



保护地黄瓜规范化栽培数学模型

中国科学院黑龙江农业现代化研究所

王立志 杨丽华 陈能泰

本试验旨在应用多元二次回归旋转正交组合设计, 择求综合栽培因子对保护地黄瓜产量、经济性状的影响, 并根据试验结果, 求出反应这种影响的数学模型, 使用电子计算机优选出产量和纯经济效益的农艺组合技术方案, 为保护地黄瓜大面积高产稳产和提高经济效益提供咨询服务, 进而实现栽培技术的规范化。

一、试验内容及设计

根据目前保护地黄瓜栽培现状, 本试验主要研究苗令(X_1)、栽培密度(X_2)、灌水量(X_3)、氮肥施用量(X_4)和磷肥施用量(X_5)五个自变量因子, 每个因子又取五个水平, 且各因子和水平都处于动

态情况下, 对黄瓜产量和经济性状指标的影响。

试验采用多元二次回归旋转正交组合设计, 按回归设计要求设置36个田间试验小区, 由三部分组成。

1. $m_c = 16$, 两水平全因子试验的1/2实施的试验总数目为 $2^{5-1} = 16$

2. $m_r = 10$, 五维空间中位于各坐标轴上, 距离坐标原点半径为 r 的试验点数(r 为根据试验要求选定的设计参数)。

3. $m_s = 10$, 五个自变量均取零水平的中心点试验次数自变量设计水平及编码表如表1。

表一 变量水平编码表

变量名称	变化区间	零水平	自变量设计水平 $r = 2$				
			-2	-1	0	1	2
X_1 苗令(天)	5	50	60	55	50	45	40
X_2 密度(株/亩)	700	3200	1800	2500	3200	3900	4600
X_3 灌水(吨/亩)	5	20	10	15	20	25	30
X_4 N肥(公斤/亩)	15	50	20	35	50	65	80
X_5 P肥(公斤/亩)	15	50	20	35	50	65	80

二、试验条件及方法

试验在本所新组装式塑料大棚中进行, 前茬作物西红柿土地平坦肥力均匀。

36个小区分成三个正交区组。第一区组由八个 m_c 与四个 m_s 组成, 第二区组由另外八个 m_c 与四个 m_s 组成, 第三区组由十个 m_2 与两个 m_s 组成。零水平中心点重复试验小区组中按等间距排列用以估计土

壤自然肥力差异趋向, 其各试验小区在三个正交区组内随机排列。

依据哈市情况确定四月二十日为最佳定植期, 按苗令的五个水平分别于2月20日、25日, 3月2日、7日、12日播种。

灌水采用人工定量给水的方法, 按设计要求将每次的亩灌水量折合成每垅需水的桶数(15公斤/桶)进行灌水。从五月十

三日即缓苗后进行第一次定量给水，全生育期共定量给水十次。七月二十八日拉秧。

氮肥根据各处理水平的施用量平均分为六次按穴追施后期改为随水追施。

磷肥用作基肥根据各处理水平的施用量一次性沟施于试验小区内。

三、产量结果的回归分析

产量性状是进行生产活动的重要目标，依据田间试验过程获得的各试验小区

的产量结果进行回归分析和模拟寻优以确定最佳农艺措施并予测到该措施组合下的产量值，是我们该项研究的重要内容之一。

(一) 回归方程的检验和预报方程的建立

将试验计划结构矩阵及试验获得的各处理小区亩产量结果列于表2。

表二、结构矩阵及产量目标结果

小区号	播 期 X_1	密 度 X_2	灌水量 X_3	氮肥量 X_4	磷肥量 X_5	总产量 y_1 (斤/亩)
1	1	1	1	1	1	12329.1
2	1	1	1	-1	-1	20160.9
3	1	1	-1	1	-1	16587.0
4	1	1	-1	-1	1	17313.8
5	1	-1	1	1	-1	15518.3
6	1	-1	1	-1	1	14774.4
7	1	-1	-1	1	1	15142.1
8	1	-1	-1	-1	-1	15800.4
9	-1	1	1	1	-1	15988.5
10	-1	1	1	-1	1	12209.4
11	-1	1	-1	1	1	13466.3
12	-1	1	-1	-1	-1	15296.0
13	-1	-1	1	1	1	13449.2
14	-1	-1	1	-1	-1	14073.3
15	-1	-1	-1	1	-1	11610.9
16	-1	-1	-1	-1	1	14877.0
17	2	0	0	0	0	15227.6
18	-2	0	0	0	0	9755.6
19	0	2	0	0	0	18510.8
20	0	-2	0	0	0	11679.3
21	0	0	2	0	0	15744.0
22	0	0	-2	0	0	11422.8
23	0	0	0	2	0	14759.0
24	0	0	0	-2	0	13808.3
25	0	0	0	0	2	14603.4
26	0	0	0	0	-2	17884.3
27	0	0	0	0	0	14039.1
28	0	0	0	0	0	14988.2
29	0	0	0	0	0	16852.1
30	0	0	0	0	0	15293.7
31	0	0	0	0	0	15980.0
32	0	0	0	0	0	15783.3
33	0	0	0	0	0	15518.3
34	0	0	0	0	0	16638.3
35	0	0	0	0	0	16800.8
36	0	0	0	0	0	13004.6

1. 区组间产量差异显著性测定
首先求区组F值, 依公式 $F_Q = \frac{S_Q/f_Q}{S_{\text{误}}/f_{\text{误}}}$

(式中 s_Q 为区组平方和, f_Q 为区组自由度,
 $s_{\text{误}}$ 和 $f_{\text{误}}$ 分别为机误平方和和机误自由度)
对各小区产量结果的统计, 得 $s_Q = 7191508$
 $s_{\text{误}} = 11986745$, $f_Q = 2$, $f_{\text{误}} = 7$, 所以:

$$F_Q = \frac{7191508/2}{11986745/7} = 2.099$$

查F检验的临界值表 $F(2, 7)$, $\alpha = 0.05$ 的临界值 $F_{\alpha} = 4.74$, $F_Q < F_{\alpha}$ 说明区组间产量差异并不显著, 无需估计区组部分回归系数, 区组部分的平方和归入误差项。

2. 计算回归系数

根据试验设计的相关矩阵 $(X'X)^{-1}$ 和试验结果的常数项矩阵 $(X'Y)$ 并依据公式 $b = (X'X)^{-1} \cdot (X'Y)$ 求得:

$$\begin{aligned} b_0 &= 15391.69 & b_1 &= 1149.98 \\ b_2 &= 907.02 & b_3 &= 293.83 \\ b_4 &= -351.68 & b_5 &= -733.16 \\ b_6 &= 137.86 & b_7 &= -158.18 \\ b_8 &= -408.26 & b_9 &= -346.28 \\ b_{10} &= -147.50 & b_{11} &= -175.29 \\ b_{12} &= -872.10 & b_{13} &= 159.25 \\ b_{14} &= -905.24 & b_{15} &= 52.37 \\ b_{16} &= -602.33 & b_{17} &= 48.53 \\ b_{18} &= -329.38 & b_{19} &= -149.82 \\ b_{20} &= 308.23 \end{aligned}$$

由此获得五元二次多项式回归方程

$$\hat{y} = 15391.69 + 1149.98X_1 + 907.02X_2 + 293.83X_3 - 351.64X_4 + 733.16X_5$$

$$\begin{aligned} &+ 137.86X_1X_2 - 158.15X_1X_3 - 408.26 \\ &X_1X_4 - 346.28X_1X_5 - 346.28X_1X_6 - \\ &147.50X_2X_3 - 175.29X_2X_4 - 872.10X_2X_5 \\ &+ 159.25X_3X_4 - 905.24X_3X_5 + 52.37X_4 \\ &X_5 - 602.33X_1^2 + 48.53X_2^2 - 329.38X_3^2 - \\ &149.82X_4^2 + 308.23X_5^2 \end{aligned}$$

3. 回归方程及回归系数的显著性检验

(1) 拟合度测定

依据公式 $F_{L1} = \frac{S_{L1}/f_{L1}}{S_{\text{误}}/f_{\text{误}}}$ (式中 S_{L1}

失拟平方和 $S_{L1} = S_{\text{剩}} - S_{\text{误}}$; f_{L1} : 失拟自由度; $S_{\text{误}}$: 误差平方和; $f_{\text{误}}$: 误差自由度)。计算求得 $F_{L1} = 2.55$, 知 $F(f_{L1}, f_{\text{误}})$ 为(6, 9)的临界值 $F_{0.05} = 3.37$, $F_{L1} < F_{0.05}$, 说明失拟不显著, 失拟平方和基本上是由试验误差等偶然因素引起, 将失拟平方和与误差平方和合并, 继续检验 $S_{\text{回}}$ 。

$$F_{L2} = \frac{S_{\text{回}}/f_{\text{回}}}{(S_{L1} + S_{\text{误}})/(f_{L1} + f_{\text{误}})}$$

计算结果 $F_{L2} = 5.2$, $F(f_{\text{回}}, f_{L1} + f_{\text{误}})$ 为 $F(20, 15)$ 的临界值 $F_{0.05} = 3.37$, 检验结果显著, 说明回归方程拟合的很好, 可以用回归方程进行预报。

(2) 方差分析

第一次方差分析见表三。

回归方程中有很多不显著(NS)项, 由于试验计划具有正交性, 消除了回归系

表三、第一次方差分析

变量来源			平方和	自由度	均方	F_{t_e}	显著性
回 归	一 次 项	X_1	31738620.1	1	31738620.1	12.78	$\alpha = 0.01$
		X_2	19744301.7	1	19744301.7	7.95	$\alpha = 0.05$
		X_3	2072112.6	1	2072112.6	0.834	$\alpha = 0.05$
		X_4	2968348	1	2968348	1.195	$\alpha = 0.05$
		X_5	12900507	1	12900507	5.194	

	变量来源		平方和	自由度	均 方	F _{0.05}	显著性
回	交互项	X ₁ X ₂	304097	1	304097	0.122	ns
		X ₁ X ₃	400309	1	400309	0.161	ns
		X ₁ X ₄	2666852	1	2666852	1.073	
		X ₁ X ₅	1918502	1	1918502	0.772	
		X ₂ X ₃	318100	1	318100	0.140	ns
		X ₂ X ₄	491611	1	491611	0.197	ns
		X ₂ X ₅	12168934	1	12168934	4.899	
		X ₃ X ₄	405768	1	405768	0.163	ns
		X ₃ X ₅	13111278	1	13111278	5.278	
		X ₄ X ₅	43890	1	43890	0.0176	ns
归	平方项	X ₁ ²	11609774	1	11609774	4.674	ns
		X ₂ ²	75362	1	75362	0.030	
		X ₃ ²	3471788	1	3471788	1.398	
		X ₄ ²	718281	1	718281	0.289	
		X ₅ ²	3040167	1	3040167	1.223	
	总 计		120198608	20	6009930	2.42	α = 0.25
剩 余		37257786	15	2483852	R = 0.88		
总 计		157456394	35				

$$\text{复相关系数 } R = \sqrt{\frac{120198608}{157456394}} = 0.88$$

数量之间的相关性，可以直接把它们从回归项，然后再进行第二次方差分析（表四）
归方程中剔除，将它们的平方和并入剩余

表四 第二次方差分析

变 量 来 源			平方和	自由度	均 方	F值	显著性
回	一 次 项	X_1	31738620	1	31728620	18.23	0.01
		X_2	19744301	1	19744301	11.34	0.01
		X_3	2072112	1	2070112	1.19	
		X_4	2968348	1	2968348	1.71	0.05
		X_5	12900507	1	12900507	7.41	0.25
交 互 项	X_1X_4	2666852	1	2666852	1.53	0.25	
	X_1X_5	1918502	1	1918502	1.10		
	X_2X_5	12168934	1	12168934	6.99	0.05	
	X_3X_5	13111278	1	13111278	7.53	0.05	
归	平 方 项	X_1^2	11609774	1	11609774	6.67	0.25
		X_2^2	2471788	1	3471788	1.99	0.25
		X_5^2	3040167	1	3040167	1.75	0.25
	共 计		117411187	12	9784265	5.16196	0.01
剩 余		40045206	23	1741095			
总 计		157456394	35				

经第二次方差分析各项偏回归系数不同程度显著, 得到回归预报方程为:

$$\hat{y}_a = 15391.69 + 1149.98X_1 + 907.02X_2 + 293.83X_3 - 351.68X_4 - 733.16X_5 - 408.26X_1X_4 - 346.27X_1X_5 - 905.24X_3X_5 - 602.33X_1^2 - 329.38X_2^2 + 308.23X_5^2 \pm \delta$$

上式中 δ 为随机误差的根方差:

$$\delta = \sqrt{\frac{\text{剩余平方和}}{\text{剩余自由度}}} = \sqrt{\frac{40045206}{23}} =$$

1320

(二) 综合农艺措施对产量目标的计算机模拟寻优以建立的回归预报方程(数学模型)为基础, 以产量性状为目标, 用“苹果 I”微机对综合农艺措施, 进行高产区域的稳定产量模拟优化。

依据预报方程, 将五个因素的水平编

码值步长系数选为 $|1.0|$, 由微机可得到 3125 个组合方案的产量值, 我们给定 $Y_a > 18000$ 斤和 $18000 \text{ 斤} > Y_a > 15000$ 斤二个高产区间值, 在这两个区间内选择稳定产量的方案。 $Y_a < 15000$ 斤的方案予以淘汰。

在给定的高产区间内进行稳定产量点的选择, 实质上是入选方案内五个变量各水平编码出现的频数的选择, 即选择水平编码出现次数多(频数高)的区间, 即为稳定产量区间。为增加求得稳定产量点的可靠性, 我们不采用选取控制变量出现最高频数的编码值方法, 而是选取各控制变量编码值的加数平均值作为寻得的稳定产量点, 代入数学模型求得稳定产量。

根据计算机模拟计算, $\hat{Y}_a > 18000$ 斤产量的共 420 个, 其各变量水平编码频率列表五。

表五、稳产区间控制变量编码值频数表

(控制水平 $Y > 18000$ 斤/亩, 入选方案 420 个)

编码值 \ 变量	X_1	X_2	X_3	X_4	X_5
- 2	13	20	28	122	248
- 1	51	30	60	96	132
0	99	67	91	79	16
1	132	121	114	66	0
2	125	182	127	57	24

由表五看出, 当 X_1, X_2, X_3 三个变量在 0、1、2 三个水平时和 X_4, X_5 在 -1、-2、二个水平时 $Y_a > 18000$ 斤产量出现的频数最高, 而且集中。说明五个因素对产量都有明显影响, 即苗令短(40—50 天), 密度大,

灌水多, N、P 肥水平低时产量高且稳定。

根据所列编码值频数对各控制变量编码值进行加数平均计算, 获得稳定产量自变量水平取值点, 并确定相应的农艺措施方案, 列于表六。

表六 变量水平取值点

统计值	X_1	X_2	X_3	X_4	X_5	产 量
变量水平平均取值	0.73	0.99	0.60	-0.38	-1.38	2100133
标准误差 S_A	0.05	0.06	0.06	0.07	0.05	
农艺措施	$50 - 0.73 \times 5 = 46$ (天)	$3200 + 0.99 \times 700 - 38935 = 23$ (株/亩)	$20 + 0.6 \times 5 = 23$ (亩)	$50 - 0.38 \times 15 = 44.3$ (公斤/亩)	$50 - 1.38 \times 15 = 293$ (公斤/亩)	

由表七知亩产18000斤以上的方案 x 取值的稳定点是 $x_1 = 0.73$, $x_2 = 0.99$, $x_3 = 0.60$, $x_4 = -0.38$, $x_5 = -1.38$, 将上述编码值变为实施方案相应的农艺措施是: 苗令46天, 密度3893株/亩, 灌水23吨/亩, N肥44.3公斤/亩, P肥29.3公斤/亩。再将 x_1 (1-5) 代入回归预报方程中得到稳定产量为21001.33斤/亩。该数值即为稳产区间最佳农艺措施的最佳产量值。

根据计算机计算 $18000 > Y_a > 15000$ 斤的产量方案共759个, 仍然是0、1、2三个水平时出现频数最高, x_4 、 x_5 也是0, -1、-2三个水平时出现频数最高, 说明欲获得 $y_a > 15000$ 斤产量, X_1 、 X_2 仍需保持大于0水平, 即短苗令, 高密度, X_4 、 X_5 仍需保持小于0水平, 即低N、P肥处理。

但是在 $18000 > Y_a > 15000$ 的稳定产量区间, x_1 、 x_2 、 x_3 的取值要小于 $y > 18000$ 产量取值水平, x_1 、 x_2 趋于零水平, x_3 则小于0水平, x_4 、 x_5 的绝对值也小于 $y > 18000$ 斤产量区间的水平, 也更接近于0水平, 即密度、灌水量变小了, N、P肥量偏大了。 x 取值的稳定区域点是: $x_1 = 0.53$, $x_2 = 0.13$, $x_3 = -0.26$, $x_4 = -0.36$, $x_5 = 0.07$, 将 x_1 代入回归预报方程, 得稳产值为16001.61斤/亩, 要求的农艺措施是苗令47天, 密度3109株/亩, 灌水量18.7吨/亩, N肥44.6公斤/亩, P肥48.95公斤/亩。

综合上述, 对高产区间变量编码值频数的分析可知, 1. 在规定高产区间内变量编码值的变化规律及确定的编码值; 2. 确定的不同产量区间要求的变量编码值不同即要求的农艺措施不同; 3. 不同的变量编码值 (即不同的农艺措施) 将产生不同的产量区间。

由上所述得知, 以数学模型为基础, 在生产实践中给出确定的变量水平数值即可得到较为准确的高产稳产区间, 并进行预

报, 给出确定的高产指标后又可确定各变量的水平数值来指导生产, 提供咨询服务。

四、纯经济收益结果的回归分析

经济性状是进行生产活动的又一主要指标, 在此对经济性状的研究主要指纯经济效益即各试验小区获得的产值与成本消耗之差, 依据试验过程获得的各试验小区的纯经济效益结果进行回归分析和模拟寻优, 以确定最佳农艺措施并予测该措施组合下的纯收益, 并且可以对产量和纯经济效益为不同目标的农艺组合进行综合评价。

(一) 回归预报方程的建立

经计算和统计分析建立的回归预报方程为:

$$\hat{y} = 1991.6 + 253.8X_1 + 136.3x_2 - 117.2x_4 - 135.9x_5 - 108.58X_1x_4 - 138.1x_1x_5 + 100.8x_2x_3 - 149.6x_2x_5 - 266.4x_3x_5 - 152.6x_1^2 - 126x_2^2$$

(二) 综合农艺措施对经济目标的模拟寻优

1. 稳定高效益经济区域的模拟优化

我们给定在 $\hat{y}_b > 2300$ 的区域内选择稳定高收益的方案, 微机模拟计算 $y_b > 2300$ 的方案共476个。

根据编码频数分析, x_1 、 x_2 、 x_3 三变量在0, 1, 2三个水平时, $y_b > 2300$ 出现的频数最高且集中。说明三因素对纯收益都有明显影响, 且当苗令短, 密度大, 灌水多N、P肥少时纯收益稳定。

根据所得编码值频数, 对各控制变量编码值进行加数平均计算, 获得稳定高效益自变量水平取值点, 并确定相应的农艺措施方案。

是: $x_1 = 0.8$, $x_2 = 0.51$, $x_3 = 0.4$, $x_4 = -0.58$, $x_5 = -0.84$, 将上述编码值变换成实施方案, 相应的农艺措施为苗令

(下转第10页)

八五年省内九个区和生产示范点的冬红陈籽也均未抽苔。

2. 密封贮藏：如果品种的冬性较弱，并且必须使用陈籽播种，可将新采的种子进行充分干燥，然后在自然温度下密封贮藏，也能收到防止抽苔的效果。

如：七二年秋将王兆红新籽干燥后分别装入不同容器并均放在自然温度下贮藏，七三年秋播后布袋种子抽苔15.25%，而用塑料袋、陶瓷罐、玻璃瓶、白铁筒密封贮藏的种子均未抽苔。

3. 延后播种：如果必须使用弱冬性品种的受潮陈籽，就只好进行延后播种。但预防抽苔的效果很不稳定，并且势必明显降低产量。

如：不同年份延后播种的抽苔率分别为

播 期	七八年	八一年	八二年
30/6	17.9	20.7	45.4
10/7	14.80	0.00	26.30

又如：八三年分期播种的产量是

播 期	30/6	5/7	10/7	15/7
平均单株重 (斤)	2.89	2.31	1.93	1.53
单株减少 (%)	—	20.07	33.22	47.06

总之，选育强冬性品种是彻底预防萝卜当年抽苔的根本途径，干燥密封贮藏只能在陈籽抽苔而新籽不抽苔的地区适用，延后播种只应当在别无办法的情况下采用。

※ ※ ※ ※ ※

(上接第6页)

46天，密度3557株/亩，灌水量22吨/亩，N肥量41.3公斤/亩，P肥量37.4公斤/亩，将 $x_2(1-s)$ 代入纯收益为目标的回归预报方程中，其稳定纯收益为265.7元/亩。

对纯经济效益的试验结果与模拟寻优得出的结论与对产量模拟寻优的结果完全一致，说明寻求的最稳定区域组合既可获高产又可获得高收益，达到一举二得。

五、结束语

应用回归试验设计和回归分析的方法，依据田间试验获得的数据，建立了不同目标的数学模型，通过计算机的模拟寻优，获得最佳实施方案的变量输出点和最

佳输出结果，把对农业栽培试验引向综合因素和数学化的方向。

获得的数学模型可作为一种工具进行模拟田间试验，在不同的假定控制条件下，用仅仅改变某一个输入变量的方法，依次进行若干次模拟试验，模型将给出一系列的反应值，这样我们可以利用模型和根据确定的目标，相应采取农艺措施，作出决策，还可以根据自然条件和采取的农艺措施对预期结果进行预报，我们相信，经过深入细致的研究一定能够建立一个产量，经济系统的预测，预控模型为生产服务。