

DOI: 10.11766/trxb201709140242

提高三元肥效模型建模成功率的主成分回归技术研究*

李娟¹ 章明清^{1†} 许文江² 孔庆波¹ 姚宝全³

(1 福建省农业科学院土壤肥料研究所, 福州 350013)

(2 福建省亚热带植物研究所, 福建厦门 361000)

(3 福建省农田建设与土壤肥料技术推广总站, 福州 350003)

摘要 三元二次多项式肥效模型在施肥实践中存在大量非典型式, 严重制约了肥效模型计量精确性和应用价值。为提高田间肥效试验建模成功率, 本研究探讨肥效模型建模新方法。根据福建省早稻 171 个“3414”设计的氮磷钾田间肥效试验结果, 研究三元二次多项式肥效模型多重共线性的危害和诊断方法以及主成分回归建模技术。结果表明, 试验设计矩阵 X 的方阵 ($X^T X$) 条件数和方差膨胀因子的诊断指标均显示, 三元二次多项式肥效模型存在严重的多重共线性, 且这种共线性主要是因模型设定本身造成的。多重共线性严重制约了普通最小二乘法 (OLS) 的有效性, OLS 回归建模的三元典型肥效模型比例仅占试验总数的 27.5%。主成分回归对设计矩阵提取互不相关的 9 个主成分, 消除了多重共线性危害。采用前 7 个主成分回归建模, 三元典型肥效模型比例提高至 43.3%, 为 OLS 法的 1.6 倍。尤其是主成分回归大幅度降低了模型系数符号不合理的非典型式, 从 OLS 建模的 14.6% 下降至零; 模型无最高产量点的非典型式比例从 OLS 建模的 36.8% 下降至 21.1%。7 个主成分回归建模利用了试验设计矩阵 X 99.9% 以上的方差信息, 其所得典型式与 OLS 法所得典型式在推荐施肥量和预测产量上几乎无差别。针对主成分回归是一种有偏估计, 采用 OLS 法结合前 7 个主成分回归建模, 三元典型肥效模型所占比例进一步提高至 55.6%, 为 OLS 法的 2.0 倍, 结果明显优于单独使用 OLS 回归或者主成分回归。因此, 主成分回归可消除三元二次多项式肥效模型的多重共线性危害, OLS 法结合前 7 个主成分回归建模是显著提高水稻氮磷钾三元典型肥效模型比例的最优建模策略。

关键词 早稻; 氮磷钾; 肥效模型; 多重共线性; 主成分回归; 普通最小二乘法 (OLS)

中图分类号 S147.21 **文献标识码** A

肥料效应函数是实现计量施肥的主要技术手段之一, 其中, 二次多项式是研究和应用最多的肥效模型^[1-4]。但在施肥实践中, 应用普通最小二乘法 (OLS) 建立的二元或三元二次多项式肥效模型出现大量的非典型式^[5-7], 诸如“最高产量”不高、“最佳产量”不佳、模型一次项系数为负数或二次项系数为正数、推荐施肥量远外推

等令人困惑的异常结果。针对肥效模型在应用中存在的问题, 国内外学者就肥效模型选择与其适用性^[8-10]、试验设计与参数估计方法^[11-13]、类特征肥效模型构建^[14-16]及非典型式推荐施肥优化方法^[13]等方面进行了深入研究, 提出了诸多改进措施, 但对肥效模型设定本身是否存在问题及其应对方法则鲜见研究报道^[4]。当前广泛应用的

* 国家自然科学基金项目 (31572203)、福建省农业科学院 PI 项目 (2016PI-31) 和农业部测土配方施肥项目 (2011-2015) 资助 Supported by the National Natural Science Foundation of China (No. 31572203) and PI Project of Fujian Academy of Agricultural Sciences (No. 2016PI-31) and Soil-Test-Based Formulated Fertilizer Project of the Ministry of Agriculture of the People's Republic of China (No. 2011-2015)

† 通讯作者 Corresponding author, E-mail: zhangmq2001@163.com

作者简介: 李娟 (1977—), 女, 硕士, 副研究员, 主要从事平衡施肥和施肥与环境研究。E-mail: lj-95@163.com

收稿日期: 2017-09-14; 收到修改稿日期: 2017-10-26; 优先数字出版日期 (www.cnki.net): 2017-11-29

“3414”设计方案的田间试验结果, 既使采用蒙特卡罗 (Monte Carlo) 建模法^[13], 二元和三元二次多项式典型肥效模型的比例也仅有56.7%和37.3%, 这种状况严重制约了肥效模型的计量精确性和应用价值。

初步研究^[4]表明, 二次多项式肥效模型存在严重的多重共线性, 制约了OLS法回归建模的有效性。为此, 本研究利用福建省近年来“3414”试验设计方案完成的早稻氮磷钾田间肥效试验结果, 探讨三元二次多项式肥效模型多重共线性诊断方法, 以及主成分回归 (PCR) 技术与其在消除多重共线性危害上的应用效果, 旨在为提高氮磷钾田间肥效试验的建模成功率提供一种新方法。

1 材料与方法

1.1 早稻氮磷钾“3414”田间肥效试验资料的收集整理

近年来, 福建省在水稻测土配方施肥工作中, 在福州市、宁德市、南平市、三明市、龙岩市、莆田市、漳州的龙海市以及平和县、泉州的南安市等水稻主产区, 土肥技术力量较强的主要项目县的早稻上先后完成了171个氮磷钾田间肥效试验。这些试验均采用“3414”设计方案^[17], 即: (1) $N_0P_0K_0$; (2) $N_0P_2K_2$; (3) $N_1P_2K_2$;

(4) $N_2P_0K_2$; (5) $N_2P_1K_2$; (6) $N_2P_2K_2$; (7) $N_2P_3K_2$; (8) $N_2P_2K_0$; (9) $N_2P_2K_1$; (10) $N_2P_2K_3$; (11) $N_3P_2K_2$; (12) $N_1P_1K_2$; (13) $N_1P_2K_1$; (14) $N_2P_1K_1$ 。其中, “0”水平表示不施肥, “2”水平为试验前的当地氮磷钾推荐施肥量, “1”水平和“3”水平的施肥量分别为“2”水平的50%和150%。根据肥效模型建模过程中可能出现的未达统计显著水平、典型肥效模型和三种不同形式的非典型肥效模型^[7]等几种类别, 本研究选择7个代表性早稻试验结果 (选择依据主要是: 根据普通最小二乘法建立三元肥效模型可能出现的几种情形, 在兼顾主要稻田土壤类型基础上, 分别选择了OLS法未通过显著性检验的1个代表性试验点、OLS法产生的3种不同类型非典型式的3个代表性试验点、OLS法为典型式的3个代表性试验点) 来探讨三元二次多项式肥效模型的多重共线性诊断方法及其主成分回归技术, 并利用171个早稻氮磷钾田间肥效试验结果评价主成分回归的应用效果。这7个试验点涵盖了灰泥田、黄泥田和灰沙田等福建稻田的主要土壤类型, 采用常规方法^[18]测定的土壤理化性状 (表1) 也大致反映了各肥力指标的主要变化范围。7个代表性试验点 $N_2P_2K_2$ 处理的施肥量见表1, 相关试验产量见表2。供试水稻品种采用当地大面积种植的良种, 试验设计方案和田间管理措施与文献 [17] 相同。

表1 早稻7个代表性试验点土壤主要理化性状及 $N_2P_2K_2$ 处理施肥量

Table 1 Main physical and chemical properties of the soils in the seven representative experiment sites and fertilizer application rate of their Treatment $N_2P_2K_2$

序号 No	试验地点 Sites	土属 Soil genus	土壤主要理化性状 Soil physical and chemical properties				$N_2P_2K_2$ 处理施肥量Application rate of Treatment $N_2P_2K_2$ (kg hm ⁻²)			
			pH	有机质 Organic matter (g kg ⁻¹)	碱解氮 Alkaline N (mg kg ⁻¹)	有效磷 Olsen-P (mg kg ⁻¹)	速效钾 Available K (mg kg ⁻¹)	N	P ₂ O ₅	K ₂ O
1	连城县 Liancheng County	灰泥田Plaster field	5.3	28.3	109	22.4	184	180	72	135
2	龙海市 Longhai City	灰泥田Plaster field	5.3	16.6	84	25.2	73	150	60	105
3	福清市Fuqing City	黄泥田Yellow clayey soil	7.1	18.1	105	21.2	18	150	63	104
4	连城县 Liancheng County	灰泥田Plaster field	4.6	24.7	144	8.0	110	180	72	135
5	仙游县 Xianyou County	黄泥田Yellow clayey soil	5.4	20.8	128	19.0	58	165	66	99
6	龙海市 Longhai City	灰泥田Plaster field	5.2	15.6	124	37.5	117	150	60	105
7	仙游县 Xianyou County	灰沙田Grey shatin soil	5.4	21.1	116	17.0	60	165	50	116

表2 早稻7个代表性试验点各施肥处理的稻谷产量

Table 2 Early rice yields in the seven representative trial sites relative to treatment (kg hm⁻²)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	5 730	4 482	5 898	6 150	7 134	7 134	6 882	6 432	6 966	6 933	6 864	6 948	6 836	7 134
2	4 539	4 605	6 126	6 344	6 639	6 726	6 486	6 026	6 554	6 266	6 135	6 116	6 000	6 681
3	5 985	5 015	7 080	7 602	8 249	8 294	7 278	7 440	7 683	6 551	8 168	7 521	7 118	7 926
4	4 250	4 875	5 709	6 500	7 376	7 625	6 759	6 875	7 359	7 280	7 526	7 125	6 626	6 875
5	4 800	5 355	6 705	6 675	7 125	7 095	7 020	6 000	6 870	6 675	7 095	7 050	7 005	7 050
6	7 470	6 225	8 599	8 300	8 825	8 876	8 351	8 576	9 026	8 576	8 450	8 675	8 475	8 745
7	4 800	5 700	6 945	6 900	7 350	7 395	7 305	6 495	6 945	7 050	6 855	6 705	6 645	6 795

注：第一列序号1~7代表试验地点；第一行序号1~14代表施肥处理，分别为：1N₀P₀K₀；2N₀P₂K₂；3N₁P₂K₂；4N₂P₀K₂；5N₂P₁K₂；6N₂P₂K₂；7N₂P₃K₂；8N₂P₂K₀；9N₂P₂K₁；10N₂P₂K₃；11N₃P₂K₂；12N₁P₁K₂；13N₁P₂K₁；14N₂P₁K₁。下同 Note: The serial number, 1~7 in the first column represents one of the seven trial sites, sequentially. The serial number, 1~14, in the first row represents 14 fertilization treatments, i.e., 1N₀P₀K₀; 2N₀P₂K₂; 3N₁P₂K₂; 4N₂P₀K₂; 5N₂P₁K₂; 6N₂P₂K₂; 7N₂P₃K₂; 8N₂P₂K₀; 9N₂P₂K₁; 10N₂P₂K₃; 11N₃P₂K₂; 12N₁P₁K₂; 13N₁P₂K₁; and 14N₂P₁K₁. The same below

1.2 多项式肥效模型的多重共线性诊断方法

假设三元二次多项式肥效模型的数学形式如下：

$$\hat{Y} = b_0 + b_1N + b_2P + b_3K + b_4N^2 + b_5P^2 + b_6K^2 + b_7NP + b_8NK + b_9PK + \varepsilon \quad (1)$$

式中，N、P、K分别表示N、P₂O₅和K₂O施肥量，

回归变量： N P K N² P² K² NP NK PK

$$X = \begin{bmatrix} N_0 & P_0 & K_0 & N_0^2 & P_0^2 & K_0^2 & N_0P_0 & N_0K_0 & P_0K_0 \\ N_0 & P_2 & K_2 & N_0^2 & P_2^2 & K_2^2 & N_0P_2 & N_0K_2 & P_2K_2 \\ N_1 & P_2 & K_2 & N_1^2 & P_2^2 & K_2^2 & N_1P_2 & N_1K_2 & P_2K_2 \\ N_2 & P_0 & K_2 & N_2^2 & P_0^2 & K_2^2 & N_2P_0 & N_2K_2 & P_0K_2 \\ N_2 & P_1 & K_2 & N_2^2 & P_1^2 & K_2^2 & N_2P_1 & N_2K_2 & P_1K_2 \\ N_2 & P_2 & K_2 & N_2^2 & P_2^2 & K_2^2 & N_2P_2 & N_2K_2 & P_2K_2 \\ N_2 & P_3 & K_2 & N_2^2 & P_3^2 & K_2^2 & N_2P_3 & N_2K_2 & P_3K_2 \\ N_2 & P_2 & K_0 & N_2^2 & P_2^2 & K_0^2 & N_2P_2 & N_2K_0 & P_2K_0 \\ N_2 & P_2 & K_1 & N_2^2 & P_2^2 & K_1^2 & N_2P_2 & N_2K_1 & P_2K_1 \\ N_2 & P_2 & K_3 & N_2^2 & P_2^2 & K_3^2 & N_2P_2 & N_2K_3 & P_2K_3 \\ N_3 & P_2 & K_2 & N_3^2 & P_2^2 & K_2^2 & N_3P_2 & N_3K_2 & P_2K_2 \\ N_1 & P_1 & K_2 & N_1^2 & P_1^2 & K_2^2 & N_1P_1 & N_1K_2 & P_1K_2 \\ N_1 & P_2 & K_1 & N_1^2 & P_2^2 & K_1^2 & N_1P_2 & N_1K_1 & P_2K_1 \\ N_2 & P_1 & K_1 & N_2^2 & P_1^2 & K_1^2 & N_2P_1 & N_2K_1 & P_1K_1 \end{bmatrix}$$

矩阵X共有14行和9列，每一行表示一个试验处理，每一列表示一个回归变量。在回归分析的经典理论中，假设式(1)模型的9个回归变量在两两之间不存在线性相关^[19]。但在实际问题中，这些

kg hm⁻²；ε为拟合残差，kg hm⁻²；Ŷ为拟合产量，kg hm⁻²。在应用OLS回归分析时，根据“3414”设计方案14个施肥处理的氮、磷、钾施肥量与式(1)模型对应的9个回归变量N、P、K、N²、P²、K²、NP、NK和PK可组成一个试验设计矩阵X：

回归变量之间总是会存在一定程度的相关性。在数学上，如果存在某些常数c₁、c₂…c₉与相应各回归变量的线性组合等于常数c₀这一关系近似成立，则表示变量间存在多重共线性。因此，一个常用但不

完全合适的共线性度量方法是样本相关系数平方即 r^2 [20]。精确共线性对应于 $r^2=1$ ，非共线性对应于 $r^2=0$ 。当 r^2 越接近于1，近似共线性就越强。

度量多重共线性严重程度的一个重要指标是方阵 $X^T X$ 的条件数 [19-20]，即：

$$k = \lambda_{\max} / \lambda_{\min} \quad (2)$$

式中， k 为条件数， λ_{\max} 、 λ_{\min} 分别为方阵 $X^T X$ 的最大特征值和最小特征值。一般应用经验认为，若 $k < 100$ ，则多重共线性的程度很小；若 $100 \leq k \leq 1000$ ，则存在中等强度的多重共线性；若 $k > 1000$ ，则存在严重的多重共线性。

多重共线性的另一个重要诊断指标是方差膨胀因子 (VIF) [20]，即：

$$VIF_j = \text{diag}(X^T X)_j^{-1} \quad (3)$$

式中， $(X^T X)^{-1}$ 表示方阵 $X^T X$ 的逆矩阵，其中， X 矩阵中的数据需事先进行标准化处理；diag表示取矩阵的对角线元素， j 表示第 j 个回归系数 ($j=1, 2, \dots, 9$)。模型中各个回归系数的VIF度量了回归变量之间的相关性对该项回归系数方差的影响，一个或更多个较大的VIF就表明存在多重共线性。一个经验法则 [20] 认为，当VIF超过5或10时，就表明回归模型存在严重的多重共线性。

1.3 多项式肥效模型的主成分回归分析

当多项式肥效模型存在严重多重共线性时，一个有效解决办法是重构回归变量，使之成为无线性相关的新变量。主成分分析为重构回归变量提供了可靠的数学方法，它将三元二次多项式肥效模型的9个回归变量转换成互不相关的9个主成分，然后选择合适的几个主成分进行OLS回归分析。

因此，从试验处理的原始施肥量数据出发，计算试验设计矩阵 X 的协方差矩阵 S ，然后求取 S 矩阵的特征值和特征向量。对式 (1) 而言，特征值共有9个，从大到小排列为 $\lambda_1, \lambda_2, \dots, \lambda_9$ ，与之对应的特征向量为 q_1, q_2, \dots, q_9 ，每个 q 均为长度为9的列向量。设向量 $x = (N, P, K, N^2, P^2, K^2, NP, NK, PK)^T$ ，则9个主成分的矩阵 Z 表达式为：

$$Z = (q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9)^T \quad (4)$$

描述某个回归变量 X_j 与某个主成分 Z_i 相关程度的样本相关系数计算式 [21-22] 为：

$$r(X_j, Z_i) = \frac{\sqrt{\lambda_i} q_{ij}}{\sqrt{S_{jj}}} \quad (5)$$

r 的平方即为决定系数。若 $r^2(X_j, Z_i)$ 数值大，

就表明第 i 个主成分主要反映了第 j 个回归变量的信息，其结果可用于解释各个主成分的专业含义。

因此，应用中可根据各个主成分对试验设计矩阵的方差贡献大小，选用适当的主成分个数结合14个试验处理产量结果进行OLS回归分析，并进行相应的统计检验，这种方法称为主成分回归。

1.4 多重共线性诊断和主成分回归的计算机实现

本研究采用MATLAB软件进行多重共线性诊断和主成分回归分析。在多重共线性诊断中，回归变量间的相关系数计算调用corrcoef功能函数；方阵 $X^T X$ 的条件数、逆矩阵和VIF计算分别调用cond、inv等功能函数； X 矩阵的协方差矩阵 S 则调用cov功能函数。在主成分回归中，主要调用了pca和regress等功能函数。上述计算的数学原理和有关功能函数使用方法参阅相关专著 [21-22]。

2 结果

2.1 多项式肥效模型的多重共线性诊断

根据表2的试验产量和表1的相关处理施肥量，利用OLS对式 (1) 进行回归建模，然后对各个回归模型进行统计检验和模型典型性判别 [7]，结果见表3。在7个代表性肥效模型中，1号试验点未达到统计显著水平，其余6个试验点的肥效模型均满足统计检验要求。但是，2号试验点的P项系数为负数，3号试验点的推荐施肥量外推，4号试验点无最高产量点，均属于非典型式，仅有第5、6、7号三个试验点的肥效模型满足了植物营养的一般肥效规律，属于典型肥效模型。

1号至4号试验点的OLS回归建模，出现了未通过显著性检验或者非典型式，是试验失误造成的还是回归方法不合理？为了明确这类问题，对表2的7个试验资料应用式 (2) 和式 (3) 进行多重共线性诊断。以2号试验点为例，与文献 [4] 的表2结果相同，9个回归变量的方差膨胀因子均远大于10，平均值达31.90；试验设计矩阵 X 的施肥量数据未经标准化处理时，矩阵条件数达7628，即使对各处理施肥量进行标准化转换，其条件数也达548.8，均显示肥效模型存在强烈的多重共线性。

进一步对9个回归变量进行简单相关性分析，表明N与 N^2 、P与 P^2 、K与 K^2 之间的相关系数高达0.942 9，N与NP或NK之间、P与NP或PK之间、K与NK或PK之间的相关系数达0.738 8， N^2 与NP或

NK之间、P²与NP或PK之间、K²与NK或PK之间的相关系数达0.732 2，存在高度线性相关。因试验设计方案和肥效模型设定相同，其他6个试验点的试验设计矩阵条件数、方差膨胀因子、自变量相关

系数等的相关结果与此一致，不再赘述。因此，三元二次多项式肥效模型存在严重的多重共线性，且这种共线性主要是由模型设定本身造成的。

表3 三元二次多项式肥效模型的OLS回归分析和典型性判别

Table 3 OLS regression analysis and typicality discrimination of TQPMs (ternary quadratic polynomial models for fertilizing effect)

序号 No.	式(1)模型参数 Parameters of the Model (1)										统计检验 Statistical test			典型性判别 Typicality discrimination		
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	F	R^2	p	PS	Max	RF
1	5 786	10.01	11.71	3.74	-0.074	-0.27	-0.059	0.14	0.067	-0.028	5.7	0.928	0.054	—	—	—
2	4 533	26.47	-3.94	8.44	-0.091	-0.12	-0.084	0.024	-0.050	0.12	298.6	0.999	0.000	N	Y	—
3	5 965	11.47	0.035	30.25	-0.073	-0.28	-0.19	0.25	0.032	-0.14	9.9	0.957	0.021	Y	Y	N
4	4 282	5.96	29.63	19.55	-0.045	-0.30	-0.034	0.19	0.026	-0.23	8.3	0.949	0.03	Y	N	—
5	4 842	10.28	18.91	25.06	-0.061	-0.20	-0.18	0.069	0.068	-0.081	13.6	0.968	0.012	Y	Y	Y
6	7 460	16.21	10.09	7.89	-0.115	-0.35	-0.072	0.21	0.058	-0.086	20.4	0.979	0.005	Y	Y	Y
7	4 780	16.41	12.26	12.43	-0.058	-0.084	-0.056	0.015	0.018	-0.030	27.4	0.984	0.003	Y	Y	Y

注：式(1)为三元二次多项式肥效模型。“Y”表示正常状态，“N”表示异常状态。“PS”表示模型参数符号，“Max”表示最高产量点，“RF”表示推荐施肥量。下同。Note: Model (1) is TQPM. “Y” stands for normal, “N” for abnormal. “PS” for parameter symbol, “Max” for site of Maximum yield, and “RF” for recommended fertilization rate. The same below

2.2 回归变量的主成分分析及其重构

上述诊断表明，三元二次多项式肥效模型存在严重的多重共线性。因此，从重构回归变量入手来消除这种共线性危害就成为提高建模成功率的关键。根据主成分分析原理和式(4)，从试验设计矩阵的原始数据出发，对肥效模型回归变量进行主

成分分析，将式(1)肥效模型的9个回归变量按照不同的线性组合重新划分为无线性相关的9个主成分，并与X矩阵特征值对应从大至小排列。同样以2号试验点为例，表4的分析表明，具有最大特征值的第1主成分对矩阵X的方差贡献率达73.82%，前7个主成分累加方差贡献率达99.9%以上。

表4 第2号试验点设计矩阵X的9个主成分及其特征值

Table 4 Nine principal components and their eigenvalues of the matrix X designed for Field Trial Site No.2

PC	PC: $Z = q_1N + q_2P + q_3K + q_4N^2 + q_5P^2 + q_6K^2 + q_7NP + q_8NK + q_9PK$									λ	VCR (%)
	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9		
第1主成分PC1	0.003 8	0.005 0	0.001 3	0.837 1	0.037 4	0.178 4	0.226 2	0.456 8	0.078 9	2.399×10^8	73.82
第2主成分PC2	-0.001 5	0.000 4	0.004 8	-0.412 9	0.035 3	0.736 9	-0.068 6	0.452 5	0.275 2	6.263×10^7	19.27
第3主成分PC3	-0.000 2	0.005 6	-0.000 5	-0.115 4	0.472 3	-0.052 8	0.699 6	-0.233 4	0.465 7	1.710×10^7	5.26
第4主成分PC4	0.001 8	-0.002 0	-0.002 3	-0.319 5	-0.025 8	-0.574 0	0.209 2	0.720 4	-0.071 4	3.924×10^6	1.21
第5主成分PC5	-0.009 9	0.001 7	0.007 5	0.115 7	0.255 1	-0.298 1	-0.605 7	0.066 2	0.679 1	1.021×10^6	0.31
第6主成分PC6	0.011 0	0.002 6	0.003 2	0.000 0	0.841 7	0.063 5	-0.209 7	0.093 6	-0.484 4	4.164×10^5	0.00
第7主成分PC7	0.824 5	0.263 4	0.500 6	-0.002 2	-0.009 2	-0.003 5	-0.001 5	-0.002 8	-0.007 0	3.297×10^2	0.00
第8主成分PC8	-0.548 3	0.154 8	0.821 7	0.000 5	-0.001 0	-0.001 4	0.007 4	0.000 2	-0.011 3	4.967×10^1	0.00
第9主成分PC9	-0.139 0	0.952 2	-0.272 1	0.000 1	-0.002 9	0.000 3	-0.002 8	0.002 8	-0.003 1	1.571×10^1	0.00

注：“PC”表示主成分， λ 表示特征值，VCR表示方差贡献率。下同。Note: “PC” stands for principal component. “ λ ” for eigenvalue, and “VCR” for variance contribution rate. The same below

为了明确9个主成分在反映氮磷钾施肥效应上的差异, 根据式(5)分别计算表4的9个主成分与肥效试验设计矩阵X的N、P、K、N²、P²、K²、NP、NK、PK等9个回归变量的决定系数r²(表5)。根据r²值大小结合式(1)肥效模型, 第1主成分主要反映了回归变量N效应, 部分反映了NP交

互效应; 第2主成分主要反映了回归变量K效应, 部分反映了回归变量N效应; 第3主成分主要反映了回归变量P效应, 部分反映了回归变量N效应; 等等。其他试验点的主成分特点亦有与此类似的结果, 不再赘述。

表5 第2号试验点各回归变量与9个主成分的决定系数(r²)

Table 5 Determination coefficient (r²) between various regression variables and 9 principal component for field trial site No.2

PC	式(1)模型回归变量Regression variables of Model(1)								
	N	P	K	N ²	P ²	K ²	NP	NK	PK
第1主成分 PC1	0.865 2	0.035 7	0.000 2	0.003 2	0.024 7	0.012 3	0.054 9	0.003 7	0.000 1
第2主成分 PC2	0.081 4	0.014 8	0.812 4	0.023 8	0.004 7	0.004 3	0.035 1	0.001 8	0.021 8
第3主成分 PC3	0.188 1	0.709 8	0.001 9	0.010 5	0.028 9	0.002 1	0.041 3	0.016 8	0.000 6
第4主成分 PC4	0.936 9	0.059 5	0.001 3	0.002 2	0.000 1	0.000 0	0.000 0	0.000 0	0.000 0
第5主成分 PC5	0.073 1	0.017 0	0.830 6	0.000 6	0.014 5	0.064 2	0.000 0	0.000 0	0.000 0
第6主成分 PC6	0.177 2	0.789 6	0.001 1	0.030 0	0.002 1	0.000 0	0.000 0	0.000 0	0.000 0
第7主成分 PC7	0.570 8	0.013 7	0.389 3	0.008 0	0.017 4	0.000 9	0.000 0	0.000 0	0.000 0
第8主成分 PC8	0.760 1	0.194 7	0.014 2	0.030 9	0.000 1	0.000 0	0.000 0	0.000 0	0.000 0
第9主成分 PC9	0.141 9	0.450 2	0.352 1	0.001 9	0.044 7	0.009 3	0.000 0	0.000 0	0.000 0

2.3 主成分回归的肥效模型及其推荐施肥量

根据表4各个主成分的N、P、K、N²、P²、K²、NP、NK、PK等回归变量的线性组合系数和试验设计矩阵X的每一行对应数值, 分别计算14个处理的各自得分值, 组成新的设计矩阵。由于前8个或前7个主成分分别反映了X矩阵方差信息的99.99%和99.90%以上, 因此, 新设计矩阵由前8个主成分或前7个主成分组成。

利用新设计矩阵与对应的14个处理产量, 进行OLS回归并转化成式(1)函数形式(表6)。由于“3414”试验的处理数为14个, 总自由度为13; 采用前8个主成分或前7个主成分进行主成分回归,

就相当于回归自由度分别为8和7; 因而, 建立的三元肥效模型的误差自由度分别为5和6。结果表明, 1号、2号和3号试验点所得三元肥效模型被转化为满足植物营养学一般肥效规律的典型式; 4号试验点采用8个主成分进行回归分析时, 其结果与OLS回归相同均属于无最高产量点的非典型式, 但采用7个主成分回归即可得到三元典型肥效模型。5号和6号试验点的OLS回归分析得到典型式, 采用前8个或前7个主成分进行回归分析仍然得到典型式。但是, 7号试验点采用前8个或前7个主成分回归, 肥效模型均由OLS的典型式变成了非典型式, 结果反而变劣。

表6 三元二次多项式肥效模型的主成分回归和典型性判别

Table 6 Principal component regression and typicality discrimination of TQPMs

主成分数 Numbers of PC	序号 No.	式(1)模型参数Parameters of Model(1)										统计检验 Statistical test			典型性判别 Typicality discrimination		
		b ₀	b ₁	b ₂	b ₃	b ₄	b ₅	b ₆	b ₇	b ₈	b ₉	F	R ²	P	PS	Max	RF
		8	1	5 807	11.30	3.23	5.91	-0.075	-0.21	-0.061	0.16	0.048	-0.006	7.7	0.925	0.019	Y
	2	4 515	25.12	5.32	5.79	-0.090	-0.15	-0.081	-0.004	0.22	0.091	134.4	0.995	0.000	Y	Y	Y
	3	5 946	9.99	10.04	27.09	-0.072	-0.31	-0.187	0.22	0.064	-0.17	13.2	0.955	0.006	Y	Y	Y
	4	4 335	9.28	7.82	25.11	-0.047	-0.25	-0.039	0.24	-0.024	-0.168	9.3	0.937	0.012	Y	N	—
	5	4 855	11.17	11.90	27.72	-0.062	-0.18	-0.183	0.09	0.046	-0.061	17.9	0.966	0.003	Y	Y	Y
	6	7 469	16.93	5.13	9.31	-0.12	-0.33	-0.074	0.23	0.043	-0.070	27.6	0.978	0.001	Y	Y	Y
	7	4 794	17.29	4.71	1390	-0.058	-0.056	-0.058	0.034	0.003	-0.007	35.3	0.983	0.000	Y	Y	N
7	1	5 802	10.52	3.39	7.01	-0.074	-0.25	-0.063	0.17	0.048	-0.018	10.5	0.925	0.005	Y	Y	Y
	2	4 489	20.63	6.59	12.52	-0.086	-0.16	-0.093	0.057	0.024	-0.002	7.3	0.988	0.000	Y	Y	Y
	3	6 006	20.09	6.77	11.98	-0.081	-0.29	-0.160	0.088	0.058	0.031	12.4	0.935	0.003	Y	Y	Y
	4	4 387	18.44	5.95	12.29	-0.054	-0.23	-0.021	0.14	-0.026	-0.023	9.6	0.918	0.007	Y	Y	Y
	5	4 943	21.31	6.68	10.57	-0.069	-0.16	-0.149	-0.027	0.036	0.165	10.8	0.927	0.005	Y	Y	Y
	6	7 466	16.47	5.26	10.00	-0.12	-0.33	-0.075	0.23	0.044	-0.079	37.7	0.978	0.000	Y	Y	Y
	7	4 804	18.87	4.41	11.40	-0.06	-0.052	-0.054	0.006	0.003	0.033	45.5	0.982	0.000	Y	Y	N

根据主成分回归建立的典型肥效模型以及每千克 N4.30元、P₂O₅ 5元、K₂O6.0元和稻谷2.0元的市场均价，用边际产量导数法计算推荐施肥量（表7）。由于选用的前7个或前8个主成分综合反映了试验设计矩阵X的99.9%以上的变异方差，1号至4号试验点主成分回归模型的最高施肥量和经济施肥

量及其预测产量，其数值大小均在试验设计施肥量和试验所得产量范围内，且与已有的研究结果^[17]基本一致。尤其是5号和6号试验点，主成分回归的模型推荐施肥量与OLS回归的模型推荐施肥量差异极小甚至相同，预测早稻产量也基本一致，表明主成分回归的推荐施肥是可靠的。

表7 三元二次多项式肥效模型不同建模方法的推荐施肥量比较

Table 7 Comparison in recommended application rate between TQPMs different in modeling method

序号No.	推荐施肥 Recommend fertilization	普通最小二乘法回归建模 OLS regression modeling				8个主成分回归建模 8 PCs regression modeling				7个主成分回归建模 7 PCs regression modeling			
		N	P ₂ O ₅	K ₂ O	产量 Yield	N	P ₂ O ₅	K ₂ O	产量 Yield	N	P ₂ O ₅	K ₂ O	产量 Yield
		1	MAR	—	—	—	—	183	65	117	7 292	184	66
	EAR	—	—	—	—	132	44	73	7 144	132	44	76	7 149
2	MAR	—	—	—	—	149	41	79	6 721	148	48	87	6 717
	EAR	—	—	—	—	133	23	49	6 636	129	37	68	6 655
3	MAR	—	—	—	—	201	65	77	8 320	173	42	73	8 319
	EAR	—	—	—	—	157	48	69	8 239	150	33	58	8 262
4	MAR	—	—	—	—	—	—	—	—	261	86	83	7 564
	EAR	—	—	—	—	—	—	—	—	246	79	27	7 453
5	MAR	166	58	88	7 346	166	59	87	7 341	165	51	83	7 313
	EAR	137	49	76	7 286	136	47	77	7 278	147	33	61	7 237
6	MAR	133	45	81	9 084	132	45	81	9 082	132	45	81	9 082
	EAR	110	38	56	9 012	110	36	58	9 012	110	36	60	9 013
7	MAR	168	66	119	7 301	—	—	—	—	—	—	—	—
	EAR	144	54	92	7 219	—	—	—	—	—	—	—	—

注：“MAR”表示最高施肥量，“EAR”表示经济施肥量Note: “MAR” stands for maximum application rate and “EAR” for economical application rate

2.4 主成分回归对三元肥效模型的建模效果

表6的代表性例子表明，主成分回归能使某些肥效试验结果由OLS法的未通过显著性检验或非典型式转化为典型肥效模型，但也可能使OLS法的典型式变成非典型式。为了评价主成分回归对三元肥效模型的建模效果，利用近年来福建省早稻测土配方施肥工作在主要项目县完成的“3414”设计的氮磷钾田间肥效试验资料，逐个进行回归分析和比较，统计结果见表8。

在171个肥效试验中，OLS回归建模的三元典

型肥效模型的比例占试验点总数的27.5%，而采用前8个主成分或前7个主成分回归建模，典型肥效模型的比例则分别提高至41.5%和43.3%，为OLS法的1.5倍和1.6倍。其中，主成分回归大幅度降低了模型系数符号不合理的非典型式，从OLS法的14.6%下降至几乎为零；模型无最高产量点的非典型式也从OLS法的36.8%分别下降至32.2%和21.1%。但是，主成分回归并非利用设计矩阵X的100%方差信息，其建模结果使推荐施肥量外推的非典型式个数反而有所增加。

表8 最小二乘法回归 (OLS) 和主成分回归 (PCR) 对三元二次多项式肥效模型建模结果的影响

Table 8 Effects of modeling using OLS and PCR on TQPMs

建模方法 Modeling method	试验数 Number of trial	NRSS (%)	通过显著性检验的模型比例 Ratio of the models up to statistical significance in Duncan test (%)			
			非典型模型 Non-typical models			典型模型 Typical models
			Max	PS	RF	
OLS	171	15.8 (27)	36.8 (63)	14.6 (25)	5.3 (9)	27.5 (47)
8个主成分回归 PCR using 8 PCs	171	9.4 (16)	32.2 (55)	0.6 (1)	16.4 (28)	41.5 (71)
7个主成分回归 PCR using 7 PCs	171	17.5 (30)	21.1 (36)	0.0 (0)	18.1 (31)	43.3 (74)
OLS结合8个主成分回归 OLS combining with PCR using 8 PCs	171	7.6 (13)	29.8 (51)	0.6 (1)	13.5 (23)	48.5 (83)
OLS结合7个主成分回归 OLS combining with PCR using 7 PCs	171	12.9 (22)	18.7 (32)	0.0 (0)	12.9 (22)	55.6 (95)

注：“NRSS”表示未达统计显著水平的肥效模型；括号内的数据为试验个数 Note: “NRSS” stands for models statistically insignificant. The digits in the bracket are numbers of trials

建模过程还表明，如同7号试验点，主成分回归有时也会失效，使OLS法的典型式变为非典型式，甚至造成模型未能通过显著性检验。因此，从实用角度出发，如果主成分回归仅针对OLS法的非典型式和统计不显著的肥效模型，即：OLS结合前8个主成分的主成分回归或者前7个主成分的主成分回归的建模策略，结果使典型肥效模型的比例分别可提高至48.5%和55.6%，明显优于单独使用OLS或者主成分回归。

3 讨论

3.1 非典型肥效模型的产生原因及其多重共线性危害

经典线性回归分析理论假设回归模型的各个自变量之间无多重共线性^[19-20]。二次多项式肥效模型在数学形式上是一种非线性模型，人们通常不重视这种模型是否违背了无多重共线性的假设。故这种模型中是否存在共线性、共线性将产生什么危害、如何处理共线性等方面问题，均鲜见相关研究报告^[4]。本研究通过方阵 $X^T X$ 条件数和方差膨胀因子^[20]等常用指标进行多重共线性诊断，表明这种模型存在严重的多重共线性问题。

王兴仁^[5-6]和章明清等^[7]曾报道，二元二次多项式肥效模型的典型式比例仅占40.2%，三元肥效模型的典型式则低至23.6%。人们对多元肥效模型出现大量非典型式的现象至今仍然困惑不解，许

多人由此直接归因于田间试验失败或者对肥效模型的应用价值失去信心。研究表明，多重共线性给多项式肥效模型的OLS回归建模和推荐施肥准确性带来严重危害，具体表现在：①在完全共线性下方阵 $X^T X$ 不可逆，正则方程无解，无法建立肥效模型。②在近似共线性下，虽然可以得到OLS参数估计量，但此时 $1/X^T X \approx 0$ ，引起方阵 $X^T X$ 主对角线上的元素显著增大，从而使参数估计值的方差膨胀，难以得到精确的参数估计值，例如3号和4号试验点出现了非典型肥效模型。③如果模型中两个回归变量具有高度线性相关性，那么它们中的一个变量可以由另一个变量近似表征，这时回归变量前的参数并不反映各自与作物产量之间的结构关系，仅能反映它们对产量的共同影响，结果使参数失去了应有的肥效含义。同时，回归系数可能变得对产量数据的微小变化异常敏感，甚至越过零值，表现出系数值的正负号出现反常现象，导致参数估计值的专业含义不合理，例如2号试验点的P项本应为正数，结果却为负数。④由于方差膨胀，模型显著性检验失效，例如1号试验点出现未通过显著性检验的假象。⑤膨胀的方差使产量预测区间过度变大，模型的预测功能失效。

因此，现在基本可以明确，非典型肥效模型大多是由多重共线性引起的，肥效模型设定本身是产生多重共线性的主要原因。

3.2 三元二次多项式肥效模型的主成分回归

多项式统计模型在农学学科中出现大量非典型

式的现象, 在其他学科诸如经济学、社会学、心理学等学科也同样大量存在。主成分回归经过近几十年的发展, 已成为这些学科对付多重共线性的有效技术^[24]。主成分回归的基本思路是利用主成分分析提取 X 矩阵的主成分, 然后选取方差贡献率达到符合要求的几个主成分, 计算各处理对应回归变量的得分值, 组成新的设计矩阵, 最后利用新设计矩阵与各处理的试验产量进行OLS回归建模。本研究表明, 这种建模方法能使1号至4号试验点由OLS法未达统计显著水平或非典型式转化为典型式。

在近年来的测土配方施肥工作中, 福建省在早稻上完成了众多“3414”设计的田间肥效试验。本研究收集整理了来自土肥技术力量较强的主要项目县完成的171个试验资料。针对这些试验结果, 选用前8个主成分或前7个主成分进行回归建模, 三元典型式出现比例分别提高至41.5%和43.3%。尤其是前7个主成分回归建模使模型系数符号不合理的非典型式出现比例下降至零, 同时大幅度降低了无最高产量点的非典型式比例(表8), 表明了主成分回归技术在提高三元肥效模型建模成功率方面具有重要的应用价值。

对比研究表明, 因各施肥处理的氮磷钾施肥量所采用的量纲相同, 数据差异不显著, 试验设计矩阵的原始数据是否进行了标准化处理, 对主成分回归结果基本无影响; 若将全部9个主成分均纳入主成分回归, 其结果与OLS法完全一致, 达不到改善建模效果的作用; 虽然前5个主成分就能反映 X 矩阵方差变异的99.87%, 但采用6个或者少于6个主成分进行回归分析, 所建立的171个早稻三元二次多项式肥效模型均未能通过显著性检验。因此, 对三元肥效模型而言, 主成分回归只能选取前8个主成分或者前7个主成分。

在消除或缓解多重共线性危害方面, 数理统计学家已经提出了诸多解决方法^[19-20, 24], 诸如剔除不重要自变量、扩大样品容量, 或者采用岭回归、主成分分析和偏最小二乘法等有偏估计方法来提高参数估计的稳健性。显然, 剔除不重要自变量和扩大样品容量的方法不适合肥效模型的构建。在有偏估计方法中, 岭回归估计的质量取决于偏倚系数的选取, 但目前尚无标准的决策准则^[24], 仅能凭经验判断。偏最小二乘回归是通过迭代算法在 X 矩阵中逐步提取主成分, 使其尽可能多地反映 X 矩

阵的变异方差, 同时又兼顾对因变量(产量)的解释能力。然而, 其迭代算法复杂, 不易使用。相对而言, 主成分回归具有扎实的统计学基础, 建模思路简洁明了。该法在目前已有MATLAB、SAS和Stata等诸多成熟软件可供使用, 是较简便和实用的肥效模型参数估计方法, 易于被广大应用人员理解和掌握。

3.3 三元二次多项式肥效模型的建模策略

尽管主成分回归具有许多优点, 但它毕竟是一种有偏估计。受回归偏倚的影响, 有时会将诸如7号试验点的OLS回归典型式变成了非典型式。该法在提取主成分时仅考虑了试验设计矩阵, 而不考虑各个主成分与各处理产量的相关性, 而且, 对什么样的试验数据会造成主成分回归失效? 目前在统计学上并无明确的答案^[24-25]。鉴于此, 从实用角度出发, 主成分回归仅针对OLS法未通过显著性检验肥效模型和非典型式肥效模型, 而对OLS回归得到典型式的肥效模型就不再做主成分回归。这个策略使三元典型式的比例进一步提高至55.6%(表8), 为单独使用OLS法的2.0倍。

4 结 论

对三元肥效模型而言, 主成分回归是解决多项式肥效模型多重共线性危害的一种有效方法, OLS建模法结合前7个主成分的主成分回归是三元二次多项式肥效模型的最佳建模策略。

参 考 文 献

- [1] Valkama E, Uusitalo R, Turtola E. Yield response models to phosphorus application: A research synthesis of Finnish field trials to optimize fertilizer P use of cereals. *Nutrient Cycling in Agroecosystems*, 2011, 91(1): 1—15
- [2] Gregoret M C, Zorita M D, Dardanelli J, et al. Regional model for nitrogen fertilization of site-specific rainfed corn in haplustolls of the central Pampas, Argentina. *Precision Agriculture*, 2011, 12(6): 831—849
- [3] 王兴仁, 陈新平, 张福锁, 等. 施肥模型在我国推荐施肥中的应用. *植物营养与肥料学报*, 1998, 4(1): 67—74
Wang X R, Chen X P, Zhang F S, et al. Application of fertilization model for fertilizer recommendation in

- China (In Chinese). *Plant Nutrition and Fertilizer Science*, 1998, 4 (1): 67—74
- [4] 章明清, 李娟, 孔庆波, 等. 作物肥料效应函数模型研究进展与展望. *土壤学报*, 2016, 53 (6): 1143—1156
Zhang M Q, Li J, Kong Q B, et al. Progress and prospect of the study on crop—response—to—fertilization function model (In Chinese). *Acta Pedologica Sinica*, 2016, 53 (6): 1143—1156
- [5] 王兴仁. 二元二次肥料效应曲面等产线图在科学施肥中的位置(一). *土壤通报*, 1985, 16 (1): 30—34
Wan X R. The positions of contour map of yield of binary quadratic fertilizer response curve in scientific fertilization (part 1) (In Chinese). *Chinese Journal of Soil Science*, 1985, 16 (1): 30—34
- [6] 王兴仁. 二元二次肥料效应曲面等产线图在科学施肥中的位置(二). *土壤通报*, 1985, 16 (2): 86—88
Wan X R. The positions of contour map of yield of binary quadratic fertilizer response curve in scientific fertilization (part 2) (In Chinese). *Chinese Journal of Soil Science*, 1985, 16 (2): 86—88
- [7] 章明清, 林仁坝, 林代炎, 等. 极值判别分析在三元肥效模型推荐施肥中的作用. *福建农业学报*, 1995, 10 (2): 54—59
Zhang M Q, Lin R X, Lin D Y, et al. Function of distinguish analysis on extreme value in recommendatory fertilization for three—fertilizer efficiency model (In Chinese). *Fujian Journal of Agricultural Sciences*, 1995, 10 (2): 54—59
- [8] Cerrato M E, Blackmer A M. Comparison of models for describing corn yield response to nitrogen fertilizer. *Agronomy Journal*, 1990, 82 (1): 138—143
- [9] 杨靖一, Wadsworth G A, Greenwood D J. 三种蔬菜N肥效应曲线的比较研究. *植物营养与肥料学报*, 1995, 1 (1): 71—78
Yang J Y, Wadsworth G A, Greenwood D J. Comparative study on three kinds of vegetables response to N fertilizer curve (In Chinese). *Plant Nutrition and Fertilizer Science*, 1995, 1 (1): 71—78
- [10] 陈新平, 周金池, 王兴仁, 等. 小麦—玉米轮作制中氮肥效应模型的选择——经济和环境效益分析. *土壤学报*, 2000, 37 (3): 346—354
Chen X P, Zhou J C, Wang X R, et al. Economic and environmental evaluation on models for describing crop yield response to nitrogen fertilizers at winter—wheat and summer—corn rotation system (In Chinese). *Acta Pedologica Sinica*, 2000, 37 (3): 346—354
- [11] 吴平, 陶勤南. 氮磷钾肥效模型的研究及其应用. *浙江农业大学学报*, 1989, 15 (4): 383—388
Wu P, Tao Q N. Study on N, P and K fertilizer response model and its application (In Chinese). *Journal of Zhejiang Agricultural University*, 1989, 15 (4): 383—388
- [12] 章明清. 米氏肥效方程的灰色建模法. *土壤通报*, 1997, 28 (1): 18—19
Zhang M Q. Gray modeling method of Mitscherlich fertilizer response equations (In Chinese). *Chinese Journal of Soil Science*, 1997, 28 (1): 18—19
- [13] 章明清, 徐志平, 姚宝全, 等. Monte Carlo 法在多元肥效模型参数估计和推荐施肥中的应用. *植物营养与肥料学报*, 2009, 15 (2): 366—373
Zhang M Q, Xu Z P, Yao B Q, et al. Using Monte Carlo method for parameter estimation and fertilization recommendation of multivariate fertilizer response model (In Chinese). *Plant Nutrition and Fertilizing Science*, 2009, 15 (2): 366—373
- [14] Colwell J D. The derivation of fertilizer recommendations for crops in non—uniform environment: Fertilizer. *Crop Quality and Economy*, 1974: 936—961
- [15] 王兴仁, 陈伦寿, 毛达如, 等. 分类回归综合法及其在区域施肥决策中的应用. *土壤通报*, 1989, 20 (1): 17—21
Wang X R, Chen L S, Mao D R, et al. Classification of regression synthesis and its application for regional fertilization decision—making (In Chinese). *Chinese Journal of Soil Science*, 1989, 20 (1): 17—21
- [16] 毛达如, 张承东. 多点肥料效应函数的动态聚类方法. *北京农业大学学报*, 1991, 17 (2): 49—54
Mao D R, Zhang C D. Cluster analysis of quadratic response function on the fertilizer dispersed experiment (In Chinese). *Acta Agriculturae Universitatis Pekinensis*, 1991, 17 (2): 49—54
- [17] 李娟, 章明清, 孔庆波, 等. 福建早稻测土配方施肥指标体系研究. *植物营养与肥料学报*, 2010, 16 (4): 938—946
Li J, Zhang M Q, Kong Q B, et al. Soil testing and formula fertilization index for early rice in Fujian Province (In Chinese). *Plant Nutrition and Fertilizer Science*, 2010, 16 (4): 938—946
- [18] 鲁如坤. *土壤农业化学分析方法*. 北京: 中国农业科学技术出版社, 2000: 146—196
Lu R K. *Analytical methods for soil and agro—chemistry (In Chinese)*. Beijing: China Agricultural Science and Technology Press, 2000: 146—196

- [19] 何晓群, 刘文卿. 应用回归分析. 第3版. 北京: 中国人民大学出版社, 2013: 158—170
He X Q, Liu W Q. Applied regression analysis (In Chinese). 3rd ed. Beijing: China Renmin University Press, 2013: 158—170
- [20] Montgomery D C, Peck E A, Vining G G. 王辰勇, 译. 线性回归分析导论. 第5版. 北京: 机械工业出版社, 2016: 91—268
Montgomery D C, Peck E A, Vining G G. Wang C Y. trans. Introduction to linear regression analysis (In Chinese). 5th ed. Beijing: China Machine Press, 2016: 91—268
- [21] 袁志发, 宋世德. 多元统计分析. 第2版. 北京: 科学出版社, 2009: 219—250
Yuan Z F, Song S D. Multivariate statistical analysis (In Chinese). 2nd ed. Beijing: Science Press, 2009: 219—250
- [22] 薛毅, 陈立萍. 实用数据分析与MATLAB软件. 北京: 北京工业大学出版社, 2015: 189—274, 377—394
Xue Y, Chen L P. Practical data analysis and MATLAB software (In Chinese). Beijing: Beijing University of Technology Press, 2015: 189—274, 377—394
- [23] Gujarati D N, Porter D C, 费剑平, 译. 计量经济学基础 (上册). 第5版. 北京: 中国人民大学出版社, 2011: 313—399
Gujarati D N, Porter D C. Fei J P. trans. Econometrics foundation (the first volume) (In Chinese). 5th ed. Beijing: China Renmin University Press, 2011: 313—399
- [24] 王惠文, 王劫, 黄海军. 主成分回归的建模策略研究. 北京航空航天大学学报, 2008, 34 (6): 661—664
Wang H W, Wang J, Huang H J. Modeling strategy of principle component regression (In Chinese). Journal of Beijing University of Aeronautics and Astronautics, 2008, 34 (6): 661—664
- [25] 叶宗裕. 主成分回归能消除多重共线性吗? 统计与信息论坛, 2012, 27 (3): 15—20
Ye Z Y. Can the principal components regression eliminate the multicollinearity (In Chinese)? Statistics & Information Forum, 2012, 27 (3): 15—20

Principal Component Regression Technology of Ternary Fertilizer Response Model for Improving Success Rate of Modeling

LI Juan¹ ZHANG Mingqing^{1†} XU Wenjiang² KONG Qingbo¹ YAO Baoquan³

(1 Soil and Fertilizer Institute, Fujian Academy of Agricultural Sciences, Fuzhou 350013, China)

(2 Fujian Institute of Subtropical Plants, Xiamen, Fujian 361000, China)

(3 Fujian Cropland Construction and Soil and Fertilizer Station, Fuzhou 350003, China)

Abstract 【Objective】 There is a large number of non-typical ones of ternary quadratic polynomial models for fertilizing effect (TQPM) used in fertilization practice, severely affecting accuracy and practicality of the fertilizing effect models. In order to improve success rate of modeling for field fertilization experiments, this study was oriented to explore new modeling methods. 【Method】 Based on 171 field fertilization experiments of “3414” in design on response of early rice to N, P and K fertilization in Fujian Province, efforts were made to explore methods to diagnose multicollinearity of TQPMs and technology for modeling with principal component regression and effects of their usage. 【Result】 Diagnoses of number of conditions of the square matrix ($X^T X$) of the matrix X designed for the “3414” field experiments and variance inflation factors (VIF) of the model parameters shows that TQPMs do have serious multicollinearity, which is mainly caused by the specification of the models per se, and severely restrains application value of the ordinary least squares (OLS). Among the 171 field experiments of early rice response to N, P and K fertilization in Fujian province, ternary typical fertilizing effect models using the OLS regression modeling method accounted for only 27.5% of the total. The proportions of the non-typical typed models, such as no maximum-yield point, unreasonable signs of monomial coefficient or

quadratic coefficient, and extrapolation of recommended fertilization rate, reached 36.8%, 14.6% and 5.3%, respectively, of the total using the OLS modeling method. Principal component regression (PCR) extracted 9 unrelated principal components from the designed matrix, thus eliminating the negative effect of multicollinearity. When the first 7 principal components were used in regression modeling, the proportion of ternary typical fertilizing effect models increased up to 43.3% or 1.6 times as high as that using the OLS method. What is more, compared with the OLS modeling method, PCR decreased the proportions of non-typical models with unreasonable coefficient symbols by a large margin from 14.6% to zero, and the proportion of non-typical models with no maximum-yield point from 36.8% to 21.1%, which fully manifested the positive effect of eliminating the hazard of multicollinearity on success rate of modeling. Seven-principal-component regression modeling made use of more than 99.9% of the variance information of the designed experimental matrix X , and all the typical models established by PCR modeling did not differ much from those by OLS modeling in recommending fertilization rate and predicting yields. As principal component regression is a kind of biased estimate, only the experimental designed matrix was taken into account in extracting principal components, but not relationship between each principal component and yield of each corresponding treatment. As affected by regression bias, typical models by OLS regression may occasionally turn into non-typical ones. Therefore, from the aspect of practical application, the use of OLS modeling in combination with modeling based on regression of the first seven principal component, increased the proportion of ternary typical fertilizing effect models up to 55.6%, or 2.0 times as high as that using OLS modeling alone, and was much better than using the PCR alone.

【 Conclusion 】 Principal components regression can eliminate the defect of multicollinearity of the ternary quadratic polynomial fertilizing effect model. The use of the ordinary least squares method in combination with the PCR modeling based on the first seven principal components is the optimal modeling strategy that can significantly increase the proportion of ternary typical fertilizing effect models to predict early rice response to N, P and K fertilization.

Key words Early rice; N, P and K, Fertilizing effect model; Multicollinearity; Principal component regression; Ordinary least squares (OLS)

(责任编辑: 陈荣府)