

空间效应模型在小麦品种比较试验中的应用

李建平, 宋喜芳, 胡希远

(西北农林科技大学农学院, 陕西杨陵 712100)

摘要: 为了提高作物品种比较试验数据分析的准确性,降低试验田内空间变异对试验结果的影响,基于 SAS 9.1 数据分析平台和 Proc mixed 过程步,对一个包含 15 个品种(系)的小麦品种比较试验的数据进行了空间效应模型和传统方差分析模型在数据拟合和统计分析等方面的对比分析。结果表明,空间效应模型在模型拟合效果、效应估计、效应误差估计和效应显著性方面都优于方差分析模型,田间试验存在的空间变异性可通过空间分析得以更好地控制。建议采用空间效应模型分析小麦品种田间试验数据,以提高试验数据分析的准确性。

关键词: 小麦;品种比较;空间效应模型

中图分类号: S512.1;S11⁺4

文献标识码: A

文章编号: 1009-1041(2008)03-0523-05

Application of Spatial Effect Model Analysis in Trials of Wheat Variety Comparison

LI Jian-ping, SONG Xi-fang, HU Xi-yuan

(Agriculture College, Northwest A & F University, Yangling, Shaanxi 712100, China)

Abstract: In order to enhance the analysis precision of field trials, and to reduce the influence of field spatial variability on the trial result, this paper compared the spatial effect model with the traditional variance analysis model in respect of data fitness, effect estimation and errors of effect estimate, based on analysis of a data set for a field trial with 15 wheat varieties using Proc mixed of SAS 9.1. The results showed that the spatial effect model was superior to the variance analysis model in data fitting, effect estimation and effect estimate errors. The spatial effect model could control field spatial variation better than the variance analysis model did. The spatial effect model was suggested to analyze field trials for crop varieties.

Key words: Wheat; Spatial analysis; Analysis model; Varieties contrast

田间试验的空间变异性,即相距较近的观测值相对于相距较远的观测值具有较大的相关性,常常会由田间地质地貌、营养成分、土壤理化性质等多种因素所引起^[1]。这将会极大地影响田间试验分析结果的可靠性。对于田间试验的空间异质性,人们常常用两种方法予以解决:1)采用试验区组化处理。区组内试验条件尽量一致,区组间不相关,使得区组间的空间变异性在分析时从随机误差中分离出来,以校正其对效应估计的影响。

然而要保证区组内试验条件的均匀性,在现实中很难达到,这在处理较多的作物品种比较试验中尤为常见。此外,相应于区组化设计采用的方差分析模型,其统计模型以假设试验单元在区组间不相关,在区组内协方差相同,而实际试验条件下试验单元在区组内和区组间都可能存在连续变化的现象。因此方差分析在处理空间变异方面受到很大的限制。2)采用空间效应模型分析试验数据。此方法为近年来依据地理统计学处理空间变

* 收稿日期:2007-11-20

修回日期:2008-03-12

基金项目:国家自然科学基金项目(30571072);教育部留学回归人员基金项目(2004)。

作者简介:李建平(1982-),男,硕士研究生,主要从事生物统计应用研究。E-mail:121396116@163.com

通讯作者:胡希远(1963-),男,博士,副教授,主要从事生物统计应用研究。E-mail:xiyuanhu@yahoo.com.cn

异性原理发展而来,可依据试验数据的空间位置信息,通过空间协方差结构灵活反映试验单元空间变异的特性,将其变异特征与混合线性模型分析相结合,用于试验误差的校正,是近些年来国际上田间品种比较试验分析方法研究中的热点之一^[2]。目前,国内很少有人将空间效应模型用于小麦品种比较分析中。

本研究基于标准化统计分析软件 SAS,利用 Proc Mixed 过程步对小麦品种比较田间试验的数据进行空间效应模型分析和传统的方差分析,比较它们在数据拟合效果、效应估计、效应误差估计和效应显著性检验等方面的差异状况,从而确定空间效应模型相对于方差分析模型在作物品种田间试验分析中的准确性和控制田间空间变异性的效果,为其实际应用提供依据。

1 材料与方 法

1.1 分析模型

分析模型采用方差分析模型和空间效应模型。

方差分析模型表示为:

$$y_{ij} = \mu + g_i + bl_j + e_{ij} \tag{1}$$

空间效应模型表示为:

$$y_{ij} = \mu + g_i + e_{ij} \tag{2}$$

模型(2)中各部分的意义与模型(1)相同,但剩余误差的分布为 $e_{ij} \sim N[0, C(h)]$, $C(h)$ 代表距离为 h 的两小区剩余误差的方差-协方差。若以 $e_{(s)}$ 和 $e_{(s+k)}$ 分别表示某小区和与该小区距离为 h 的小区的剩余误差,依据空间随机场理论^[3,4], $C(h)$ 被定义为:

$$C(h) = \begin{cases} \sigma_0^2 + \sigma^2, & \text{当 } h = 0; \\ \sigma_{e_s, e_{(s+h)}} = \sigma^2[f(h)], & \text{当 } h > 0 \end{cases}$$

其中, σ_0^2 和 σ^2 分别表示随机性和结构性误差方差组分,两者之和构成剩余误差的方差; $\sigma^2[f(h)]$ 是剩余误差间的协方差,表明两个剩余误差的协方差是其所在小区间距离的函数。不同的田间空间变异表现在 $f(h)$ 的形式(模型)及其参数不同。农业田间试验中经常遇到且在 SAS 软件中可供使用的 $f(h)$ 形式,有指数模型、球状模型、高斯模型和线性模型等,有关这些模型的形式可参考文献[2]和[3]。由于区组设置本身就是对空间变异的一种特定形式的处理,模型(2)通过剩余误差间的协方差考虑了空间变异的因素,所以该模型中不再单独出现区组效应组分。

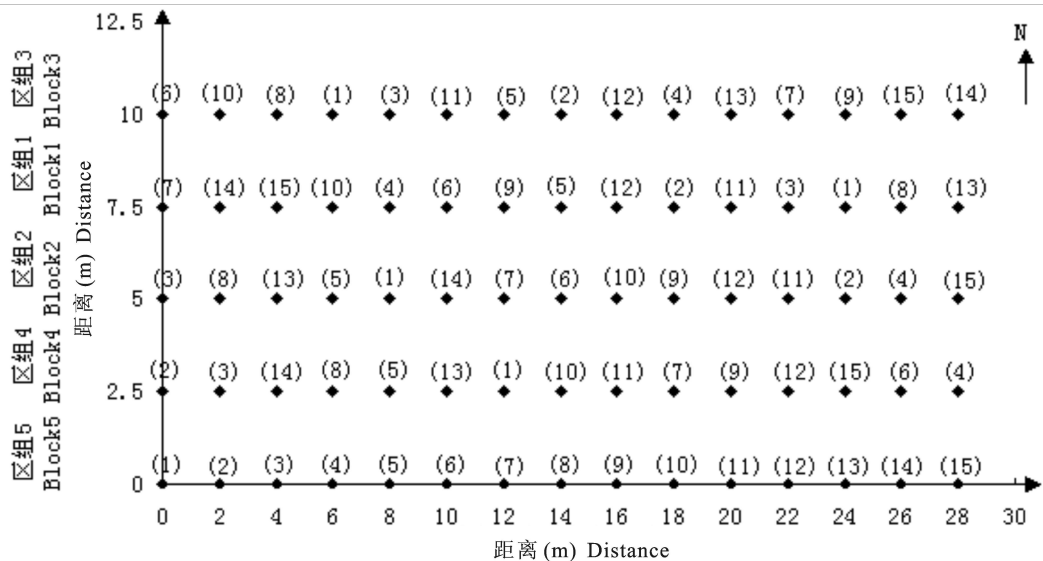


图 1 小麦品种(系)试验设计示意图(图中括号内数字为小麦品种编号,方形点为小区中心位置的坐标)

Fig. 1 A layout of the wheat experiment(the number in parentheses is variety code. the square dot denotes coordinate of plot center)

1.2 试验设计

为了展示应用 SAS 进行空间效应模型分析的方法,和其分析结果与传统方差分析结果的差异,本文分析的试验数据为小麦品种(系)产量构

成要素的千粒重(g)。试验于 2006~2007 年在西北农林科技大学试验农场进行,共有 15 个小麦品种(系),依次为:陕农 78、西农 979、陕 512、陕 698、远丰 139、h:80:33-88-3-3、普冰 151:hu:78、普冰

201、hu-06-3号、小偃22、97121F2-31-2-4-6、h:620-2、hu-06-10、小偃216、小偃6号,分别用V1、V2、V3.....V15表示。采用随机完全区组设计,重复5次。小区面积为 $2.5\text{ m} \times 2\text{ m} = 5\text{ m}^2$,每小区种植10行,行长2.5 m,行距0.2 m,株距0.025 m,采用人工点播。为了进行空间统计分析,需要计算试验小区的距离,因此数据中包含有小区中心点在田间的位置坐标信息。本文以x和y轴分别表示正东和正北方向。以区组5中第1小区中心点为坐标原点(0,0),区组5中第2小区中心点坐标为(2,0),区组4中第1小区为(0,2.5),依此类推,整个试验布局和小区位置编码如图1所示。

1.3 分析程序

对于以上试验数据,若以variety表示品种(系)处理,以weight表示小区品系千粒重,block表示区组变量,小区坐标变量为x和y,利用这些变量及其对应的数据建立SAS数据集,则空间效应模型分析的SAS程序^[3-5]为:

```
Proc mixed method = REML ;class variety ;
model weight = variety / ddfm ;repeated / sub-
ject = intercept local type = sp (exp) ( x y ) ;
lsmeans variety / pdiff ;run ;
```

Proc语句中method = REML定义方差-协方差参数估计方法为限制性极大似然法。Class语句指明试验处理因素变量为variety。语句Model中编码ddf m = Satterthwaite指出自由度估计方法为Satterthwaite。Repeated语句中编码subject = intercept表示观测值间存在空间变异。编码type(exp)定义空间变异模型为指数模型。编码(x y)定义数据空间位置坐标变量的名称为

x和y。type后括号内的编码为空间变异模型名称的缩写,可灵活改变。若要采用球状模型、高斯模型或线性模型描述空间变异特性,则该编码可分别变为type(sph)、type(gau)或type(lin)。常规方差分析的SAS程序见有关文献^[3]。

一般情况下,具有较多参数的模型比具有较少参数的模型拟合效果好。因此,在比较模型拟合效果时应考虑参数个数对数据拟合的影响。满足此要求的常用模型拟合指标有AIC(Akaike's Information Criterion)、AICC(Akaike's Information Criterion Corrected)和BIC(Schwarz's Bayesian Criterion)等。它们都是以数据拟合极大似然负对数值的2倍(-2lnL)减去参数个数的特定函数来计算其指标值^[2,6]。AIC、AICC和BIC值越小,表明相应的模型对数据拟合效果越好。

2 结果与分析

从计算结果(表1)看,三个信息量指标在空间效应模型分析中均小于在方差分析模型(信息量指标越小表明模型对数据的拟合度越高),表明试验数据存在一定的空间变异,用空间效应模型对试验数据分析很有必要。线性空间效应模型参数寻优时未能达到收敛,因而信息量指标无结果,这表明线性空间模型不能描述该试验数据空间变异的特征。其他三个空间效应模型中以指数模型信息量指标值最小,表明指数模型能更准确的描述试验数据空间变异的特征。因此下面用空间效应模型中的指数模型和方差分析模型进行统计分析结果的比较。

表1 不同模型拟合小麦品种(系)试验千粒重数据的信息量指标值

Table 1 Information criteria value from different models

信息量指标 Infromatin Criteria	方差分析模型 ANOVA	空间效应模型 Spatial model			
		指数模型 Exponential model	高斯模型 Gaussian model	球状模型 Spherical model	线性模型 Linear model
AIC	316.8	314.2	314.4	314.6	-
AICC	317.1	314.6	314.8	315.0	-
BIC	321.1	320.5	320.6	320.9	-

" - "表示拟合过程不收敛而无结果。

" - " denotes no results because of convergence problem.

在方差分析中所有品种(系)效应估计的误差均为1.584,而空间模型分析中所有品种(系)效应估计的误差平均为1.519。后者较前者的误差小,说明后者较前者分析效率有所提高。为了体现两种模型方法分析效率的差异,表2给出了两

种模型分析方法F检验和部分品种(系)效应差异t检验的结果。F检验结果在方差分析法和空间效应模型分析法中均达到极显著水平(P < 0.01)。但概率值(P = 0.0007)在空间效应模型分析中小于方差分析中的概率(P = 0.0029),第

二自由度也明显不同。说明在 0.01 水平时品种(系)千粒重差异在两种分析法中都能检验出来。也就是说两种模型分析法对该试验品系千粒重效应差异 F 检验在 0.01 概率水平的结论虽然相同,但其测验效率并不相同,只是因为试验处理效应差异很大,即使两模型分析法 F 测验效率不同也都能将该差异在 0.01 的显著水平鉴别出来。在进一步对所选的 7 对品种(系)千粒重效应比较的 t 检验中(表 2 所示),没有一对品系的千粒重

差异在方差分析法中达到显著水平,而空间模型分析法中则有二对品种(系)效应的差异(V1V3 和 V4V9)达到显著水平。这表明,在品种(系)间差异性检验方面,方差分析和空间效应模型分析存在着明显的差异。而空间效应模型分析法的效应估计误差小于方差分析法,说明空间分析法可在一定程度上控制由于空间异质性对试验分析结果的影响。

表 2 应用方差分析和空间效应模型分析对小麦品种(系)千粒重 F 检验和部分 t 检验的结果

Table 2 F and t test for thousand grain weight of some wheat varieties (lines) using ANOVA and Spatial effect model, respectively

项目 Item	第一自由度 NDR	方差分析 ANOVA			空间效应指数模型 Exponential model		
		第二自由度 DDF	F	P	第二自由度 DDF	F	P
所有品种(系) Total varieties (lines)	14	56	2.83	0.0029	46.5	3.45	0.0007
品种(系)对比 Contrast between varieties (lines)			t	P		t	P
V1 V3	1	56	- 1.98	0.0524	48	- 2.10	0.0411
V3 V5	1	56	1.21	0.2306	54.6	1.49	0.1424
V4 V9	1	56	- 1.86	0.0687	52.5	- 2.38	0.0207
V7 V10	1	56	- 1.76	0.0847	49.7	- 1.64	0.1078
V9 V12	1	56	- 0.97	0.3351	50.2	- 1.32	0.1922
V11 V15	1	56	- 0.15	0.8801	51.8	0.72	0.4774
V13 V14	1	56	- 1.12	0.2659	44.2	- 0.86	0.3945

V1、V2...V15 分别代表陕农 78、西农 979、陕 512、陕 698、远丰 139、h:80:33-88-3-3、普冰 151:hu:78、普冰 201、hu-06-3 号、小偃 22、97121F2-31-2-4-6、h:620-2、hu-06-10、小偃 216、小偃 6 号等 15 个小麦品种(系)。下表同。

V1, V2...V15 refer to the 15 varieties of Shaannong78, Xinong 979, Shaan 512, Shaan698, Yuanfeng139, h:80:33-88-3-3, Pubing151, hu:78, Pubing 201, hu-06-3, Xiaoyan 22, 97121F2-31-2-4-6, h:620-2, hu-06-10, Xiaoyan 216, Xiaoyan 6, respectively.

表 3 应用方差分析和空间效应模型分析对 15 个品种(系)千粒重的估计值和大小排序

Table 3 Estimate and rank of thousand grain weight for 15 wheat varieties(lines) based on ANOVA and spatial model analysis

品种(系) Variety (lines)	方差分析 ANOVA		空间效应指数模型 Exponential model	
	千粒重估计值 Weight estimate	位次 Rank	千粒重估计值 Weight estimate	位次 Rank
V10	45.22	1	45.48	1
V14	44.16	2	44.24	3
V15	43.86	3	43.80	4
V11	43.62	4	44.93	2
V3	43.06	5	43.42	6
V12	42.48	6	43.80	5
V7	42.44	7	43.09	7
V13	42.38	8	43.02	8
V8	41.98	9	41.57	10
V5	41.14	10	41.09	11
V9	40.94	11	41.83	9
V6	40.52	12	40.44	13
V2	40.40	13	41.06	12
V1	39.92	14	40.35	14
V4	38.00	15	39.10	15

表 3 为方差分析和空间效应模型分析对所试验 15 个品种(系)千粒重的估计值及其大小排序结果。可以看出两种分析方法的差异在千粒重估计值结果和大小排序结果都存在着明显的差异。

这些差异显然是由于两种分析方法对试验误差控制效果不同所造成的。按照千粒重的高低对品系进行选择的话,从 15 个品种(系)中选择出 20% 的优良品种(系),即应选 3 个品种(系)。若用方

差分析法则所选出的 3 个品种(系)为 V10、V14 和 V15, 而用空间效应模型分析法选择的 3 个品种(系)为 V10、V14 和 V11。空间效应模型分析中高千粒重的 3 个品种(系)有 1 个品种(系)(V11)被方差分析所排除。V9 在方差分析中位居第 11 位, 而在空间效应模型分析中位居第 9 位。这些差异说明, 不同的分析方法会导致不同的品种评价和选择结果。为了保障试验分析的精确度和提高品种选择的可靠性, 应选用适宜的方法进行试验数据分析。

3 讨论

本研究结果表明, 空间效应模型有较好的数据拟合效果和较高的分析率, 此分析结果与文献^[7~9]中的随机模拟方法的研究结果相一致。模拟研究^[7~9]表明: 方差分析模型法和空间效应模型法在品种比较误差估计值相对于经验误差值的偏差不同, 随着空间变异的增加, 方差分析模型分析田间试验数据的误差偏差增大, 分析准确性趋于降低, 而空间效应模型的误差偏差始终保持在很低的水平, 分析准确率一直较高。

由空间效应模型分析过程可知, 试验数据的空间协方差直接关系到品种(系)的效应估计误差、品种(系)的产量性状优劣排序和品系的差异性检验。空间效应分析的关键是准确地确定试验数据空间协方差的结构模型及其参数, 充分地挖掘试验数据的信息, 这和 Antti Taskinen 的研究结果相一致^[10]。采用空间效应模型分析试验数据, 可以降低试验设计的难度, 不用测定田间试验区组内土壤的性质(如, pH、水分和有机质含量等)来进行区组设计, 以保证区组内试验条件相同或尽量相同, 以及区组间不相关的条件。因为空间分析过程中已经考虑到试验条件的空间异质性而且给予校正。虽然空间分析法可以使试验设计简单化, 减少工作量, 但并不能说就不需要进行试验设计。根据 Stroup 和胡希远的模拟研究^[11, 12], 在无随机区组设计时, 空间效应模型的分析效率会降低, 因为空间相关模型能在一定程度上降低, 而非完全消除空间变异对效应值估计的影响, 所以区组化设计还是有必要的。

空间效应模型的应用可利用 SAS 统计分析软件中的 Proc mixed 来完成, 其运行程序本文已经给出, 只要对小区的空间位置的坐标进行编码,

就可很容易进行试验数据的空间模型分析。模型拟合度的优劣可由信息量指标(AIC、BIC 和 AICC)和效应估计误差来判定。这些信息也在 Proc mixed 的运行结果中自动给出, 不需要试验分析者进行复杂的计算, 应用方便。分析者可依据信息指标值越小其对试验数据模拟效果越好的原理, 在处理试验数据时对数据进行最佳分析模型的选择, 从而保证数据分析的准确性和结论的正确性。另外, 空间效应模型法也可用在品种(系)的优系选择、品种(系)产量的比较等方面, 能够更精确分离出品种(系)间特征差异和由于肥水等条件的不同而引起的品种(系)特征差异^[2, 12]。

参考文献:

- [1] Stroup W W. Power analysis based on spatial effects mixed models: a tool for comparing design and analysis strategies in the presence of spatial variability [J]. *Journal of Agricultural Biological and Environmental Statistics*, 2002, 7: 491 - 498.
- [2] 胡希远, Joachim SPIL KE. 田间试验的空间变异性及其统计控制[J]. *作物学报*, 2007, 33(4): 620 - 624.
- [3] 胡希远. SAS 与统计分析[M]. 陕西杨凌: 西北农林科技大学出版社, 2007.
- [4] Piepho H P. Analysis of randomized block design with unequal subclass numbers[J]. *Agronomy Journal*, 1997, 89: 718 - 723.
- [5] 张群远, 孔繁玲. 作物品种区域试验统计分析模型的比较[J]. *中国农业科学*, 2002, 35(4): 365 - 371.
- [6] 魏峰远, 郭继发, 李卫贤. 基于 AIC 准则的回归方法在建筑物变形分析中的应用[J]. *工程勘察*, 2007(7): 45 - 48.
- [7] Cullis B R, Gleeson A C. Spatial analysis of field experiments an extension of two dimensions [J]. *Biometrics*, 1991, 47: 1449 - 1460.
- [8] Wu T, Dutilleul P. Validity and efficiency of neighbor analyses in comparison with classical complete and incomplete block analysis of field experiments [J]. *Agronomy Journal*, 1999, 91: 721 - 731.
- [9] 胡希远. 数据非平衡性对试验分析结果的影响[J]. *西北农林科技大学学报(自然科学版)*, 2004, 32(1): 103 - 107.
- [10] Antti Taskinen. Modelling effects of spatial variability of saturated hydraulic conductivity on autocorrelated overland flow data: linear mixed model approach [J]. *Stoch Environ Res Ris Assess*, 2007, 4: 77 - 99.
- [11] 胡希远, 高金锋, 刘建军. 空间效应模型分析田间试验的方法与效果[J]. *西北农林科技大学学报(自然科学版)*, 2007, 35(3): 87 - 92.
- [12] Kravchenko A N, Robertson G P, Snap S S, et al. Using information about spatial variability to improve estimates of total soil carbon [J]. *Agronomy Journal*, 2006, 98: 823 - 829.