

# Logistic 回归模型在红壤地区早稻 推荐施肥中的应用\*

施建平 鲁如坤 时正元 孙 波

(中国科学院南京土壤研究所, 南京 210008)

**摘 要** 本文介绍了 Logistic 回归模型原理及其在红壤地区早稻推荐施肥中的应用, 并给出了上述模型与现存的二次模型在红壤地区水稻土上的比较结果。

**关键词** 红壤地区, 推荐施肥, logistic 回归

**中图分类号** S157.4

作物产量限制因素是土壤理化性质和生物有效性等综合因素所决定的。自 80 年代以来, 我国在推荐施肥技术方面取得了长足的进步, 目前我国的施肥推荐中, 主要内容是科学地确定肥料用量。当前应用的各种方法仍有其局限性, 有待不断完善<sup>[1,2]</sup>。1976 年, 联合国粮农组织率先应用建立产量—肥料效应方程法方法。80 年代以来, 我国引入基于田间试验统计的产量—肥料效应函数方法, 取得较好的效果。产量—肥料效应函数法一般以单因素或二因素多水平试验设计为基础, 将不同处理得到的产量进行回归统计, 求得产量与施肥量之间的函数关系, 建立肥料效应方程<sup>[3]</sup>。但该法只考虑地力产量, 在模型中没有代表不同土壤肥力的土壤测试值, 需要在不同类型土壤上布置多点试验, 在推广应用上均受到多种条件的限制。

由于土壤供氮量的预测目前只能达到半定量的水平<sup>[4]</sup>, 氮肥施用量的推荐仍存在着不确定性, 土壤综合肥力、施肥量与作物产量的定量关系至今还难以用一个通用的数学方程式表达。由于氮肥用量和土壤测定值之间尚未找到适宜的关系, 广泛使用于推荐施肥的肥力指标方法将土壤肥力和氮肥用量分级, 将连续的变量分解为三值响应变量。例如, 肥力的高、中、低对应于不同的氮肥施用量范围。即“高”不需施肥, “中”适量施肥, “低”大量施肥<sup>[5]</sup>。现存的肥力指标法只考虑土壤中各单因素影响, 没有考虑土壤各种养分综合影响, 而客观地评价土壤肥力水平是定量推荐施肥的关键之一。耿兴元等提出模糊定量综合评判的方法, 将地力产量、有机质、全氮、碱解氮、速效磷、速效钾等 6 个因素进行土壤综合肥力等级综合判断<sup>[6]</sup>。李仁岗借助正交多项式趋势系数与地力产量和土壤有效磷之间的回归关系, 建立综合肥料效应函数并确定最佳施肥量<sup>[7]</sup>。上述模型由于试验和计算

\* 中国科学院知识创新工程项目 (KZCX2-413) 和中国科学院南京土壤研究所知识创新工程领域前沿项目 (ISSASIP0110) 资助

收稿日期: 2001-04-02; 收到修改稿日期: 2002-02-25

复杂, 而不便于推广使用。

随着计算机技术的普及, 人们越来越多地从不同的角度去考虑问题。由于自然环境、土壤肥力及品种的不同, 其高产或低产的产量指标是相对的。事实上, 由于土壤肥力与作物产量的空间差异, 施肥模型估算的肥料用量往往具有很大的不确定性, 即使得到精确的统计模型仍然只限局部使用。然而在施肥决策中, 人们需要了解该地区某一土壤肥力下氮肥用量为多少时相对高产的概率。此时模型的因变量是定性的, 或者高产稳产, 或者低产。但在这种情况下, 多元线性回归使用定性因变量违反了其本身的假设条件。而 Logistic 回归可用于研究事件发生的概率与一组自变量间的关系, 除了模型的因变量是定性的以外, 其他与多元回归类似。Logistic 回归模型已被成功地用于与遥感、GIS 相结合的生态学研究。例如, 相关生物物理参数与社会经济变量来预测全球土地覆盖模式<sup>(1)</sup>, 野生动物保护区生境特性持续性分析<sup>(2)</sup>。本文介绍了 Logistic 回归施肥模型的原理、在红壤地区水稻土早稻试验的应用、以及与肥效模型比较结果。

## 1 模型的建立

当因变量只能取值 0 和 1, 并作为发生与不发生两种情况来处理时, 数学上通常使用 Logistic 回归来研究离散响应与一组自变量间的关系。考虑施肥模型的因变量为二值响应变量时, 某一水平的产量可能发生为一类, 不发生为另一类。在二值响应模型中, 响应  $Y$  代表一个个体或实验单元, 它的取值有两种可能: 0 和 1。如高产  $Y=1$ , 低产  $Y=0$ 。它也可以表示为某个特定值的百分率或概率。假定  $X$  是一个自变量向量, 且  $p = Pr(Y=1|x)$  是建模的响应概率, 则线性 Logistic 模型有如下的形式<sup>[9]</sup>:

$$\text{Logit}(p) = \log(p/(1-p)) = a + bx \text{ 或 } p = \exp(a + bx) / (1 + \exp(a + bx))$$

其中  $a$  是截距参数,  $b$  是斜率参数。

对于 Logistic 回归模型, 观测到的不是  $p$  本身, 而是在自变量的各个不同取值时所关心的属性变量出现的某一结果发生的频率(百分率)。拟合 Logistic 回归模型不采用通常的最小二乘法, 而采用最大似然法。Logistic 模型用非线性概率模型代替线性概率模型, 满足肥料-产量收益递减规律。

其次, 考虑到大面积推荐施肥中获得基础肥力(地力产量)的困难, 而现存大量的土壤普查养分数据可以作为宏观指导施肥的依据。那么, 能否应用土壤肥力指标分级的思想, 将土壤养分分级参数引入到施肥模型之中?

为了在施肥统计模型加入分级的土壤养分参数, 引入分类变量(Categorical variable)。分类变量数值可以只是一个符号, 数值之间不存在实际意义的量的关系, 如土壤图中土壤类型的编码; 也可以不仅具有分类作用, 而且存在量的关系, 如肥力的“高”和“低”。当自

(1) Iftikhar Uddin Sikder, Land Cover Modelling: A Spatial Statistical Approach, <http://pages.hotbot.com/edu/geoinformatics/fis.html>

(2) Noemi De La Ville, et. Al. Habitat suitability analysis using logistic regression and GIS to outline potential areas for conservation of the Grey Wolf (*Canis lupus*). [http://www.cranfield.ac.uk/public/ec/ec918/htm\\_dir/paper.htm](http://www.cranfield.ac.uk/public/ec/ec918/htm_dir/paper.htm)

变量为二值的分类变量时, Logistic 回归中的分析方法与一般自变量相同<sup>[8]</sup>。选用土壤养分分级参数和氮肥用量作为自变量, 产量出现的概率作为因变量, 进行多元 Logistic 回归, 可得到作物产量概率出现的 Logistic 模型:

$$\text{logit}(P) = a + bX_1 + cX_2 + dX_3$$

式中  $X_1, X_2, X_3$  为各种土壤养分和氮肥用量的分类变量或连续变量,  $P$  为在不同土壤养分条件和施肥条件下作物获得某种产量的概率。当产量作为推荐施肥的主要目标时, 高产出现的概率为因变量; 当追求环境保护和相对稳产时, 中等产量出现的概率为因变量。

## 2 田间试验验证

### 2.1 材料与方法

选择余江县洪湖乡母质为红砂岩的红砂泥田和母质为第四纪红色粘土的黄泥田做早稻氮肥用量的田间试验, 以确定该地区水稻田的基础地力, 探索土壤养分测定值、氮肥用量与产量的关系, 并对施肥模型进行比较评估。

选择不同土壤肥力水稻田试验地六块, 共设 5 个处理, 选用当地早稻品种, 在施磷 ( $P_2O_5$ )  $60 \text{ kg hm}^{-2}$ , 钾 ( $K_2O$ )  $150 \text{ kg hm}^{-2}$  的基础上, 施用 0、45、90、135、180  $\text{kg hm}^{-2}$  氮肥(尿素, 折纯)。小区面积红砂泥田  $24 \text{ m}^2$  黄泥田  $33 \text{ m}^2$ , 重复三次, 随机区组排列。共 90 个小区结果。供试土壤理化性质见表 1, 试验结果见表 2。

表 1 供试土壤理化性质

Table 1 Some properties of the experimental soils

试验地编号 Site No.	供试土壤 Soil type	有机质 O. M. ( $\text{g kg}^{-1}$ )	全氮 Total N ( $\text{g kg}^{-1}$ )	有效磷 Available P ( $\text{Olsen P mg kg}^{-1}$ )	速效钾 Rapidly available K ( $\text{K mg kg}^{-1}$ )
98-05	红砂泥田	28.0	1.63	20.9	53
98-08	红砂泥田	29.4	1.53	8.9	49
98-09	红砂泥田	20.4	1.10	7.7	58
98-12	黄泥田	31.2	1.62	7.0	43
98-13	黄泥田	31.8	1.73	8.7	110
98-16	黄泥田	30.7	1.43	9.8	78

表 2 不同氮肥用量下的早稻产量 ( $\text{kg hm}^{-2}$ )

Table 2 Rice yield with different N rates

试验地编号 Site No.	氮肥用量 ( $\text{kg hm}^{-2}$ )				
	N rate				
	0	45	90	135	180
98-05	3015	3750	4125	4575	4155
98-05	2910	3750	2910	4575	4365
98-05	3015	3960	4170	4575	4170
98-08	4125	4320	4635	4635	4740

续表

试验地编号 Site No.	氮肥用量( kg hm <sup>-2</sup> )				
	N rate				
	0	45	90	135	180
98- 08	4125	4635	4320	4530	4850
98- 08	4320	4320	4635	5055	4440
98- 09	3630	4500	4260	4770	4860
98- 09	3420	4185	4560	4545	5070
98- 09	3420	3045	4710	4575	4455
98- 12	3300	4500	5040	5040	5160
98- 12	3900	4740	5040	5490	5430
98- 12	3510	4530	4500	4770	4710
98- 13	3360	4410	5070	5460	5400
98- 13	3210	4200	4950	5310	5250
98- 13	2610	3600	4170	5070	5250
98- 16	3960	4890	5670	5730	5910
98- 16	3750	4680	5370	5670	5580
98- 16	3510	4470	5325	5730	6000

## 2.2 试验结果和讨论

2.2.1 土壤肥力、氮肥用量与产量的关系 对 90 个小区试验数据进行多因素方差分析,把造成产量变化的可知因素与未知因素加以划分,并对它们作用下产生的变异程度进行估算与比较,由此选择 Logistic 回归模型中的自变量参数。造成产量差异的主要因素为氮肥用量和土壤肥力,土壤有机质的储量在一定程度上反映了该土壤肥力的水平,是农田能否获得高产的重要标志。本试验中红砂泥田有机质平均值为 25.9 g kg<sup>-1</sup>,平均产量为 4 260 kg hm<sup>-2</sup>;黄泥田有机质平均值为 31.3 g kg<sup>-1</sup>,平均产量为 4 694 kg hm<sup>-2</sup>,土壤有机质与产量存在着明显的相关关系,土壤有机质可作为土壤肥力的主要代表参数。其次,红壤地区水稻田钾素亏缺很大,母质为红砂岩的红砂泥田速效钾均值为 51 mg kg<sup>-1</sup>,母质为第四纪红色粘土的黄泥田速效钾均值为 79 mg kg<sup>-1</sup>,速效钾可作为土壤肥力的又一表征参数。此外,本次试验土壤有效磷变化范围为 7~ 20 mg kg<sup>-1</sup>,对水稻而言,均在磷肥肥效没有显著影响范围之内,且不满足不同因子组合的试验设计条件。同时,土壤全氮与土壤有机质存在着相关关系,不符合统计学的基本假设。故选用土壤有机质、土壤速效钾和氮肥用量作为模型的自变量参数。

参照全国土壤普查办公室对土壤肥力水平的划分及有关文献<sup>[11,12]</sup>,考虑试验地土壤养分范围并满足随机区组设计 2×2 模型条件,设计分类变量与水平如下:

表3 分类变量与水平

Table 3 Categorical variable and levels

分类变量 Categorical variable	水平 Levels				
	1	2	3	4	5
氮肥用量(N kg hm <sup>-2</sup> )	0	45	90	135	180
有机质(g kg <sup>-1</sup> )	20~ 29	≥30			
速效钾(K mg kg <sup>-1</sup> )	< 50	50~ 110			

由表3, 可得到分类变量为有机质与速效钾时观测次数不等的 2×2 模型。以氮肥用量、有机质和速效钾等级为自变量, 产量为因变量, 进行交互作用下的非均衡方差分析(表4)。结果表明, 氮肥用量、土壤有机质等级对产量的影响显著 ( $Pr < 0.05$ ), 土壤速效钾等级对产量的影响达到 90% 的显著水平 ( $Pr < 0.1$ )。氮肥用量与有机质等级的交互作用、氮肥用量与速效钾等级的交互作用、及有机质水平与速效钾等级的交互作用, 对产量的影响均显著 ( $Pr < 0.05$ )。

表4 试验各变量的方差分析<sup>(1)</sup>

Table 4 Variable analysis of experimental parameters

差异来源 Source	自由度 DF	F 统计量 F Value
N	4	38.31**
O.M.	1	26.23**
avK	1	3.33+
N× avK	4	2.90*
avK× O.M.	1	12.47**
N× O.M.	4	4.86**

\*\*  $Pr < 0.01$ , \*  $Pr < 0.05$ , +  $Pr < 0.1$

应用 Duncan's 多重比较, 考察分类变量水平对产量均值的影响(表5)。土壤有机质水平对产量均值的影响差异显著 ( $Pr < 0.05$ ), 土壤速效钾水平影响差异未达到 95% 的显著水平。此外, 氮肥用量在水平 135 kg 和水平 180 kg 时对产量的差异不显著, 其他水平显著。

表5 氮肥用量、有机质和速效钾水平对产量均值的影响(kg hm<sup>-2</sup>)

Table 5 Effects of N rate, O.M. and rapidly available K on yield

水平 Level	氮肥用量 N rate	土壤有机质 O.M.	土壤速效钾 Rapidly available K
1	3505.0d	4215.7b	4578a
2	4249.2c	4738.3a	4426.8a
3	4636.7b		
4	5005.8a		
5	4988.3a		

注: 同一列数据中含有相同字母者表示差异不显著 ( $Pr > 0.05$ )

(1) 本文所涉及的统计计算, 均由 SAS 统计软件完成

2.2.2 模型的计算和比较 根据 6 块地 90 个小区试验数据,应用肥效法和 Logistic 模型,计算比较两种模型的误差。

肥效方程法(Quadratic equation)

选择 90 个小区试验数据进行回归统计,得到早稻产量与氮肥用量的一元二次统计模型。

$$Y = 3512.09 + 18.06X - 0.054X^2 \quad Pr = 0.0001 \quad R^2 = 0.55$$

式中:  $Y$  产量( $\text{kg hm}^{-2}$ )

$X$  氮肥用量( $\text{N kg hm}^{-2}$ )

Logistic 模型(Logistic regression)

据统计,试验所在地早稻产量一般在  $4\,500 \sim 5\,250 \text{ kg hm}^{-2}$  之间,本次试验产量平均值为  $4\,477 \text{ kg hm}^{-2}$ 。为比较不同的土壤肥力对产量的影响,分别考虑目标产量  $\geq 4\,500 \text{ kg hm}^{-2}$ 、 $\geq 4\,800 \text{ kg hm}^{-2}$  和  $\geq 5\,250 \text{ kg hm}^{-2}$  出现的概率条件。

选择氮肥用量、土壤有机质和速效钾等级为自变量,目标产量出现的概率为因变量。当产量  $\geq$  目标产量时,设出现该产量的事件值为 1,否则为 0。Logistic 回归模型中自变量可以是连续变量,也可以是分类变量,分类变量采用虚拟变量方法处理<sup>[10]</sup>。设土壤有机质为  $20 \sim 29 \text{ mg kg}^{-1}$  时,其虚拟变量编码为 0;  $\geq 30 \text{ mg kg}^{-1}$  时,编码为 1。设土壤速效钾  $< 50 \text{ mg kg}^{-1}$  时,虚拟变量编码为 0;速效钾为  $50 \sim 110 \text{ mg kg}^{-1}$  时,编码为 1。得到用于 Logistic 回归的编码表(表 6)。通过 Logistic 回归计算,选择最大似然预测,并设置信度 95%,得到产量  $\geq 4\,500 \text{ kg hm}^{-2}$  和  $\geq 4\,800 \text{ kg hm}^{-2}$  时的 Logistic 模型( $n = 90$ ):

$$\text{logit}(P_{4500}) = -2.95 + 0.027N + 1.65OM - C$$

$$\text{logit}(P_{4800}) = -7.62 + 0.039N + 4.43OM - C$$

表 6 Logistic 回归编码表

Table 6 The code of Logistic regression

田块编号 Site No.	氮肥用量 N rate ( $\text{N kg hm}^{-2}$ )	速效钾 Rapidly available K		有机质 O.M.		产量 Yield ( $\text{kg hm}^{-2}$ )	事件 Event $\geq 4500$	事件 Event $\geq 4800$	事件 Event $\geq 5250$
		含量	编码	含量	编码				
		Content	Code	Content	Code				
		( $\text{mg kg}^{-1}$ )	av. K_C	( $\text{mg kg}^{-1}$ )	OM_C				
1	0	53	1	2.8	0	3015	0	0	0
1	0	53	1	2.8	0	2910	0	0	0
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
6	90	110	1	3.18	1	4950	1	1	0
6	90	110	1	3.18	1	4170	0	0	0
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
6	180	110	1	3.18	1	5250	1	1	1

式中:

$P_{4500}$ 产量  $\geq 4\ 500\ \text{kg}\ \text{hm}^{-2}$  的概率

$P_{4800}$ 产量  $\geq 4\ 800\ \text{kg}\ \text{hm}^{-2}$  的概率

$N$  氮肥用量( $\text{N}\ \text{kg}\ \text{hm}^{-2}$ )

$OM\_C$  有机质虚拟变量编码

两个模型的显著性和各参数效应均满足显著性检验 ( $Pr < 0.05$ )。由 logist 模型可以看出, 此时目标产量概率条件与土壤有机质等级和氮肥用量正相关。两个模型中有机质等级和氮肥用量的系数不同, 后者大于前者, 说明有机质等级和氮肥用量更显著地影响产量  $\geq 4\ 800\ \text{kg}\ \text{hm}^{-2}$  出现的概率。

由  $P = \exp(a + bx) / (1 + \exp(a + bx))$ , 可计算出不同土壤有机质等级和不同氮肥用量下的早稻产量  $\geq 4\ 500\ \text{kg}\ \text{hm}^{-2}$  和  $\geq 4\ 800\ \text{kg}\ \text{hm}^{-2}$  的概率分布(图1)。概率随着氮肥用量的增加而增加, 但出现不同产量的概率曲线形状不同。有机质等级增加, 同一产量的概率曲线向上移动。当有机质  $\geq 30\ \text{mg}\ \text{kg}^{-1}$  时, 氮肥用量达到  $135\ \text{kg}\ \text{hm}^{-2}$  后, 概率均接近 0.9; 当氮肥用量达到  $180\ \text{kg}\ \text{hm}^{-2}$  时, 概率趋近 1, 符合肥料—产量报酬递减规律。当有机质为  $20\sim 29\ \text{mg}\ \text{kg}^{-1}$  时, 氮肥用量达到  $180\ \text{kg}\ \text{hm}^{-2}$  时,  $\geq 4\ 500\ \text{kg}\ \text{hm}^{-2}$  的概率为 0.87,  $\geq 4\ 800\ \text{kg}\ \text{hm}^{-2}$  的概率为 0.35。

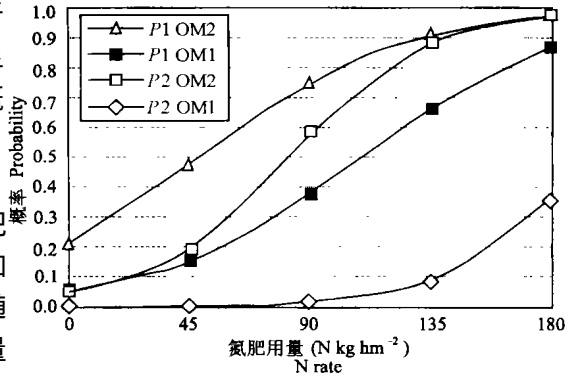


图1 早稻产量  $\geq 4\ 500\ \text{kg}\ \text{hm}^{-2}$  和  $\geq 4800\ \text{kg}\ \text{hm}^{-2}$  的概率分布

P1: 产量  $\geq 4\ 500\ \text{kg}\ \text{hm}^{-2}$  概率; P2: 产量  $\geq 4\ 800\ \text{kg}\ \text{hm}^{-2}$  概率。OM1: 有机质  $20\sim 29\ \text{mg}\ \text{kg}^{-1}$ ; OM2: 有机质  $\geq 30\ \text{mg}\ \text{kg}^{-1}$ 。

Fig.1 The probability distribution of different rice yields

同理, 可计算得到早稻单位产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  时的 Logistic 模型

$$\text{logist}(P_{5250}) = -47.06 + 0.039N + 1.27OM + 3.18avK\_C$$

式中:

$P_{5250}$ 产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  的概率

$N$  氮肥用量( $\text{kg}\ \text{hm}^{-2}$ )

$OM$  有机质含量( $\text{g}\ \text{kg}^{-1}$ )

$avK\_C$  速效钾虚拟变量编码

模型的显著性和各参数效应均满足显著性检验 ( $Pr < 0.05$ )。模型中土壤肥力参数对产量的影响与前式有所不同, 有机质的含量和速效钾等级更显著地影响高产出现的概率, 而有机质等级由于对高产概率没有显著效应从模型中除去, 说明现存有机质等级的划分对高产发生的概率不敏感, 如何合理地划分土壤养分等级尚需进一步研究。

不同土壤有机质、速效钾等级和不同氮肥用量下早稻产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  的概率分布如图2。从图中可以看出, 早稻产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  的概率不仅与土壤有机质含量和氮

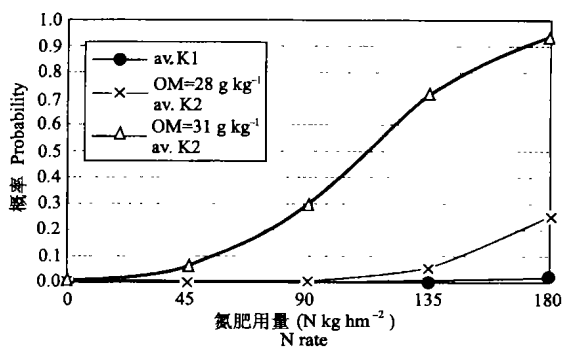


图 2 早稻产量  $\geq 250 \text{ kg hm}^{-2}$  的概率分布

av. K1: 速效钾  $< 50 \text{ mg kg}^{-1}$ ;

av. K2: 速效钾  $51 \sim 110 \text{ mg kg}^{-1}$ 。

Fig. 2 The probability distribution of early rice yield  $\geq 250 \text{ kg hm}^{-2}$

肥用量相关,而且与土壤速效钾含量等级相关。当土壤速效钾低于  $50 \text{ mg kg}^{-1}$  时严重缺钾,钾素的缺乏成为早稻产量  $\geq 250 \text{ kg hm}^{-2}$  的障碍因子,发生该产量的概率近似为 0。当速效钾含量在  $51 \sim 110 \text{ mg kg}^{-1}$  范围内,土壤供钾水平中等,概率曲线随土壤有机质含量增加而改变。有机质为  $31 \text{ g kg}^{-1}$ ,氮肥用量为  $135 \text{ kg hm}^{-2}$  时,概率为 0.71;氮肥用量为  $180 \text{ kg hm}^{-2}$  时,高产概率达到 0.93。有机质为  $28 \text{ g kg}^{-1}$  且氮肥用量为  $180 \text{ kg hm}^{-2}$  时,概率仅为 0.25。显然,氮肥的投入对增产仍起举足轻重的作用,而培肥土壤,

增加钾肥的投入,提高土壤速效钾和有机质是高产的主要途径。

2.2.3 模型的比较 肥效方程法通过大田实验本身统计而得出,预测产量具有较高的精度。肥效方程法相关系数  $R^2$  为 0.55, 该法较准确地反映了肥料用量与作物产量的关系,但由于未引入土壤肥力参数,不同土壤肥力的差异无法体现。Logistic 模型引入土壤肥力参数,得到的是不同施肥量和土壤肥力下早稻高产事件发生的概率。在线性回归中,确定系数  $R^2$  表示在因变量的变化中能够由外生变量所解释的比例,但是在 Logistic 回归中没有对应的统计指标。评价 Logistic 回归的一种方法是通过回归模型对样本数据因变量进行估计,比较样本中事件发生和不发生的观察结果与估计结果,以及在这两类估计正确的百分比和整体估计正确的百分比。为了比较 Logistic 模型与肥效法模型,将模型预测值事件发生次数与实际观测事件发生次数进行比较,见表 7 和表 8。

表 7 早稻产量  $4500 \sim 4800 \text{ kg hm}^{-2}$  下不同模型预测与观测的比较

Table 7 Comparison of the estimation with the observation for different models in yield  $4500 \sim 4800 \text{ kg hm}^{-2}$

观测 Observation	Logistic 法预测 Logistic regression estimation			肥效法预测 Quadratic method estimation		
	一致性 Concordant			一致性 Concordant		
	$\geq 4500$	$< 4500$		$\geq 4500$	$< 4500$	
$\geq 4500$	42	8	84%	43	7	86%
$< 4500$	4	36	90%	11	29	75%
总体一致性	87%			80.5%		

观测 Observation	Logistic 法预测 Logistic regression estimation			肥效法预测 Quadratic method estimation		
	一致性 Concordant			一致性 Concordant		
	$\geq 4800$	$< 4800$		$\geq 4800$	$< 4800$	
$\geq 4800$	21	5	78%	20	8	71%
$< 4800$	1	63	94%	16	46	74%
总体一致性	87%			72.5%		



表 8 早稻产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  时 Logistic 法预测与观测的比较Table 8 Comparison between estimation and observation for Logistic model in yield  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$ 

观测 Observation	Logistic 法 Logistic regression estimation		一致性 Concordant
	$\geq 5250$	$< 5200$	
$\geq 5250$	13	3	81%
$< 5250$	0	74	100%
总体一致性		88.6%	

产量  $\geq 4\ 500\ \text{kg}\ \text{hm}^{-2}$  的观测事件中, Logistic 模型预测正确的占 84%, 肥效法预测正确的占 86%。产量  $< 4\ 500\ \text{kg}\ \text{hm}^{-2}$  的观测事件中, Logistic 法预测正确的占 90%, 肥效法预测正确的占 75%。总体一致性: Logistic 模型 87%, 肥效法 80.5%。产量  $\geq 4\ 800\ \text{kg}\ \text{hm}^{-2}$  的观测事件中, Logistic 模型预测正确的占 78%, 肥效法预测正确的占 71%。产量  $< 4\ 800\ \text{kg}\ \text{hm}^{-2}$  的观测事件中, Logistic 模型预测正确的占 94%, 肥效法预测正确的占 74%。总体一致性: Logistic 模型 87%, 肥效法 72.5%。在一般中等产量情况下, Logistic 模型中包含氮肥用量和有机质为代表的基础地力, 预测结果与肥效法结果接近。

产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  的观测事件中, Logistic 模型预测正确的占 81%; 产量  $< 5\ 250\ \text{kg}\ \text{hm}^{-2}$  的观测事件中, Logistic 模型预测正确的占 100%。模型包含氮肥用量、土壤有机质、土壤速效钾等级, 说明在高产情况下, 不但肥料和土壤有机质起作用, 土壤速效钾也起较大的作用。模型中土壤速效钾的引入, 使得预测值的误差大大减少。而肥效法模型是通过回归统计得到一平均的氮肥与产量效应模型, 预测模型上的各点距各观测点距离最短, 不能全面地反映由于土壤肥力不同而造成产量的差异状态, 根据肥效法模型计算的最大产量值  $< 5\ 250\ \text{kg}\ \text{hm}^{-2}$ , 无法预测产量  $\geq 5\ 250\ \text{kg}\ \text{hm}^{-2}$  出现的情况。只有分别统计高产或中产肥效方程, 才能得到准确的产量预测值。

综合比较上述两种模型, 由于肥料用量与产量的关系近似二次曲线, 基于统计的肥效方程法能真实地反映客观规律, 但缺少土壤肥力参数, 不同土壤肥力的差异无法体现。Logistic 回归模型引入土壤肥力分类变量, 反映出土壤肥力的差异, 将复杂连续过程简化为离散过程, 有利于试验结果进行推广, 对定量研究土壤肥力、施肥量和产量之间的关系是一个有益的尝试。Logistic 模型与 GIS 结合的施肥决策具有广泛的应用前景。

## 参 考 文 献

1. 陆允甫等. 中国测土施肥工作的进展和展望. 土壤学报, 1995, 32: 241~ 251
2. 周鸣铮等. 浙江省水稻土“因产定氮”基本公式及其有关参数的探讨. 土壤学报, 1987, 24: 127~ 133
3. 刘光崧等. 大面积经济施肥的建模与施肥推荐研究. 土壤, 1991, 23(4): 61~ 81
4. 朱兆良. 土壤氮素有效性指标与土壤供氮量的预测. 土壤, 1990, 22: 177~ 180
5. 鲁如坤. 土壤-植物营养学原理和施肥. 北京: 化学工业出版社, 1998. 360~ 366
6. 耿兴元等. 土壤肥力模糊定量评价(判)系统的建立. 北京农业大学学报, 1995, 21: 23~ 28
7. 李仁岗等. 冬小麦施肥模式的初步研究. 土壤, 1990, 22(4): 210~ 211

8. 郭志刚等. 社会调查研究的量化方法. 北京: 中国人民大学出版社, 1989
9. 高惠璇等编译. SAS 系统 SAS/STAT 软件使用手册. 北京: 中国统计出版社, 1997. 458~ 472
10. 郭志刚主编. 社会经济统计分析方法- SPSS 软件应用. 北京: 中国人民大学出版社, 1999. 55~ 61
11. 谢为民等. 高产土壤条件及其配肥管理. 江西农业学报, 1995, 7(增刊): 67~ 74
12. 谢建昌. 钾与中国农业. 南京: 河海大学出版社, 2000. 81~ 83

## APPLICATION OF LOGISTIC REGRESSION FOR EARLY RICE RECOMMENDED FERTILIZATION IN RED SOIL AREA

Shi Jian-ping Lu Ru-kun Shi Zheng-yuan Sun Bo  
(*Institute of Soil Science, Chinese Academy of Sciences, Nanjing 200008*)

### Summary

This paper outlines an approach to the logistic regression for early rice recommended fertilization in red soil area. Logistic Regression model are used to relate nitrogen fertilizer with soil fertility variables to predict the high yield probability. The comparison between Logistic model and existing quadratic model is given on paddy soil of the red soil area.

**Key words** Red soil area, Recommend fertilization, Logistic regression