ ,说明回归曲线拟合程度非常好( $R^2 = 1$ 为最高),对比于直接采用逐步回归( $R^2 = 0.969$ )模型精度更高。根据表 4、表 5 可以

表 4 1997—2010 年 13 个主因子数据

FAC_1	FAC_2	FAC_3	FAC_4	FAC_5	FAC_6	FAC_7	FAC_8	FAC_9	FAC_10	FAC_11	FAC_12	FAC_13
-1.268	-1.327	-1.454	1.005	0.034	1.861	-1.098	-0.803	0.107	0.187	0.448	0.119	0.038
-1.247	1.270	-1.576	0.977	-0.931	-1.826	-0.124	0.514	-0.489	-0.648	-0.369	-0.429	-0.081
-1.204	0.098	0.128	-0.734	1.283	0.319	0.261	2.219	1.642	0.326	-0.576	0.339	0.266
-0.932	0.652	0.213	-0.977	-0.575	0.367	1.457	-0.163	-1.307	1.637	1.360	0.813	-0.513
-0.738	-0.882	0.442	-1.356	-0.657	-0.045	0.787	-1.315	-0.099	-0.196	-2.152	-1.024	0.455
-0.470	-0.947	0.752	-0.663	1.085	-0.721	-0.014	-0.147	-0.744	-2.341	1.467	0.146	-0.145
-0.271	1.965	1.198	-0.095	0.937	-0.025	-1.699	-1.413	0.494	0.540	0.215	-0.569	0.165
0.223	-1.715	1.134	1.094	-0.355	-1.739	-0.451	-0.005	0.625	1.406	0.092	0.315	-0.878
0.386	0.552	1.108	0.629	-1.797	0.760	-0.018	0.100	0.575	-1.108	-0.314	1.960	0.824
0.725	-0.021	0.936	0.485	-0.789	1.108	-0.369	1.629	-1.083	-0.038	0.197	-2.133	-0.176
0.890	0.590	-0.254	1.309	0.927	0.571	1.991	-0.748	0.859	-0.534	-0.233	-0.370	-1.494
1.120	-0.235	-0.555	0.588	0.882	-0.474	0.757	-0.150	-0.359	0.653	0.380	-0.134	2.791
1.275	0.054	-0.687	-0.493	0.991	0.081	-0.977	0.378	-1.651	0.167	-1.518	1.354	-0.866
1.511	-0.055	-1.383	-1.768	-1.035	-0.239	-0.503	-0.095	1.432	-0.051	1.003	-0.388	-0.388

表 5 模型概述输出结果

模型	R	R <sup>2</sup>	调整 R <sup>2</sup>	R <sup>2</sup> 更改
1	0.956 <sup>a</sup>	0.914	0.907	0.914
2	0.972 <sup>b</sup>	0.945	0.934	0.030
3	0.987 <sup>c</sup>	0.974	0.966	0.029
4	0.996 <sup>d</sup>	0.991	0.988	0.017
5	0.998 <sup>e</sup>	0.995	0.992	0.004
6	0.999 <sup>f</sup>	0.998	0.996	0.003
7	1.000 <sup>g</sup>	0.999	0.998	0.001

看出对万元 GDP 用水量有影响的因素有主成分因子 FAC\_1, FAC\_2, FAC\_3, FAC\_4, FAC\_6, FAC\_7, FAC\_11。回归系数输出结果见表 6。

根据表 6 以及主因子的得分系数矩阵,将回归方程中的 FAC\_1, FAC\_2, FAC\_3, FAC\_4, FAC\_6, FAC\_7, FAC\_11 换算为 21 个影响因素:

$$y = -0.956f_1 - 0.172f_2 - 0.174f_3 + 0.063f_4 +$$

$$0.132f_6 - 0.050f_7 + 0.036f_{11} = -0.956(a_1x_1 + \dots + a_{21}x_{21}) - 0.172(b_1x_1 + \dots + b_{21}x_{21}) \quad (3)$$

表 6 回归系数输出结果

模型	非标准化系数		标准系数
	B	标准误差	Beta
7(常量)	0.000	0.011	
因子得分 1	-0.956	0.012	-0.956
因子得分 3	-0.174	0.012	-0.174
因子得分 2	-0.172	0.012	-0.172
因子得分 6	0.132	0.012	0.132
因子得分 4	0.063	0.012	0.063
因子得分 7	-0.050	0.012	-0.050
因子得分 11	0.036	0.012	0.036

通过换算影响计算结果,可以得出应变量江苏省万元 GDP 用水量与其 21 个影响因素之间的响应回归方程为

$$y = -0.114x_1 + 0.028x_2 - 0.098x_3 - 0.008x_4 + 0.1x_5 - 0.029x_6 + 0.093x_7 - 0.086x_8 + 0.122x_9 + 0.302x_{10} + 0.026x_{11} + 0.107x_{12} - 0.028x_{13} - 0.139x_{14} + 0.009x_{15} + 0.011x_{16} - 0.097x_{17} - 0.232x_{18} + 0.165x_{19} - 0.052x_{20} + 0.092x_{21} \quad (4)$$

从回归方程可以看出:人均灌溉面积(IC -  $x_{10}$ )、工业废水达标排放率(ID -  $x_{18}$ )、教育支出占财政支出比重(EF -  $x_{19}$ )、第三产业产值占GDP比重(TG -  $x_{14}$ )为影响江苏省万元GDP用水量最重要的4个因素。

## 5 结 语

本文通过模型构建与理论实际相结合,对影响江苏省万元GDP用水量的因子进行筛选,得到的5个主要因素对因变量的影响方式与现实情况相符,并根据其作用方式提出以下结论和建议。

(1)人均灌溉面积对用水效率影响最大,随着该指标增长,万元GDP用水量增加。这是由于:江苏省不仅是沿海经济发达省,又是农业大省,农业经济的快速发展需要大量的灌溉用水作为支撑,造成了一定程度的水资源浪费。基于江苏省的用水量总量控制指标,建议配套政策优惠措施,如对实施“万顷良田”建设的村庄给予适当财政奖励;加大灌溉技术改造以达到水资源高效以及重复利用;对污染严重的灌区周围水系进行治理,通过“质”的提高,达到“量”的提升。

(2)工业废水达标排放率是工业技术指标,达标排放率越高,万元GDP用水量越小。该结果确切地表明了水质与用水量的关系,废水治理越成功、水环境保护的越好,越能促进水的重复利用,提高用水效率,同时越能激发用水单位或人的节水意识,从而走上良性循环的发展道路,故建议对废水治理较好的工业企业进行奖励或成本补偿。

(3)教育支出占财政支出比重是经济指标。近年来,江苏省教育支出的比重不断减少,伴随着的是科学技术支出的比重不断加大。这说明提高用水效率,已经不能单一地通过宣传教育。在节水意识培养和人们素质提高的同时,应运用科技的力量从源头上提高工农业用水大户的效率问题,如高产水稻、低压管灌、喷灌等新科技都有效提高了用水效率。

(4)第一产业是用水大户,第三产业相对用水量较少。第三产业产值占GDP比重越高,总用水量越小,可以理解为相对于第一、第二产业,第三产业用水效率较高。从江苏省1997—2010年三大产业的比例数据可以看出江苏省第一产业的比例在逐渐减少,而第三产业比例在逐渐增加;另一方面,建议政府需同时加大第一产业、第二产业的用水管理,进一步提高用水效率。

## 参考文献:

- [1] 马静,陈涛,申碧峰,等. 水资源利用国内外比较与发展趋势[J]. 水利水电科学进展, 2007, 27(1):6-13.
- [2] 中国科学院可持续发展战略研究组. 2007中国可持续发展战略报告——水:治理与创新[M]. 北京:科学出版社, 2007.
- [3] 李世祥,成金华,吴巧生. 中国水资源利用效率区域差异分析[J]. 资源与环境, 2008, 18(3):215-220.
- [4] 林定恕. 重庆市西部水危机与水资源配置结构性矛盾[J]. 重庆环境科学, 1997, 19(6).
- [5] 向东进,李宏伟,刘小雅. 实用多元统计分析[M]. 武汉:中国地质大学出版社, 2005:54-66.
- [6] 佟金萍,马剑锋,刘高峰. 基于完全分解模型的中国万元GDP用水量变动及因素分析[J]. 资源科学, 2011, 33(10):1870-1876.