

广义遗传模型与数量遗传分析新方法*

朱 军

浙江农业大学生物数学研究中心, 杭州 310029

摘要 根据广义遗传模型的建模原理, 提出了一些扩展的广义遗传模型。这些模型可以分析种子数量性状, 也可以分析基因型与环境的互作效应。采用混合线性模型的统计分析方法, 可以有效地估算方差分量和协方差分量、预测遗传效应值。介绍了我们提出的一些遗传模型及相应的统计分析方法。

关键词 遗传模型; 基因型与环境互作; 方差和协方差分量估算方法; 遗传效应值预测方法;

General genetic models and new analysis methods for quantitative traits

(Research Center of Biomathematics, Zhejiang Agricultural University, Hangzhou 310029)

Abstract Based on the theory of general genetic model, several extended models were proposed for analyzing seed quantitative traits or for genotype by environment interaction. Mixed model approaches are applicable for estimating variance and covariance components, and for predicting genetic effects. Some new genetic models and their analysis methods were also presented.

Key words genetic models; genotype by environment interaction; methods of estimating variances and covariances; methods of predicting gene effects.

自 Fisher 提出方差分析方法(ANOVA 方法)^[1]以来, 数量遗传学家运用方差分析的原理, 发展了许多实用的遗传模型。至今仍有不少遗传模型被遗传育种工作者广泛使用, 比如北卡罗来纳设计 I 和 II (NC I 和 NC II)^[2,3]、双列杂交遗传模型^[4,5,6]。方差分析方法在数量遗传研究中的广泛运用, 曾经极大地推动了数量遗传学的发展。但是 ANOVA 方法存在一些固有的局限性, 它不能无偏分析有不规则缺失的非平衡数据、也不能分析系数出现分数的线性模型。因此, 数量遗传分析对 ANOVA 方法的依赖性, 成为阻碍数量遗传进一步发展的制约因素。

数量遗传学家根据生物体的实际遗传规律, 曾提出了一些具有生物学意义的遗传模型, 以突破简单的加性—显性遗传假设的限制。Eisen 等提出了一个分析动物 F₁ 双列杂交的遗传模型^[7], 包括加性、显性、母体效应以及性连锁效应。Cockerham 和 Weir 提出了双列杂交的生物模型(Bio-model)^[8], 把 F₁ 的遗传效应值分解为加性、显性、母体、父体等四项分量。这些模型虽然发展了 Griffing 的配合力模型^[5], 但无法用 ANOVA 方法分析。

七十年代初统计学家创立了一套崭新的统计分析方法—混合线性模型分析方法(Mixed Linear Model Approaches)。这些统计分析方法克服了 ANOVA 方法的局限性, 不但可以无偏分析有不规则缺失的非平衡数据, 而且还能分析各种复杂的遗传模型。统计分析方法的突破极大地推动了数量遗传学科的新发展。朱军用蒙特卡罗模拟证明^[9], 运用混合线性模型分析方法可以无偏地估算 Eisen 等人的动物双列杂交遗传模型^[7]以及

Cockerham 和 Weir 的双列杂交生物模型[8]中的各项遗传方差分量。1980 年 Cockerham 提出了广义遗传模型的建模原理^[10]，为数量遗传学家建立各种复杂的遗传模型奠定了理论基础。

近年来我们结合我国的育种实践，提出了一些新的遗传模型和新的统计分析方法，推动了数量遗传学科的进一步发展。本文简要介绍 Cockerham 提出的广义遗传模型原理^[10]以及我们对原模型的扩展，还介绍了相应的统计分析方法；概述我们近年来提出的数量遗传分析的一些新方法。

1. 广义遗传模型

当遗传试验在一个环境条件下实施，某遗传群体的性状平均观察值 y 可以由以下遗传模型表示，

$$y = \mu + G + e$$

其中 μ 是群体平均数， G 是基因的遗传效应值， e 是环境机误。

由一组随机的纯合亲本衍生的各世代个体的遗传效应值(G)都可以用一个广义遗传模型分解其不同的基因效应^[10]。如果一个遗传群体的遗传变异归因于若干不同的纯合亲本来源的基因效应，并且遗传效应只包括基因的加性效应和显性效应，由这个遗传群体衍生的任何个体的平均基因型值 G 可由以下式子表示，

$$G = \sum_i \alpha_i A_i + \sum_i \sum_{j \geq i} \delta_{ij} D_{ij} \quad (1)$$

其中 α_i 是加性效应值 A_i 的系数值 ($\sum_i \alpha_i = 2$)，

δ_{ij} 是显性效应值 D_{ij} 的系数值 ($\sum_i \sum_{j \geq i} \delta_{ij} = 1$)。

根据这一广义遗传模型原理，我们提出了利用亲本和 F_1 不分离世代的双列杂交试验资料预测 F_2 基因型值的分析方法^[11, 12]。对于由纯合亲本 i 与 j 交配的后代($i=j$ 时为亲本， $i \neq j$ 时为杂交种)，可由以下遗传模型作双列分析，

$$y_{ijk} = \mu + G_{ij} + B_k + e_{ijk}$$

其中 B_k 是区组效应。

亲本($i=j$)和杂种 $F_1(i \neq j)$ 的总遗传效应值(G_{ij}) 由以下加性效应和显性效应遗传分量组成，

$$G_{ij} = A_i + A_j + D_{ij}$$

经遗传分析获得加性、显性遗传效应值的无偏预测后，便可用下式预测杂交组合($i \times j$)的 F_2 基因型值，

$$y_{ijk} = \mu + A_i + A_j + \frac{1}{4} D_{ii} + \frac{1}{4} D_{jj} + \frac{1}{2} D_{ij} + B_k + e_{ijk}$$

在棉花品种间杂种优势遗传分析中，我们已采用这一分析方法评价了 F_2 组合的利用价值^[13]。

对一些难以获得 F_1 杂交种子的禾谷类作物(如小麦、大麦等)，也可采用 F_2 模型分析双列杂交的 F_2 世代平均数资料，然后预测 F_1 的基因效应值。

对于某些数量性状，遗传效应分量除了有加性效应和显性效应以外，还可能包括上位性效应(Epistatic Effect)。Cockerham 把上位性效应分解为加性与加性的上位性效应 (AA)、

*本研究由国家教委《跨世纪优秀人才计划》基金资助。

加性与显性的上位性效应(AD)和显性与显性的性上位效应(DD)^[14]。因而总遗传效应值可以进一步扩展,包括加×加上位、加×显上位和显×显上位三种上位性效应^[10]。由于加×显上位性和显×显上位性比较复杂,而且在动植物遗传分析中一般均比其它效应弱,因此常可忽略不计。加×加上位性是可被选择固定的遗传效应,常有必要与加性和显性效应一齐分析。包括加性、显性和加×加上位性效应的遗传分量为^[10],

$$G = \sum_i \alpha_i A_i + \sum_i \sum_{j \geq i} \delta_{ij} D_{ij} + (\sum_i \alpha_i A_i)^2 \quad (2)$$

总遗传效应 加性 显性 加×加上位性

根据广义遗传模型(2),朱军提出了利用双列杂交的三个世代(亲本、F¹和F²),分析加性、显性和加×加上位性效应的遗传模型^[15]。遗传效应G_{ij}的分解,因组合及世代而异。

$$\text{亲本和杂种一代: } G_{ij} = (A_i + A_j) + D_{ij} + (AA_{ii} + AA_{jj} + 2AA_{ij})$$

$$\text{交种二代: } G_{ij} = (A_i + A_j) + \frac{1}{4}D_{ii} + \frac{1}{4}D_{jj} + \frac{1}{2}D_{ij} + (AA_{ii} + AA_{jj} + 2AA_{ij})$$

徐绍英等运用加性-显性-加×加上位性遗传模型,分析了大麦7个亲本双列杂交的三个熟期性状^[16]。

农作物品质育种的深入发展,促进了种子数量性状的遗传研究。种子的营养物质由母体植株所提供,因此种子的数量性状表现可能会同时受到种子核基因和母体植株核基因两套遗传体系的控制。另外细胞质基因也可能通过控制叶绿体(或线粒体)的合成而影响植株的光合(或呼吸)作用,从而间接控制胚乳性状的表现。虽然Cockerham和Weir曾提出了可以分析加性、显性、母体和父体四项遗传效应的生物模型(Bio-model)[8]。但是在植物遗传中一般并不存在父体效应,这时采用生物模型估算遗传方差分量会导致母体方差的有偏估算^[9]。朱军等已提出了一个包括加性、显性和母体效应的ADM模型^[12],适用于二倍体种子或幼苗性状的分析。对于一组亲本及其F₁植株自交所结的P和F₂种子的三倍体胚乳性状,可采用以下的ADM模型分析直接加性、直接显性和母体效应的遗传表现:

$$\text{亲本 } P_i \text{ 遗传效应分量} = 3A_i + 3D_{ii} + M_{ii}$$

$$\text{杂种二代 } F_{2ij} \text{ 遗传效应分量} = 1.5A_i + 1.5A_j + D_{ii} + D_{jj} + D_{ij} + M_{ij}$$

其中A_i或A_j是胚乳核基因的直接加性效应,D_{ii}、D_{jj}、D_{ij}是胚乳核基因的直接显性效应,M_{ii}或M_{ij}是母体植株基因效应。

采用ADM模型分析种子性状,只需分析亲本及其F₁植株上的自交种子,当代不必做杂交,但是母体效应M_{ii}或M_{ij}可能包括母体植株的核基因效应和细胞质基因效应。因此尚不能区分母体和细胞质遗传效应。朱军和Weir进一步扩展了Cockerham的广义遗传模型^[10],提出了包括种子、母体和细胞质遗传效应的广义遗传模型^[17],

$$G = G_O + G_C + G_M \quad (3)$$

上式中的三项遗传效应可进一步分解为,

$$\text{直接遗传效应 } G_O = \sum_i \tau_i A_i + \sum_i \sum_{j \geq i} \delta_{ij} D_{ij}$$

其中A_i是种子核基因的直接加性效应,D_{ij}是种子核基因的直接显性效应;

$$\text{细胞质遗传效应 } G_C = \sum_i \gamma_i C_i$$

其中C_i是细胞质基因效应;

$$\text{母体遗传效应 } G_M = \sum_i \tau_{m_i} A_{m_i} + \sum_i \sum_{j \geq i} \delta_{m_{ij}} D_{m_{ij}}$$

其中 A_{m_i} 是母体植株核基因的母体加性效应, $D_{m_{ij}}$ 是母体植株核基因的母体显性效应。

种子性状的广义遗传模型, 为发展具有生物学意义的双子叶作物(大豆、油菜、棉花等)二倍体种子模型^[15,17]和单子叶作物(水稻、小麦、玉米等)三倍体胚乳模型^[15,18,19]奠定了基础。

作物二倍体种子模型可以包括三个遗传世代, 各组合世代遗传效应的分解为,

$$\text{亲本 } P_i \text{ 遗传效应分量} = 2A_i + D_{ii} + C_i + 2Am_i + Dm_{ii}$$

$$\text{杂种一代 } F_{1ij} \text{ 遗传效应分量} = A_i + A_j + D_{ij} + C_i + 2Am_i + Dm_{ii}$$

杂种二代 F_{2ij} 遗传效应分量

$$= A_i + A_j + \frac{1}{4}D_{ii} + \frac{1}{4}D_{jj} + \frac{1}{2}D_{ij} + C_i + Am_i + Am_j + Dm_{ij}$$

也可以采用三个遗传世代分析作物三倍体胚乳性状的遗传表现, 各组合世代遗传效应的分解为,

$$\text{亲本 } P_i \text{ 遗传效应分量} = 3A_i + 3D_{ii} + C_i + 2Am_i + Dm_{ii}$$

$$\text{杂种一代 } F_{1ij} \text{ 遗传效应分量} = 2A_i + A_j + D_{ii} + 2D_{ij} + C_i + 2Am_i + Dm_{ii}$$

$$\text{杂种二代 } F_{2ij} \text{ 遗传效应分量} = \frac{3}{2}A_i + \frac{3}{2}A_j + D_{ii} + D_{jj} + D_{ij} + C_i + Am_i + Am_j + Dm_{ij}$$

采用三倍体胚乳性状的遗传模型, Yu 等分析了 9 个大麦品种部分双列杂交的蛋白质等性状^[20], 石春海和朱军分析了 6 个籼型不育系与 3 个恢复系杂交组合的碾磨、外观和蒸煮等品质性状^[21,22,23]。

控制动物或昆虫数量性状的遗传效应除了有常染色体的基因加、显性效应和母体效应以外, 还可能包括 X(或 Z)性染色体的基因效应。Eisen 等提出的动物双列杂交遗传模型^[7]

虽然可以分析加性、显性、母体效应以及性连锁效应, 其应用受制于模型的二个隐含假设 (1) 哺乳动物雌性个体 XX 或鳞翅目昆虫雄性个体 ZZ 的二条性染色体都具有活性; (2) 所有组合的雌雄个体总体表现不存在遗传差异。当这两个遗传假设不成立时, 可以采用以下遗传模型分析亲本及其 F_1 雌雄个体的平均表现。母本 $i \times$ 父本 j 的异配性别后代(XY 或 ZW, $s = 1$) 和同配性别后代(XX 或 ZZ, $s = 2$)的基因型值 G_{ij_s} 定义为,

$$G_{ij1}^{XY} = A_i + A_j + D_{ij} + L_{i1} + M_i$$

$$\text{或 } G_{ij1}^{ZW} = A_i + A_j + D_{ij} + L_{j1} + M_i$$

$$G_{ij2}^{XX/ZZ} = A_i + A_j + D_{ij} + \frac{1}{2}L_{i2} + \frac{1}{2}L_{j2} + M_i$$

其中 A_i 或 A_j 是常染色体基因的加性效应, D_{ij} 是常染色体基因的显性效应, M_i 是母体基因效应, L_{i1} 或 L_{j1} 是异配性别的性连锁效应, L_{i2} 或 L_{j2} 是同配性别的性连锁效应。

2. 包括基因型与环境互作效应的广义遗传模型

以上所介绍的广义遗传模型及其扩展模型都基于遗传假设: 不存在基因型与环境的互作效应。由于生物体的基因是在环境条件作用下表达的, 环境条件的不同可能导致基因表达方式或程度的差异。已有许多研究表明, 不少数量性状存在基因型与环境的互作效应。如果互作确实存在, 则需要不同的环境条件下实施遗传实验, 并采用包括互作效应的遗传

模型，才能无偏地分析基因效应以及基因与环境的互作效应。

因此，广义遗传模型需要在原有的基因效应基础上扩展，进一步包括基因各项效应与环境的互作效应。当遗传试验在多个环境条件下实施时，某遗传群体在一种环境下的性状平均观察值 y 可以由以下遗传模型表示，

$$y = \mu + E + G + GE + e$$

其中 μ 是群体平均数， E 是环境效应， G 是基因的遗传效应值， GE 是基因与环境的互作效应值， e 是机误。

Cockerham 的广义遗传模型^[10]可扩展为 $G + GE$ ，基因遗传效应值 G 的定义如模型(1)所示，基因与环境互作效应值 GE 的定义为，

$$GE = \sum_h \sum_i \alpha_{hi} AE_{hi} + \sum_h \sum_i \sum_{j \geq i} \beta_{hij} DE_{hij} \quad (4)$$

其中 AE_{hi} 是加性×环境互作效应， DE_{hij} 是显性×环境互作效应。

根据加性-显性遗传模型，亲本、 F_1 和 F_2 在环境 h 中的基因型与环境互作效应为，

$$\text{亲本和杂种一代: } GE_{hij} = AE_{hi} + AE_{hj} + DE_{hij}$$

$$\text{交种二代: } GE_{hij} = AE_{hi} + AE_{hj} + \frac{1}{4}DE_{hii} + \frac{1}{4}DE_{hjj} + \frac{1}{2}DE_{hij}$$

当广义遗传模型(2)扩展为 $G + GE$ 时，基因遗传效应值 G 的定义如模型(2)所示，基因与环境互作效应值 GE 可分解为加性×环境互作(AE)、显性×环境互作(DE)和加×加×环境互作(AAE)等效应的遗传分量，

$$GE = \sum_h \sum_i \alpha_{hi} AE_{hi} + \sum_h \sum_i \sum_{j \geq i} \beta_{hij} DE_{hij} + \sum_h (\sum_i \alpha_{hi} AE_{hi})^2 \quad (5)$$

根据广义遗传模型(5)，可以利用双列杂交的三个世代(亲本、 F_1 和 F_2)在若干环境下的遗传实验资料，分析加性×环境、显性×环境和加×加×环境的各项遗传效应。亲本、 F_1 和 F_2 的遗传效应 G_{ij} 的分解已有定义^[15]，基因效应与环境互作效应 GE_{hij} 的分解因组合及世代而异，

$$\text{亲本和杂种一代: } GE_{ij} = (AE_{hi} + AE_{hj}) + DE_{hij} + (AAE_{hii} + AAE_{hjj} + 2AAE_{hij})$$

$$\text{交种二代: } GE_{hij}$$

$$= (AE_{hi} + AE_{hj}) + \frac{1}{4}DE_{hii} + \frac{1}{4}DE_{hjj} + \frac{1}{2}DE_{hij} + (AAE_{hii} + AAE_{hjj} + 2AAE_{hij})$$

对于分析二倍体种子性状的 ADM 模型，亲本和 F_2 种子在环境 h 中的基因型与环境互作效应为，

$$\text{亲本 } P_i: GE_{hij} = 3AE_{hi} + DE_{hii} + ME_{hii}$$

$$\text{杂种二代 } AE_{hi} + AE_{hj} + DE_{hii} + DE_{hjj} + DE_{hij} + ME_{hij}$$

其中 AE_{hi} 或 AE_{hj} 是胚核基因的直接加性×环境互作效应， DE_{hii} 、 DE_{hjj} 、 DE_{hij} 是胚核基因的直接显性×环境互作效应， ME_{hii} 或 ME_{hij} 是母体植株基因×环境互作效应。

在分析胚乳性状的 ADM 模型中，亲本和 F_2 种子在环境 h 中的基因型与环境互作效应定义为，

$$\text{亲本 } P_i: GE_{hij} = 3AE_{hi} + 3DE_{hii} + ME_{hii}$$

$$\text{杂种二代 } 1.5AE_{hi} + 1.5AE_{hj} + \frac{1}{4}DE_{hii} + \frac{1}{4}DE_{hjj} + \frac{1}{2}DE_{hij} + ME_{hij}$$

其中 AE_{hi} 或 AE_{hj} 是胚乳核基因的直接加性×环境互作效应, DE_{hii} 、 DE_{hij} 、 DE_{hji} 是胚乳核基因的直接显性×环境互作效应, ME_{hii} 或 ME_{hij} 是母体植株基因×环境互作效应。

如分析种子性状的广义遗传模型(3)作相应的扩展($G + GE$), 包括基因型与环境的互作效应 GE , 互作总效应 GE 是三项分量的组成,

$$GE = G_O E + G_C E + G_M E \quad (6)$$

上式中的三项互作效应可进一步分解为,

$$\text{直接互作效应 } G_O E = \sum_h \sum_i \alpha_{hi} AE_{hi} + \sum_h \sum_i \sum_{j \geq i} \beta_{hij} DE_{hij}$$

其中 AE_{hi} 是种子核基因的直接加性×环境互作效应, DE_{hij} 是种子核基因的直接显性×环境互作效应;

$$\text{细胞质互作效应 } G_C E = \sum_h \sum_i \lambda_{hi} CE_{hi}$$

其中 CE_{hi} 是细胞质基因×环境互作效应;

$$\text{母体互作效应 } G_M E = \sum_h \sum_i \alpha_{mhi} AmE_{hi} + \sum_h \sum_i \sum_{j \geq i} \beta_{m hij} DmE_{hij}$$

其中 AmE_{hi} 是母体植株核基因的母体加性×环境互作效应, DmE_{hij} 是母体植株核基因的母体显性×环境互作效应。

对于双子叶作物种子数量性状, 母本 i 与父本 j 的第 k 种交配类型在环境 h 内的基因与环境互作效应值 GE_{hijk} 的分解如下式所示。

亲本 P_i 互作效应分量 ($k = 0$)

$$GE_{hii0} = 2AE_{hi} + DE_{hii} + CE_{hi} + 2AmE_{hi} + DmE_{hii}$$

杂种一代 $F_{1ij}(P_i \times P_j)$ 互作效应分量 ($k = 1$)

$$GE_{hij1} = AE_{hi} + AE_{hj} + DE_{hij} + CE_{hi} + 2AmE_{hi} + DmE_{hii}$$

杂种二代 F_{2ij} 互作效应分量 ($k = 2$)

$$GE_{hij2} = AE_{hi} + AE_{hj} + \frac{1}{4}DE_{hii} + \frac{1}{4}DE_{hjj} + \frac{1}{2}DE_{hij} + CE_{hi} \\ + AmE_{hi} + AmE_{hj} + DmE_{hij}$$

而对于单子叶作物胚乳数量性状, 各遗传世代基因与环境互作效应值 GE_{hijk} 的分解如下式所示。

亲本 P_i 互作效应分量 ($k = 0$)

$$GE_{hij0} = 3AE_{hi} + 3DE_{hii} + CE_{hi} + 2AmE_{hi} + DmE_{hii}$$

杂种一代 $F_{1ij}(P_i \times P_j)$ 互作效应分量 ($k = 1$)

$$GE_{hij1} = 2AE_{hi} + AE_{hj} + DE_{hii} + 2DE_{hij} + CE_{hi} + 2AmE_{hi} + DmE_{hii}$$

杂种二代 F_{2ij} 互作效应分量 ($k = 2$)

$$GE_{hij2} = \frac{3}{2}AE_{hi} + \frac{3}{2}AE_{hj} + DE_{hii} + DE_{hjj} + DE_{hij} + CE_{hi} \\ + AmE_{hi} + AmE_{hj} + DmE_{hij}$$

根据广义遗传模型(6)所定义的各项遗传参数, 利用若干环境下遗传实验 (在田间种植亲本和 F_1 , 由其自交产生亲本种子和 F_2 种子, 亲本间杂交可获得 F_1 种子)的资料, 便

可有 效地分析控制种子性状的三套遗传体系的基因效应及其与环境的互作效应。

当动物遗传模型包括基因型与环境互作效应时,母本 $i \times$ 父本 j 的异配性别后代(XY 或 ZW, $s = 1$)和同配性别后代(XX 或 ZZ, $s = 2$)在环境 h 中的互作效应值 GE_{hij_s} 定义为,

$$GE_{hij1}^{XY} = AE_{hi} + AE_{hj} + DE_{hij} + LE_{hi1} + ME_{hi}$$

$$\text{或 } GE_{hij1}^{ZW} = AE_{hi} + AE_{hj} + DE_{hij} + LE_{hj1} + ME_{hi}$$

$$GE_{hij2}^{XX/ZZ} = AE_{hi} - AE_{hj} - DE_{hij} - \frac{1}{2}LE_{hi2} + \frac{1}{2}LE_{hj2} + ME_{hi}$$

其中 AE_{hi} 或 AE_{hj} 是常染色体基因的加性 \times 环境互作效应, DE_{hij} 是常染色体基因的显性 \times 环境互作效应, ME_{hi} 是母体基因 \times 环境互作效应, LE_{hi1} 或 LE_{hj1} 是异配性别的性连锁基因 \times 环境互作效应, LE_{hi2} 或 LE_{hj2} 是同配性别的性连锁基因 \times 环境互作效应。

3. 广义遗传模型的分析方法

在广义遗传模型中遗传效应可以定义为随机效应,也可以定义为固定效应。如果遗传材料是一组随机样本,试验目的主要是分析性状的遗传变异表现,则应将遗传效应定义为随机效应。这时可以估算各项随机效应的方差分量或协方差分量。如果试验材料是一组特定的遗传材料,试验目的主要是分析不同遗传材料的特定遗传表现,可将遗传效应定义为固定效应,从而可以估算各项固定效应的数值。

任何广义遗传模型都可以用一个混合线性模型(Mixed Linear Model)表示,

$$\mathbf{y} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 + \cdots + \mathbf{X}_n\mathbf{b}_n + \mathbf{U}_1\mathbf{e}_1 + \mathbf{U}_2\mathbf{e}_2 + \cdots + \mathbf{U}_m\mathbf{e}_m$$

$$= \mathbf{X}\mathbf{b} + \sum_{u=1}^m \mathbf{U}_u\mathbf{e}_u$$

其中 \mathbf{b} 是固定效应向量, \mathbf{X} 是固定效应的系数矩阵; \mathbf{e}_u 是第 u 项独立随机变量向量,具有平均数零、方差 $\sigma_u^2 \mathbf{I}$, \mathbf{U}_u 是第 u 项随机效应的系数矩阵。

采用不同的统计分析方法,可以估算混合线性模型中的方差分量。七十年代以来发展的一系列混合线性模型分析方法,包括最大似然法(Maximum Likelihood,简称 ML 法)、限制性最大似然法(Restricted Maximum Likelihood,简称 REML 法)、最小范数二阶无偏估算法(Minimum Norm Quadratic Unbiased Estimation,简称 MINQUE 法)等。在分析混合线性模型的各种方法中,Rao 提出的 MINQUE 法^[24]比 ML 法和 REML 法等更为简便和优越。MINQUE 法不需要进行迭代运算,对线性模型也没有正态分布的限定。

在运用 MINQUE 法时,如设定所有的先验值为常数 1,这种方法称为 MINQUE(1)法,所获得的方差分量的估计量是 MINQUE(1)估计量。朱军提出了估算成对性状之间遗传协方差分量的 MINQUE(1)算法^[15]。

种子性状的遗传模型包括了三套遗传体系,比形态性状的遗传模型更为复杂。由于种子核基因部分来自其母体植株,因而直接遗传效应与母体遗传效应不是相互独立的。直接加性效应与母体加性效应、直接显性效应与母体显性效应可能存在协方差。采用朱军提出的 MINQUE(0/1)法^[15],即设先验协方差为 0,先验方差为 1 的一种 MINQUE 法,可以无偏地估算以上各项方差和协方差分量。对于成对性状,也可以采用 MINQUE(0/1)法无偏地估算各项遗传协方差分量^[15]。

研究种子营养品质性状的遗传表现与植株形态性状的遗传表现的相关性,对于提高种子品质的育种改良和栽培管理均具有重要的指导意义。但是,种子和植株具有不同的遗传模型,它们的遗传交配设计的矩阵是不等的。在混合线性模型研究领域,国际上尚未提出不

等矩阵成对性状协方差分量估算的实用统计分析方法。朱军在这一研究领域已取得突破，首次提出了估算不同性状遗传协方差分量的新方法^[25]。运用该方法。可以无偏估算具有不同遗传模型(如种子和植株)性状之间的遗传协方差分量。

遗传模型中的各项遗传效应通常是不可估计的，但是采用混合线性模型的一些分析方法，模型中的随机效应却是可预测的。Henderson 提出^[26]，如果随机因素的方差分量 σ_u^2 是已知的，可以获得随机效应的最佳线性无偏预测(Best Linear Unbiased Prediction, 简称 BLUP)。在运用混合线性模型分析具体试验数据时，方差分量 σ_u^2 是需要被估算的。因而实际上并不可能获得随机效应值的最佳线性无偏预测。在实际应用时，常采用方差分量的估计值代替其参数。但是这时的预测值既不是依变量的线性函数，也不能保证是无偏的。

朱军建议采用运算简便的 MINQUE 法，预测随机效应值^[15]。如果先验值的选择取决于所分析的数据，该预测值是观察值向量 y 的线性函数，也是随机效应的无偏预测值。所以这一方法称为线性无偏预测法(Linear Unbiased Prediction, 简称 LUP)。采用 LUP 法虽然可以获得随机效应的无偏预测值，但这些预测值的变异幅度往往不够大。朱军进一步发展了调整无偏预测(Adjusted Unbiased Prediction, 简称 AUP)的分析方法^[11]，可以使随机效应预测值的均值和方差都具有无偏性。LUP 法^[15]和 AUP 法^[19]也同样适用于预测种子数量性状的遗传效应值。

5. 结语

方差分析统计分析方法的不断完善和线性模型在数量遗传分析中的运用，曾极大地推动了数量遗传学的发展。五十年代至七十年代，各国遗传育种学家广泛地研究了动、植物的一些数量性状。但是这些研究大多基于简单的遗传模型。由于传统的数量遗传分析方法不能分析具有不规则缺失的非平衡数据，也无法分析具有生物学意义的复杂遗传模型，因此制约了数量遗传学的进一步发展。

七十年代初诞生的混合线性模型统计分析方法，可以克服传统的方差分析方法的局限性，开辟了分析非平衡数据和复杂线性模型的新途径。八十年代初 Cockerham 提出的广义遗传模型原理^[10]为数量遗传学家发展各种复杂的新遗传模型奠定了遗传学基础。现在数量遗传发展的两大制约因素已不复存在。数量遗传学家有可能根据生物遗传的特殊规律，创立各种复杂的遗传模型；动、植物育种工作者可以按他们自己的实际需要设计遗传试验，分析育种试验中积累的大量非平衡资料。数量遗传学的发展已经走出了低谷，即将进入一个蓬勃发展的新时期。

在数量遗传学科发展的转折关头，我们将不失时机地开展数量遗传学的理论研究，不断发展新的遗传模型和新的分析方法，使我国的数量遗传研究达到国际领先水平。

参考文献

- 1 Fisher, R A. Statistical Methods for Research Workers. 1st ed. Oliver & Boyd. Edinburgh and London. 1925
- 2 Comstock, R E. Robinson, H F. Estimation of average dominance of genes. In Heterosis. Gowan, J. W. (editor), Iowa State University Press, Ames, Iowa. 1952
- 3 Hallauer, A R. Miranda, J B. Quantitative Genetics in Maize Breeding. Iowa State University Press, Ames, Iowa. 1981
- 4 Yates, F. Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity, 1947, 1, 287~301

- 5 Griffing, B. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* 1956, 9: 463~493
- 6 Gardner, C O. Eberhart, S A. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics*, 1966,22, 439~452
- 7 Eisen, E J. Bohren, B B. McKean, H E. Sex-linked and maternal effects in the diallel cross. *Aust. J. Biol. Sci.* 1966, 19: 1061~1071
- 8 Cockerham, C C. Weir, B S. Quadratic analysis of reciprocal crosses. *Biometrics*, 1977, 33: 187~203
- 9 朱军. Estimation of genetic variance components in the general mixed model. Ph.D. Dissertation. North Carolina State Univ. Raleigh, N.C. 1989.
- 10 Cockerham, C C. Random and fixed effects in plant genetics. *Theoretical and Applied Genetics*, 1980, 56: 119~131
- 11 朱军. 作物杂种后代基因型值和杂种优势的预测方法. *生物数学学报*, 1993, 8(1):32~44
- 12 朱军、季道藩、许馥华. 作物品种间杂种优势遗传分析的新方法. *遗传学报*, 1993, 20(3): 262~271
- 13 吴吉祥、朱军、许馥华、季道藩. 陆地棉 F2 产量性状杂种优势的遗传分析及其预测. *北京农业大学学报*, 1993, 19 增刊(下):95~99
- 14 Cockerham, C C. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics*, 1954, 39: 859~882
- 15 朱军. Mixed Model Approaches for Estimating Genetic Variances and Covariances. *生物数学学报* 1992, 7(1): 1~11
- 16 徐绍英、郭银燕. 大麦熟期性状的遗传分析. *生物数学学报*, 1992, 7(4):172~178
- 17 朱军、Weir B S. Analysis of cytoplasmic and maternal effects: I. a genetic model for diploid plant seeds and animals. *Theoretical and Applied Genetics*, 1994, (in press).
- 18 朱军、Weir B S. Analysis of cytoplasmic and maternal effects: II. genetic models for triploid endosperms. *Theoretical and Applied Genetics*, 1994, (in press)
- 19 朱军、许馥华. 1994. 禾谷类作物胚乳品质性状的遗传模型及其分析方法. *作物学报*, 20(3): 264~270
- 20 Yu, S. Zhang, Q. Shen Q. Genetic control of endosperm protein content and some kernel traits in barley (*Hordeum vulgare* L.). *J. Biomath.*, 1992, 7(4): 11~21
- 21 石春海、朱 军. 粳型杂交稻碾磨品质的遗传分析. *生物数学学报*, 1992, 7(4): 37~45
- 22 石春海、朱 军. 粳型杂交稻外观品质的遗传分析. *北京农业大学学报*, 1993, 19(增刊上):69~74
- 23 石春海、朱 军. 粳稻米蒸煮品质的种子和母体遗传效应分析. *中国水稻* 1994(3):129~134
- 24 Rao, C R. Estimation of variance and covariance components - MINQUE theory. *J. Multivar. Anal.*,1971, 1: 257~275
- 25 朱军. Mixed model approaches for estimating covariances between two traits with unequal design matrices. *生物数学学报*, 1993, 8(3) 24~30
- 26 Henderson, C R. Selection index and expected genetic advance. In: Hanson, W. D., Robinson, H. F. (eds.) *Statistical Genetics and Plant Breeding*. Natl. Acad. Sci. -N R C, 1963, 982: 141~163