

DOI: 10.3880/j.issn.1004-6933.2016.04.009

改进的偏最小二乘法在青海省农业用水预测中的应用

王 洁, 章恒全

(河海大学商学院, 江苏 南京 211100)

摘要:根据 2000—2009 年影响青海省农业用水的 11 个因子的基础数据, 建立偏最小二乘回归模型, 考虑到模型的实用性和准确性, 运用后退法对偏最小二乘法进行改进, 剔除了 5 个不需要的变量, 得到了拟合精度更高的结果。选取 2010—2013 年数据进行模型检验。结果表明: 运用偏最小二乘法预测的结果与实际情况贴近, 并且改进的模型的贴近度更高。通过模型的应用, 可以看到偏最小二乘法在青海省农业用水预测中有较好的应用价值, 并且改进后的偏最小二乘法简化了模型, 提高了预测精度, 为青海省的农业用水预测提供了依据。

关键词:农业用水; 用水预测; 改进的偏最小二乘法; 青海省

中图分类号: TV211.1

文献标志码: A

文章编号: 1004-6933(2016)04-0055-05

Application of improved partial least squares method to prediction of agricultural water consumption in Qinghai Province

WANG Jie, ZHANG Hengquan

(Business School, Hohai University, Nanjing 211100, China)

Abstract: Based on the basic data of 11 factors that influenced the agricultural water consumption of Qinghai Province from 2000 to 2009, a partial least squares regression model was established. Considering the practicality and accuracy of the model, the backward method was used to improve the partial least squares method. Five unnecessary variables were excluded, and a higher prediction accuracy was obtained. Data from 2010 to 2013 were selected to verify the model. The results show that the prediction using the partial least squares method agreed with the actual situation, and the improved model showed even greater agreement. Therefore, the partial least squares method is highly applicable to the prediction of the agricultural water consumption of Qinghai Province. The improved partial least squares method greatly simplifies the original model, improves the prediction accuracy, and provides a basis for the prediction of the agricultural water consumption of Qinghai Province.

Key words: agricultural water consumption; water consumption prediction; improved partial least squares method; Qinghai Province

水资源问题是整个人类社会面临的一个严峻的问题, 中国作为世界第一人口大国, 水资源问题尤为严重, 人均水资源可利用量为 $2\ 200\ \text{m}^3$, 仅为全球平均水平的 $1/4$ ^[1]。青海省位于青藏高原东北部, 是长江、黄河、澜沧江和黑河的源头, 是我国重要的水源地, 素有“中华水塔”之称。青海省水资源总量相对丰富, 2013 年全省的平均降水量为 $298.8\ \text{mm}$, 水

资源总量为 $645.60\ \text{亿}\ \text{m}^3$, 其中地表水资源量为 $629.55\ \text{亿}\ \text{m}^3$, 地下水资源量为 $290.77\ \text{亿}\ \text{m}^3$ 。但其水资源存在较强的季节性和地域性, 全年降水主要集中在 6—9 月份, 且降水总量由东南向西北递减, 再加上其水资源的利用率和利用效率低下, 2013 年全省水资源开发利用率为 4.4% , 这些都加剧了水资源的供需矛盾^[2-3]。

作者简介: 王洁(1990—), 女, 硕士研究生, 研究方向为管理科学与工程。E-mail: 1174764572@qq.com

通信作者: 章恒全(1957—), 教授, 博士。E-mail: hqzhang630@163.com

青海省是农业大省,而水资源是粮食生产的基础,2013年青海省农业用水占总用水量的81.15%^[2],因此掌握农业用水情况对青海省的持续发展起到了至关重要的作用。截至2013年,我国的农田灌溉水有效利用系数为0.52^[4],而青海省的农田灌溉水利用系数为0.448^[2],低于全国平均水平,水资源浪费情况十分严重,严重威胁了青海省的粮食安全和经济发展。因此有必要对青海省的农业用水进行预测,从而合理分配各行业的用水,加强农业用水效率,使得农田灌溉水有效利用系数达到0.55以上^[5]。

目前,对于农业用水的分析主要集中在节水灌溉^[3,6-7]以及对作物用水的测算^[8-9]方面,而对农业用水的整体预测的研究相对较少^[10-11],尤其是对多影响因子下的农业用水量预测。笔者筛选了11个影响青海省农业用水的因素,构建了偏最小二乘法模型,通过MATLAB软件进行模型的模拟运算。为了提高模型的精度并且简化模型,对偏最小二乘模型进行了改进,并将预测结果与实际观测结果进行对比检验。结果表明,所构建的模型有较高的准确性,是合理的。所构建模型在一定程度上为青海省预测农业用水提供了依据,并且对调节农业用水,提高农业用水效率提供了定量的标准,具有现实意义。

1 偏最小二乘回归模型

1.1 偏最小二乘法的概念

偏最小二乘算法(partial least squares algorithm, PLSA)最早出现在20世纪70年代,Wold在研究经济学的过程中创建了非线性迭代偏最小二乘法^[12]。偏最小二乘法同一般的回归方法相比,不仅具备了主成分分析、典型相关性分析和多元线性回归分析的优点^[13],还能弥补这些方法的缺陷,解决自变量的多重共性问题,并且保证了主成分对因变量的解释力度,在解决实际问题时更具备优势。目前,PLSA在已被广泛运用于物理、生物、管理学等多领域,并在不断地改进中。

1.2 PLSA的计算步骤

本文考虑单因变量的PLSA的回归模型,设有一个因变量 y , p 个自变量 x_1, x_2, \dots, x_p ,样本数为 n ,将 \mathbf{X} 和 \mathbf{Y} 表示成矩阵 $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$,将原始数据进行标准化处理,记为 \mathbf{E}_0 和 \mathbf{F}_0 。按下列步骤建立回归方程:

1.2.1 提取主成分

求矩阵 $\mathbf{E}^T \mathbf{F} \mathbf{F}^T \mathbf{E}$ 最大特征值所对应的单位向量 \mathbf{w}_1 ,由此得到自变量的第一主成分 $\mathbf{t}_1 = \mathbf{E}_0 \mathbf{w}_1$,其中

$$\mathbf{w}_1 = \frac{\mathbf{E}_0^T \mathbf{F}_0}{\|\mathbf{E}_0^T \mathbf{F}_0\|} = \frac{1}{\sqrt{\sum_{j=1}^p r^2(x_j, y)}} \begin{bmatrix} r(x_1, y) \\ r(x_2, y) \\ \vdots \\ r(x_p, y) \end{bmatrix} \quad (1)$$

$$\mathbf{t}_1 = \mathbf{E}_0 \mathbf{w}_1 = \frac{\sum_{j=1}^p [r(x_j, y) \mathbf{E}_{0j}]}{\sqrt{\sum_{j=1}^p r^2(x_j, y)}} \quad (2)$$

式中: \mathbf{E}_0^T 为 \mathbf{E}_0 的转置矩阵; \mathbf{E}_{0j} 为 \mathbf{E}_0 的第 j 列; $r(x_j, y)$ 为第 j 个自变量 x_j 与 y 的相关系数,相关性越强则其值越大。

然后求出 \mathbf{E}_0 对 \mathbf{t}_1 的回归方程及回归系数 \mathbf{p}_1 :

$$\mathbf{E}_0 = \mathbf{t}_1 \mathbf{p}_1^T - \mathbf{E}_1 \quad \mathbf{p}_1 = \frac{\mathbf{E}_0^T \mathbf{t}_1}{\|\mathbf{t}_1\|^2} \quad (3)$$

则回归方程的残差方程为:

$$\mathbf{E}_1 = \mathbf{E}_0 - \mathbf{t}_1 \mathbf{p}_1^T \quad (4)$$

令 $\mathbf{E}_1 = \mathbf{E}_0$,重复下面的步骤(2),直到满足精度要求,则成分提取完毕,现假设最终对自变量提取了 k 个主成分 $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_k (k \leq p)$ 。

1.2.2 确定成分提取的个数——交叉有效性

在偏最小二乘回归模型中,只需引入对模型精度有改进作用的成分,并且每引入一个成分都需要检验其是否提高模型预测的精度。

此时需要交叉有效性 Q_h^2 :

$$Q_h^2 = 1 - \frac{P_h}{S_{h-1}} \quad (5)$$

$$\text{其中} \quad P_h = \sum_{i=1}^n \left(y_i - \hat{y}_{h(-i)} \right)^2$$

$$S_{(h-1)} = \sum_{i=1}^n \left(y_i - \hat{y}_{(h-1)i} \right)^2 \quad (6)$$

式中: $\hat{y}_{h(-i)}$ 为在样本点集合中舍掉第 i 个样本点后,用 h 个主成分拟合所得的预测值; $\hat{y}_{(h-1)i}$ 为 $(h-1)$ 个主成分的回归模型所得的拟合值; P_h 是指从所有样本中舍去第 i 个样本点后,得到 h 个主成分的回归模型,再进行预测所得的预测误差平方和; $S_{(h-1)}$ 是指用所有样本拟合得到含 $(h-1)$ 个主成分的回归模型,再进行预测所得的预测误差平方和。

当 $Q_h^2 \geq 0.0975$ 时,认为新增主成分对模型有改进,则引入;反之,则停止。

1.2.3 变量投影重要性分析

在偏最小二乘回归建模中,用变量投影重要性指标 I 来说明自变量对因变量的解释能力,即自变量表达式为

$$I_j = \sqrt{\frac{p}{\sum_{h=1}^k r^2(y, t_h)} \sum_{h=1}^k r^2(y, t_h) w_{hj}^2} \quad (j = 1, 2, \dots, p) \quad (7)$$

式中: I 指在解释因变量时,自变量的作用的重要性; $r(y, t_h)$ 是因变量 y 与提取的第 h 个主成分的相关系数; w_{hj} 是第 h 个主成分对应的单位向量的第 j 个指标值。通常情况下, I 越大,表明该变量对因变量的影响越大,该变量越重要。一般 $I \geq 1$ 时,认为它为重要因素; $I < 0.5$ 时,认为它为不重要因素。

1.2.4 建立回归模型

按照以上步骤,确定可以提取 k 个主成分,得到:

$$E_0 = t_1 p_1^T + t_2 p_2^T + \dots + t_k p_k^T \quad (8)$$

$$F_0 = t_1 r_1 + t_2 r_2 + \dots + t_k r_k + F_k \quad (9)$$

式中: p_1, p_2, \dots, p_k 为 E_0 分别对 t_1, t_2, \dots, t_k 的回归系数; r_1, r_2, \dots, r_k 分别是因变量 y 与 t_1, t_2, \dots, t_k 的相关系数; F_k 为残值。

因为 t_1, t_2, \dots, t_k ,可以表示为 E_{0j} 的线性组合,所以对标准化后的自变量 $x_j^* (j=1, 2, \dots, p)$ 的回归方程 y^* :

$$y^* = \alpha_1 x_1^* + \alpha_2 x_2^* + \dots + \alpha_p x_p^* + F_k \quad (10)$$

其中

$$\alpha_j = \sum_{h=1}^k r_h w_{hj}^*$$

$$w_{hj}^* = \prod_{m=1}^{h-1} (E - w_m p_m^T) w_{hj} \quad h=1, 2, \dots, k \quad (11)$$

式中: E 为单位矩阵。

1.3 改进的PLSA

PLSA的目的是求解使方差 $V(t_i)$ 和相关系数 $C(t_i, y)$ 较大的成分,事实上,PLSA的结果是求解使协方差 $C(t_i, y)$ 较大的成分。这是由于在选择自变量时,无法剔除与因变量不相关的信息,因此要改进偏最小二乘回归模型,将解释性不强的变量剔除,使回归方程更加的简洁、准确。

本文在原有偏最小二乘回归模型的基础上采用后退法^[14],对模型进行改进,具体步骤为:

a. 构建原始变量的偏最小二乘回归方程。

表1 青海省2000—2013年相关因子原始数据

| 年份 | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | x_{11} | y |
|------|-------|--------|---------|-------|--------|--------|-------|-------|--------|----------|----------|-------|
| 2000 | 516.5 | 669.16 | 263.68 | 0.15 | 2047.2 | 612.70 | 27.91 | 20.04 | 211.42 | 631.86 | 100 | 21.27 |
| 2001 | 523.1 | 652.50 | 300.10 | 0.15 | 1947.6 | 580.02 | 27.23 | 19.23 | 208.33 | 615.30 | 100 | 20.52 |
| 2002 | 528.6 | 604.67 | 340.70 | 0.14 | 2053.5 | 558.23 | 27.03 | 19.05 | 193.55 | 656.17 | 100 | 20.36 |
| 2003 | 533.8 | 555.21 | 390.20 | 0.12 | 2146.7 | 634.66 | 29.01 | 19.64 | 181.73 | 720.46 | 100 | 22.76 |
| 2004 | 538.6 | 542.05 | 466.10 | 0.13 | 2094.9 | 606.80 | 30.17 | 19.65 | 180.33 | 726.48 | 100 | 22.82 |
| 2005 | 543.2 | 542.25 | 543.30 | 0.12 | 2430.4 | 768.72 | 31.46 | 19.66 | 176.54 | 742.35 | 620 | 22.87 |
| 2006 | 547.7 | 542.10 | 648.50 | 0.10 | 2052.7 | 569.00 | 32.20 | 19.66 | 176.32 | 743.39 | 3200 | 22.82 |
| 2007 | 551.6 | 542.20 | 797.35 | 0.10 | 2283.4 | 661.62 | 31.11 | 17.72 | 176.59 | 669.05 | 3600 | 21.47 |
| 2008 | 554.3 | 542.72 | 1018.62 | 0.10 | 2317.2 | 657.78 | 31.56 | 18.95 | 177.24 | 712.62 | 3740 | 23.44 |
| 2009 | 557.3 | 565.40 | 1081.27 | 0.10 | 2654.2 | 868.58 | 34.96 | 18.55 | 175.43 | 704.87 | 5032 | 23.40 |
| 2010 | 563.2 | 567.30 | 1350.43 | 0.10 | 2454.4 | 741.11 | 36.23 | 18.40 | 177.69 | 690.41 | 7684 | 24.31 |
| 2011 | 569.1 | 588.32 | 1670.44 | 0.09 | 2417.5 | 733.12 | 36.60 | 18.71 | 179.44 | 695.00 | 8370 | 24.61 |
| 2012 | 573.2 | 588.53 | 1893.54 | 0.09 | 2653.0 | 895.22 | 27.62 | 16.30 | 186.69 | 582.00 | 11000 | 22.48 |
| 2013 | 572.6 | 588.64 | 2307.86 | 0.09 | 2134.4 | 645.60 | 28.09 | 16.18 | 187.19 | 576.28 | 21192 | 22.52 |

$$\hat{y} = \left(\bar{y} - \sum_{j=1}^p \alpha_j \frac{s_y \bar{x}_j}{s_j} \right) + \sum_{j=1}^p \alpha_j \frac{s_y}{s_j} x_j \quad (12)$$

式中: \bar{x}_j 和 s_j 分别为 x_j 的均值及标准差; \bar{y} 和 s_y 分别为 y 的均值及标准差,记该回归模型的平均拟合误差为 δ_0 。

b. 依次删去变量 $x_j (j=1, 2, \dots, p)$,用其余的 $(p-1)$ 个变量构建回归模型 $\hat{y}_{lj} = f_{lj}(x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p)$,则平均拟合误差为 $\delta_{lj} (j=1, 2, \dots, p)$ 。若 $\min_{j=1, 2, \dots, p} \delta_{lj} < \delta_0$ 则认为去掉变量 x_j 后模型精度有所提高,应删掉该变量,进行下一步;反之保留,以前一步的回归模型为最终回归模型。

c. 重复步骤b,假设前 $(l-1)$ 步删掉了 $(l-2)$ 个变量,回归方程的拟合误差为 $\delta_{(l-1)j} (j=l-1, \dots, p)$ 。若 $\min_{j=l-1, 2, \dots, p} \delta_{(l-1)j} \geq \min_{j=l-2, \dots, p} \delta_{(l-2)j}$ 时,则第 $(l-1)$ 步所得的回归方程的预测精度是最佳的,停止删除变量,所得模型为最终模型。

2 实证分析

本文根据青海省的实际状况,选取了11个影响农业用水(y ,亿 m^3)的因子,分为社会经济系统因子:总人口(x_1 ,万人)、耕地面积(x_2 , 10^3hm^2)、GDP(x_3 ,亿元)、第一产业占GDP比例(x_4 ,%) ;水资源系统因子:年降水总量(x_5 ,亿 m^3)、水资源总量(x_6 ,亿 m^3)、供水量(x_7 ,亿 m^3)、农田灌溉用水量(x_8 ,亿 m^3)、有效灌溉面积(x_9 , 10^3hm^2)、农田灌溉单位面积用水量(x_{10} , m^3) ;管理因子:水资源费征收(x_{11} ,万元),其中2000—2009年的数据作为训练样本,2010—2013年的数据作为检验样本(表1),统计的数据来自历年青海省统计年鉴、青海省水资源公报、中国水利年鉴和黄河年鉴。

2.1 原始数据的偏最小二乘模型

通过MATLAB编程,得到各变量的相关系数,

发现变量之间存在着严重的多重相关性。为了克服变量之间的多重相关性在建模过程中的消极影响,运用 PLSA 提取主成分,得到交叉有效性 $Q_3^2 = -0.2207 < 0.0975$,所以只需要提取 3 个主成分 t_1, t_2, t_3 即可。提取完主成分后,最终确定含 3 个主成分的偏最小二乘回归模型,并转化为原始数据的回归模型: $y = -2.4683 + 0.0047x_1 - 0.0016x_2 + 0.0006x_3 - 4.29x_4 + 0.0006x_5 + 0.0019x_6 + 0.0941x_7 + 0.7088x_8 - 0.0041x_9 + 0.0072x_{10} + 0.00002x_{11}$,各变量的拟合值如表 2 所示,计算可得平均拟合误差为 1.2559%,并且回归模型的系数符号与相关系数一致,由此可得方程的拟合效果较好。

2.2 改进的偏最小二乘模型

由上述可知,原始数据的回归模型的精度较高,但解释变量较多,模型较为复杂,因此可以在提高精度的基础上,将一些解释性相对弱的变量剔除,由此得到更为简洁的回归模型。

根据后退法的原理,运用 MATLAB 编写程序并运行,变量剔除结果如表 3 所示,由表 3 可知,前 5 轮剔除变量时,每一轮的可以通过删除变量提高模型的精度,依次删除变量 $x_{10}, x_9, x_{11}, x_5, x_1$ 。在第 6 轮变量删除时,由表可知,当剔除变量 x_7 后,方程的拟合误差最小,但此时的拟合误差 $\delta_{s_1} = 0.9843\% < \min_{j=1, \dots, p} \delta_{s_j}$,此时删除解释变量不提高模型精度,所以前一轮剔除变量 x_1 后所得模型即为最终模型。

因此,在剔除了变量 $x_{10}, x_9, x_{11}, x_5, x_1$ 后所得的模型为最终模型,模型精度在原始模型的基础上提高了 21.63%,回归方程为: $y = 0.4447 - 0.0067x_2 + 0.0021x_3 - 8.7907x_4 + 0.0018x_5 + 0.073x_6 + 1.1455x_7$,各变量的系数与相关系数的符号一致,变量的剔除是有效的。

对检验样本进行检验,其拟合值如表 4 所示,从表 4 中可以得出,原始回归方程的平均拟合误差 $\delta = 4.6101\%$,改进后的平均拟合误差 $\delta = 1.4178\%$ 。

表 2 回归方程拟合值

| 样本点 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|--------|---------|
| 拟合值 | 21.2122 | 20.4306 | 20.8318 | 22.3664 | 22.4988 | 23.3771 | 23.0923 | 21.5059 | 22.897 | 23.5178 |
| 拟合误差(%) | 0.2719 | 0.4355 | 2.3173 | 1.7294 | 1.4074 | 2.2173 | 1.1933 | 0.1673 | 2.3164 | 0.5034 |

表 3 各轮变量剔除拟合误差值

| 轮次 | 剔除变量 | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | x_7 | x_8 | x_9 | x_{10} | x_{11} |
|----|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|----------|
| 1 | 拟合误差(%) | 1.2456 | 1.2207 | 1.3580 | 1.2333 | 1.2453 | 1.2317 | 1.2189 | 1.9273 | 1.2295 | 1.1581 | 1.2431 |
| 2 | 拟合误差(%) | 1.1690 | 1.2525 | 1.2846 | 1.1564 | 1.1649 | 1.1782 | 1.1302 | 2.2697 | 1.1065 | — | 1.1568 |
| 3 | 拟合误差(%) | 1.1309 | 1.1647 | 1.2964 | 1.1537 | 1.0841 | 1.1720 | 1.0622 | 2.2192 | — | — | 1.0289 |
| 4 | 拟合误差(%) | 1.0531 | 1.2647 | 1.2347 | 1.0294 | 1.0102 | 1.1293 | 1.0368 | 2.2310 | — | — | — |
| 5 | 拟合误差(%) | 0.9843 | 1.2446 | 1.2887 | 1.0294 | — | 1.0347 | 1.0323 | 2.1173 | — | — | — |
| 6 | 拟合误差(%) | — | 1.3438 | 1.2943 | 1.0313 | — | 1.0346 | 0.9844 | 2.1214 | — | — | — |

注:表中打线段的单元格表示本轮次已删除变量。

从表 4 中可以直接看到,改进后的模型所得的每一个拟合值的拟合误差均小于原始模型的拟合误差,进一步说明了改进后的偏最小二乘模型的预测精度更高,与观测值更贴近,因此对原始模型进行改进是必要的、有效的。

表 4 检验样本拟合值

| 观察值 | 原始回归模型拟合值 | 原始的拟合误差/% | 改进回归模型拟合值 | 改进的拟合误差/% |
|-------|-----------|-----------|-----------|-----------|
| 24.31 | 23.3148 | 4.0938 | 23.6893 | 2.5532 |
| 24.61 | 23.8115 | 3.2446 | 24.6768 | 0.2714 |
| 22.48 | 21.0705 | 6.2700 | 22.0249 | 2.0245 |
| 22.52 | 20.6956 | 4.8319 | 22.3349 | 0.8219 |

通过观察观测值、原始拟合值以及改进后的拟合值的贴近程度,如图 1 所示,发现两者与观测值都有较好的贴近度,但改进后的模型所得的拟合值与观测值的贴近度更高,因此更加证明了对模型进行改进的准确性,并且大大简化了模型,使模型运用起来更方便。通过建立模型来预测青海省的农业用水量,为合理分配各行用水提供了保障,并且通过分析不同解释变量对农业用水的影响的大小,为提高农业用水效率、减少浪费提供了依据。模型改进后,各变量对青海省农业用水的解释能力都较好(表 5),其中农田灌溉用水量和供水量最为重要,农业用水中农田灌溉为主要耗水部分,并且青海省为农业大省,供的水越多,农业所分配的水量越多,因此这两个变量对农业用水量最为重要,符合实际情况,说明

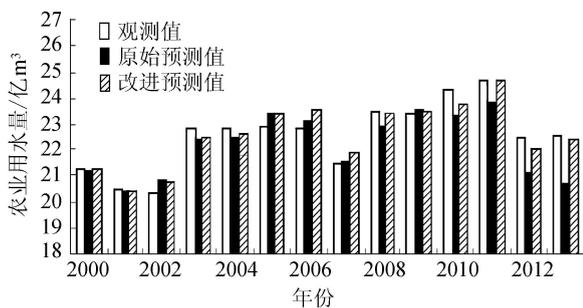


图 1 模型拟合值

表5 各因数对农业用水量的重要程度

| 因数 | 因数符号 | <i>t</i> |
|------------|-------|----------|
| 农田灌溉用水量 | x_8 | 1.2162 |
| 供水量 | x_7 | 1.0752 |
| 耕地面积 | x_2 | 0.9844 |
| GDP | x_3 | 0.9595 |
| 第一产业占GDP比例 | x_4 | 0.9535 |
| 水资源总量 | x_6 | 0.7522 |

拟合模型与事实相符。此外,青海省水资源总量较为稳定,年际变化不大,对农业用水的影响相对于其他因素较小。其余因素对农业用水的重要程度较为接近,主要集中在社会经济因子中,影响程度接近1,说明经济的发展、社会的改变对农业用水有较大的影响。近年来,青海省政府不断引导社会向工业化发展,工业在国民生产中比重不断上升,青海省的产业结构在改变,农业的比例在减小,农业用水也会受到较大的影响,这与实际情况一致,进一步说明了模型的有效性。

3 结论

a. 从社会经济系统、水资源系统和管理方面筛选11个影响青海省农业用水的解释变量,运用MATLAB软件对2000—2009年的数据建立偏最小二乘回归模型,得到平均拟合误差为1.2559%,说明了PLSA的拟合精度较高。

b. 偏最小二乘模型无法剔除与因变量不相关的因子,通过改进模型,删减对农业用水量解释性不强的总人口、年降水总量、有效灌溉面积、农田亩均灌溉用水量、水资源费征收5个变量,得到拟合精度最高为0.9843%的回归模型,极大地简化了模型,并提高了模型精度。

c. 选取2010—2013年的数据为检验样本,在对检验样本进行检验后,得到较为贴近观测值的拟合值,表明模型的可靠性,并且改进后的模型的贴程度更高,说明了本文所得的预测模型与实际符合,具有现实意义。

参考文献:

[1] SUN Caizhi, ZHAO Liangshi, ZOU Wei, et al. Water resource utilization efficiency and spatial spillover effects in China[J]. Journal of Geographical Sciences, 2014, 24(5):771-788.

[2] 青海省水利厅. 2013年青海省水资源公报[EB/OL]. [2015-05-03]. <http://www.qhsl.gov.cn/uploadfile/2013年青海省水资源公报.pdf>.

[3] 青海省统计局. 青海省统计年鉴2014[M]. 北京:中国统计出版社,2014:24-30.

[4] 袁寿其,李红,王新坤. 中国节水灌溉装备发展现状、问

题、趋势与建议[J]. 排灌机械工程学报,2015,33(1):78-92. (YUAN Shouqi, LI Hong, WANG Xinkun. Status, problems, trends and suggestions for water-saving irrigation equipment in China[J]. Journal of Drainage and Irrigation Machinery Engineering, 2015, 33(1):78-92. (in Chinese))

[5] 中华人民共和国中央人民政府. 国家农业节水纲要(2012—2020)[EB/OL]. [2015-05-03]. http://www.gov.cn/zwgg/2012-12/15/content_2291002.htm.

[6] JENSEN M E. Beyond irrigation efficiency[J]. Irrigation Science, 2007, 25(3):233-245.

[7] 李保国,黄峰. 1998—2007年中国农业用水分析[J]. 水科学进展, 2010, 21(4):575-583. (LI Baoguo, HUANG Feng. Trends in China's agricultural water use during recent decade using the green and blue water approach[J]. Advances in Water Science, 2010, 21(4):575-583. (in Chinese))

[8] ZHANG Yongqin, MIAO Qilong, PENG Buzhuo. Calculation and analysis on change of agricultural water consumption in the Changjiang delta[J]. Journal of Geographical Sciences, 2001, 11(4):321-325.

[9] FU Yuanyuan, YANG Guijun, WANG Jihua, et al. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyper spectral measurements[J]. Computers and Electronics in Agriculture, 2014, 100(2):51-59.

[10] 田丝,张永丽. 主成分回归模型在农业需水量预测中的应用[J]. 资源开发与市场, 2012, 38(7):580-582. (TIAN Si, ZHANG Yongli. Agricultural water demand forecast based on principal component regression model[J]. Resource Development and Market, 2012, 38(7):580-582. (in Chinese))

[11] 刘迪,胡彩虹,吴泽宁. 基于定额定量分析的农业用水需求预测研究[J]. 灌溉排水学报, 2008, 27(6):88-91. (LIU Di, HU Caihong, WU Zening. Predicting method for demand of agriculture water based on quantitative analysis[J]. Journal of Irrigation and Drainage, 2008, 27(6):88-91. (in Chinese))

[12] 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,2006:56.

[13] 王琛娇. 低碳视角下的城市交通发展路径研究[D]. 无锡:江南大学,2012.

[14] 杨国栋. 基于变量筛选的偏最小二乘回归方法及其应用[D]. 长沙:中南大学,2013.

(收稿日期:2015-09-08 编辑:徐娟)

