MIXED LINEAR MODEL APPROACHES FOR ANALYZING GENETIC MODELS OF COMPLEX QUANTITATIVE TRAITS [*]

ZHU Jun

Research Center of Biomathematics   Huajiachi Campus of Zhejiang University   Hangzhou 310029   China

Abstract   New approaches based on general mixed linear models were presented for analyzing complex quantitative traits in animal models seed models and QTL quantitative trait locus mapping models. Variances and covariances can be appropriately estimated by MINQUE minimum norm quadratic unbiased estimation approaches. Random genetic effects can be predicted without bias by LUP linear unbiased prediction or AUP adjusted unbiased prediction methods. Mixed-model based composite interval mapping MCIM methods are suitable for efficiently searching QTLs along the whole genome. Bayesian methods and Markov Chain Monte Carlo MCMC methods can be applied in analyzing parameters of random effects as well as their variances.

Key words   mixed model approaches genetic models estimation of variances and covariances   prediction of genetic effects QTL mapping Bayesian methods.

Document code   A   **CLC number**   R69

INTRODUCTION

Many genetic models based on the approach of ANOVA analysis of variance were developed by Fisher 1925 . Some of these models e.g. NC design I and II Comstock et al. 1952 Hallauer et al. 1981   diallel models Yates 1947 Griffing 1956 Gardner and Eberhart 1966   are still widely used by plant and animal breeders. But ANOVA approaches have some deficiencies in analyzing complex genetic models for quantitative traits. Many genetic models cannot be appropriately analyzed by the ANOVA approach if they have more complicated effects other than additive and dominance effects e.g. Eisen's animal model with sex-linkage and maternal effects Eisen et al. 1966 and a bio-model including maternal and paternal effects Cockerham and Weir 1977 .

In the 1970's statisticians developed some methods for analyzing mixed linear models which can be applied in quantitative genetics. Mixed linear model approaches can overcome the shortcomings of ANOVA methods for handling unbalanced data and complicated models. Development of mixed linear model approaches and their application in quantitative genetics will create enormous challenges for quantitative geneticists in dealing with complicated genetic problems.

In the present paper we will review some of our recent works in extending the mixed linear model approaches and constructing complicated genetic models for analyzing complex quantitative traits. Methods recently developed for mixed linear models and their applications will be given to show the ways for solving the real complicated problems in quantitative genetics.

MIXED LINEAR MODEL APPROACHES

General mixed linear models

Many genetic models with biological meanings for different generations are complicated and have no integer coefficients or even correlated effects. Parameters in these genetic models are not manageable by traditional ANOVA methods but can be analyzed by mixed linear model approaches.

Most genetic models can be expressed by a general form of the mixed linear model

$$y = X_1 b_1 + X_2 b_2 + \cdots + X_n b_n + U_1 e_1$$
$$+ U_2 e_2 + \cdots + U_{m-1} e_{m-1} + e_m$$

$$= Xb + \sum_{u=1}^{m} U_u e_u$$

$$\sim N(Xb, V) \qquad\qquad 1$$

where y is the vector ($n \times 1$) of phenotype value with mean Xb and variance V; b is the vector ($p \times 1$) of fixed effects; X is the known incidence matrix ($n \times p$) relating to the fixed effects; e_u ($u = 1, 2, \ldots, m-1$) is the vector ($q_u \times 1$) of the u-th random factor, $e_u \sim (O, \sigma_u^2 R_u)$, R_u is a constant matrix describing the relationship of e_u; U_u is the known coefficient matrix relating to the random vector e_u; e_m is the vector ($n \times 1$) of the residual random effects with $e_m \sim (O, \sigma_m^2 I)$ and $R_m = I$.

If random factors are independent with cov$(e_u, e_v^T) = O$, $V = \sum_{u=1}^{m} \sigma_u^2 V_u$ with $V_u = U_u R_u U_u^T$ ($u = 1, 2, \ldots, m-1$) and $V_m = I$. When random effects of e_u are also independent ($R_u = I$), then $V_u = U_u U_u^T$.

Variance and covariance estimation

The variances in model (1) can be analyzed by mixed linear model approaches such as maximum likelihood (ML) method (Hartley and Rao, 1967), restricted maximum likelihood (REML) method (Patterson and Thompsom, 1971) or minimum norm quadratic unbiased estimation (MINQUE) method (Rao, 1971).

When random factors are independent, estimated variance components can be obtained by the methods of ML in (2), REML in (3) or MINQUE in (4). ML estimates of variances can be calculated by numerically solving the following normal equations

$$\mathrm{tr}(\hat{V}_h^{-1} V_u \hat{V}_h^{-1} V_v) \hat{\sigma}_{u(h+1)}^2$$
$$= y^T \hat{Q}_h V_u \hat{Q}_h y \qquad\qquad 2$$

where

$$\hat{Q}_h = \hat{V}_h^{-1} - \hat{V}_h^{-1} X (X^T \hat{V}_h^{-1} X)^+ X^T \hat{V}_h^{-1}$$
$$\hat{V}_h = \sum_n \hat{\sigma}_{u(h)}^2 V_u = \sum \hat{\sigma}_{u(h)}^2 U_u R_u U_u^T$$

$\hat{\sigma}_{u(h)}^2$ is the estimate of σ_u^2 by the h-th iteration.

Estimated variances obtained by the ML method tend to be influenced by the fixed effects, so ML is rarely used. The REML estima-

tion method can overcome the influence of the fixed effects. The normal equations for obtaining REML estimates of variances are very similar to equation (2) except that $\hat{V}_h$ in the left-hand side of equation (2) is replaced by $\hat{Q}_h$.

$$\mathrm{tr}(\hat{Q}_h V_u \hat{Q}_h V_v) \hat{\sigma}_{u(h+1)}^2$$
$$= y^T \hat{Q}_h V_u \hat{Q}_h y \qquad\qquad 3$$

The MINQUE method does not need the assumption of normal distribution for observed data and can give estimated variances without iteration. The MINQUE equations for estimating variance components are

$$\mathrm{tr}(Q_\alpha V_u Q_\alpha V_v) \hat{\sigma}_u^2 = y^T Q_\alpha V_u Q_\alpha y_2 \qquad 4$$

where

$$Q_\alpha = V_\alpha^{-1} - V_\alpha^{-1} X (X^T V_\alpha^{-1} X)^+ X^T V_\alpha^{-1}$$
$$V_\alpha = \sum_u \alpha_u V_u = \sum_u \alpha_u U_u R_u U_u^T$$

Covariance components $\sigma_{u/u}$ between two traits y_1 and y_2 can also be easily estimated without bias by solving the following equations (Zhu and Weir, 1996)

$$\mathrm{tr}(Q_\alpha V_u Q_\alpha V_v) \hat{\sigma}_{u/u} = y_1^T Q_\alpha V_u Q_\alpha y_2 \qquad\qquad 5$$

The MINQUE method uses the prior values α_u that may be chosen from prior experiments, from iterations or theoretical considerations. MINQUE(1) as a MINQUE method with $\alpha_u = 1$ can be employed for estimating variance components when random factors are independent.

For certain genetic models, some of the random factors are correlated. Variance matrix V consists of variance components σ_u^2 for random factors and also covariance components $\sigma_{u,v}$ between correlated random factors. Zhu and Weir (1994a, 1994b) proposed genetic models for diploid seed and triploid endosperm, which can be written by a mixed linear model

$$y = Xb + U_A e_A + U_D e_D + U_C e_C + U_{Am} e_{Am}$$
$$+ U_{Dm} e_{Dm} + U_B e_B + e_\varepsilon$$
$$= Xb + \sum_u^7 U_u e_u \qquad\qquad 6$$

with variance matrix

$$V = \sigma_A^2 V_1 + \sigma_D^2 V_2 + \sigma_C^2 V_3 + \sigma_{Am}^2 V_4 + \sigma_{Dm}^2 V_5$$
$$+ \sigma_B^2 V_6 + \sigma_{A.Am} V_7 + \sigma_{D.Dm} V_8 + \sigma_\varepsilon^2 V_9$$

$$= \sum_{u=1}^{9} \theta_u \boldsymbol{V}_u$$

where $\boldsymbol{V}_u = \boldsymbol{U}_u \boldsymbol{U}_u^T$ $u = 1$ 2 … 6 $\boldsymbol{V}_7 = \boldsymbol{U}_1 \boldsymbol{U}_4^T + \boldsymbol{U}_4 \boldsymbol{U}_1^T$ $\boldsymbol{V}_8 = \boldsymbol{U}_2 \boldsymbol{U}_5^T + \boldsymbol{U}_5 \boldsymbol{U}_2^T$ $\boldsymbol{V}_9 = \boldsymbol{I}$.

MINQUE 0/1 was suggested by Zhu and Weir 1994a for unbiased estimation of variances and covariances for one traits and covariances between two traits . MINQUE 0/1 procedure is a MINQUE method setting 0 for all the prior covariances $\alpha_{u.v}$ and 1 for all the prior variances α_u . Variances and covariances for one trait $\boldsymbol{y}_1 = \boldsymbol{y}_2$ and covariances between two traits $\boldsymbol{y}_1 \neq \boldsymbol{y}_2$ can be estimated by the following MINQUE 0/1 equations

$$\text{tr} \ \boldsymbol{Q}_{0/1} \ \boldsymbol{V}_u \boldsymbol{Q}_{0/1} \ \boldsymbol{V}_v \quad \hat{\theta}_{u/u}$$
$$= \ \boldsymbol{y}_1^T \boldsymbol{Q}_{0/1} \ \boldsymbol{V}_u \boldsymbol{Q}_{0/1} \ \boldsymbol{y}_2 \qquad 7$$

where

$$\boldsymbol{Q}_{0/1} = \boldsymbol{V}_{0/1}^{-1} - \boldsymbol{V}_{0/1}^{-1} \ \boldsymbol{X} \ \boldsymbol{X}^T \boldsymbol{V}_{0/1}^{-1} \ \boldsymbol{X} \ ^+$$
$$\cdot \ \boldsymbol{X}^T \boldsymbol{V}_{0/1}^{-1}$$

$$\boldsymbol{V}_{0/1} = \sum_{u=1}^{6} \boldsymbol{U}_u \boldsymbol{U}_u^T + \boldsymbol{I}$$

Monte Carlo simulations showed that both variances and covariances of seed traits can be unbiasedly estimated by the MINQUE 0/1 procedure Zhu and Weir 1994a 1994b .

In plant breeding breeders usually want to improve seed quality traits while still keeping the genetic merit of yield traits . Therefore understanding the genetic relationship between seed quality traits and plant yield traits is of importance . Seed models and plant models have unequal design matrices . Zhu 1993b developed a new method for estimating genetic covariance components between seed traits $\boldsymbol{y}_s$ and plant traits $\boldsymbol{y}_p$. For seed model 6 the corespondent plants bearing the seeds will have the following mixed linear model

$$\boldsymbol{y}_P = \boldsymbol{Xb}_P + \boldsymbol{U}_C \boldsymbol{e}_{C\ P} + \boldsymbol{U}_{Am} \boldsymbol{e}_{Am\ P}$$
$$+ \ \boldsymbol{U}_{Dm} \boldsymbol{e}_{Dm\ P} + \boldsymbol{U}_B \boldsymbol{e}_{B\ P} + \boldsymbol{e}_{\varepsilon\ P}$$
$$= \boldsymbol{Xb}_P + \sum_{u}^{5} \boldsymbol{U}_u \boldsymbol{e}_{u\ P}$$

There are covariances between random factors of seed traits and those of plant traits $\sigma_{A/Am} =$ covariance between seed direct additive effects and

plant additive effects $\sigma_{D/Dm} =$ covariance between seed direct dominance effects and plant dominance effects $\sigma_{C/C} =$ covariance between seed cytoplasm effects and plant cytoplasm effects $\sigma_{Am/Am} =$ covariance between seed maternal additive effects and plant additive effects $\sigma_{Dm/Dm} =$ covariance between seed maternal dominance effects and plant dominance effects $\sigma_{B/B} =$ covariance between seed block effects and plant block effects $\sigma_{\varepsilon/\varepsilon} =$ covariance between seed residual effects and plant residual effects .

If we define $\boldsymbol{F}_1 = \boldsymbol{U}_A \boldsymbol{U}_{Am}^T + \boldsymbol{U}_{Am} \boldsymbol{U}_A^T$ $\boldsymbol{F}_2 = \boldsymbol{U}_D \boldsymbol{U}_{Dm}^T + \boldsymbol{U}_{Dm} \boldsymbol{U}_D^T$ $\boldsymbol{F}_3 = 2\boldsymbol{U}_C \boldsymbol{U}_C^T$ $\boldsymbol{F}_4 = 2\boldsymbol{U}_{Am} \boldsymbol{U}_{Am}^T$ $\boldsymbol{F}_5 = 2\boldsymbol{U}_{Dm} \boldsymbol{U}_{Dm}^T$ $\boldsymbol{F}_6 = 2\boldsymbol{U}_B \boldsymbol{U}_B^T$ and $\boldsymbol{F}_7 = 2\boldsymbol{I}$ covariance components between seed trait and plant trait can then be estimated by the following equations

$$\text{tr} \ \boldsymbol{Q}_{0/1} \ \boldsymbol{F}_u \boldsymbol{Q}_{0/1} \ \boldsymbol{F}_v \quad \hat{\sigma}_{u/u}$$
$$= \ 2\boldsymbol{y}_s^T \boldsymbol{Q}_{0/1} \ \boldsymbol{F}_u \boldsymbol{Q}_{0/1} \ \boldsymbol{y}_p \qquad 8$$

where

$$\boldsymbol{Q}_{0/1} = \boldsymbol{V}_{0/1}^{-1} - \boldsymbol{V}_{0/1}^{-1} \ \boldsymbol{X} \ \boldsymbol{X}^T \boldsymbol{V}_{0/1}^{-1} \ \boldsymbol{X} \ ^+$$
$$\cdot \ \boldsymbol{X}^T \boldsymbol{V}_{0/1}^{-1}$$

$$\boldsymbol{V}_{0/1} = 2 \ \boldsymbol{U}_C \boldsymbol{U}_C^T + \boldsymbol{U}_{Am} \boldsymbol{U}_{Am}^T + \boldsymbol{U}_{Dm} \boldsymbol{U}_{Dm}^T$$
$$+ \ \boldsymbol{U}_B \boldsymbol{U}_B^T + \boldsymbol{I}$$

For time-dependable traits the phenotype data observed at time t $t = 1$ 2 … has the following mixed linear model

$$\boldsymbol{y}_t = \boldsymbol{Xb}_t + \sum_{u=1}^{m} \boldsymbol{U}_u \boldsymbol{e}_{u\ t}$$
$$\sim N \ \boldsymbol{Xb}_t \quad \boldsymbol{V}_t = \sum_{u=1}^{m} \sigma_{u\ t}^2 \ \boldsymbol{U}_u \boldsymbol{R}_u \boldsymbol{U}_{u\ t}^T$$

Given the observed phenotype vector $\boldsymbol{y}_{t-1}$ measured at time $t-1$ the conditional random variables of $\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}$ at time t have conditional distribution

$$\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1} = \boldsymbol{Xb}_{t|t-1} + \sum_{u=1}^{m} \boldsymbol{U}_u \boldsymbol{e}_{u\ t|t-1}$$
$$\sim N \ \boldsymbol{Xb}_{t|t-1} \quad \boldsymbol{V}_{t|t-1}$$
$$= \sum_{u=1}^{m} \sigma_{u\ t|t-1}^2 \ \boldsymbol{U}_u \boldsymbol{R}_u \boldsymbol{U}_{u\ t|t-1}^T$$
$$9$$

Since conditional $\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}$ is independent of $\boldsymbol{y}_{t-1}$ conditional random effects $\boldsymbol{e}_{t|t-1}$ and

conditional variance components $\sigma^2_{t|t-1}$ contain extra variation not explainable by the accumulated effects from the initial time to time t-1. Zhu 1995 proposed a mixed model approach for analyzing conditional variance components and conditional random effects which had been used in developmental quantitative genetic analysis Atchley and Zhu 1997 Yan et al. 1998 .

When phenotypic vector y has large size and multivariate normal distribution variance estimates will have asymptotic multivariate normal distribution

$$\hat\sigma^2_u \sim N \quad \sigma^2_u \quad 2H^{-1} \qquad 10$$

where $H = $ tr $V^{-1}V_u V^{-1}V_v$ $^{-1}$ for ML estimates Searle 1970 or $H =$ tr $QV_u QV_v$ $^{-1}$ for REML estimates Searle 1970 and MINQUE estimates Rao and Kleffe 1988 .

Hypothesis test for linear combination of variances can be conducted by a χ^2 test. For null hypothesis H_0 $\sum_{u=1}^{m} c_u\sigma^2_u = \rho$ vs. alternative hypothesis H_1 $\sum_{u=1}^{m} c_u\sigma^2_u \neq \rho$ the statistic χ^2_{cal} will asymptotically have χ^2 distribution with 1 degree of freedom when H_0 is true

$$\chi^2_{cal} = \frac{\left(\sum_{u=1}^{m} c_u\hat\sigma^2_u - \rho\right)^2}{2\left(\sum_{u=1}^{m} c_u{}^2 \hat{H}^{-1}_{uu} + 2\sum_{u=1}^{m-1}\sum_{v>u}^{m} c^2_u c^2_v \hat{H}^{-1}_{uv}\right)}$$

$$\underset{\sim}{H_0} \chi^2_{df=1} \qquad 11$$

If $\chi^2_{cal} > \chi^2_{df=1}$ H_0 will be rejected and H_1 be accepted otherwise H_0 is not rejectable.

Zhu and Weir 1994a 1996 suggested obtaining estimates and their standard errors by resampling genetic entries or experimental blocks with the Jackknife method Miller 1974 .

If $\hat\theta$ is an estimate of a genetic parameter from a sample of K entries and $\hat\theta_k$ $k=1$ 2 $\cdots$ K is the estimate resulting when entry k is omitted then the Jackknife estimate $\hat\theta_J$ and its sampling variance var $\hat\theta_J$ are

$$\hat\theta_J = K\hat\theta - K-1 \overline{\hat\theta}$$

$$var \hat\theta_J = \frac{K-1}{K}\sum_{k=1}^{K} \hat\theta_k - \overline{\hat\theta} {}^2$$

where $\overline{\hat\theta} = \frac{1}{K}\sum_{k=1}^{K}\hat\theta_k$. If K is not large $\hat\theta_J - \theta$ / $\sqrt{var \hat\theta_J}$ is approximately distributed as a t-distribution with $K-1$ degrees freedom under the null hypothesis H_0 .

Monte Carlo simulation showed that statistical test for variances by χ^2 test in equation 11 is much more conservative than that by a t-test with the Jackknife procedure Zhu 1989 .

Random effects prediction

Except for the residual effects e_m random effects in mixed linear model 1 can be predicted by the best linear unbiased prediction BLUP method Henderson 1963 . Henderson 1988 proposed a general approach which can result in both unbiased estimation for fixed effects and unbiased prediction for random effects without calculating the inverse of V. For model 1 the estimation and prediction can be obtained by solving the following equations

$$\begin{bmatrix} X^TR^{-1}_m X & X^TR^{-1}_m U_1 & X^TR^{-1}_m U_2 & \cdots \\ U_1^TR^{-1}_m X & U_1^TR^{-1}_m + R^{-1}_1/\sigma^2_1 & U_1^TR^{-1}_m U^2 & \cdots \\ U_2^TR^{-1}_m X & U_2^TR^{-1}_m Z_1 & U_2^TR^{-1}_m U_2 + R^{-1}_2/\sigma^2_2 & \cdots \end{bmatrix}$$

$$\cdot \begin{bmatrix} \hat{b} \\ \hat{e}_1 \\ \hat{e}_2 \end{bmatrix} = \begin{bmatrix} X^TR^{-1}_m y \\ U_1^TR^{-1}_m y \\ U_2^TR^{-1}_m y \end{bmatrix} \qquad 12$$

Normal equations 12 can be expressed simply as $W\beta = d$.

Since the solution of equations 12 requires the unknown variances Eq. 12 's estimates still depend on the calculation of $\hat{V}^{-1}$ by the REML method.

If fixed effects and random effects of some factors are not of interest we can obtained BLUP for specific random effects e_u $u=1$ 2 $\cdots$ $m-1$ Henderson 1963

$$\hat{e}_u {}_{\sigma^2} = \sigma^2_u R_u U_u^T V^{-1} \quad y - X\hat{b}$$

$$= \sigma^2_u R_u U_u^T Q y \qquad 13$$

Since the true variances are always unknown in practice estimated variances are usually used in prediction

$$\hat{e}_u {}_{\hat\sigma^2} = \hat\sigma^2_u R_u U_u^T \hat{Q} y \qquad 14$$

With such prediction by using estimates only a

so-called BLUP which may have lost linearity and unbiasedness is obtainable.

Instead of using parameters 13 or their estimates 14 for predicting random effects Zhu and Weir 1996 suggested choosing prior values α_u as in the case of the MINQUE method

$$\hat{\boldsymbol{e}}_{u\ \alpha}\ =\ \alpha_u \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_\alpha \boldsymbol{y} \qquad 15$$

If the choice of prior value is not based on the observed data the predictor $\hat{\boldsymbol{e}}_u$ is a linear unbiased prediction LUP for random vector $\boldsymbol{e}_u$.

Monte Carlo simulations revealed that both BLUP and LUP will give prediction with unbiased mean but under-estimated variance $E\ \hat{\boldsymbol{e}}^T\hat{\boldsymbol{e}}/df\ < \sigma^2$ for random variables Zhu and Weir 1996 . In order to solve this problem a method of adjusted unbiased prediction AUP was suggested for predicting random variables Zhu 1993a Zhu and Weir 1996

$$\hat{\boldsymbol{e}}_{u\ \alpha}^A\ =\ \kappa_u\ \alpha_u \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_\alpha \boldsymbol{y} \qquad 16$$

where $\kappa_u = \sqrt{\ q_u - 1\ \hat{\sigma}_u^2/\ \alpha_u^2 \boldsymbol{y}^T \boldsymbol{Q}_\alpha \boldsymbol{V}_u \boldsymbol{Q}_\alpha \boldsymbol{y}}$ is an adjusted coefficient to insure $\hat{\boldsymbol{e}}_{u\ \alpha}^{A^T}\ \hat{\boldsymbol{e}}_{u\ \alpha}^A /\ q_u - 1\ = \hat{\sigma}_u^2$ and set $\hat{\sigma}_u^2 = 0$ when $\hat{\sigma}_u^2 < 0$.

Monte Carlo simulations proved that AUP gives unbiased mean and variance for predicted effects Zhu and Weir 1996 .

For mixed linear models with correlated random variables cov $\boldsymbol{e}_u\ \boldsymbol{e}_v^T\ = \sigma_{u.v}\ \boldsymbol{R}_{u.v}$ the BLUP for random effects $\boldsymbol{e}_u$ is given by

$$\hat{\boldsymbol{e}}_{u\ \theta}\ =\ \sigma_u^2 \boldsymbol{R}_u \boldsymbol{U}_u^T + \sigma_{u.v} \boldsymbol{R}_{u.v} \boldsymbol{U}_v^T\ \boldsymbol{Q}\boldsymbol{y} \qquad 17$$

When MINQUE 0/1 is used for estimating variances and covariances for seed traits Zhu and Weir 1994a LUP can be used for predicting random genetic effects $\boldsymbol{e}_u\ u = 1\ 2\ \ldots\ m - 1$.

$$\hat{\boldsymbol{e}}_{u\ 0/1}\ =\ \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_{\ 0/1}\ \boldsymbol{y} \qquad 18$$

And AUP is obtainable by

$$\hat{\boldsymbol{e}}_{u\ 0/1}^A\ =\ \kappa \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_{\ 0/1}\ \boldsymbol{y} \qquad 19$$

where $\kappa = \sqrt{\ q_u - 1\ \hat{\sigma}_u^2/\ \boldsymbol{y}^T \boldsymbol{Q}_{\ 0/1}\ \boldsymbol{V}_u \boldsymbol{Q}_{\ 0/1}\ \boldsymbol{y}}$.

When random effects are predicted by BLUP or LUP the sampling variances for predicted effects can be calculated by

$$\text{var}\ \hat{\boldsymbol{e}}_{u\ \sigma^2}\ =\ \text{var}\ \sigma_u^2 \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}\boldsymbol{y}$$

$$=\ \sigma_u^2\ ^2 \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q} \text{var}\ \boldsymbol{y}\ \boldsymbol{Q}\boldsymbol{U}_u \boldsymbol{R}_u$$

$$=\ \sigma_u^2\ ^2 \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}\boldsymbol{V}\boldsymbol{Q}\boldsymbol{U}_u \boldsymbol{R}_u$$

$$=\ \sigma_u^2\ ^2 \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}\boldsymbol{U}_u \boldsymbol{R}_u \qquad 20$$

for BLUP and

$$\text{var}\ \hat{\boldsymbol{e}}_{u\ \alpha}\ =\ \text{var}\ \alpha_u \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_\alpha \boldsymbol{y}$$

$$=\ \alpha_u\ ^2 \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_\alpha \text{var}\ \boldsymbol{y}\ \boldsymbol{Q}_\alpha \boldsymbol{U}_u \boldsymbol{R}_u$$

$$=\ \alpha_u^2 \boldsymbol{R}_u \boldsymbol{U}_u^T \boldsymbol{Q}_\alpha \boldsymbol{V} \boldsymbol{Q}_\alpha \boldsymbol{U}_u \boldsymbol{R}_u \qquad 21$$

for LUP.

Hypothesis test for linear combination of genetic effects can be conducted by a χ^2 test in the same way as the variance test. But Zhu and Weir 1994a 1996 suggested use of the Jack-knife resampling technique in conducting a t-test for predicted genetic effects because it is more powerful for detecting non-zero parameters.

QTL mapping approaches

For mapping quantitative trait loci QTLs of plants putative QTL within two flanking markers M_{i-} and M_{i+} is searched along the whole genome. The markers linked to other QTLs are often included as fixed effects in a regression model for controlling background noise Jansen 1993 Zeng 1994 . For mapping QTLs of animal populations QTLs effects are often treated as random effects which will fit the framework of mixed linear models Grignola et al. 1997 Wang et al. 1998 . The QTL effects for these animal models are predicted by Henderson's BLUP approach as in equation 12 .

An approach of mixed-model-based composite interval mapping MCIM Zhu 1998 1999 Zhu and Weir 1998 Wang et al 1999 by which the marker effects are treated as random effects can be constructed for handling QTL × environment interaction or other complicated effects. If the MCIM approach is used QTL mapping models can all be expressed in the general mixed linear model 1 . The likelihood function L for the parameters of fixed effects $\boldsymbol{b}$ and variance components σ_u^2 is

$$L\ \boldsymbol{b}\ \boldsymbol{V}\ =\ 2\pi\ ^{-\frac{n}{2}}\ |\ \boldsymbol{V}\ |^{-\frac{1}{2}} \exp\ -\frac{1}{2}\ \boldsymbol{y}$$

$$-\ \boldsymbol{X}\boldsymbol{b}\ ^T \boldsymbol{V}^{-1}\ \boldsymbol{y}\ -\ \boldsymbol{X}\boldsymbol{b} \qquad 22$$

with the log of the likelihood function l as

$$l\ \boldsymbol{b}\ \boldsymbol{V}\ =\ -\frac{n}{2}\ln\ 2\pi\ -\frac{1}{2}\ln\ |\ \boldsymbol{V}\ |$$

$$-\frac{1}{2}(y - Xb)^T V^{-1}(y - Xb)$$

$$(23)$$

If variance components of the model are known, the maximum likelihood estimates of QTL effects in b can be obtained by the maximum likelihood estimation

$$\hat{b} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \qquad (24)$$

with sampling variance matrix

$$\mathrm{var}(\hat{b}) = (X^T V^{-1} X)^{-1}.$$

We can search QTL within two flanking markers M_{i-} and M_{i+} for the whole genome by setting a prior value for recombination fraction $r_{M_{i-Q}}$ between marker and locus Q. For each prior value $\hat{r}_{M_{i-Q}}$ the likelihood ratio statistic (LR) can be calculated by

$$LR = 2l_1(\hat{b}, \hat{V}, \hat{r}_{M_{i-Q}}) - 2l_0(\hat{b}, \hat{V}, r_{M_{i-Q}} = 0.5)$$

$$(25)$$

where variance components in V can be replaced by their unbiased estimates

$$\hat{V} = \sum_u \hat{\sigma}_u^2 U_u R_u U_u$$

The likelihood ratio statistic can also be used for testing the null hypothesis H_0: $r_{M_{i-Q}} = 0.5$ vs. the alternative hypothesis H_1: $r_{M_{i-Q}} < 0.5$. LR approximately has χ^2 distribution.

When the null presentation of no QTL within two flanking markers M_{i-} and M_{i+} is rejected $\hat{r}_{M_{i-Q}}$ infers the position of the QTL while $\hat{b}$ gives the estimates of additive and dominance effects of this QTL. Hypothesis for additive and dominance effects can be conducted by a t-test in a general way for

$$H_0: c'b = m \text{ vs. } H_1: c'b \neq m$$

If the statistic $|c'(\hat{b} - b)|/\sqrt{c'(X'\hat{V}^{-1}X)c|} > z_{\alpha/2}$ the null hypothesis is then rejected.

Bayesian analysis

Bayesian methods and Markov Chain Monte Carlo (MCMC) methods were recently applied in QTL mapping (Bink et al. 1998). The conditional distribution for generating the data is

$$y | (b, e_1, e_2, \cdots, e_{m-1}, \sigma_m^2)$$
$$\sim N\left(Xb + \sum_{u=1}^{m-1} U_u e_u, \sigma_m^2 I \right) \qquad (26)$$

The prior distributions for the unknowns in the model can be assigned for b, e_u ($u = 1, 2, \cdots, m-1$) and σ_m^2 (Wang et al. 1994).

Let $\boldsymbol{\beta}^T = (b^T, e_1^T, e_2^T, \cdots, e_{m-1}^T)$
$$= (\beta_1, \beta_2, \cdots, \beta_N)$$
$$v^T = (\sigma_1^2, \sigma_2^2, \cdots, \sigma_{m-1}^2)$$
$$\boldsymbol{\alpha}^T = (\alpha_1, \alpha_2, \cdots, \alpha_{m-1}, \alpha_m)$$
$$\boldsymbol{v}^T = (\nu_1, \nu_2, \cdots, \nu_{m-1}, \nu_m)$$
be β^T without β_i

where $N = p + \sum_{u=1}^{m-1} q_u = p + q$ and

$$\boldsymbol{\beta}^T_{-i} = (\beta_1, \beta_2, \cdots, \beta_{i-1}, \beta_{i+1}, \cdots, \beta_N).$$

And let

$$\boldsymbol{v}^T_{-u} = (\sigma_1^2, \sigma_2^2, \cdots, \sigma_{u-1}^2, \sigma_{u+1}^2, \cdots, \sigma_{m-1}^2)$$

be $\boldsymbol{v}^T$ without variance component u. Then the joint posterior density is in the normal-gamma form (Macedo and Gianola 1987)

$$\Pr(\boldsymbol{\beta}, \boldsymbol{v}, \sigma_m^2 | y, \boldsymbol{\alpha}, \boldsymbol{v}) \propto$$

$$(\sigma_m^2)^{-(n+v_m+2)/2} \exp\left\{ -\frac{1}{2\sigma_m^2}\left[\left(y - Xb - \sum_{u=1}^{m-1} U_u e_u \right)^T \right.\right.$$

$$\cdot \left. \left(y - Xb - \sum_{u=1}^{m-1} U_u e_u \right) + v_m \alpha_m \right] \right\}$$

$$\times \prod_{u=1}^{m-1} (\sigma_u^2)^{-(q_u+v_u+2)/2} \exp\left[-\frac{1}{2\sigma_u^2}(e_u^T R_u^{-1} e_u \right.$$

$$\left. + v_u \alpha_u) \right] \qquad (27)$$

Inference about each of the unknowns ($\boldsymbol{\beta}$, $\boldsymbol{v}$, σ_m^2) are based on their marginal densities respectively.

Gibbs sampling can be implemented for constructing the fully conditional posterior densities of all unknowns in (27).

The conditional posterior densities of each of the location parameters in $\boldsymbol{\beta}$ is normal with mean $\tilde{\beta}_i$ and variance $\tilde{s}_i^2$

$$\boldsymbol{\beta}_i | y, \boldsymbol{\beta}_{-i}, \boldsymbol{v}, \sigma_m^2, \boldsymbol{\alpha}, \boldsymbol{v} \sim N(\tilde{\beta}_i, \tilde{s}_i^2)$$
$$i = 1, 2, \cdots, N \qquad (28)$$

where $\tilde{\beta}_i = \left(d_i - \sum_{j=1, j\neq i}^{N} w_{ij}\beta_j \right)/w_{ii}$ and $\tilde{s}_i^2 = \sigma_m^2/w_{ii}$. w_{ij} is the element of ith row and jth column of matrix W in the left hand side of normal equations (12) and d_i is the ith element of vector d in the right hand side of normal equation (12).

The conditional posterior density of residual varianceis σ_m^2 is in the scaled inverted chi-square form

$$\Pr\left(\sigma_m^2 \mid y, \beta, v, \alpha, \upsilon\right) \propto$$

$$\left(\sigma_m^2\right)^{-(n+v_m+2)/2}\exp\left\{-\frac{1}{2\sigma_m^2}\Big[\Big(y - Xb - \sum_{u=1}^{m-1} U_u e_u\Big)^T\right.$$

$$\left.\cdot\Big(y - Xb - \sum_{u=1}^{m-1} U_u e_u\Big) + v_m\alpha_m\Big]\right\}$$

or

$$\sigma_m^2 \mid y, \beta, v, \alpha, \upsilon \sim \frac{\tilde{v}_m\tilde{\alpha}_m}{\chi^2_{df=\tilde{v}_m}} \qquad 29$$

with parameters $\tilde{v}_m = n + v_m$ and $\tilde{\alpha}_m =$ $\Big[\big(y - Xb - \sum_{u=1}^{m-1} U_u e_u\big)^T\big(y - Xb - \sum_{u=1}^{m-1} U_u e_u\big) + v_m\alpha_m\Big]/\tilde{v}_m$

The conditional posterior density of the uth variance σ_u^2 is also in the scaled inverted chi-square form

$$\Pr\left(\sigma_u^2 \mid y, \beta, v_{-u}, \sigma_m^2, \alpha, \upsilon\right) \propto$$

$$\left(\sigma_u^2\right)^{-(q_u+v_u+2)/2}\exp\left\{-\frac{1}{2\sigma_u^2}\Big(e_u^T R_u^{-1} e_u + v_u\alpha_u\Big)\right\}$$

or

$$\sigma_u^2 \mid y, \beta, v_{-u}, \sigma_m^2, \alpha, \upsilon \sim \frac{\tilde{v}_u\tilde{\alpha}_u}{\chi^2_{df=\tilde{v}_u}} \qquad 30$$

with parameters $\tilde{v}_u = q_u + v_u$ and $\tilde{\alpha}_u = \big(e_u^T R_u^{-1} e_u + v_u\alpha_u\big)/\tilde{v}_u$.

A set of the $N + m$ conditional posterior distributions (28 – 30) is called the Gibbs sampler. Flat priors for all variance components $\Pr(v, \sigma_m^2) \propto$ constant can be set for the Gibbs sampler. The degree of belief parameters can be set as $v_u = -2$ for $u = 1, 2, \ldots, m$.

Bayesian inference can be obtained for the marginal distributions through Gibbs sampling (Gelfan and Smith 1990). Generating random samples from the joint posterior distribution (27) can be achieved through successively drawing samples from and updating the Gibbs sampler (28 – 30).

Gibbs sampling works as follows:

A. set unbiased predictors and estimates as initial values for β, v and σ_m^2

B. generate β_i from (28) and update β_i ($i = 1, 2, \ldots, N$)

C. generate σ_m^2 from (29) and update σ_m^2

D. generate σ_u^2 from (30) and update σ_u^2 ($u = 1, 2, \ldots, m-1$)

Repeat B. – D. for k length of the chain times. As $k \to \infty$ this will create a Markov chain with equilibrium distribution. After running initial iterations as "warm-up" samples are stored every d iterations with the total number of samples saved to be s.

If the Gibbs sampler converges to the equilibrium distribution the s samples are randomly drawn from the joint posterior distribution with density (27). The kth sample

$$\big(\beta_k, v_k \text{ and } \sigma_m^2\big)_i \quad k = 1, 2, \ldots, s \qquad 31$$

is then an $N + m$ vector each elements of which is a drawing from the appropriate marginal distribution. The s samples in (31) is called Gibbs samples for reference. The features of the posterior distribution $P(x)$ can be estimated by

$$\hat{c} = \frac{1}{s}\sum_{k=1}^{s} g(x_k)$$

where $g(x_k)$ can be any feature of $P(x)$ such as its mean or variance.

Bayesian inference can also be made on the functions of the original parameters.

GENETIC MODELS

Animal genetic models

Genotype × environment (GE) interactions have been detected for quantitative traits of many plants and animals. With genetic experiments conducted in multiple environments the average phenotypic performance of a genetic entriy in one environment can be expressed by the following genetic model

$$y = \mu + E + G + GE + \varepsilon \qquad 32$$

where μ = population mean, E = environment effect, G = total genotypic effect, GE = genotype × environment interaction effect, ε = residual effect.

But most animal models (or reduced animal models) consist of only simple genetic effects such as additive dominance and/or maternal effects (Lynch and Walsh 1998). Recently Zhu and Weir (1996) proposed an animal model which is a modification of Eisen's model

Eisen et al. (1966). The genetic model for the phenotypic mean y_{ijsk} of sex s in block k within environment h from the cross between maternal line i and paternal line j is

$$y_{hijsk} = \mu + E_h + G_{ijs} + GE_{hijs} + B_{k(h)} + \varepsilon_{hijsk} \qquad 33$$

where μ is the population mean, E_h is the environment effect, $B_{k(h)}$ is the block effect, $\varepsilon_{hijsk} \sim (0, \sigma_e^2)$ is the residual effect. The total genotype effect is further partitioned into three components $G =$ additive effect $(A) +$ dominance effect $(D) +$ sex-linked effect $(L) +$ maternal effect (M), the same as in the partitioning of the genotype $\times$ environment interaction effect $GE = AE + DE + LE + ML$, for heterogametic progeny (XY or ZW, $s = 1$) and for homogametic progeny (XX or ZZ, $s = 2$) (Zhu and Weir 1996):

$$G_{ij1}^{XY} + GE_{hij1}^{XY} = A_i + A_j + D_{ij} + L_{i1} + M_i$$
$$+ AE_{hi} + AE_{hj} + DE_{hij}$$
$$+ LE_{hi1} + ME_{hi}$$

$$\text{or } G_{ij1}^{ZW} + GE_{hij1}^{ZW} = A_i + A_j + D_{ij} + L_{j1}$$
$$+ M_i + AE_{hi} + AE_{hj}$$
$$+ DE_{hij} + LE_{hj1} + ME_{hi}$$

$$G_{ij2}^{XX/ZZ} + GE_{hij2}^{XX/ZZ} = A_i + A_j + D_{ij} + \frac{1}{2}L_{i2}$$
$$+ \frac{1}{2}L_{j2} + M_i + AE_{hi}$$
$$+ AE_{hj} + DE_{hij} + \frac{1}{2}LE_{hi2}$$
$$+ \frac{1}{2}LE_{hj2} + ME_{hi} \qquad 34$$

The phenotypic mean of this animal model with sex-linked and maternal effects can be expressed by a mixed linear model as

$$y = Xb + U_A e_A + U_D e_D + U_L e_L + U_M e_M$$
$$+ U_{AE} e_{AE} + U_{DE} e_{DE} + U_{LE} e_{LE} + U_{ME} e_{ME}$$
$$+ U_B e_B + e_\varepsilon$$
$$= Xb + \sum_{u}^{10} U_u e_u \qquad 35$$

with variance-covariance matrix

$$\text{var}(y) = \sigma_A^2 U_A U_A^T + \sigma_D^2 U_D U_D^T + \sigma_L^2 U_L U_L^T$$
$$+ \sigma_M^2 U_M U_M^T + \sigma_{AE}^2 U_{AE} U_{AE}^T$$
$$+ \sigma_{DE}^2 U_{DE} U_{DE}^T + \sigma_{LE}^2 U_{LE} U_{LE}^T$$

$$+ \sigma_{ME}^2 U_{ME} U_{ME}^T + \sigma_B^2 U_B U_B^T + \sigma_\varepsilon^2 I$$
$$= \sum_{u=1}^{10} \sigma_u^2 V_u \qquad 36$$

Unbiased estimation of variances and covariances can be obtained by REML or MINQUE (1) approaches. Random effects can be predicted by the BLUP (LUP or AUP) method. Mouse body weight and tail length of a 7×7 diallel cross were analyzed by this animal model (Zhu and Weir 1996; Atchley and Zhu 1997). Silkworm is a heterogametic female species with ZZ sex chromosomes for males and ZW sex chromosomes for females. Data for cocoon weight and fibroin content for a 5×5 diallel cross in two seasons were analyzed for evaluating genetic variance components as well as genotype $\times$ environment interaction variance components (Zhu and Weir 1996).

Seed genetic models

One of the important breeding goals now is improvement of crop quality. Creating seed genetic models in biological meaning with applicable statistical methods is of importance for efficient analysis of seed quantitative traits. By extending Cockerham's general genetic model (Cockerham 1980), Zhu and Weir (1994a) partitioned the total genotype effect G into seed direct gene effect G_O, cytoplasm gene effect G_C, and maternal nuclear gene effect G_M ($G = G_O + G_C + G_M$). Further partitioning was also proposed

$$G_O = \sum_i \tau_i A_i + \sum_i \sum_{j \geq i} \delta_{ij} D_{ij}$$
$$G_C = \sum_i \gamma_i C_i \qquad 37$$
$$G_M = \sum_i \tau_{mi} Am_i + \sum_i \sum_{j \geq i} \delta_{mij} Dm_{ij}$$

where $A_i =$ direct additive effect, $D_{ij} =$ direct dominance effect, $C_i =$ cytoplasm gene effect, $Am_i =$ maternal additive effect, $Dm_{ij} =$ maternal dominance effect.

Genetic models were proposed for quantitative traits of diploid seeds and animals (Zhu and Weir 1994a) and of triploid endosperm (Zhu and Weir 1994b). Usually means of only three generations (P's, F_1's and F_2's) are required for analyzing seed traits.

The total genotype $\times$ environment interac-

tion GE can also be partitioned into three terms $GE = G_O E + G_C E + G_M E$ Zhu 1994 which can be further partitioned into its components

direct $\times$ environment interaction $G_O E =$
$$\sum_h \sum_i \alpha_{hi} AE_{hi} + \sum_h \sum_i \sum_{j \geq i} \beta_{hij} DE_{hij}$$

cytoplasm $\times$ environment interaction $G_C E$
$$= \sum_h \sum_i \lambda_{hi} CE_{hi} \qquad\qquad 38$$

maternal $\times$ environment interaction $G_M E =$
$$\sum_h \sum_i \alpha_{m_{hi}} Am E_{hi} + \sum_h \sum_i \sum_{j \geq i} \beta_{m_{hij}} Dm E_{hij}$$

where $AE_{hi} = A_i \times E_h$ interaction effect $DE_{hij} = D_{ij} \times E_h$ interaction effect $AE_{hi} = A_i \times E_h$ interaction effect $AmE_{hi} = Am_i \times E_h$ interaction effect $DmE_{hij} = Dm_{ij} \times E_h$ interaction effect.

Based on the extension of the general genetic model for seeds experiments of a diallel cross with three generations P's F$_1$'s and F$_2$'s in multiple environments can be appropriately analyzed by the mixed linear approaches Zhu 1996 . The phenotype mean of seed models can be expressed by a mixed linear model as

$$y = Xb + U_A e_A + U_D e_D + U_C e_C + U_{Am} e_{Am}$$
$$+ U_{Dm} e_{Dm} + U_{AE} e_{AE} + U_{DE} e_{DE} + U_{CE} e_{CE}$$
$$+ U_{AmE} e_{AmE} + U_{DmE} e_{DmE} + U_B e_B + e_\varepsilon$$
$$= Xb + \sum_u^{12} U_u e_u \qquad\qquad 39$$

with variance-covariance matrix

$$\text{var } y = \sigma_A^2 V_1 + \sigma_D^2 V_2 + \sigma_C^2 V_3 + \sigma_{Am}^2 V_4$$
$$+ \sigma_{Dm}^2 V_5 + \sigma_{AE}^2 V_6 + \sigma_{DE}^2 V_7 + \sigma_{CE}^2 V_8$$
$$+ \sigma_{AmE}^2 V_9 + \sigma_{DmE}^2 V_{10} + \sigma_B^2 V_{11}$$
$$+ \sigma_{A.Am} V_{12} + \sigma_{D.Dm} V_{13} + \sigma_{AE.AmE} V_{14}$$
$$+ \sigma_{DE.DmE} V_{15} + \sigma_\varepsilon^2 V_{16}$$
$$= \sum_{u=1}^{16} \theta_u V_u$$

where $V_u = U_u U_u^T$ $u = 1$ 2 $\cdots$ 11 $V_{12} = U_1 U_4^T + U_4 U_1^T$ $V_{13} = U_2 U_5^T + U_5 U_2^T$ $V_{14} = U_6 U_9^T + U_9 U_6^T$ $V_{15} = U_7 U_{10}^T + U_{10} U_7^T$ $V_{16} = I$.

Unbiased estimation of variances and covariances can be obtained by MINQUE 0/1 approaches. Random effects can be predicted by LUP or AUP method. Quantitative traits of seed

nutrition were studied by using the seed models for rice Shi et al. 1997 cotton Zhu et al. 1997 barley Yan X. F. et al. 1998 and corn Lou et al. 1998 .

QTL mapping models

Mixed linear model approaches were often used for mapping QTLs of animal populations based on animal model or reduced animal model Grignola et. al. 1997

$$y = Xb + U_A e_A + U_Q e_Q + e_\varepsilon$$
$$\sim N Xb V = \sigma_A^2 U_A R_A U_A^T + \sigma_Q^2 U_Q R_Q U_Q^T$$
$$+ \sigma_\varepsilon^2 R_\varepsilon \qquad\qquad 40$$

where y is the phenotype vector b is a vector of fixed effects X is the design/covariate matrix relating to b $e_A \sim N O \sigma_A^2 R_A$ is the vector of additive effects $e_Q \sim N O \sigma_Q^2 R_Q$ is the vector of QTL allelic effects $e_\varepsilon \sim N O \sigma_\varepsilon^2 R_\varepsilon$ is the vector of residual effects.

Until now there is no appropriate animal model for analyzing QTLs with genetic main effects as well as GE interaction effects. The mixed linear model approaches can be used indirectly for searching QTLs with genetic main effects and QTL $\times$ environment interaction effects Zhu 1998 1999 Yan et al. 1998 .

If QTL mapping experiments are conducted in several environments years and/or locations the phenotype means of genetic entries can be fitted by linear model 32 with random effects for environment effects E genotype effects G genotype $\times$ environment interaction effects GE and residual errors. The matrix form of the mixed linear model is

$$y = I_\mu + U_E e_E + U_G e_G + U_{GE} e_{GE} + e_\varepsilon$$
$$\sim N I\mu V = \sigma_E^2 U_E U_E^T + \sigma_G^2 U_G U_G^T$$
$$+ \sigma_{GE}^2 U_{GE} U_{GE}^T + \sigma_\varepsilon^2 I \qquad\qquad 41$$

The random effects E G and GE are predicted by the AUP method Zhu 1993 Zhu and Weir 1996 . and then used to predict main effect data $y_{j\ G} = \mu + G_j$ on the genotype j across environments and genotype $\times$ environment interaction data $y_{hj\ GE} = \mu + E_h + GE_{hj}$ on the genotype j in environment h respectively.

The composite interval mapping CIM method Zeng 1994 is then applied to analyze the predicted $\hat{y}_{j\ G}$ for searching QTLs with ge-

netic main effects

$$\hat{y}_{j\ G} = \beta_{0\ G} + \beta^*_G X^*_j + \sum_i \beta_{i\ G} X_{ij} + \varepsilon_{j\ G}$$

$$42$$

where $\beta_{0\ G}$ is the population mean β^*_G is the QTL main effect X^*_j is the coefficient for QTL effect $\beta_{i\ G}$ is the main effect for the i-th marker X_{ij} is the coefficient for the i-th marker effect and $\varepsilon_{j\ G}$ is the residual error of the j-th genotype.

The predicted $\hat{y}_{hj\ GE}$ are analyzed also by the CIM method Zeng 1994 for dissecting QTLs with QTL $\times$ E interaction effects in the h-th environment

$$\hat{y}_{hj\ GE} = \beta_{0\ GE_h} + \beta^*_{GE_h} X^*_{hj} + \sum_i \beta_{i\ GE_h} X_{hij}$$
$$+ \varepsilon_{j\ GE_h} \qquad 43$$

where $\beta_{0\ GE_h}$ is the population mean of environment h $\beta^*_{GE_h}$ is the QTL $\times$ E interaction effect of environment h with coefficient X^*_{hj} $\beta_{i\ GE_h}$ is the effect for the i-th marker $\times$ environment h with coefficient X_{hij} and $\varepsilon_{j\ GE_h}$ is the residual error of the j-th genotype in environment h.

Mixed-model-based composite interval mapping MCIM Zhu and Weir 1998 Zhu 1998 can be applied to directly search QTLs with genetic main effects and QTL $\times$ environment interaction effects Zhu 1998 1999 Wang et al. 1999 . When experiments for QTL mapping are conducted in multiple environments the phenotype value of the j-th genetic entry in environment h can be expressed by a mixed linear model

$$y_{hj} = \mu + ax_{A_j} + dx_{D_j} + u_{E_{hj}} e_{E_h} + u_{AE_{hj}} e_{AE_h}$$
$$+ u_{DE_{hj}} e_{DE_h} + \sum_f u_{M_{fj}} e_{M_f} + \sum_l u_{ME_{hlj}} e_{ME_{hl}}$$
$$+ \varepsilon_{hj} \qquad 44$$

where μ is the population mean a and d are the fixed additive and dominance effects of QTL respectively x_{A_j} and x_{D_j} are coefficients for genetic main effects $e_{E_h} \sim N\ 0\ \sigma^2_E$ is the effect of environment h with coefficient $u_{E_{hj}}$ $e_{AE_h} \sim N$ $0\ \sigma^2_{AE}$ is the additive $\times$ environment interac-

tion effect with coefficient $u_{AE_{hj}}$ $e_{DE_h} \sim N\ 0$ σ^2_{DE} is the dominance $\times$ environment interaction effect with coefficient $u_{DE_{hj}}$ $e_{M_f} \sim N\ 0$ σ^2_M is the marker main effect with coefficient u_{M_f} $e_{ME_{hl}} \sim N\ 0\ \sigma^2_{ME}$ is the marker $\times$ environment interaction with coefficient $u_{ME_{hlj}}$ $\varepsilon_{hj} \sim$ $N\ 0\ \sigma^2_\varepsilon$ is the random residual effects.

This model 44 can be expressed in matrix form

$$y = Xb + U_E e_E + U_{AE} e_{AE} + U_{DE} e_{DE} + U_M e_M$$
$$+ U_{ME} e_{ME} + e_\varepsilon$$
$$= Xb + \sum_{u=1}^6 U_u e_u$$
$$\sim N\ Xb\ V = \sum_{u=1}^6 \sigma^2_u U_u R_u U^T_u \ . \qquad 45$$

where y is the phenotype vector b is the fixed parameter vector for population mean and QTL effects X is the known incidence matrix of the fixed parameters $e_1 = e_E \sim N\ 0\ \sigma^2_E I$ is the vector of environment effects $e_2 = e_{AE} \sim N\ 0$ $\sigma^2_{AE} I$ is the vector of $A \times E$ interaction effects $e_3 = e_{DE} \sim N\ 0\ \sigma^2_{DE} I$ is the vector of $D \times E$ interaction effects $e_4 = e_M \sim N\ O\ \sigma^2_M R_M$ is the vector of marker main effects $e_5 = e_{ME} \sim N$ $O\ \sigma^2_{ME} R_{ME}$ is the vector of $M \times E$ interaction effects $e_6 = e_\varepsilon \sim N\ O\ \sigma^2_\varepsilon I$ is the vector of residual effects $U_u\ u = 1\ 2\ \dots\ 6$ are the known incidence matrix of the random effects and $U_6 = I$.

QTLs with epistasis main effects and epistasis $\times$ environment interaction effects can also be analyzed by the MCIM method Zhu 1998 1999 . If DH or RIL populations are used for mapping QTLs with additive and additive $\times$ additive epistasis effects as well as their environment interaction effects the phenotype value of the j-th genetic entry in environment h can be expressed as the following mixed linear model

$$y_{hj} = \mu + a_1 x_{A_{1j}} + a_2 x_{A_{2j}} + aax_{AA_j} + u_{E_{hj}} e_{E_h}$$
$$+ u_{A_1 E_{hj}} e_{A_1 E_h} + u_{A_2 E_{hj}} e_{A_2 E_h}$$
$$+ u_{AAE_{hj}} e_{AAE_h} + \sum_f u_{M_{fj}} e_{M_f} + \sum_l u_{MM_{lj}} e_{MM_l}$$
$$+ \sum_p u_{ME_{hpj}} e_{ME_{hp}} + \sum_q u_{MME_{hqj}} e_{MME_{hq}} + \varepsilon_{hj}$$
$$46$$

where μ is the population mean a_1 and a_2 are the additive effects of loci Q_1 and Q_2 respectively aa is the additive $\times$ additive epistasis effect of loci Q_1 and Q_2 $x_{A_{1j}}$ $x_{A_{2j}}$ and x_{AA_j} are coefficients of these genetic main effects e_{E_h} are the random effects of environment h with coefficient $u_{E_{hj}}$ $e_{A_1 E_h}$ or $e_{A_2 E_h}$ are the additive $\times$ environment interaction effects with coefficient $u_{A_1 E_{hj}}$ or $u_{A_2 E_{hj}}$ for Q_1 or Q_2 e_{AAE_h} is the epistasis $\times$ environment interaction effect with coefficient $u_{AAE_{hj}}$ e_{M_f} is the marker main effect with coefficient u_{M_f} e_{MM_l} is the marker $\times$ marker interaction effect with coefficient u_{MM_l} $e_{ME_{hp}}$ is the marker $\times$ environment interaction effect with coefficient $u_{ME_{hpj}}$ $e_{MME_{hq}}$ is the marker $\times$ marker $\times$ environment interaction effect with coefficient $u_{MME_{qhj}}$ ε_{hj} is the residual effect.

This model 46 can be expressed in matrix form

$$y = Xb + U_E e_E + U_{A_1 E} e_{A_1 E} + U_{A_2 E} e_{A_2 E}$$
$$+ U_{AAE} e_{AAE} + U_M e_M + U_{MM} e_{MM}$$
$$+ U_{ME} e_{ME} + U_{MME} e_{MME} + e_\varepsilon$$
$$= Xb + \sum_{u=1}^{9} U_u e_u$$
$$\sim N \ Xb \ V = \sum_{u=1}^{9} \sigma_u^2 U_u R_u U_u^T \ . \qquad 47$$

where y is the phenotype vector b is the fixed parameter vector for population mean and QTL effects X is the known incidence matrix of the fixed parameters $e_1 = e_E \sim N \ O \ \sigma_E^2 I$ is the vector of environment effects $e_2 = e_{A_1 E} \sim N \ O \ \sigma_{A_1 E}^2 I$ is the vector of $A_1 \times E$ interaction effects $e_3 = e_{A_2 E} \sim N \ O \ \sigma_{A_2 E}^2 I$ is the vector of $A_2 \times E$ interaction effects $e_4 = e_{AAE} \sim N \ O \ \sigma_{AAE}^2 R_{AAE}$ is the vector of $AA \times E$ interaction effects $e_5 = e_M \sim N \ O \ \sigma_M^2 R_M$ is the vector of marker main effects $e_6 = e_{MM} \sim N \ O \ \sigma_{MM}^2 R_{MM}$ is the vector of interaction marker main effects $e_7 = e_{ME} \sim N \ O \ \sigma_{ME}^2 R_{ME}$ is the vector of $M \times E$ interaction effects $e_8 = e_{MME} \sim N \ O \ \sigma_{MME}^2 R_{MME}$ is the vector of $MM \times E$ interaction effects $e_9 = e_\varepsilon \sim N \ O \ \sigma_\varepsilon^2 I$ is the vector of residual effects $U_u \ u = 1 \ 2 \ \dots \ 8$

are the known incidence matrix of the random effects and $U_9 = I$.

DISCUSSION

Complex quantitative traits consist of genetic effects more than simple additive and dominance effects. As genetic models become more and more complicated for fitting the biological situations total phenotypic variance can be partitioned into various ways. Therefore some definitions of traditional quantitative parameters might need to be updated too.

Heritability is an important parameter widely used in quantitative genetics as well as plant and animal breeding but different definitions should be assigned in dissimilar genetic models. Quantitative traits can be controlled by genetic main effects and GE interaction effects. Accordingly the total heritability h^2 can be partitioned into general heritability h_G^2 and interaction heritability h_{GE}^2 Zhu 1997 . General heritability which is applicable to multiple environments is defined as the ratio of variances of accumulated inheritable genotypic effects to phenotypic variance. Interaction heritability which is only applicable to specific environments is defined as the ratio of variances of accumulated inheritable GE interaction effects to phenotypic variance. In seed genetic models variances with accumulated effects consist of direct additive variance V_A cytoplasm variance V_C maternal additive variance V_{Am} as well as variances due to gene by environment interactions. General heritability consists of direct general heritability h_O^2 cytoplasm general heritability h_C^2 and maternal general heritability h_M^2 Zhu 1996

$$h_G^2 = h_O^2 + h_C^2 + h_M^2$$
$$= \ V_A + C_{A.Am} \ /V_p + V_c/V_p + \ V_{Am}$$
$$+ C_{A.Am} \ /V_p$$

Interaction heritability includes direct interaction heritability h_{OE}^2 cytoplasm interaction heritability h_{CE}^2 and maternal interaction heritability h_{ME}^2

$$h_{GE}^2 = h_{OE}^2 + h_{CE}^2 + h_{ME}^2$$
$$= \ V_{AE} + C_{AE.AmE} \ /V_p + V_{CE}/V_p + \ V_{AmE}$$

$$+ C_{AE.AmE} / V_p$$

with total phenotypic variance calculated as

$$
\begin{aligned}
V_p &= V_{Go} + V_C + V_{Gm} + 2C_{Go.Gm} + V_{GoE} \\
&\quad + V_{CE} + V_{GmE} + 2C_{GoE.GmE} + V_e \\
&= V_A + V_D + V_C + V_{Am} + V_{Dm} \\
&\quad + 2C_{A.Am} + 2C_{D.Dm} + V_{AE} + V_{DE} \\
&\quad + V_{CE} + V_{AmE} + 2C_{AE.AmE} + 2C_{DE.DmE} \\
&\quad + V_e
\end{aligned}
$$

Heritability is often used in predicting selection response. Since heritability consists of several components for some complicated models the definition for selection response should also be changed. For seed models with *GE* interaction total selection response $R = ih^2 \sqrt{V_p}$ can be partitioned into several components Zhu 1997

$$
\begin{aligned}
R &= R_G + R_{GE} \\
&= R_O + R_C + R_M + R_{OE} + R_{CE} + R_{ME}
\end{aligned}
$$

where $R_G = ih_G^2 \sqrt{V_p}$ is general response which consists of direct general response $R_O = ih_O^2 \sqrt{V_p}$ cytoplasm general response $R_C = ih_C^2 \sqrt{V_p}$ and maternal general response $R_M = ih_M^2 \sqrt{V_p}$ $R_{GE} = ih_{GE}^2 \sqrt{V_p}$ is interaction response which consists of direct interaction response $R_{OE} = ih_{OE}^2 \sqrt{V_p}$ cytoplasm interaction response $R_{CE} = ih_{CE}^2 \sqrt{V_p}$ and maternal interaction response $R_{ME} = ih_{ME}^2 \sqrt{V_p}$.

In order to analyze complicated genetic effects such as epistasis effects endosperm effects etc. segregating generations of F_2 BC1 and BC2 can be included in the genetic models which have non-integer coefficients for some effects and even correlation between factors. These kinds of genetic models are usually analyzable not by ANOVA approaches but by mixed linear model approaches. Most animal geneticists use the REML method Patterson and Thompsom 1971 instead of MINQUE method Rao 1971 for analyzing animal models. Monte Carlo simulations Zhu 1989 Zhu and Weir 1994a 1994b 1996 showed that MINQUE has advantages over REML due to its A simple computation without iterations B no requirement for normality distribution and C

unbiased estimation.

The jackknife resampling method is efficient for calculating estimates or predictors and their standard errors. Resampling technique is based on resampling unit. For genetic experiments with randomized complete block design in multiple environments blocks within environments can serve as resampling units. If there are only replications but not blocks genetic entries can serve as resampling units.

References

Atchley W. R. and Zhu J. 1997. Developmental quantitative genetics conditional epigenetic variability and growth in mice. *Genetics* **147** 765-776.

Bink M. C. A. M. Quass Q. L. and Van Arendonk J. A. M. 1998. Bayesian estimation of dispersion parameters with a reduced animal model including polygenic and QTL effects. *Genet. Sel. Evol.* **30** 103 – 125.

Cockerham C. C. and Weir B. S. 1977. Quadratic analysis of reciprocal crosses. *Biometrics* **33** 187 – 203.

Cockerham C. C. 1980. Random and fixed effects in plant genetics. *Theor Appl Genet* **56** 119 – 131.

Comstock R. E. and Robinson H. F. 1952. Estimation of average dominance of genes. *In* Heterosis. Gowan J. W. ed. Iowa State University Press Ames Iowa.

Eisen E. J. Bohren B B. and McKean H E. 1966. Sex-linked and maternal effects in the diallel cross. *Aust. J. Biol. Sci.* **19** 1061 – 1071.

Fisher R. A. 1925. Statistical Methods for Research Workers. 1st ed. Oliver & Boyd. Edinburgh and London.

Gardner C. O. and Eberhart S. A. 1966. Analysis and interpretation of the variety cross diallel and related populations. *Biometrics* **22** 439 – 452.

Gelfan A. E. and Smith A. F. M. 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat Assoc* **85** 398 – 409.

Griffing B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.* **9** 463 – 493.

Grignola F. E. Zhang Q. and Hosechele I. 1997. Mapping linked quatitative trait loci via residual maximum likelihood. *Genetics Seletion Evolution* **29** 6 529 – 544.

Hallauer A. R. and Miranda J. B. 1981. Quantitative Genetics in Maize Breeding. Iowa State University Press Ames Iowa.

Hartley H. D. and Rao J. N. K. 1967. Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54** 93 – 108.

Henderson C. R. 1963. Selection index and expected genetic advance. *In* Statistical Genetics and Plant Breeding. Hanson W. D. and Robinson H. F. eds.

Washington. National Academy of Science National Research Council. Publication 982 p.141 – 163.

Henderson C. R. 1988. Progress in statistical methods applied to quantitative genetics since 1976. *In* Procceedings of the Second Interactional Conference on Quantitative Genetics. Weir B. S. Eisen E. J. Goodman M. M. et al. eds. Sinouer Assoc. Sunderland MA. p.85 – 90.

Jansen R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* **135** 205 – 211.

Lou X. Y. Zhu J. Zhang Q. D. et al. 1998. A genetic study on the effects of Opaque – 2 and polygenes on kernel quality traits in maize. *In* Advanced Topics in Biomathematics Procceedings of International Conference on Mathematical Biology. Chen L. S. Ruan S. G. and Zhu J. eds. World Scientific Publishing Co. Singapore. p.143 – 148.

Lynch M. and Walsh B. 1998. Genetics and Analysis of Quantitative Traits. Sinouer Assoc. Sunderland MA.

Macedo F. W. M. and Gianola D. 1987. Bayesian analysis of univariate mixed models with informative priors. Eur. Assoc. Anim Prod 38th annu Meet Lisbon Portugal.

Miller R. G. 1974. The jackknife a review. *Biometrika* **61** 1 – 15.

Patterson H. D. and Thompson R. 1971. Recovery of inter – block information when block sizes are unequal. *Biometrika* **58** 545 – 554.

Rao C. R. 1971. Estimation of variance and covariance components MINQUE theory. *Journal of Multivariate Analysis* **1** 257 – 275.

Rao C. R. and Kleffe J. 1980. Estimation of variance components. *In* Handbook of Statistics. Krishnaiah P. R. ed.. North-Holland New York. Vol.1 p.1 – 40.

Searle S. R. Casella G. and McCulloch C. E. 1992. Variance Components. Wiley New York.

Shi. C. H. Zhu J. Zeng R. C. et al. 1997. Genetic and heterosis analysis for cooking quality traits of indica rice in different environments. *Theor. Appl. Genet.* **95** 294 – 300

Wang T. Fernando R. L. and Grossman M. 1998. Genetic evaluation by best lineaar unbiased prediction using marker and trait information in a multibreed population *Genetics* **148** 1 507 – 515.

Wang C. S. Rutledge J. J. and Gianola D. 1994. Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* **26** 91 – 115.

Wang D. L. Zhu J. Li Z. K. et al. 1999. Mapping QTLs with epistatic effects and QTL × environment interactions by mixed linear model approaches. *Theor. Appl. Genet.* **99** 1255 – 1264.

Wu J. X. Zhu J. Ji D. F. et al. 1995. Genetic

analysis for heterosis of fiber traits in highland cotton. *Acta Gossypii Sinica* **7** 4 217 – 222. in Chinese.

Yan X. F. Xu S. Y. Xu Y. H. et al.. 1998. Genetic investigation of contributions of embryo and endosperm genes to malt kolbach index alpha-amylase activity and wort nitrogen content in barley. *Theor. Appl. Genet.* **96** 5 709 – 715.

Yan J. Q. Zhu J. He C. X. et al. 1998. Molecular dissection of developmental behavior of plant height in rice Oryza sativa L.. *Genetics* **150** 1257 – 1265.

Yates F. 1947. Analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity* **1** 287 – 301.

Zeng Z. B. 1994. Precision mapping of quantitative trait loci. *Genetics* **136** 1457 – 1468.

Zhu J. 1989. Estimation of genetic variance components in the general mixed model. Ph. D. Dissertation NC State University Raleigh U.S.A.

Zhu J. 1993a. Methods of predicting genotype value and heterosis for offspring of hybrids. *Journal of Biomathematics* **8** 1 32 – 44. in Chinese.

Zhu J. 1993b. Mixed model approaches for estimating covariances between two traits with unequal design matrices. *Journal of Biomathematics* **8** 3 24 – 30. in Chinese.

Zhu J. 1994. General genetic models and new analysis methods for quantitative traits. *Journal of Zhejiang Agricultural University* **20** 6 551 – 559. in Chinese.

Zhu J. 1996. Analysis methods for seed models with genotype × environment interactions. *Acta Genetica Sinica* **23** 1 56 – 68. in Chinese.

Zhu J. 1998. Mixed model approaches of mapping genes for complex quantitative traits. *In* Procceedings of Genetics and Crop Breeding of China. L. Z. Wang and J. R. Dai eds. Chinese Agricultural Science and Technology Publication House Beijing p.19 – 20.

Zhu J. 1999. Mixed model approaches of mapping genes for complex quantitative traits. *Journal of Zhejiang University Natural Science* **33** 3 327 – 335. in Chinese.

Zhu J. and Weir B. S. 1994a. Analysis of cytoplasmic and maternal effects I. a genetic model for diploid plant seeds and animals. *Theor Appl Genet* **89** 153 – 159.

Zhu J. and Weir B. S. 1994b. Analysis of cytoplasmic and maternal effects II. genetic models for triploid endosperm. *Theor Appl Genet* **89** 160 – 166.

Zhu J. and Weir B. S. 1996. Diallel analysis for sex-linked and maternal effects. *Theor Appl Genet* **92** 1 1 – 9.

Zhu J. Wang G. J. and Zhang R. C. 1997. Genetic analysis on gene effects and GE interaction effects for kernel nutrient quality traits of highland cotton. *Journal of Biomathematics*. **12** 2 111 – 120. in Chinese