

# 基于主成分和聚类分析的大豆品种生物学性状的比较研究<sup>\*</sup>

张玉革<sup>1,2</sup> 胡绪彬<sup>2</sup>

(1. 沈阳农业大学土地与环境学院, 沈阳 110161; 2. 沈阳市农业技术推广站, 沈阳 110034)

**摘要** 采用主成分分析和聚类分析方法研究了 10 个大豆品种在沈阳地区种植的适应性。结果表明, 不同大豆品种的主要生物学性状可通过主成分分析归纳成单株个体因子、株形因子及致病虫害因子 3 个主成分; 聚为 3 类后, 每一类的单位面积产量的差异极显著, 高产品种一般具有较多的有效分枝、单株荚数、单株粒数、单株粒重, 生育期相对较长, 株高和主茎节数的数值相对较低; 主成分分析和聚类分析方法是用于区域大豆品种筛选的十分有效的统计学方法。

**关键词** 大豆; 主成分分析; 聚类分析

中图分类号 S 565. 101 文献标识码 A 文章编号 1000—9841(2004)03—0178—06

## 0 引言

研究多个数量性状之间的关系及对受多个性状影响的群体进行分类时, 日益广泛地使用聚类分析和主成分分析。聚类分析是在不同品种的各性状综合为相似距离的基础上进行的。主成分分析可以将多个相互关联的数量性状综合为少数几个主成分。通过对变量之间的相关系数矩阵内部结构的研究, 找出数目较少能控制所有变量的主成分。关于对大豆品种生物学性状进行主成分分析和聚类分析的研究, 前人已有不少报道。如胡立成等<sup>[1]</sup>在黑龙江大豆品种基因库中选取 50 个品种, 对 6 个主要生物学性状进行聚类分析, 判别大豆品种的多样性和遗传分歧的多向性。周述明等<sup>[2]</sup>在对四川省 74 个大豆地方品种资源的遗传距离进行测定的基础上, 按遗传距离聚为 8 类, 考查了类间的遗传差异和各品种间地理差异。李向华和常汝镇<sup>[3]</sup>对 89 个中国春大豆品种按“遗传距离”进行聚类分析, 在分为 6 大类的基础上选取 11 个生物学性状进行主成分分析, 最后选取 6 个综合性状对品种进行综合评价。Zhang 等<sup>[4]</sup>对大豆的遗传多样性进行分析, 利用聚类和主成分分析, 国内一些研究者对地方大豆品种遗传特

性进行了研究<sup>[4—7]</sup>。这些研究主要是针对各地区不同区域分布的主要大豆品种, 对大豆的遗传学性状进行聚类 and 主成分分析, 找出各品种遗传距离间的差异。由于品种的遗传分歧和地理分布的差异, 自然选择的结果导致品种遗传分歧的多向性。这些研究为大豆的杂交育种和亲本选择提供科学依据。本文在借鉴以往研究方法的基础上, 选择 10 个的当地育成和引种的大豆品种, 借助聚类分析和主成分分析的数学方法, 对这些品种在沈阳地区种植的适应性进行评价, 从而为区域性大豆品种更新和筛选工作提供理论依据, 为作物品种的区域适应性评价提供方法参考。

## 1 材料与方法

### 1.1 供试材料

供试大豆品种有铁岭市农科院育成的铁丰 29 (CK)、铁丰 31 (高油)、开原市农示场的开 8157、法库县种子示范场从吉林长春引进的相文 1 (相文 88—A)、相文 2 (相文 88—8)、相文 3 (相文 88—9)、辽宁省农科院大豆所育成的辽豆 11、辽豆 13、辽豆 14、辽豆 15。

### 1.2 试验设计及田间管理

<sup>\*</sup> 收稿日期: 2004—01—12

基金项目: 辽宁省自然科学基金项目(20031008)和沈阳市“十五”重大科技项目(1031020—3—02)

作者简介: 张玉革(1968—), 女, 高级农艺师, 博士研究生, 从事大豆栽培研究及技术推广。E-mail: zhangyvg@sina.com

试验在沈阳市新城子区虎石台镇中古城子村和法库县法库镇大泉眼村。试验地选择平肥地, 前茬为玉米。小区试验采用随机区组排列, 3 次重复。6 行区, 行长 10 m, 行距 0.6 m, 小区面积 36 m<sup>2</sup>, 各小区间隔 1 m。

根据各品种特性, 种植密度为 120000—150000 株/hm<sup>2</sup>, 底施优质农家肥 30 t/hm<sup>2</sup>, 复合肥 187 kg/hm<sup>2</sup>, KCl 75 kg/hm<sup>2</sup>, 施种前一次施入, 药剂灭草。2003 年 5 月 9 日播种, 6 月 10 日铲地, 6 月 15 日用吡虫啉 1500 倍液防蚜虫, 7 月 29 日和 8 月 10 日分别用 15 kg/hm<sup>2</sup> 敌敌畏防食心虫, 10 月初收获。

1.3 统计方法

主成分分析、聚类分析(按欧氏距离进行聚类)、相关分析等采用 SPSS 软件包。选择单位产量、株高、结荚高度、主茎节数、有效分枝、单株荚数、单株

粒数、每荚粒数、单株粒重、百粒重、紫斑粒数、褐斑粒数、虫蚀粒数等 13 个生物学性状及产量指标进行主成分分析和聚类分析。

2 结果与讨论

2.1 不同大豆品种主要生物学性状及产量状况

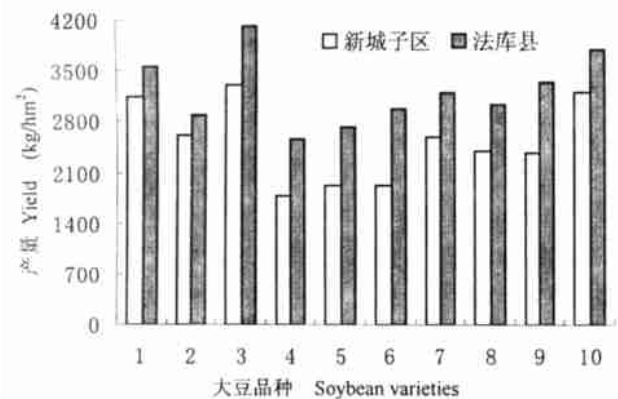
不同大豆品种各生物学性状数值在两个试验点基本表现出相同的趋势(表 1)。通过对两试验点各生物学性状数值进行相关性分析的结果表明, 两点间的生育期之间、株高之间、有效分枝之间、单株粒重之间、百粒重之间等均呈极显著正相关, 结荚高度之间和主茎节数之间呈显著正相关, 其他性状在两个试验点之间呈正相关, 但未达到显著水平。

新城子区试验点单位产量最高的 3 个品种为

表 1 大豆品种主要生物学性状  
Table 1 Biological characters of soybean varieties

品种 Varieties	生育期 (d) Days of growth	株高 (cm) Plant height	结荚高度 (cm) Height of pod	主茎节数 (个) No. of nodes in main stem	有效分枝 (个) No. of branches	单株荚数 (荚) No. of pods per plant	单株粒数 (粒) No. of seeds per plant	每荚粒数 (粒) No. of seeds per pod	单株粒重 (g) Seed weight per plant	百粒重 (g) 100-seed weight
新城子区										
铁丰 29	133	78.3	16.1	17.9	3.2	75.7	128.6	1.7	29.6	23.1
铁丰 31	127	92.5	13.8	22.3	2.7	92.2	161.6	1.8	28.8	17.9
8157	132	88.1	17.6	17.5	4.8	96.5	159.3	1.7	37.8	23.7
相文 1	114	81.3	14.6	20.4	1.9	46.7	114.7	1.6	22.6	19.7
相文 2	114	91.2	14.0	22.0	1.2	59.8	84.4	1.4	16.5	19.6
相文 3	113	88.0	14.2	22.2	2.5	87.4	131.1	1.6	25.0	18.9
辽豆 10	127	112.8	11.5	25.1	2.7	73.6	138.3	1.9	28.6	20.7
辽豆 11	128	109.0	12.6	21.3	2.1	64.7	135.9	2.1	27.3	20.1
辽豆 14	128	104.7	16.2	24.2	3.1	80.0	129.2	1.6	19.4	15.0
辽豆 15	132	91.3	19.1	19.5	3.6	82.7	163.1	2.0	35.7	22.1
法库县										
铁丰 29	128	95.3	13.6	13.8	3.7	125.0	246.1	2.0	26.0	25.6
铁丰 31	127	115.0	14.0	18.0	1.8	56.2	103.6	1.8	26.0	17.8
8157	126	106.0	16.5	17.8	5.3	141.3	298.0	2.1	37.0	25.1
相文 1	111	110.0	13.9	17.0	0.0	34.8	76.0	2.2	16.0	20.6
相文 2	111	116.0	14.8	21.0	0.5	40.6	86.0	2.1	17.0	20.5
相文 3	121	110.0	14.5	20.0	1.0	40.1	90.4	2.3	20.0	21.5
辽豆 10	126	123.0	12.6	20.8	1.6	61.8	131.2	2.1	20.0	22.4
辽豆 11	126	123.0	13.2	17.8	1.0	39.4	110.5	2.8	28.0	22.1
辽豆 14	126	124.0	14.4	18.7	1.0	52.0	128.8	2.5	18.0	17.2
辽豆 15	127	118.0	14.5	19.0	2.0	69.2	157.2	2.3	26.0	25.0

8157、辽豆 15 和铁丰 29, 产量最低的 3 个品种为相文 1、相文 2 和相文 3; 法库县试验点单位产量最高



编号: 1. 铁丰 29、2. 铁丰 31、3. 8157、4. 相文 1、5. 相文 2、6. 相文 3、7. 辽豆 11、8. 辽豆 13、9. 辽豆 14、10. 辽豆 15

图 1 不同品种大豆的单位产量

Fig. 1 Yields of soybean varieties

品种的也是 8157、铁丰 29 和辽豆 15, 产量最低的是

表 2 大豆品种主要生物学性状的主成分分析

Table 2 Principal analysis for biological characters of soybean varieties

成分 Comp- onent	初始化特征值			因子提取后特征值			成分 Comp- onent	初始化特征值			因子提取后特征值			
	Initial eigenvalues			Extraction sums of squared badings				Initial eigenvalues			Extraction sums of squared badings			
	总方差	因子贡	累计贡	总方差	因子贡	累计贡		总方差	因子贡	累计贡	总方差	因子贡	累计贡	
	Total	献率	献率	Total	献率	献率		Total	献率	献率	Total	献率	献率	
	% of	Cumul		% of	Cumul		% of	Cumul		% of	Cumul		% of	Cumul
	variance	ative %		variance	ative %		variance	ative %		variance	ative %		variance	ative %
新城子区							法库县							
1	6.124	47.110	47.110	6.124	47.110	47.110	1	7.459	57.373	57.373	7.459	57.373	57.373	
2	2.343	18.022	65.132	2.343	18.022	65.132	2	2.346	18.043	75.416	2.346	18.043	75.416	
3	1.660	12.773	77.905	1.660	12.773	77.905	3	1.490	11.459	86.875	1.490	11.459	86.875	
4	1.161	8.928	86.833				4	0.688	5.289	92.164				
5	0.900	6.923	93.756				5	0.420	3.227	95.391				
6	0.373	2.868	96.625				6	0.362	2.784	98.176				
7	0.291	2.236	98.861				7	0.194	1.491	99.666				
8	0.109	0.836	99.696				8	0.036	0.277	99.943				
9	0.039	0.304	100.000				9	0.007	0.057	100.000				

大豆品种主要生物学性状的主成分矩阵显示了各主要性状在各个主成分矩阵中的权重系数。第一主成分中,新城子区和法库县试验中的单株粒重、有效分枝、单株粒数、单株荚数等性状的权重系数均较大,这些主要是反映大豆单株个体性状的参数,表明大豆单株个体间的差异是造成不同品种间差异的主导因子,可以称之为“单株个体”因子。第一主成分中的主茎节数和株高的权重系数均为负值。由于两试验点在地理位置和自然条件方面存在一定的差异,因此各性状在第一主成分中的权重也有一定的差异。如紫斑粒数和每荚粒数的权重系数不仅在数量上存在一定差异,而且在正负向上是相反的(表 3)。

相文 1、相文 2 和铁丰 31,相文 3 的产量略高于铁丰 31。各品种在两个试验点之间的产量表现基本一致(图 1)。

2.2 不同大豆品种主要生物学性状的主成分分析

从主成分分析的初始化特征值和因子提取后的特征值可看出,新城子区和法库县初始化特征值中 9 个成分的累计贡献率都已达到 100%,提取前 3 个主成分可分别解释 77.91%和 86.88%的方差。如果主成分分析中所提取主成分的特征值能达到 70%以上的贡献率,就可以用这几个主成分对事物的属性进行概括性分析,基本可以得出影响事物性质的主要因素<sup>[8]</sup>,因此提取 3 个主成分可以概括不同大豆品种生物学性状的绝大部分信息(表 2)。

第二主成分中,新城子区以株高、主茎节数、每荚粒数、结荚高度的权重系数较大,主要是“株形”因子。但法库县以结荚高度、紫斑粒数、褐斑粒数、虫蚀粒数的权重系数较大,这这几个性状主要表现了大豆病虫害状况,可以归纳为“致病虫”因子。第三主成分中,新城子区的紫斑粒数、褐斑粒数、虫蚀粒数的权重系数较大,主要为“致病虫”因子,而法库县的株高、主茎节数、结荚高度、每荚粒数的权重系数较大,主要是“株形”因子(表 3)。

2.3 不同大豆品种主要生物学性状的聚类分析

根据所选择的生物学性状对 10 个品种进行聚类的树状图可看出,在新城子区和法库县不同大豆品种的聚类结果略有差异。在法库县的试验中,铁

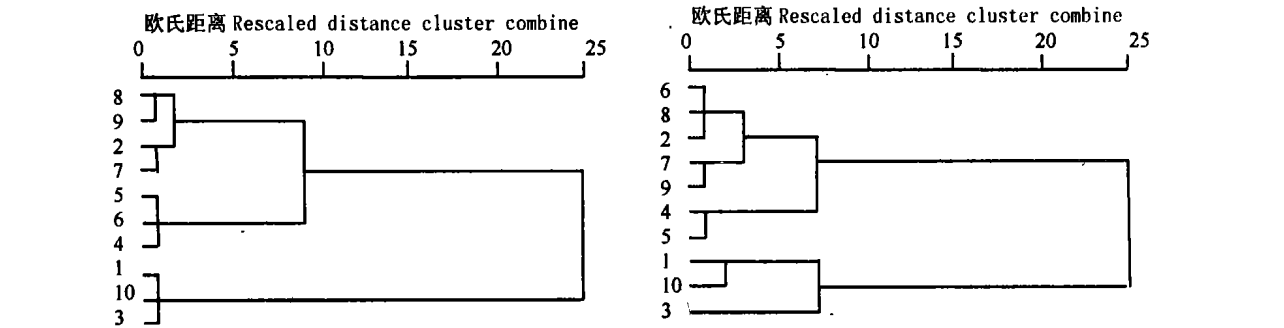
表 3 大豆品种主要生物学性状的主成分矩阵

Table 3 Principal component matrix for biological characters of soybean varieties

	主成分 Principal component					
	1	2	3	1	2	3
		新城子区			法库县	
单株粒重 Seed weight per plant	0. 928	0. 174	0. 218	0. 854	0. 000	0. 312
单位产量 Yield	0. 916	0. 000	0. 128	0. 820	0. 216	0. 417
有效分枝 No. of branches	0. 861	0. 000	0. 000	0. 963	0. 200	0. 111
单株粒数 No. of seeds per plant	0. 827	0. 470	0. 000	0. 955	0. 222	0. 000
单株荚数 No. of pods per plant	0. 684	0. 254	－0. 291	0. 956	0. 224	0. 000
结荚高度 Height of pod	0. 684	－0. 433	－0. 401	0. 297	0. 762	0. 387
主茎节数 No. of nodes	－0. 664	0. 628	－0. 310	－0. 582	0. 000	0. 646
紫斑粒数 No. of purple seeds	－0. 634	0. 198	0. 548	0. 666	－0. 634	0. 286
百粒重 100—seed weight	0. 617	－0. 298	0. 533	0. 659	0. 334	－0. 123
褐斑粒数 No. of brown seeds	0. 571	0. 109	－0. 568	0. 764	－0. 616	0. 126
株高 Plant height	－0. 280	0. 901	0. 000	－0. 529	－0. 518	0. 641
每荚粒数 No. of seeds per pod	0. 460	0. 691	0. 223	－0. 812	0. 296	0. 301
虫蚀粒数 No. of pest—bitted seeds	0. 474	0. 000	0. 539	0. 690	－0. 568	－0. 146

丰 29 和辽豆 15 聚在一类,这两个品种又与 8157 聚在一类,相文 1 和相文 2 之间、辽豆 11 和辽豆 14 之间的欧氏距离也较近(图 2)。在新城子区的试验中,铁丰 29、8157、辽豆 15 聚在第 1 类,铁丰 31、辽豆 11、辽豆 13、辽豆 14 聚在第 2 类,相文 1、2、3 号聚在第 3 类(图 2)。从新城子区的试验结果看,聚在第 1 类的 3 个品种的单位产量水平排在 10 个品种的前 3 位(图 1),这 3 个品种的共同特点是生育期较长,百粒重较重,有效分枝数较多,单株粒数多,单株粒重较重(表 1),这 3 个品种的单株个体因子在第一主成分中的权重系数较大,是影响不同大豆品种单位产量的主导因素;株形因子(第二主成分)中,株高、主

茎节数的实测值都相对较低,而结荚高度相对较高(表 1),表明株形对不同品种单位产量的影响。聚在第 3 类的相文 1、2、3 号大豆品种的共同特点是生育期较短(仅为 114d),有效分枝数较少,每荚粒数较少,单株粒数较少,单株粒重较轻(表 1),而这 3 个品种的结荚高度和主茎节数的实测值相对较高,但有效分枝相对较少,这与第 1 类的 3 个品种在方面的表现正好相反。由于相文系列为引种品种,通过对产量及生物学性状的聚类分析,表明其在沈阳地区种植基本不适应。聚类的结果反映了不同品种间产量构成因子及大豆产量之间的差异,这为正确选用品种提供了依据。



A. 新城子区 B. 法库县

品种编号: 1. 铁丰 29、2. 铁丰 31、3. 8157、4. 相文 1、5. 相文 2、6. 相文 3、7. 辽豆 11、8. 辽豆 13、9. 辽豆 14、10. 辽豆 15

图 2 大豆品种的欧氏距离聚类树状图

Fig. 2 Dendrograms of soybean varieties using squared Eudidean distances

对新城子区和法库县不同品种大豆 13 个性状 (含单位产量)作为变量进行聚类的结果是一致的,  
?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

株高、单株英数、单株粒数等 3 个性状被聚在第 1 个性状被聚在第 2 类。  
类,不同大豆品种的单位产量被聚在第 3 类,其它 9

表 4 大豆产量与生物学性状的相关性

Table 4 Correlations of soybean yield to its biological characters

生育期 (d) Days of growth	株高 (cm) Plant height	结荚高度 (cm) Height of pods	主茎节数 (个) No. of nodes	有效分枝 (个) No. of branches	单株英数 (英) No. of pods per plant	单株粒数 (粒) No. of seeds per plant	每荚粒数 (粒) No. of seeds per pod	单株粒重 (g) Seed weight per plant	百粒重 (g) 100seed weight
新城子区									
产量 Yield	0.914 **	-0.078	0.583	-0.549	0.849 **	0.614	0.708 *	0.465	0.847 **
法库县									
产量 Yield	0.677 *	-0.245	0.480	-0.242	0.859 **	0.829 **	0.855 **	-0.017	0.752 *

注: \*, \*\* 分别为 0.05 和 0.01 水平相关 \*, \*\* are significant at 0.05 and 0.01 levels

根据主成分分析和聚类分析的结果,对不同品种大豆的产量与第一、第二主成分中权重系数较大的主要生物学性状之间的相关性进行分析的结果见表 4。产量与有效分枝、单株英数、单株粒数、单株粒重等“单株个体因子”呈显著或极显著正相关,与品种的生育期也是显著或极显著正相关。产量与株高及主茎节数等“株形因子”呈负相关,但未达到显著水平,这个结果可能与结荚习性有关。一般情况下,大豆的产量应与株高、主茎节数呈正相关,但本研究中,主茎节数少的有效分枝相对较多,有效分枝对产量构成的影响超出了主茎节数的影响。大豆的产量一般与单位面积株数、单株英数、单荚粒数、百粒重密切相关,大豆各产量因子优越,则高产潜力越大;同一地区同一生态类型品种,荚粒和粒重变化幅度差异不会很大,而单位面积结荚数等影响产量的提高<sup>[9]</sup>,因此,大豆单位面积内的结荚数及与之相关的“单株个体因子”是主要的产量因子。

学性状间的相关性分析表明,产量与有效分枝、单株英数、单株粒数、单株粒重等“单株个体因子”之间呈显著或极显著正相关,与品种的生育期之间呈显著或极显著正相关,与株高及主茎节数等“株形因子”之间呈负相关关系,但相关性不显著。主成分分析结果反映了大豆生物学性状的主导影响因子,聚类结果反映了产量的差异,这为正确评价大豆品种的适应性及科学地选择品种提供了依据。

参 考 文 献

1 胡立成,姚远,李秀兰,等.黑龙江省大豆品种聚类分析初探[J].大豆科学,1991,10(1):10-16.

2 周述明,谢林,林文君.四川大豆地方品种资源初步研究Ⅲ.遗传距离测定及聚类分析[J].四川农业大学学报,1994,12(1):37-41.

3 李向华,常汝镇.中国春大豆品种聚类分析及主成分分析[J].作物学报,1998,24(3):325-332.

4 Zhang J, Arelli PR, Skper DA, et al. Genetic diversity of soybean germplasm resistant to *Heterodera glycines* [J]. Euphytica, 1999, 107: 205-216.

5 游明安,盖钧镒.长江下游大豆地方品种的聚类分析[J].中国油料,1994,16(4):36-45.

6 刘灿洪,林荣辉,刘德金.福建省大豆地方品种多元统计分析[J].福建农业学报,1999,14(4):11-15.

7 陈荣江,王莹.大豆农艺性状的因子分析及品种的聚类分析[J].河南职业技术学院学报,1998,26(2):20-24.

8 徐克学.生物数学[M].北京:科学出版社,1999.12-102.

9 董丽华.大豆产量构成因素与产量[J].黑龙江农业科学,1996,1:40-41.

3 结 语

主成分分析表明,不同大豆品种的生物学性状可分为单株个体因子、株形因子和致病虫因子 3 个主成分,各个生物学性状在各个主成分中的权重系数决定其在主成分中的影响度;按生物学性状对不同品种聚成 3 类,分别表达了具有单株个体和株形优势的为高产品种类、占劣势的为低产品种类、处于二者之间的为中产品种类。不同品种产量与各生物

## A COMPARATIVE STUDY ON BIOLOGICAL CHARACTERS OF SOYBEAN VARIETIES USING PRINCIPAL COMPONENT ANALYSIS AND CLUSTER ANALYSIS

Zhang Yuge<sup>1, 2</sup> Hu Xubin<sup>2</sup>

(1. College of Land and Environment, Shenyang Agricultural University, Shenyang, 110161;  
2. Shenyang Agro-technique Extension and Service Center, Shenyang, 110034)

**Abstract** The adaptability of 10 soybean varieties in Shenyang region of Liaoning province was studied using principal component analysis and cluster analysis. The results showed that the main biological characters of soybean varieties could be divided into plant individual factors, plant shape factors and factors related to plant diseases and insect pests through principal component analysis. The differences of soybean yields among each group were significant as clustered into 3 groups. Varieties with high yield had comparatively more effective branches, number of pods and number of seeds per plant, heavier of seed weight per plant and longer of growth periods while their plant heights were shorter and nodes in main stems were less. Principal component analysis and cluster analysis were proven to be effective statistical methods in selecting soybean varieties in a specifically region.

**Key words** Soybean (*Glycine max* L.); Principal component analysis; Cluster analysis

## 欢迎订阅 2005 年《作物学报》

《作物学报》是中国科学技术协会主管、中国作物学会和中国农业科学院作物育种栽培研究所共同主办、科学出版社出版的有关作物科学的全国性学术刊物。主要刊登农作物遗传育种、耕作栽培、生理生化、生态、种质资源、谷物化学、贮藏加工以及与农作物有关的生物技术、生物数学、生物物理、农业气象等领域以第一手资料撰写的学术论文、研究报告、简报以及专题综述、评述等。读者对象从事农作物科学研究的科技工作者、大专院校师生和具有同等水平的专业人士。

《作物学报》为月刊, 2005 年 136 页/期, 定价: 25 元/册, 全年 300 元。可通过全国各地邮局订阅, 刊号: ISSN0496-3490, CN 11-1809/S, 邮发代号: 82-336。也可向编辑部直接订阅。

编辑部地址: 北京市海淀区中关村南大街 12 号中国农科院作物所《作物学报》编辑部(邮编 100081)。  
联系电话: 010-68918548, 传真: 010-68978616。

银行汇款: 交通银行北京分行农科院分理处, 户名: 中国作物学会, 帐号: 060435018001069607

E-mail: xzbw@chinajournal.net.cn, zwx301@mail.caas.net.cn