Theor Appl Genet (2008) 117:65–73
DOI 10.1007/s00122-008-0753-z

ORIGINAL PAPER

REML approach for adjusting the *Fusarium* head blight rating to a phenological date in inoculated selection experiments of wheat

K. Emrich · F. Wilde · T. Miedaner · H. P. Piepho

Received: 3 July 2006 / Accepted: 17 March 2008 / Published online: 5 April 2008

Abstract *Fusarium* head blight is one of the most important wheat diseases causing grain yield and quality losses as well as mycotoxin contamination all over the world. Since *Fusarium* cannot be reliably controlled with fungicides, breeding has become a favorable tool to decrease the infection severity. In most cases, selection for *Fusarium* resistance is done by artificial infection in the field. However, there is a risk in preferring late heading genotypes, because heading of wheat is negatively correlated to head blight severity. Because an indirect selection for late maturity is not intended, we considered a statistical approach to avoid this problem. In this paper, we propose a mixed model to analyze extensive *Fusarium* head blight rating in resistance breeding experiments of wheat. The objective of the analysis was to select for *Fusarium* resistance, while at the same time ensuring that late heading genotypes, which show less head blight over the shorter vegetation period, are not preferred. Thus, selection was to be done such that genetic variability for heading date was retained. Therefore, the statistical model contained a covariate to adjust for differences in the heading date. The use of covariate adjustment is an easily handled alternative to a bivariate analysis. Covariate adjustment will in practice often work almost equally well as bivariate analysis. Any statistical software with powerful mixed model analysis tools can be used for this type of analysis. We propose an ad hoc method to obtain heritability estimates and a form of LSD (least significance difference) as a measure of accuracy on the basis of the proposed model and under special consideration of the experimental design. The ad hoc LSD was used as a rough measure to judge rankings of genotypic means (BLUPs). Friedman's super smoother was used to compare smoothed rank estimates for adjusted and unadjusted genotypes against increasing smoothed heading dates. Traits were transformed to meet the model assumptions, especially homogeneity of errors and normality, and back-transformation of means and standard errors was conducted by using the delta method.

Communicated by F. van Eeuwijk.

K. Emrich (✉) · H. P. Piepho
Bioinformatics Unit (340c), Universität Hohenheim,
Fruwirthstrasse 23, 70599 Stuttgart, Germany
e-mail: emrich@uni-hohenheim.de

F. Wilde · T. Miedaner
State Plant Breeding Institute (720),
University of Hohenheim,
Fruwirthstrasse 21, Stuttgart, Germany

Introduction

Fusarium head blight (FHB), primarily caused by *Fusarium graminearum* and *F. culmorum*, is one of the most important wheat diseases leading to grain yield and quality losses as well as mycotoxin contamination. FHB can be only partially controlled by fungicide spraying whereas the timing and application are critical (McMullen et al. 1997). The best treatments may reduce FHB severity up to 50% associated with the reduction of damaged kernels, but mycotoxin contamination forms a constant threat to the food chain. Planting resistant cultivars is an alternative method to control the disease. Phenotypic selection of superior progeny in the field is a common resistance breeding procedure. It is well documented that environmental conditions, directly after inoculation, are likely to have a high influence on the occurrence and severity of FHB (Miedaner et al. 2001). The time of maximal

susceptibility is mid-flowering. However, flowering time among large progeny usually varies and the length of flowering period depends on the origin of the material and the differences between the crossing parents. Especially when using exotic resistance donors, the span of flowering might become quite wide. In early generations, flowering cannot be taken into account because of the high segregational variation within plots. This may lead to misinterpretation of the results because in many cases genotypes with early heading dates will show more symptoms than genotypes with later heading dates, leading to a correlation between FHB severity and flowering time ($r = -0.19$, Buerstmayr et al. 2000) or heading date ($r = -0.43$, Schmolke et al. 2005). Because mid-flowering is difficult to assess, the heading date may be used as an indicator of the developmental stage of the plant. Some authors try to avoid this problem by inoculating each genotype individually at its respective flowering date (e.g., Buerstmayr et al. 2000). This technique is laborious and cost intensive, especially when selection is conducted in large populations. Additionally, this method results in several inoculation dates and, consequently, different weather conditions at the time of inoculation. Instead, we propose an approach to inoculate and score all genotypes several times at the same date and adjust statistically for differences in heading date, when all genotypes are planted simultaneously. With the same method, covariation between FHB severity and plant height could also be adjusted if necessary. The latter covariation often occurs in experimental studies amounting to correlations of up to $r = -0.37$ (Buerstmayr et al. 2000; Schmolke et al. 2005). In experiments with exotic resistance donors, even a correlation of $r = -0.55$ was reported in barley (Buerstmayr et al. 2004).

As the heading date (HD) of wheat is (negatively) correlated with *Fusarium* head blight (FHB), a selection on the basis of unadjusted average FHB rating will favor varieties with a late heading date. It should be considered that the FHB severity, heading date and plant height are in most cases inherited by different quantitative trait loci (QTL, e.g., Buerstmayr et al. 2003; Schmolke et al. 2005). Their relationships are mainly based on developmental or epidemiological conditions that should not bias the selection decision of the breeders. We will discuss an experiment designed to breed for FHB-resistant genotypes by selection across several environments. Selection is done on the basis of ranking of estimated means, which are adjusted for an average heading date. This ensures that favoring of late heading varieties does not take place in a regular way. Essentially, selection is based on the genotypic deviation from a regression on heading date, thus ensuring that for each flowering time, genotypes with the best resistance are retained. The approach could be used to compare genotypes between all earliness classes, but this could result in the use of unplausible predicted combinations between earliness and resistance. Therefore, in practical applications where this kind of unplausibility may occur, we recommend that only genotypes in similar earliness (i.e., covariate) classes are compared.

The methods proposed in this paper can be easily adapted to other selection experiments. This is an important issue when the methods are to be used broadly in practical selection work, which usually involves large plant populations and often needs to be performed under time pressure. Traits of interesting properties can be correlated with other quantitative traits. An analysis of genotypes on the basis of the "main" trait may result in selection of varieties from a less desirable subpopulation (e.g., late heading resistant varieties) of the secondary trait. This undesirable side-effect can be avoided by using the method described here.

This paper is organized as follows. In the following section, we will introduce the genetical architecture and randomization structure of a large breeding experiment. The second part of "Materials and methods" presents the statistical methodology used to analyze this experiment. "Results" discusses the analysis of the experiment and is followed by the discussion of the methods introduced.

Materials and methods

Field trials

We will discuss a series of field experiments conducted in four environments in the north and south of Germany. All crops were artificially inoculated by spraying conidia suspensions of *F. culmorum* at a density of 500,000 conidia ml^{-1} by a machine-driven field sprayer. Inoculation was performed three to five times for all genotypes of an experiment, according to their length of flowering period. For more details see Miedaner et al. (2006). Rating was done on a plot basis by visually estimating the percentage of diseased spikelets (0–100%) at three to four times during pathogenesis and calculating an average FHB rating.

A large population of about 1,200 winter wheat genotypes was tested. The genotypes were derived from a recurrent selection program based on a double cross [RIL1 (G16-92/Hussar)/Brando//RIL2 (Dream/Lynx)/LP235.1] and consisted of three subpopulations. One subpopulation (R) consisted of 50 unselected progenies. The second subpopulation (P) underwent one cycle of phenotypic selection and comprised about 1,000 genotypes. The third subpopulation (M) of 113 genotypes resulted from marker-assisted selection (MAS) based on three diallelic QTL. The three QTL gave rise to eight marker classes (groups).

Marker class M1 has all three tested QTL combined, M2–M4 have two QTL, M5–M7 have one and M8 has none. For the first two subpopulations (R and P) there was no such subdivision, so each of these may be regarded as a single group. Thus, overall, there were ten genetically different groups of genotypes of common origin.

The randomization of genotypes was hierarchical in two steps. In the first step, the 1,200 genotypes were subdivided into 6 sets of similar size. The first set of 200 genotypes contained subpopulations R and M and standard varieties. Subpopulation P, which comprised 1,000 genotypes, was split into five sets of 200. Each set was randomized separately according to an α-lattice with block size ten (Patterson and Williams 1976; Paterson and Patterson 1984). The first set had four replicates, while the other five sets had two replicates. A replicate was randomized as a main plot in a split-plot design. The total of 14 ($1 \times 4 + 5 \times 2$) main plots resulting from the six α-lattices were completely randomized. The different α-lattices were connected by four common standards (the four parents of the original double cross) planted five times per replication. The design is not a standard one, but variations of it are often used in practical breeding applications (Piepho et al. 2006b). In our case one reason for the choice of this design was that unequal replication was desired for the subpopulations and that the hierarchical randomization was easy to implement using standard procedures. It is stressed, here, however, that powerful software is available to randomize very complex breeding trials according to various experimental designs (Whitaker et al. 2002).

The main purpose of this experiment was to compare the eight class means for the marker classes of subpopulation M among one another and with the other two subpopulation means (R and P). A further objective was to obtain good point estimates of the genetic values of individual genotypes within subpopulations for selection.

The mixed model approach discussed in the following section can accommodate the unbalance, which results from heterogeneous replication of genotypes under the full hierarchical lattice design.

Statistical methods

Overview of methods

Standard analyses of breeding experiments usually yield least significant differences (LSD) and heritabilities. We will discuss methods to calculate these measures for mixed models. The methods are illustrated using examples from resistance breeding, where the main aim is the selection of FHB-resistant genotypes. Nonetheless, all methods discussed here can be generalized easily to other situations.

Statistical model

Following Piepho et al. (2003), we partition the model into block and treatment components. The experimental design is included in the model for statistical analysis.

The statistical model is presented for lattice/hierarchical lattice design and is generalized easily to other situations, as done exemplarily in the second part of the example. The model is

$$y_{ijkmn} = \mu + d_0 \times g_i + a_0 \times x_{ikm} + \tau_j + l_k + d_0 \times k_{ik} + r_{km} + b_{kmn} + e_{ijkmn} \quad (1)$$

y_{ijkmn}	observation of (transformed) trait
μ	general mean
τ_j	fixed effect for standards, coding standard and non-standard genotypes ($j = 1, \ldots, s$ for $s - 1$ standards and one level for non-standards)
d_0	dummy with 1 for non-standards and 0 for standards
g_i	random genotypes effect with variance σ^2_{GT}, standard genotypes are blocked out by regression on the dummy d_0
a_0	regression coefficient
x_{ikm}	covariate for heading dates
l_k	random effect of location k, with variance σ^2_{L}, ($k = 1, \ldots, l$)
k_{ik}	random interaction for genotypes × location with variance $\sigma^2_{\mathrm{GT\times L}}$, standard genotypes are blocked out by regression on the dummy d_0
r_{km}	random effect for replication/large (incomplete) block effect for hierarchical lattice m in location k with variance σ^2_{B}
b_{kmn}	random incomplete block effect (in location k and replicate/large block m) with variance σ^2_{IB}
e_{ijkmn}	random error, $\sim N\left(0, \sigma^2_{\mathrm{e}}\right)$.

The purpose of the dummy variable d_0 is to block out standards from the random part of the model (Piepho et al. 2006b). The random effects have expectation zero and the subscripts on the variance components are described as follows: GT, genotype; GT × L, genotype in locations; B, replication block; IB, incomplete block; e, random error.

For an RCBD, r_{km} is the complete block effect and b_{kmn} is dropped.

Delta method for back-transformation of standard errors in original scale

The genotype means and pairwise comparisons of genotype means with standard errors are back-transformed to the original scale using the delta method, also referred to as

method of statistical differentials, based on Taylor approximation (cf. Hinkelmann and Kempthorne 1994; Stuart and Ord 1994 and formulas in "Appendix"). For clarity, however, the graphic presentation of means with LSD is done on the transformed scale and shown in combination with results on a back-transformed scale. It is stressed that especially in the case of non-linear transformations like the logit transformation, the back-transformation of standard errors of pairwise comparisons of means by delta method may result in upwardly biased standard errors on the original scale and should be handled with care.

Ad hoc least significant difference (LSD)

Breeders have used LSD in standard analysis over many years as a measure to decide if genotypic values differ significantly (Miedaner et al. 2004). On this basis, a ranking of genotypic values is done and the LSD serves as a simple measure of accuracy of genotypic effect estimates and provides a "rule of thumb" for comparison of genotypes. The LSD is based on an analysis with fixed genotypic effects, with estimation by the best linear unbiased estimation (BLUE). However, it is emphasized, that the best linear unbiased predictors (BLUPs) are optimized for ranking and selection (Piepho and Möhring 2006). In general, it is reasonable to make inferences for BLUP in much the same way as for BLUE (Kackar and Harville 1984). In either case, it must be realized, however, that the null hypothesis of a test of equality of genotypic effects can usually be ruled out on a priori grounds. Thus, for two treatment means A and B, say, it is usually known that $A \neq B$. Clearly, when genotypic effects are assumed to come from a normal distribution, the probability of two genotypes being exactly equal is zero. The real question is whether $A > B$ or $B > A$ (Tukey 1991). When a test rejects H_0: $A = B$, there is a basis for deciding among these two alternatives, otherwise the data are inconclusive (Hsu 1996).

For a large number of genotypes, conducting all pairwise tests is neither practical nor useful, mainly due to the multiplicity problem involved (Hochberg and Tamhane 1987). The LSD may simply be taken as a measure of accuracy to aid selection decisions in borderline cases. For BLUP we propose ad hoc LSD computed as two times the average standard error of differences. Let $\boldsymbol{A}$ be the variance–covariance matrix of least square means or of BLUPs. Then the average variance of genotype differences may be computed as (Bueno and Gilmour 2003):

$$\bar{v} = \frac{t \times \text{trace}(\mathbf{A}) - \mathbf{1}'\mathbf{A}\mathbf{1}}{t(t-1)/2} \tag{2}$$

where t is the number of treatments. We also apply this formula for calculation of the ad hoc heritability described below. The ad hoc LSD is presented on the transformed scale where normality and homoscedasticity are assumed to apply.

Heritability in unbalanced designs with covariates

Heritability is used in breeding experiments as a measure to assess the power of the selection process across locations and years. Broad-sense heritability should be calculated on the scale where normality applies, i.e., if the data is transformed to fulfill the assumptions, heritability should be calculated on the transformed scale. This ensures that standard formula such as the selection gain equation can be applied.

A conventional block design with r blocks that are replicated at l locations has a heritability of:

$$h^2 = \frac{\sigma^2_{\text{GT}}}{\sigma^2_{\text{GT}} + \sigma^2_{\text{GT}\times\text{L}}/l + \sigma^2_{\text{e}}/(l \times r)} \tag{3}$$

with $\sigma^2_{\text{GT}}, \sigma^2_{\text{GT}\times\text{L}}, \sigma^2_{\text{e}}$, l and r are defined as in (1). However, this formula includes insufficient information about the randomization scheme if used in more complex designs like lattice designs and if we include covariates.

Until recently, no formal method has been developed to include other randomization structures adequately into the calculation of heritabilities, but some such procedures have been suggested recently (Oakey et al. 2006; Piepho and Möhring 2007; Cullis et al. 2006). We propose an ad hoc method to calculate heritabilities for mixed models with covariates and complex experimental designs. The method is based on the fact that the difference of denominator and numerator in the usual equation for h^2 in (3) equals half the squared standard error of a difference of two genotype means based on a mixed model with fixed genotype effects.

Consider a randomized complete block design (RCBD), replicated at l locations, with r blocks and t fixed genotypes (GT). The model (without covariate) is given as:

$$y_{ijkm} = \mu + \gamma_i + \tau_j + l_k + d_0 \times k_{ik} + b_{km} + e_{ijkm}. \tag{4}$$

Effects are defined as before, but genotype effects γ_i are taken as fixed. Then the heritability is calculated as in (3). The variance of differences of genotype means, divided by two, is exactly equal to

$$\text{Var}(\hat{\gamma}_i - \hat{\gamma}_j)/2 = \sigma^2_{\text{GT}\times\text{L}}/l + \sigma^2_{\text{e}}/(l \times k),$$
$$(i,\ j \in \{1,\ldots,t\},\ i \neq j).$$

Therefore, the heritability can be expressed as

$$h^2 = \frac{\sigma^2_{\text{GT}}}{\sigma^2_{\text{GT}} + \overline{\text{Var}(\hat{\gamma}_i - \hat{\gamma}_j)/2}}.$$

Our models are unbalanced, because we use covariates in a lattice design with (or without) standards and a complex

randomization and genetic structure. Nonetheless, we may use the average of $\mathrm{Var}(\hat{\gamma}_i - \hat{\gamma}_j)/2$ in equivalent models to define an ad hoc measure of heritability for mixed model (1) with covariate and a block or (hierarchical) lattice design with random genotypes. $\overline{\mathrm{Var}(\hat{\gamma}_i - \hat{\gamma}_j)/2}$ is calculated as mean of the squared standard error of differences of least squares means of (fixed) genotypes. Our approach differs from the recently proposed by Oakey et al. (2006) and Cullis et al. (2006), who define heritability measures based on $\overline{\mathrm{Var}(\hat{\gamma}_i - \hat{\gamma}_j)/2}$ for BLUP.

Comparison of genotype ranking with and without adjustment for large experiments

Friedman and Stuetzle (1981, 1982) propose a local averaging smoother (running lines smoother), called Super Smoother, to combine a graphical model fit with a data-dependent automated choice of appropriate (and variable) spans by cross validation. The Super Smoother shows good properties for highly variable scatter plots (Friedman and Silverman 1989) and is used in the example to account for high variability of ranks of genotypes over heading dates. It produces smoothed rank estimates for adjusted and unadjusted genotypes against increasing smoothed heading dates.

Results

Genotypic analysis. Percent values for FHB were logit-transformed by $\mathrm{logit}(\mathrm{FHB}/100) = \log[\mathrm{FHB}/100/(1 - \mathrm{FHB}/100)]$. The formula of back-transformation of means and average standard errors ($\mathrm{se_a}$) is given in "Appendix". We provide a graphical presentation of means on the transformed and original scale.

The ad hoc LSD of genotype comparisons is calculated as LSD = 0.3. Ad hoc heritability on the basis of genotypic information is calculated as $h^2 = 0.90$. Both measures are calculated on the transformed (logit) scale.

Table 1 gives the variance components for the random effects, which are all significant. The *F* tests for standards and the covariate are significant (*P* values not shown). A test for heterogeneity of regression slopes was not significant.

Figure 1 compares the smoothed ranks of genotype means per increasing smoothed average heading date for the adjusted and unadjusted approach. It shows distinctly that adjustment results in genotype ranks per heading date, which decrease for early heading genotypes and increase for late heading genotypes. Figure 1 shows exactly the tendency that was observed in the estimated genotypic means of the adjusted and unadjusted approach (the two sets of 1,200 genotype means and their ranks are not shown): nearly all ranks of adjusted genotypes of early heading varieties had smaller ranks than the unadjusted genotypes, and equivalently higher ranks for nearly all late heading genotypes were observed.

Table 1 Variance components of *Fusarium* head blight (FHB) rating (0–100%) of 1,000 winter wheat genotypes inoculated with *Fusarium culmorum* across four locations in hierarchical lattice experiment (experiment a)

Covariance parameters	Estimates
loc	0.028
loc × b	0.00012
loc × b × b2	0.000021***
gt × s2	0.231***
loc × gt × s2	0.029***
Residual	0.1844

*** *P* value < 0.0001

A histogram in Fig. 2 gives an impression of the variability of adjusted genotype means on the transformed and original scale. The ad hoc LSD is calculated on the transformed scale and the limits of the histogram classes are back-transformed in the original scale. The ad hoc LSD of 0.5 is quite high. It is easily checked that, starting with ratings of 5% FHB, comparison-wise significance for genotype means results in differences of around 3% for small FHB ratings and up to 13% for large ratings (with a maximum for ratings around 50%). Spearman's rank

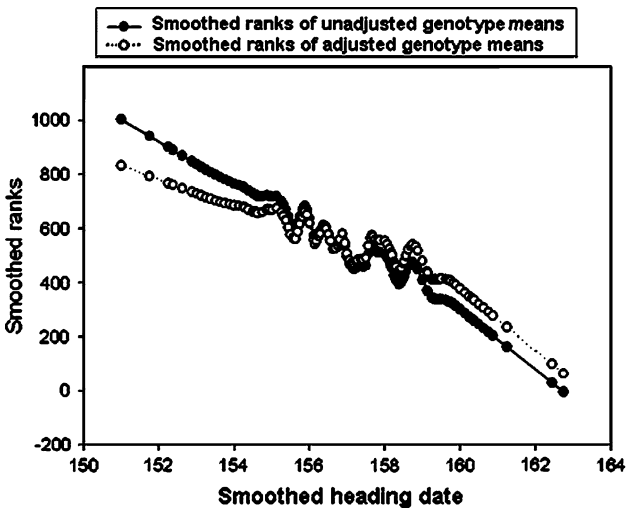

Fig. 1 Smoothed ranks of means of 1,000 winter wheat genotypes per increasing average heading date (HD, days after January 1) inoculated with *Fusarium culmorum* across four locations (genotypic analysis). Smoothing of ranks of genotypic FHB means per HD by cross validation with Friedman's Supersmoother (Friedman and Silverman 1989)

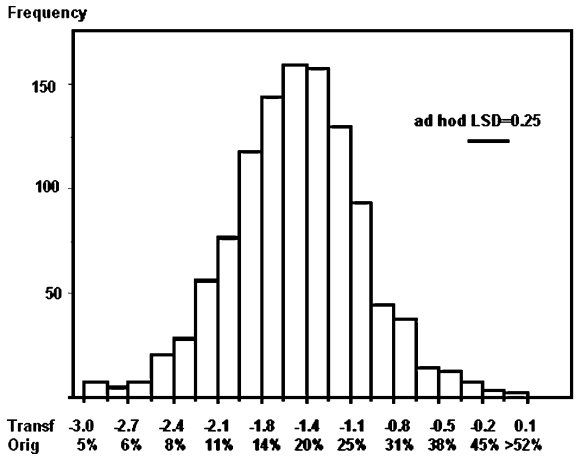

Fig. 2 Distribution of adjusted means of 1,000 winter wheat genotypes for *Fusarium* head blight (FHB) rating (0–100%) inoculated with *Fusarium culmorum* across four locations (genotypic analysis) on the transformed and original scale. Ad hoc LSD is calculated in the transformed scale (*Transf*) and the limits of the histogram classes are back-transformed in the original scale (*Orig*)

correlation coefficient between genotype means and heading date was reduced from −0.38 to −0.22 by the adjustment.

Analysis of groups of genotypes. As described in "Field trials", the genotypes originate from three subpopulations (R, M and P cf. "Field trials") and standards. The standards are modeled as levels of a fixed factor. In addition, standards are assigned to a common level (0) of the grouping factor for the classes of genotypes and are blocked out by dummy coding from the covariance structure of the mixed model (Piepho et al. 2006a).

The other levels of the grouping factor for the comparison of classes of genotypes consist of ten distinct groups of genotypes from the three subpopulations, because the interest lays in comparing eight marker classes with each other and to compare the best marker class with phenotypic selected (P) and unselected population (R).

Therefore, we analyze the variation between these ten classes of genotypes and genotypic variation within the classes of genotypes. This analysis requires a different covariance structure for the mixed model. The fixed part of the model contains the HD and the standard factor. In the random part of the model, the genotype effect and the genotype × location interaction are substituted by the group effect, an interaction between group × genotype, interactions between location × group and between location × group × genotype.

Table 2 gives the variance components for the random effects. The Wald *F* tests for fixed covariate HD are significant in the discussed examples (test statistics not shown). The average standard error for comparisons between groups (i.e., differences of group means) is

Table 2 Variance components of *Fusarium* head blight (FHB) rating (0–100%) of ten classes of winter wheat genotypes inoculated with *Fusarium culmorum* across four locations analyzed within and between groups of genotypes in hierarchical lattice design (example b)

Variance components	Estimated variance
loc	0.0285
loc × b	0.000085
loc × b × b2	0.000021***
group × s2	0.0025
loc × group × s2	0.0000046
group × gt × s2	0.1722***
loc × group × gt × s2	0.0005***
Residual	0.2014

Effects and model: cf. Table 1

*** *P* value < 0.001

calculated as $se_a = 2.82$. Heritability is calculated as $h^2 = 0.75$ and this ad hoc heritability is now based on the variance between groups on the logit scale as described above.

Figure 3 compares the covariate-adjusted and unadjusted means (BLUP) for ten classes, giving the transformed and original scale. The ad hoc LSD in the original scale is given as 0.3, the range of means is much smaller than in other examples (not shown) and the adjustment is not very pronounced. The reason for this is

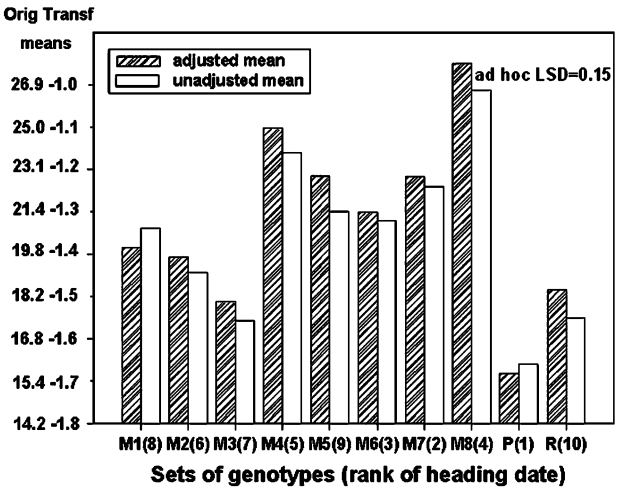

Fig. 3 Comparison of means (best linear unbiased predictors) of *Fusarium* head blight (FHB) rating (0–100%) with and without adjustment for heading date (ranks in brackets behind class means) and ad hoc LSD for ten classes of winter wheat genotypes inoculated with *Fusarium culmorum* across four locations (analysis of groups of genotypes). Means and ad hoc LSD are presented in transformed scale (*Transf*) and original scale (*Orig*) to enable a rough back-transformation of genotype means

the large variation of genotypes within the investigated groups of genotypes and lack of extremely susceptible genotypes. As expected, the adjusted and unadjusted means (BLUPs) follow a similar pattern. However, the adjustment for heading date does not have a large effect on such large groups of genotypes. The reason for this is that the observed average heading dates per class are close to the overall average heading date of the full data set [range of class means (156.5; 158.5), overall mean: 156.9, check ranks of heading dates in Fig. 3, in brackets behind the group names]. Note that the classes contain between 144 and 8,000 observations. Smaller differences in heading date are typical when testing adapted genotypes only. However, four of five marker genotype classes with the smallest heading date (ranks 1–5) are adjusted in the expected direction. Spearman's rank correlation for the group means and heading date for the analysis with and without covariate is calculated as $r_s = 0.79$ ($n = 11$, $P = 0.0037$).

Discussion

In this paper, we have assumed independent homoscedastic main effects and genotype–environment interaction effects, corresponding to the compound symmetry model for genetic correlation across environments. This model is restrictive, in that it assumes homogeneous variance and covariances. We are aware that heterogeneous models may result in a better selection result (cf. Kelly et al. 2007), especially if the genotype × environment interaction is large. In our case, the interaction variance was rather smaller than that of the genotypic main effect, so gross differences in selection results are not expected. The analysis for this paper was done as part of a cooperation project in SAS (SAS Institute Inc., 2004) and major resource problems were encountered for all checked covariance structures except the homogenous compound symmetry model. To evaluate whether this model would result in a major bias, we fitted a factor-analytic model with one factor (Piepho 1998a; Smith et al. 2005) in SAMM (Butler et al. 2003) and found that similar variance estimates did not differ much, even if the AIC indicated a worse fit (difference between AIC of both models was quite high with approximately 200). However, two of four factor loadings of this solution were estimated at the boundary of the parameter space and the genotypic main effect was dropped from the analysis (Heywood case, cf. Mardia et al. 1988). Further higher dimensional models than compound symmetry (heterogeneous compound symmetry, heterogeneous variance, first-order autoregressive, unstructured covariance model) were not fitted at all by SAMM (mostly convergence problems and singularities, in a lesser degree than in SAS). Therefore, we decided to use the compound symmetry model mainly to enable prediction of main effects, which were of special interest in the discussed application. Kuchel et al. (2006) used a similar (homogeneous) model to calculate BLUPs for heading scores.

Recently, independently from this work, Oakey et al. (2006) presented a generalized heritability measure for mixed models on the basis of contrasts of true and predicted genetic effects. Cullis et al. (2006) presented a further generalized heritability measure dependent on the average prediction error variance and the vector of genetic variance parameters. Piepho and Möhring (2007) proposed a simulation-based measure of heritability and related quantities. A comparison of these measures of heritability would be worthwile.

Smith et al. (2001) propose an alternative to LSDs for crop variety evaluation trials. On the basis of the correlation between the true variety effects and those predicted from the mixed model analysis (Cullis et al. 2000), they calculate the probability that a variety is truly greater if its BLUP of yield is greater than that of a standard. Similar suggestions have been made for fixed effects models (Eskridge and Mumm 1992; Piepho 1998b; Piepho and van Eeuwijk 2002; Piepho and McCulloch 2004).

A bivariate model for average FHB and heading date would be another way to analyze the *Fusarium* data, resulting in BLUPs of both traits per genotype. One could then select genotypes with relatively small FHB among candidates with comparable heading date. Thus, selection would be conditionally on genotypes with about the same heading date. This is expected to yield similar results as the analysis of covariance adjustment proposed in this paper.

One referee expressed concern that our covariate adjustment to a common heading date essentially creates genotypes that do not exist and that for this reason analysis of covariance may not be appropriate. As stated above, for such applications we recommend, therefore, to base the selection of genotypes for further trials on comparisons between genotypes in similar earliness classes. This avoids the use of adjusted genotype means with unplausible combinations of earliness and susceptibility that do not exist in reality. For early heading genotypes, this will result in similar choices as the bivariate model under similar restrictions and will not result in selection of early heading varieties as expected in the other approaches. While this type of adjustment may not always be desirable in other applications of covariance adjustment (Smith 1957), it seems particularly suitable for the case at hand. The objective of the breeding program was to select genotypes with good resistance to *Fusarium*, while at the same time

avoiding selection for late heading date. Resistance is a complex trait. It may be assumed that resistance is governed by many genes, some of which are also related to heading date. The purpose of our adjustment was to avoid selection pressure on pleiotropic genes that influence both heading date and resistance to *Fusarium*. A regression-based correction of resistence traits for dependencies on maturity is common practice in plant breeding applications (Bormann et al. 2004; Bradshaw et al. 2004).

A pleiotropic (or at least partly overlapping) gene for heading date and resistance to *Fusarium* was reported in Schmolke et al. (2005). The covariance adjustment is intended to block these genes from selection, such that selection for resistance exerts pressure mainly on genes determining resistance, but not heading date. By way of analogy, consider composite interval mapping (CIM) for a complex trait. CIM involves a covariate adjustment for important QTL, when scanning for other putative QTL. While the adjustment will create genotypes that do not exist, it blocks out other QTL (analogous to genes governing heading date), thus allowing unbiased assessment of a putative QTL (analogous to a resistance gene with no pleiotropic effect for heading date). Other examples of ANCOVA are found in animal trials, as e.g., the use of weaning weight as covariate to analyses the average daily gain in weight in pig feeding trials (Leibbrandt et al. 1975).

The present study allows further conclusions for resistance screening. The biologically occurring correlation between heading date and FHB severity preferring later heading genotypes was already reduced by our multiple inoculation system, i.e., three to four inoculation dates during the whole flowering period within each experiment ($r = -0.38$ in the discussed example). Even this rather low correlation could be substantially reduced further by the adjustment procedure described here. Figure 1 clearly shows that the adjustment favors early heading genotypes and penalizes later heading genotypes, thus preventing a preferential selection of the latter. This is especially important for selection of FHB-resistant progeny, because an indirect selection for late maturity is not desired. Our objective was to select for *Fusarium* resistance, while retaining the variability in heading date. Note that for the very early genotypes, the (smoothed) rank difference between the adjusted and non-adjusted treatment is about 200.

In QTL studies on FHB resistance, linkage between QTL for morphological traits, especially heading date and plant height, and resistance QTL is often found across environments (e.g., Schmolke et al. 2005). Applying the proposed adjustment procedure could help clarify whether these coincidences really have a genetic basis (linkage or pleiotropy) or are caused by epidemiological or physiological factors.

Appendix

Back-transformation of means and standard errors

A genotype mean $\bar{x}_i$ on the logit scale with standard error se_i is back-transformed in the original scale by

$$h^{-1}(\bar{x}_i) \cong \frac{\exp(\bar{x}_i)}{[1+\exp(\bar{x}_i)]} + \frac{1}{2} \times \mathrm{Var}(\bar{x}_i)\left[\frac{\partial^2}{\partial \bar{x}_i{}^2} h^{-1}(\bar{x}_i)\right]$$

and

$$\mathrm{Var}(h^{-1}(\bar{x}_i)) \cong se_i^2 \left[\frac{\partial}{\partial \bar{x}_i} h^{-1}(\bar{x}_i)\right]^2.$$

In case of differences of genotype means (i.e., pairwise comparisons of genotype means) $d = \bar{x}_1 - \bar{x}_2$ we back-transformed the expectations and variances by

$$E(h^{-1}(d)) = h^{-1}(\bar{x}_1) - h^{-1}(\bar{x}_2)$$

and

$$\begin{aligned}\mathrm{Var}(h^{-1}(d)) = {} & \mathrm{Var}(\bar{x}_1)\left[\frac{\partial}{\partial \bar{x}_1} h^{-1}(\bar{x}_1)\right]^2 \\ & + \mathrm{Var}(\bar{x}_2)\left[\frac{\partial}{\partial \bar{x}_2} h^{-1}(\bar{x}_2)\right]^2 - 2\ \mathrm{Cov}\ (\bar{x}_1, \bar{x}_2) \\ & \times \left[\frac{\partial}{\partial \bar{x}_1} h^{-1}(\bar{x}_1)\right]\left[\frac{\partial}{\partial \bar{x}_2} h^{-1}(\bar{x}_2)\right].\end{aligned}$$

$\mathrm{Var}(\bar{x}_1)$, $\mathrm{Var}(\bar{x}_2)$ and $\mathrm{Cov}(\bar{x}_1, \bar{x}_2)$ are the variances and covariances of genotype means and covariances of genotype means on the transformed scale.

For back-transformation from the logit scale, the formulas for first and second-order derivatives are:

$$\frac{\partial}{\partial \bar{x}} h^{-1}(\bar{x}) = \frac{\exp(\bar{x})}{[1+\exp(\bar{x})]^2} \quad \text{and}$$
$$\frac{\partial^2}{\partial \bar{x}^2} h^{-1}(\bar{x}) = \frac{\exp(\bar{x})\ (1-\exp(\bar{x}))}{[1+\exp(\bar{x})]^3}.$$

References

Bormann CA, Rickert AM, Castillo Ruiz RA, Paal J, Lübeck J, Strahwald J, Buhr K, Gebhardt C (2004) Tagging quantitative trait loci for maturity-corrected late blight resistance in tetraploid potato with PCR-based candidate gene markers. MPMI 17:1126–1138

Bradshaw JE, Pande B, Bryan GJ, Hackett CA, McLean K, Stewart HE, Waugh R (2004) Interval mapping of quantitative trait loci for resistance to late blight [*Phytophthora infestans* (Mont.) de Bary], height and maturity in a tetraploid population of potato (*Solanum tuberosum* subsp. tuberosum). Genetics 168:983–995

Bueno JSD, Gilmour SG (2003) Planning incomplete block experiments when treatments are genetically related. Biometrics 59:375–381

Buerstmayr H, Steiner B, Lemmens M, Ruckenbauer P (2000) Resistance to *Fusarium* head blight in winter wheat: heritability and trait associations. Crop Sci 40:1012–1018

Buerstmayr H, Steiner B, Hartl L, Griesser M, Angerer N, Lengauer D, Miedaner T, Schneider B, Lemmens M (2003) Molecular mapping of QTLs for *Fusarium* head blight resistance in spring wheat. II. Resistance to fungal penetration and spread. Theor Appl Genet 107:503–508

Buerstmayr H, Legzdina L, Steiner B, Lemmens M (2004) Variation for resistance to *Fusarium* head blight in spring barley. Euphytica 137:279–290

Butler D, Cullis BR, Gilmour AR, Gogel BJ (2003) Spatial analysis with mixed models with S-Plus. SAMM reference manual. Training series QE02001, NSW 24 Agriculture (pdf version)

Cullis BR, Smith A, Hunt C, Gilmour A (2000) An examination of the efficiency of Australian crop variety evaluation programmes. J Agric Sci 135:213–222

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:1–13

Eskridge KM, Mumm RF (1992) Choosing plant cultivars based on the probability of outperforming a check. Theor Appl Genet 84:894–900

Friedman JH, Silverman BW (1989) Flexible parsimonious smoothing and additive modeling. Technometrics 31:3–39

Friedman JH, Stuetzle W (1981) Projection pursuit regression. J Am Stat Assoc 76:817–823

Friedman JH, Stuetzle W (1982) Smoothing of scatterplots. Technical report Orion 003. Department of Statistics, Stanford University

Hinkelmann K, Kempthorne O (1994) Design and analysis of experiments. Introduction to experimental design, vol 1. Wiley, New York

Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. Wiley, New York

Hsu JC (1996) Multiple comparisons. Chapman & Hall, London

Kackar RN, Harville DA (1984) Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J Am Stat Assoc 79:853–862

Kelly A, Smith AB, Eccleston J, Cullis BR (2007) The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci 47:1063–1070

Kuchel H, Hollamby G, Langridge P, Williams K, Jefferies SP (2006) Identification of genetic loci associated with ear-emergence in bread wheat. Theor Appl Genet 113:1103–1112

Leibbrandt VD, Ewan RC, Speer VC, Zimmerman DR (1975) Effect of weaning and age at weaning on baby pig performance. J Anim Sci 40:1077–1080

Mardia KV, Kent JT, Bibby JM (1988) Multivariate analysis. Academic Press, London

McMullen MP, Schatz B, Stover R, Gregoire T (1997) Studies of fungicide efficacy, application timing, and application technologies to reduce *Fusarium* head blight and deoxynivalenol. Cereal Res Commun 25:779–783

Miedaner T, Reinbrecht C, Lauber U, Schollenberger M, Geiger HH (2001) Effects of genotype and genotype × environment interaction on deoxynivalenol accumulation and resistance to *Fusarium* head blight in rye, triticale, and wheat. Plant Breed 120:97–105

Miedaner T, Heinrich N, Schneider B, Oettler G, Rohde S, Rabenstein F (2004) Estimation of deoxynivalenol (DON) content by symptom rating and exoantigen content for resistance selection in wheat and triticale. Euphytica 139:123–132

Miedaner T, Wilde F, Steiner B, Buerstmayr B, Korzun V, Ebmeyer E (2006) Stacking quantitative trait loci (QTL) for *Fusarium* head blight resistance from non-adapted sources in an European elite spring wheat background and assessing their effects on deoxynivalenol (DON) content and disease severity. Theor Appl Genet 112:562–569

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H (2006) Joint modelling of additive and non-additive genetic line effects in single field trials. Theor Appl Genet 113:809–819

Paterson LJ, Patterson HD (1984) An algorithm for generating alpha-lattice designs. ARS Combin 16A:87–98

Patterson HD, Williams ER (1976) A new class of resolvable incomplete block designs. Biometrika 63:83–92

Piepho HP (1998a) Empirical best linear unbiased prediction in cultivar trials using factor analytic variance–covariance structures. Theor Appl Genet 97:195–201

Piepho HP (1998b) Methods for comparing the yield stability of cropping systems—a review. J Agron Crop Sci 180:193–213

Piepho HP, Möhring J (2006) Selection in cultivar trials is it ignorable? Crop Sci 146:193–202

Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881–1888

Piepho HP, McCulloch CE (2004) Transformations in mixed models: application to risk analysis for a multienvironment trial. J Agric Biol Environ Stat 9:123–137

Piepho HP, van Eeuwijk FA (2002) Stability analyses in crop performance evaluation. In: Kang M (ed) Crop improvement: challenges in the twenty-first century. Haworth Press, New York, pp 315–351

Piepho HP, Büchse A, Emrich K (2003) A hitchhiker's guide to the mixed model analysis of randomized experiments. J Agron Crop Sci 189:310–322

Piepho HP, Büchse A, Truberg B (2006a) On the use of multiple lattice designs and a-designs in plant breeding trials. Plant Breed 125:523–528

Piepho HP, Williams ER, Fleck M (2006b) A note on the analysis of designed experiments with complex treatment structure. HortScience 41:446–452

SAS Institute Inc (2004) SAS OnlineDoc® 9.1.2. SAS Institute Inc, Cary

Schmolke M, Zimmermann G, Buerstmayr H, Schweizer G, Miedaner T, Korzun V, Ebmeyer E, Hartl L (2005) Molecular mapping of *Fusarium* head blight resistance in the winter wheat population Dream/Lynx. Theor Appl Genet 111:747–756

Smith HF (1957) Interpretation of adjusted treatment means and regressions in analysis of covariance. Biometrics 13:282–308

Smith AB, Cullis BR, Gilmour AR (2001) The analysis of crop variety evaluation data in Australia. Aust N Z J Stat 43:129–145

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci Camb 143:449–462

Stuart A, Ord JK (1994) Kendall's advanced theory of statistics, vol 1, 6th edn. Arnold Publishers, London

Tukey JW (1991) The philosophy of multiple comparisons. Stat Sci 6:100–116

Whitaker D, Williams ER, John JA (2002) CycDesigN: a package for computer generation of experimental designs. CSIRO Forestry and Forest Products, Canberra