

**TRANSFORMATIONS IN MIXED MODELS: APPLICATION TO RISK ANALYSIS
FOR A MULTI-ENVIRONMENT TRIAL**

BU-1460-MA

MAY, 2000

**Hans-Peter Piepho
Institut fuer Natzpflanzenkunde
Universitaet Kassel
Witzenhausen, Germany**

and

**Charles E. McCulloch
Department of Biometrics
College of Agriculture and Life Sciences
Cornell University
Ithaca, NY 14853**

Keywords: Box-Cox transformation, nonlinear mixed model, genotype \times environment interaction, biadditive model, test of normality.

Abstract

An important trait in crop cultivar evaluation stability of performance across environments. There are many different measures of stability, most of which are related to variance components of a mixed model. We believe that stability measures assessing yield risk are of particular relevance, because they integrate location and scale parameters in a meaningful way. A prerequisite for obtaining valid risk estimates is an appropriate model for the distribution of yield across environments. Multi-environmental trials (MET) are often analyzed by mixed linear models, assuming that environments are a random sample from a target population, and that random terms in the model are normally distributed. The normality assumption may not always be tenable, and consequently, risk estimates may be biased. In this paper, we suggest three different approaches to cope with non-normality in mixed models. The first uses a Box-Cox transformation of the response, while the second and third are based on nonlinear mixed models, which transform either an effect or a larger part of the predictor. Transforming an effect offers a nice way of checking the normality assumption. The methods are exemplified using an international wheat yield trial. The importance of accounting for non-normality in risk analyses based on MET is emphasized. We suggest that transformations should be routinely considered in analyses to assess risk.

Transformations in mixed models: application to risk analysis for a multi-environment trial

Hans-Peter Piepho¹, Charles E. McCulloch²

¹Institut fuer Nutzpflanzenkunde, Universitaet Kassel, Witzenhausen, Germany

(piepho@wiz.uni-kassel.de)

²Cornell University, College of Agriculture and Life Sciences, Department of Biometrics, Ithaca, NY, USA

Summary: An important trait in crop cultivar evaluation is stability of performance across environments. There are many different measures of stability, most of which are related to variance components of a mixed model. We believe that stability measures assessing yield risk are of particular relevance, because they integrate location and scale parameters in a meaningful way. A prerequisite for obtaining valid risk estimates is an appropriate model for the distribution of yield across environments. Multienvironment trials (MET) are often analyzed by mixed linear models, assuming that environments are a random sample from a target population, and that random terms in the model are normally distributed. The normality assumption may not always be tenable, and consequently, risk estimates may be biased. In this paper, we suggest three different approaches to cope with non-normality in mixed models. The first uses a Box-Cox transformation of the response, while the second and third are based on nonlinear mixed models, which transform either an effect or a larger part of the predictor. Transforming an effect offers a nice way of checking the normality assumption. The methods are exemplified using an international wheat yield trial. The importance of accounting for non-normality in risk analyses

based on MET is emphasized. We suggest that transformations should be routinely considered in analyses to assess risk.

Key words: Box-Cox transformation, nonlinear mixed model, genotype \times environment interaction, biadditive model, test of normality.

1. Introduction

Multienvironment trials (MET) are conducted to assess the yield performance of crop cultivars under varying environmental conditions. One purpose of such trials is to derive cultivar recommendations for a given target region. In order to ensure that recommendations are representative for the target region, it is desirable to choose a random sample of test locations. Usually, it is not just the mean performance that determines the relative merit of different cultivars, but also the yield stability. Specifically, measures associated with yield risk are of particular relevance in this context. For both plant breeders and farmers, an important question is how a given cultivar will perform in a newly selected environment from the target. Due to cultivar \times environment interaction, any yield prediction obtained from a statistical analysis of the MET is subject to uncertainty. For example, cultivar A may have a higher mean yield than cultivar B, but still it is possible that in a given environment, cultivar B outperforms cultivar A. Thus, aside from an estimate of the expected performance, it is useful to quantify the probability that A outperforms B (Eskridge and Mumm, 1992). Similarly, we may compute the probability that a cultivar outperforms all other cultivars in a given set of cultivars (Piepho and van Eeuwijk,

1999). Another quantity of interest is the probability (risk) that yield falls below a certain level (Mead et al., 1986; Eskridge, 1990; Piepho, 1996).

A prerequisite for estimating these probabilities is an appropriate model for the random variation among environments. MET are often analyzed using mixed model procedures (Gilmour et al., 1995; Cullis et al., 1996; Denis et al., 1997), assuming the response (usually yield) follows a multivariate normal distribution. The normality assumption is not usually critical when means parameters are of primary interest, provided the first two moments are specified correctly (Vonesh and Chinchilli, 1997). When probabilities need to be estimated, however, the distributional assumption can be quite important (Taylor et al., 1996; Piepho and van Eeuwijk, 1999). The present paper explores three ways to cope with non-normality in mixed model analyses of MET data: (i) transform the response (e.g. using the Box-Cox family), (ii) transform one (or more) of the random effects in the model, and (iii) transform part of the predictor. The transform-the-response approach (i) tries to achieve two goals simultaneously, i.e. normality and additivity. By contrast, the transform-an-effect (ii) approach assumes that additivity holds on the original scale. The sole purpose of the transformation is to obtain normality. The transform-part-of-the-predictor approach (iii) takes an intermediate position between (i) and (ii). In addition, (i) and (ii) or (iii) can be applied simultaneously, giving yet more flexibility in modeling nonnormality. With all of these approaches, heteroscedasticity at various levels, including trial errors, can be accommodated by using a suitable mixed model (Frensham, Cullis, and Verbyla, 1997; Piepho, 1997). The environmental and genetic factors governing the distribution of the response on MET are usually numerous, interacting in complex and largely unpredictable ways. Thus, it is difficult to decide on *a priori* grounds, what are the sources of nonnormality, if any, and which

approach/model preferable. We therefore suggest to fit a number of models and compare their fits. The approach is exemplified using data from a subset of a larger set of MET on wheat.

2. Data

To exemplify the methods proposed in this paper, we use data from international wheat trials conducted by CIMMYT (Centro Internacional del Mejoramiento del Maiz y Trigo, Mexico) in seven South American countries. We use an unbalanced subset of a larger data set that has been analyzed by Piepho and van Eeuwijk (1999). It comprises a total of 839 observations of 15 cultivars tested across 25 locations and 12 years. There were a total of 119 year \times location combinations. Year \times location combinations will be considered as independent environments for simplicity of exposition. For analyses of the wheat data taking the three-way structure into account see Piepho and van Eeuwijk (1999).

3. Initial model

For the analysis of a MET, we may start from a simple two-way mixed model of the form

$$y_{ij} = \mu_i + u_j + f_{ij}, \tag{1}$$

where y_{ij} is the yield of the i -th cultivar ($i = 1, \dots, I$) in the j -environment ($j = 1, \dots, J$), μ_i is the expected value of the i -th cultivar, u_j is the random main effect of the j -th environment, and f_{ij} is a residual corresponding to y_{ij} , which comprises both cultivar x environment interaction and unexplained interaction. Random effects u_j and f_{ij} are taken to be independent normal deviates with variances σ_u^2 and σ^2 .

The model (1) is fairly restrictive with respect to the variance-covariance structure. More flexibility is available if we assume that the vector y_j of observed yields in the j -th environment follows a multivariate normal distribution with unstructured variance-covariance matrix. The unstructured model is certainly the most flexible and can serve as a benchmark against which to judge other models. However it may be a wasteful use of parameters (Diggle, 1988).

Accordingly a number of models have been suggested which are intermediate, including heteroscedastic (e.g., Shukla, 1972; Frensham et al., 1997) and factor-analytic structures. For more details see Denis et al. (1997), Piepho (1997, 1999) and Piepho et al. (1998).

Our main model will be the factor-analytic model

$$y_{ij} = \mu_i + \lambda_i w_j + f_{ij} \tag{2}$$

where λ_i is the factor loading of the i -th cultivar and w_j is a latent environmental variable. This model was first proposed in a fixed effects model context by Yates and Cochran (1938) and further elaborated by Finlay and Wilkinson (1963) and Eberhart and Russell (1966). Here, we will look at the model (2) from a mixed model perspective (Oman, 1991; Gogel et al., 1995;

Piepho, 1997). For the data set at hand, model (2) was found to fit well relative to a number of different heteroscedastic mixed models (Piepho and van Eeuwijk, 1999). It can be assumed without loss of generality that w_j has unit variance and is independent of f_{ij} . Model (2) induces a factor-analytic structure for the variance-covariance structure of y_j , namely $\Sigma = \Lambda\Lambda' + I\sigma^2$, where Λ is a vector of factor loadings λ_i .

4. Transformed models

In the following we will introduce a number of modifications of the basic models considered in Section 3. The main purpose of these modifications is to derive a set of models, which can account for various kinds of nonnormality. Since for our example, we have no *a priori* reason to favor a particular candidate among these models, we will fit a number of models and select good ones based on the Schwarz Bayesian Criterion (see Section 5).

4.1 Transforming the response y_{ij}

There are a number of data transformations that can be considered (Heuts, 1984; Atkinson, 1985; Sakia, 1992; DeSilva et al., 1997). Here, we concentrate on the power family of transformations suggested by Box and Cox (1964) which has been extensively studied by many authors (e.g. Solomon, 1985; Sakia, 1990; Gianola et al., 1990; Solomon and Cox, 1992; Taylor et al.; 1996). For a recent review see Sakia (1992). The Box-Cox family is given by

$$t(y; \phi) = \begin{cases} (y^\phi - 1)/\phi & \text{if } \phi \neq 0 \\ \log(y) & \text{if } \phi = 0 \end{cases} \quad (3)$$

The objective of the transformation is to obtain data that follow a normal distribution more closely than the untransformed data. This objective can only be approximately achieved since for $\phi \neq 0$, $t(y; \phi)$ cannot be strictly normal, as it is bounded either above or below (Seber and Wild, 1989). We fit the factor-analytic model (2) to the transformed data, i.e. the model is

$$t(y_{ij}; \phi) = \mu_i + \lambda_i w_j + f_{ij} \quad (4)$$

Fitting of the Box-Cox model by maximum likelihood (ML) is described in Appendix A.

4.2 Transforming the environmental effect w_j

In the preceding section, a data transformation was sought so that a mixed model assuming normality for the response could be used to analyze the data on the transformed scale.

Alternatively, we may start with a mixed model and transform an effect or a larger part of the predictor, leading to a nonlinear mixed model (Davidian and Giltinan, 1995; Vonesh and Chinchilli, 1997). Due to limitations of available software, we restrict our attention to models with functions of normal random effects. For example, we can replace w_j in the factor-analytic model (2) by an exponential term, i.e. $w_j = [\exp(z_j) - \theta_1]/\theta_2$, where z_j has standard normal

distribution and θ_1 and θ_2 are chosen so that w_j has zero mean and unit variance. In this model, w_j follows a log-normal distribution, so that y_{ij} will have a skewed distribution. We will discuss extensions of the exponential transformation. Wherever possible, w_j is scaled to have zero mean and unit variance, as in model (2), to facilitate the search for sensible starting values of model parameters. The approach of transforming parts of the predictor is applicable to any term in a mixed model, but here we will mainly focus on the factor-analytic model (2).

The first idea to introduce a log-normal distribution in the linear predictor led us to consider a more flexible class of distributions, i.e. the Johnson (1949) family, which is based on three transformations of standard normal deviates, denoted as S_L , S_U , and S_B . For every combination of skewness and kurtosis, measured by the standardized third and fourth moments, there is exactly one member in the Johnson family, which provides flexibility for modeling non-normal data. All three transformations have the normal distribution as a limiting case.

We now consider a number of models derived from the Johnson family. Our attention will be restricted to modifications of the random effect w_j in model (2), though any other mixed model can be modified in a similar way. Also, in this section we only consider models where the shape of the distribution for w_j is the same for each cultivar, but it is emphasized that the models may be extended so that the distribution shape of w_{ij} depends on the cultivar.

The lognormal model: We have not exploited all the potential flexibility by just setting $w_j = [\exp(z_j) - \theta_1]/\theta_2$. Specifically, since z_j has unit variance, the shape of the distribution is fixed. More flexibility is achieved by allowing the variance of z_j to vary freely. We will express the model in a form so that the resulting w_j has zero mean and unit variance, using results given in Johnson (1949) and Johnson and Kotz (1970). The model for w_j is

$$w_j = [\exp(\psi z_j) - \exp(\frac{1}{2}\psi^2)] / \sqrt{\omega(\omega - 1)} \quad (5)$$

where z_j is a standard normal deviate and $\omega = \exp(\psi^2)$. The limiting case of normality corresponds to $\psi \rightarrow 0$. When fitting the model, we therefore choose a small starting value for ψ , e.g. $\psi = 0.1$.

The S_U -model: Again standardizing to zero mean and unit variance, using Johnson (1949) and Johnson and Kotz (1970), the model for w_j is

$$w_j = (\sinh(\nu + \psi z_j) - \omega^{0.5} \sinh(\nu)) / \sqrt{\left(\frac{1}{2}(\omega - 1)[\omega \cosh(2\nu) + 1]\right)} \quad (6)$$

where z_j is a standard normal deviate and w_j has zero mean and unit variance. Normality is approached as $\psi \rightarrow 0$. We use $\nu = 0$ as a starting value for ν , corresponding to symmetry, and a small starting value for ψ , e.g. 0.1.

The S_B - model: Consider the model

$$w_j = \left(\frac{\exp(\nu + \psi z_j)}{1 + \exp(\nu + \psi z_j)} - \theta_1 \right) / \theta_2 \quad (7)$$

where θ_1 and θ_2 are the mean and standard deviation of the S_B transformation. Unfortunately, explicit expressions for θ_1 and θ_2 in terms of ν and ψ are very complicated (Johnson, 1949), involving infinite series, and are not convenient for standardizing w_j . We deliberately choose fixed values for θ_1 and θ_2 , however, so that the choice of sensible starting values for the model parameters is facilitated. Pearson and Hartley (1972) tabulated θ_1 and θ_2 for different values of $\gamma = -\nu\psi$ and $\delta = 1/\psi$. We choose values for θ_1 and θ_2 corresponding to a nearly normal model for w_j , i.e. $\nu = 0$, so that the distribution is symmetric with $\theta_1 = 0.5$, and small ψ , since the S_B distribution approaches normality as $\psi \rightarrow 0$. The values of θ_1 and θ_2 remain fixed during model fitting iterations. The tabulated value of $1/\psi$ closest to normality is $1/\psi = 4.241$, for which $\theta_2 = 0.0582$ and $\psi = 0.2358$. With these fixed values for θ_1 and θ_2 and these starting values for ν and ψ , w_j has (to the accuracy of the tabulation) a mean of zero and unit variance, and it is reasonable to use as starting values for μ_i , λ_i and σ^2 estimates obtained under the normal model with $w_j \sim N(0, 1)$. Note that for the estimated S_B -model, w_j will not usually have zero mean and unit variance, since θ_1 and θ_2 are fixed at their starting values. This needs to be taken into account when interpreting model parameters, but otherwise does not affect the fit.

To use the Johnson family we separately fit each of the transformations given in (5), (6), and (7).

We then select the one that gives the best model in an information criterion sense (see below).

This gives us the flexibility to model random effects with any skewness and kurtosis, while leaving the first and second moment structure unchanged.

4.3 Transforming part of the linear predictor

In Section 4.1, we have transformed y_{ij} , while in 4.2 the random term w_j was transformed. Here we consider a transformation of the term $\mu_i + \lambda_i w_j$. The model is

$$y_{ij} = \psi g(\mu_i + \lambda_i w_j) + f_{ij}, \quad (8)$$

where $g(\cdot)$ is some monotonically increasing function and w_j has a standard normal distribution. This can be regarded as a generalized biadditive mixed model with biadditive term $\lambda_i w_j$ and an extended link function involving the parameter ψ . We use the Johnson system as candidates for $g(u_{ij})$ with $u_{ij} = \mu_i + \lambda_i w_j$, i.e. $g(u_{ij}) = \exp(u_{ij})$ for S_L , $g(u_{ij}) = \exp(u_{ij})/[1+\exp(u_{ij})]$ for S_B , and $g(u_{ij}) = \sinh(u_{ij})$ for S_U . Obviously, the shape of the resulting distribution for y_{ij} depends on the cultivar. For the log-normal model, we set $\psi = 1$ without loss of generality. For S_B and S_U , we chose starting values $\psi = 100$ (just above the maximum of observed yields) and $\psi = 1$, respectively. For all three models, μ_i and λ_i were initialized as the sample mean and standard deviation of $g^{-1}(y_{ij}/\psi)$. We could extend (8), e.g. by letting ψ depend on the cultivar, but this is not pursued here.

5. Model selection

5.1 Information criteria

Our models clearly use quite different numbers of parameters and are non-nested. We thus need a criterion for choosing amongst the models (Wolfinger, 1993; Vonesh and Chinchilli, 1997).

We have used the Schwarz Bayesian Criterion – *SBC* (Schwarz, 1978) which is defined as

$$SBC = -2 \log L + p \log(n), \tag{9}$$

where L is the likelihood function, evaluated by substituting the maximum likelihood estimates of the parameters, p is the number of parameters and n is the number of observations. We have chosen *SBC* since it provides a consistent estimate of the true order of the model, assuming that a true model exists and is low dimensional (Buckland et al., 1997). It should be noted, however, that *SBC* was developed for linear models with a single error term (Schwarz, 1978), so it is appropriate when the variance-covariance structure of \mathbf{y}_j is $I\sigma^2$. With an unstructured variance-covariance matrix and complete data, we are in a multivariate linear regression framework, where n needs to be replaced by the number of subjects (McQuarrie and Tsai, 1998, p.149). It is not clear how to adapt *SBC* in the general mixed model case, where the variance-covariance structure has an intermediate complexity and data may be unbalanced (Pauler, 1999). Thus, our use of *SBC* here is of an *ad hoc* nature, and there may be a tendency to over-penalize the likelihood. Nevertheless, we prefer *SBC* to the Akaike Information Criterion (*AIC*), since *AIC* is known to overfit in small samples when the true underlying model is simple (McQuarrie and

Tsai, 1998). A small simulation study on the performance of *SBC* is described in Appendix B. We considered simple random samples from the normal and the Johnson system. The results show that *SBC* selects the correct model among two alternatives a good proportion of the time, provided models differ notably.

5.2 Checking normality

For the untransformed data, normality may be checked by studying the scaled residuals

$$\mathbf{r}_j = (\hat{\Sigma}_j)^{-0.5} (\mathbf{y}_j - \hat{\boldsymbol{\mu}}_j) \quad (10)$$

where Σ_j and $\boldsymbol{\mu}_j$ are the variance-covariance matrix and expectation of \mathbf{y}_j under model (2) and model parameters are estimated under normality. When there are many cultivars, non-normality in \mathbf{y}_j will tend to be masked, however, since \mathbf{r}_j is a weighted average across cultivars in an environment. For this reason, we considered residuals

$$\mathbf{r}_j = [\text{diag}(\hat{\Sigma}_j)]^{-0.5} (\mathbf{y}_j - \hat{\boldsymbol{\mu}}_j) \quad (11)$$

These residuals have unit variance, but may be strongly correlated. Despite this problem, they should be useful for revealing non-normality. For the Box-Cox model, we may study the same residuals on the transformed scale.

We may also fit a member of the Johnson family to w_j on the transformed scale to check if the Box-Cox transformation has achieved approximate normality. To simultaneously fit the Box-Cox transformation for y_{ij} and the Johnson transformation for w_j , we use (4) and perform a grid search over ϕ .

6. Estimating probabilities

Let Y_i be a random variable denoting the yield of the i -th cultivar in a random environment.

Then, the three probabilities of interest are:

- (a) $\Pr(Y_i < \tau)$, the probability that the i -th cultivar falls below a critical threshold τ
- (b) $\Pr(Y_i > Y_{i'})$, the probability the i -th cultivar outperforms the i' -th cultivar
- (c) $\Pr(Y_i > Y_{i'} ; i \neq i')$, the probability that the i -th cultivar outperforms all others in the set.

Under a Box-Cox model, probabilities (a) and (b) can be computed from the estimated model, while (c) is easiest to compute by simulation. For the nonlinear mixed models, we computed all three probabilities by simulation.

7. Numerical example

To exemplify the methods, we use the wheat yield data described in Section 2. All models were fitted by ML, using the procedures MIXED and NLMIXED of the SAS System (Version 7). NLMIXED uses adaptive Gaussian quadrature as described by Pinheiro and Bates (1995) to

integrate the likelihood over the random effects. Prior to comparing different transformations of the data or model we investigated the variance-covariance structure. We fitted the simple model (2) as well as a number of heteroscedastic and factor-analytic models. We were not able to fit an unstructured variance-covariance matrix due to convergence problems. Using the *SBC* criterion, the simple factor-analytic model (2) fitted second best ($SBC = 5878.4$), closely following a model with variance-covariance structure $\mathbf{J}\sigma_u^2 + \mathbf{A}\mathbf{A}' + \mathbf{I}\sigma^2$ ($SBC = 5876.1$). For this reason and because of the relative simplicity, we considered modifications of model (2) to allow for non-normality.

A natural transformation to try in this case is the log transformation. Perhaps surprisingly, it performs worse than no transformation at all (Table 1). The ML estimate of the Box-Cox transformation parameter was $\hat{\phi} = 0.528$, which corresponds approximately to a square-root transformation. The associated drop in $-2 \log L$ was dramatic (from 5669.7 to 5542.0), and the Box-Cox model had the most favorable value for *SBC* (see Table 1). Fig. 1 shows a normal plot of residuals r_j (eq. 11) based on a fit of model (2), assuming normality. The plot shows a marked departure from normality. The Box-Cox transformation seems to well remove non-normality (see Fig. 2). We also checked normality on the transformed scale by fitting the S_L model for w_j . The resulting value for $-2 \log L$ was 5540.7 (Table 1), which is only a minor change compared the Box-Cox model, and the shape parameter estimate ($\hat{\psi} = 0.11$) was small compared to its standard error (0.087) (Table 2).

The nonlinear mixed models using the Johnson system for w_j as well as for $\mu_i + \lambda_i w_j$ were not quite as successful in coping with non-normality as the Box-Cox transformation, but the

improvement relative to the untransformed model was noticeable (Table 1). The models transforming $\mu_i + \lambda_i w_j$ were better than models just transforming w_j , but inferior to Box-Cox, which transforms $\mu_i + \lambda_i w_j + f_{ij}$. In terms of *SBC*, there was little to choose between different models for w_j , with a slight advantage for S_L . It is interesting to study the estimates of transformation parameters for w_j in Table 2. With all three Johnson models, $\hat{\psi}$ was notably larger than twice its standard error, which is indication of non-normality. In the S_U model, ν was poorly estimated, but large, and $\hat{\psi}$ was the same as in the S_L model. The likelihoods for S_U and S_L were also the same. Thus, S_U converged to the log-normal line dividing the area between S_U and S_B in the plane spanned by the third and fourth standardized moments (Johnson and Kotz, 1970). From this result, it is expected that S_B fits slightly better than S_U , and this is in fact the case.

Table 3 compares the estimated means and three risk parameters under four models: the untransformed data, the Box-Cox transformed data, and the S_L models (applied to both w_j and $\mu_i + \lambda_i w_j$). Fig. 3 plots the risk versus critical value τ for some cultivars. The mean estimates are stable under the four different fitted models but the risk parameters show appreciable variation between the untransformed model and different transformed models, while differences between transformations are relatively small (Table 3; Fig. 3). The estimated probabilities to fall below the threshold of 10 dt/ha are smaller for all the transformations, while for the threshold 30 dt/ha they are smaller for the untransformed model. Differences are more pronounced for the latter. For example, $\Pr(Y_i < 30)$ is 0.54 for the untransformed model and 0.64 for the Box-Cox model. At critical levels near zero, the fit to the untransformed data gives nonsensical results for a number of cultivars, while the transformations produce more reasonable fits (Fig. 3), which

underscores the need for a transformation. In summary, for $\Pr(Y_i < \tau)$ the untransformed model gives results, which are not in the range of the transformed models, though the differences vary in magnitude. Considering the results, choice of transformation does not seem critical for assessment of risk to fall below τ . The important choice is whether or not to transform at all. Our results suggest that for the wheat data a transformation is clearly indicated.

Cultivar 12 has a probability of 0.29 to outperform the others under the untransformed model, while with the Box-Cox model the value is 0.23. Thus, under the Box-Cox model, the relative advantage of cultivar 12 is less clear-cut, and cultivars 13 and 4 appear more competitive. The other two models (Johnson S_B for w_j and for $\mu_i + \lambda_i w_j$) give a value for $\Pr(Y_{12} > Y_i)$, which is similar to that under the untransformed model, while rather different compared to the Box-Cox model. This example shows that the choice of transformation may be important in some cases.

Based on the fitting information in Table 1 and the risk analyses, we prefer the Box-Cox model.

To check for residual heteroscedasticity under this model, we plotted residuals

$y_{ij}^{0.528} - (\hat{\mu}_i + \hat{\lambda}_i \hat{w}_j)$ vs. predicted values $(\hat{\mu}_i + \hat{\lambda}_i \hat{w}_j)$ (see Fig. 4). The residual plot reveals no heteroscedasticity. Thus, the model gives a satisfactory fit, and it seems unnecessary to further model the residual.

8. Discussion and extensions

We have shown for a wheat yield example that non-normality can be a problem and failure to

accommodate the non-normality can give misleading estimates of risk parameters. The example has demonstrated that the particular choice of transformation was not critical regarding the resulting risk estimates. The important decision is whether or not to transform. All transformed models yielded substantially better fits and risk estimates than the untransformed. We therefore suggest that transformations be routinely considered in analyses to assess risk.

Among the options studied, the transform-the-response approach (Box-Cox-model) has the virtue of simplicity. Also, it is straightforward to extend to several random sources of variation, while approaches to transform an effect or part of the predictor can be tedious because of the multitude of potential models to consider when there are many random effects. The Box-Cox-transformation has the additional advantage that standard linear mixed model software can be used. Nonlinear mixed model software such as NLMIXED is currently limited to repeated measurement data with one subject level, so it is not yet possible to handle more complex models such as a three-way model for cultivars \times locations \times years with random years and environment, which involves several crossed effects. However, the transform-a-random-effect approach gives a straightforward way to check normality of the random effects. This option, which is easily implemented using NLMIXED, is very useful, especially in light of the failure of plotting empirical BLUPs as a diagnostic: as shown by Verbeke and LeSaffre (1996), the distribution of empirical BLUPs can look normal even in cases when the distribution is highly non-normal.

To fit the Box-Cox parameter ϕ , we profiled the likelihood by searching a grid for ϕ (i.e. treating ϕ as known at each step on the grid) and estimating all other parameters by ML (see Appendix). Note that our approach is valid only with ML, not with REML estimation. ML is known to

produce more biased estimates of variance components than REML. Thus, in practice, one could use ML to find the optimal ϕ and then analyze the data by REML using this value for ϕ .

We have used the Johnson family for transforming an effect or part of the predictor and the Box-Cox family to transform the data. The Johnson family could also be used to transform the data, with the complication that for S_U and S_B , a two-dimensional grid search over ν and ψ is needed. Conversely, a power transformation of the form $(\nu + \psi z_j)^\phi$ where z_j is standard normal could be contemplated to transform an effect or part of the predictor. This approach is not feasible, however, since the resulting density for the data is undefined for non-integer ϕ . In order to accommodate non-integer ϕ , we need to avoid the problem that $\Pr(\nu + \psi z_j < 0) > 0$ for standard normal z_j . This can be achieved by replacing the normal distribution for $\nu + \psi z_j$ by some nearly normal distribution, which is bounded below at zero, for example a log-normal distribution with small scale parameter. This leads to a model of the form $[\exp(\nu + \psi z_j)]^\phi = \exp(\phi\nu + \phi\psi z_j)$, but this is equivalent to the log-normal model of the Johnson family, suggesting that the power transformation is approximately covered by the Johnson family.

For the wheat yield data, we had no covariate information on the environments. Thus, only mixed models with a simple expectation structure were used. The use of covariates is potentially very useful in that part of the cultivar \times environment interaction can be explained and thus removed from the random part of the mixed model. The explained part can be integrated into the fixed part of a mixed model allowing predictions of expected yield conditional on a set of covariates (rainfall etc.) (Piepho et al., 1998a). To account for the remaining unexplained residual variation, probability statements can then be made conditional on covariates (Piepho,

1998). The expected value, together with probability statements provide valuable guidelines for making informed cultivar recommendations and for judging the relative merit of different cultivars in a specific environment, for which covariate information is available.

Model selection in mixed models with transformations is a difficult problem, for there are many modeling options. There are no simple general rules, and some intuition will always be needed. We have chosen to first fit a number of different variance-covariance structures without transformation and to select a model that fits relatively well. Transformations were then considered only for the selected model. Alternatively, one might first seek a suitable transformation, using the most general model for the variance-covariance structure and subsequently reduce the model. For the wheat data the Box-Cox transformation parameter is quite stable across different variance-covariance structures, allowing different kinds of heteroscedasticity, and both strategies lead to essentially the same answers (Piepho and van Eeuwijk, 1999). The situation is more complex, when transformations of model effects are considered. Here, starting with a flexible variance-covariance structure for y_j , e.g.

$\Sigma = J\sigma_u^2 + A_R A_R' + D$, where A_R is a matrix of R sets of factor loadings and D is a diagonal matrix, there are a large number of options for transformation, all of which cannot possibly be checked. Reducing the model first and then looking for a suitable transformation partly avoids this dilemma, but entails the risk of missing better fitting models. In any case, it is important to check model fit, e.g. by residual plots, to see whether a selected model agrees well with the data at hand.

Acknowledgements

We are grateful to the CIMMYT Regional Office for Latin America and INIA-Uruguay for allowing us to use the data. Special thanks go to Sergio Ceretta (INIA-Uruguay) for compiling the data we have used. Three referees are thanked for helpful comments. This paper was written while the first author was visiting the Department of Biometrics, College of Agriculture and Life Sciences, Cornell University, Ithaca, New York, USA. Support of the Heisenberg Programm of the Deutsche Forschungsgemeinschaft (DFG) and NSF DMS 9625476 is thankfully acknowledged.

References

- Atkinson, A. C. (1985). Plots, transformations, and regression. Clarendon Press, Oxford.
- Box, G. E. P., and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society B* **26**, 211-246.
- Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53**, 603-618.
- Cullis, B. R., Thomson, F. M., Fisher, J. A., Gilmour, A. R., and Thompson, R. (1996). The analysis of the NSW wheat variety data base. II. Variance component estimation. *Theoretical and Applied Genetics* **92**, 28-39.
- Davidian, M., and Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. London: Chapman and Hall.

- Denis, J. B., Piepho, H. P., and van Eeuwijk, F. A. (1997). Modelling expectation and variance for genotype by environment data. *Heredity* **79**, 162-171.
- DeSilva, H. N., Lai, C. D., and Ball, R. D. (1997). Fitting S_B distributions to fruit sizes with implications for prediction methods. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 333-346.
- Diggle, P. (1988). An approach to the analysis of repeated measurements. *Biometrics* **44**, 959-971.
- Eberhart, S. A., and Russell, W. A. (1966). Stability parameters for comparing varieties. *Crop Science* **6**, 36-40.
- Eskridge, K. M., (1990). Selection of stable cultivars using a safety first rule. *Crop Science* **30**, 369-374.
- Eskridge, K. M., and Mumm, R. F. (1992). Choosing plant cultivars based on the probability of outperforming a check. *Theoretical and Applied Genetics* **84**, 894-900.
- Finlay, K. W., and Wilkinson, G. N. (1963). The analysis of adaptation in a plant breeding programme. *Aust. J. Agric. Res.* **14**, 742-754.
- Frensham, A. B., Cullis, B. R., and Verbyla, A. P. (1997). Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* **53**, 1373-1383.
- Gianola, D., Im, S., Fernando, R. L., and Foulley, J. L. (1990). Mixed model methodology and the Box-Cox theory of transformations: A Bayesian approach. In *Advances in Statistical Methods for Genetic Improvement of Livestock*. In D. Gianola, and K Hammond (eds.), 15-40. Berlin: Springer Verlag.
- Gilmour, A. R., Thompson, R., Cullis, B. R. (1995). Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440-1450.

- Gogel, B. J., Cullis, B., and Verbyla, A. (1995). REML estimation of multiplicative effects in multienvironment variety trials. *Biometrics* **51**, 744-749.
- Heuts, R. M. J. (1984). The use of nonlinear transformations in ARIMA-models when the data are non-Gaussian distributed. *Computational Statistics* **1**, 155-177.
- Johnson, N. L. (1949). System of frequency curves generated by method of translation. *Biometrika* **36**, 149-176.
- Johnson, N. L., Kotz, S. (1970). *Continuous univariate distributions*. Volume 1. New York: Wiley.
- Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996). *SAS System for mixed models*. Cary: SAS Institute.
- McQuarrie, A. D. R., and Tsai, C. L. (1998). *Regression and time series model selection*. Singapore: World Scientific Publishers.
- Mead, R., Riley, J., Dear, K., and Singh, S. P. (1986). Stability comparison of intercropping and monocropping systems. *Biometrics* **42**, 253-266.
- Oman, S. D. (1991). Multiplicative effects in mixed model analysis of variance. *Biometrika* **78**, 729-739.
- Pauler, D.K. (1999). *The Schwarz criterion for mixed effects models*. Ph. D. thesis, Department of Statistics, Carnegie Mellon University, Pittsburgh.
- Pearson, E. S., and Hartley, H. O. (1972). *Biometrika tables for statisticians*. Vol. II. Cambridge: Cambridge University Press.
- Piepho, H. P. (1996). A simplified procedure for comparing the stability of cropping systems. *Biometrics* **52**, 378-383.
- Piepho, H. P. (1997). Analyzing genotype-environment data by mixed models with

multiplicative effects. *Biometrics* **53**, 761-766.

Piepho, H. P. (1998). Methods for comparing the yield stability of cropping systems - A review. *Journal of Agronomy and Crop Science* **180**, 193-213.

Piepho, H. P. (1999). Fitting a regression model for genotype-by-environment data on heading dates in grasses by methods for nonlinear mixed models. *Biometrics* **55**, 1120-1128:

Piepho, H. P., Denis, J. B., and van Eeuwijk, F. A. (1998a). Predicting cultivar differences using covariates. *Journal of Agricultural, Biological and Environmental Statistics* **3**, 151-162.

Piepho, H.P., Denis, J.B., and van Eeuwijk, F.A. (1998b). Mixed badditive models. *In* Proceedings of the XIX International Biometric Conference, Cape Town, South Africa, 14-18 December 1998, pp 79-89. Invited papers.

Piepho, H. P., and van Eeuwijk, F. A. (1999). Stability Analysis in Crop Performance Evaluation. *In* *Crop improvement for the 21st century*. Vol. 2. : M. S. Kang (ed),???-???. New York, Haworth. to appear

Pinheiro, J. C., and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics* **4**, 12-35.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1989). *Numerical recipes in Pascal*. Cambridge, Cambridge University Press.

Rode, R. A., and Chinchilli, V. M. (1988). The use of Box-Cox transformations in the development of multivariate tolerance regions with applications to clinical chemistry *American Statistician* **42**, 23-30.

Sakia, R. M. (1990). Retransformation bias: A look at the Box-Cox transformation to linear balanced mixed ANOVA models. *Metrika* **37**, 345-351.

- Sakia, R. M. (1992). The Box-Cox transformation technique: a review. *Statistician* **41**, 169-178
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461-464.
- Seber, G. A. F., and Wild, C. J. (1989). *Nonlinear regression*. New York: Wiley.
- Shukla, G. K. (1972). Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity* **29**, 237-245.
- Solomon, P. J. (1985). Transformations for components of variance and covariance. *Biometrika* **72**, 233-239.
- Solomon, P. J., and Cox, D. R. (1992). Nonlinear components of variance model. *Biometrika* **79**, 1-12.
- Spitzer, J. J. (1982). A fast and efficient algorithm for the estimation of parameters in models with the Box-and-Cox transformation. *Journal of the American Statistical Association* **77**, 760-766.
- Taylor, J. M. G., Cumberland, W. G., Xiangyi M. (1996). Components of variance models with transformations. *Australian Journal of Statistics* **38**, 183-191.
- Verbeke, G., and LeSaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association* **91**, 217-221.
- Vonesh, E. F., and Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. New York: Marcel Dekker.
- Wingo, D. R. (1984) Fitting three-parameter lognormal models by numerical global optimization - an improved algorithm. *Computational Statistics & Data Analysis* **2**, 13-25.
- Wolfinger, R. D. (1993). Covariance structure selection in general mixed models. *Communications in Statistics A* **22**, 1079-1106.

Yates, F., and Cochran, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science* **28**, 556-580.

Appendix A: Fitting the Box-Cox model

Let $\mathbf{y} = (y_1, \dots, y_n)'$ be the vector of the untransformed data assumed to follow some multivariate distribution with probability density function (p.d.f.) $f(\mathbf{y})$ and $\mathbf{t} = \mathbf{t}(\mathbf{y}; \phi)$ so that \mathbf{t} has a multivariate p.d.f. $g(\mathbf{t})$. A mixed model assuming multivariate normality can be fitted for $g(\mathbf{t})$. This is the approach taken in the present paper. From standard results it follows that

$$\log[f(\mathbf{y})] = \log[g(\mathbf{t})] + \sum_i \sum_j \log \left(\frac{\partial t_{ij}}{\partial y_{ij}} \right) \quad (4)$$

where the summation on the right-hand side corresponds to the Jacobian of the transformation from \mathbf{y} to \mathbf{z} . Note that if the data are normalized with the geometric mean of the observations (Atkinson, 1985, p.86) then the Jacobian term vanishes which makes it easier to fit using standard software. For given ϕ , $\log[f(\mathbf{y})]$ may be maximized by maximum likelihood (ML). This can be done by fitting a mixed model to the transformed data \mathbf{z} using software such as PROC MIXED of the SAS System (Littell et al., 1996). Thus, a simple way to maximize $\log[f(\mathbf{y})]$ with respect to all parameters is through a grid search over ϕ (Atkinson, 1985, p. 86). Alternatively, some numerical routine for maximizing a nonlinear function (Press et al., 1989) can be used, for example a Newton-Raphson procedure (Spitzer, 1982; Rode and Chinchilli, 1988). We use a modified grid search strategy, in which initially a triplet of starting values is chosen that brackets

the maximum of the likelihood. The triplet is then updated by the value of ϕ maximizing the quadratic polynomial fitted to the current triplet (see Press et al., 1989, for an account of similar strategies). In the examples we considered, the procedure usually took only a few updates (typically five) to find the maximum.

Appendix B: Simulation

Extensive simulation based on nonlinear mixed models such as (2) with transformed w_j was not feasible due to high computational burden and convergence problems. To gain some insight into the performance of our model selection strategy, we therefore simulated i.i.d. data following a distribution of the Johnson system or a normal when $\psi \rightarrow 0$ was assumed. The likelihood under either model was maximized using the NLPNRR and NLPNMS routines of SAS/IML. The S_U and S_B systems have a total of four parameters, while S_L has three (Johnson, 1949). For the Johnson system we profiled the likelihood using explicit expressions for the maximum likelihood estimators of shape parameters ψ and ν , conditional on the remaining parameters (location and scale). For S_L and S_B , we used numerically advantageous reparameterizations of Wingo (1984) and DeSilva et al. (1997). A grid search was performed to find good starting values for the location parameter of S_L . We used *SBC* to select among the normal distribution and the Johnson distribution employed to generate the data. The results are shown in Table 4. For many of the cases investigated, *SBC* selects the correct model among normal and Johnson a good proportion of the time. As expected, the performance of *SBC* improves as difference between models becomes more pronounced.

Table 1: Fitting information for two-way mixed models with variance-covariance structure

$\Sigma = \mathcal{A}\mathcal{A}' + I\sigma^2$ and a transformation to attain normality. Wheat data.

Transformation	What was transformed?	p^\S	$-2 \log L$	SBC
None	-	31	5669.7	5878.4
Box-Cox §	y_{ij}	32	5542.0	5757.4
Logarithmic	y_{ij}	31	5758.4	5967.1
S_L	w_j	32	5647.8	5863.2
S_B	w_j	33	5646.4	5868.6
S_U	w_j	33	5647.8	5870.0
S_L	$\mu_i + \lambda_i w_j$	31	5623.9	5832.6
S_B	$\mu_i + \lambda_i w_j$	32	5618.9	5834.3
S_U	$\mu_i + \lambda_i w_j$	32	5623.8	5839.2
Box-Cox/ S_L	y_{ij}/w_j	33	5540.7	5762.6

\S : p = number of parameters. \S : $\hat{\phi} = 0.528$.

Table 2: Estimates of shape parameters under the S_L , S_B and S_U models for w_j in $y_{ij} = \mu_i + \lambda_i w_j + e_{ij}$. and under SL model for w_j in $y_{ij}^\phi = \mu_i + \lambda_i w_j + e_{ij}$

Model	$\hat{\psi}$	s.e.($\hat{\psi}$)	$\hat{\nu}$	s.e.($\hat{\nu}$)
$S_L(w_j)$	0.42	0.094	–	–
$S_B(w_j)$	0.77	0.238	–1.33	0.450
$S_U(w_j)$	0.42	0.094	6.41	156.37
Box-Cox [#] (y_{ij})/ $S_L(w_j)$	0.11	0.087 [§]		

#: $\hat{\phi} = 0.543$. §: Standard error based on assumption that $\hat{\phi}$ is fixed.

Table 3: $E(Y_i)$ and various risk probabilities under four models for cultivars 1, 12, 13 and 14.

	Untransformed	Box-Cox	Johnson S_L (w_j)	Johnson S_L ($\mu_i + \lambda_i w_j$)
Means				
$E(Y_1)$	29.41	29.41 ^(#,\$)	29.40	29.95 [#]
$E(Y_{12})$	32.18	32.45 ^(#,\$)	32.39	33.71 [#]
$E(Y_{13})$	27.55	26.98 ^(#,\$)	27.64	27.65 [#]
$E(Y_{14})$	31.76	31.80 ^(#,\$)	31.94	33.00 [#]
Risks				
$\Pr(Y_1 < 10)$	0.14	0.12	0.12 [#]	0.12 [#]
$\Pr(Y_{12} < 10)$	0.17	0.14	0.15 [#]	0.12 [#]
$\Pr(Y_{13} < 10)$	0.055	0.052	0.026 [#]	0.041 [#]
$\Pr(Y_{14} < 10)$	0.15	0.13	0.12 [#]	0.097 [#]
$\Pr(Y_1 < 30)$	0.51	0.57	0.59 [#]	0.61 [#]
$\Pr(Y_{12} < 30)$	0.46	0.53	0.54 [#]	0.57 [#]
$\Pr(Y_{13} < 30)$	0.54	0.64	0.64 [#]	0.63 [#]
$\Pr(Y_{14} < 30)$	0.47	0.53	0.54 [#]	0.56 [#]
$\Pr(Y_1 > Y_i, i \neq 4)$	0.082 [#]	0.10 [#]	0.084 [#]	0.081 [#]
$\Pr(Y_{12} > Y_i, i \neq 12)$	0.29 [#]	0.23 [#]	0.27 [#]	0.27 [#]
$\Pr(Y_{13} > Y_i, i \neq 13)$	0.15 [#]	0.15 [#]	0.16 [#]	0.16 [#]
$\Pr(Y_{14} > Y_i, i \neq 14)$	0.19 [#]	0.18 [#]	0.18 [#]	0.19 [#]

#: Estimates based on 100,000 simulations.

\$: Simulated transformed values < 0 were set equal to zero before backtransforming to original scale.

Table 4: Simulation (10,000 runs) to assess performance of model selection procedures. Comparison of distributions of Johnson system with normal distribution.

Distribution	$\psi^{\$}$	ν	n	$\pi^{\&}$
S_L	0	0	20	0.9479
	0	0	100	0.9828
	0.2	0	20	0.7697
	0.2	0	100	0.4573
	0.5	0	20	0.3183
	0.5	0	100	0.0009
	0.8	0	20	0.0726
	0.8	0	100	0.0000
	1	0	20	0.0265
	1	0	100	0.0000
S_U	0	0	20	0.8797
	0	0	100	0.9484
	0.5	0	20	0.7598
	0.5	0	100	0.6069
	1	0	20	0.3440
	1	0	100	0.0321
	1	-0.5	20	0.3403
	1	-0.5	100	0.0086
	2	-0.5	20	0.0496
	2	-0.5	100	0.0000
S_B	0	0	20	0.5448
	0	0	100	0.8745
	0.5	0	20	0.4271
	0.5	0	100	0.7387
	1	0	20	0.2138
	1	0	100	0.0723
	1	-0.5	20	0.1666
	1	-0.5	100	0.0175
	2	-0.5	20	0.0081
	2	-0.5	100	0.0000

$\&$: π = probability of selecting normal distribution based on *SBC*.

$\$$: For $\psi = 0$ data were simulated from a normal distribution

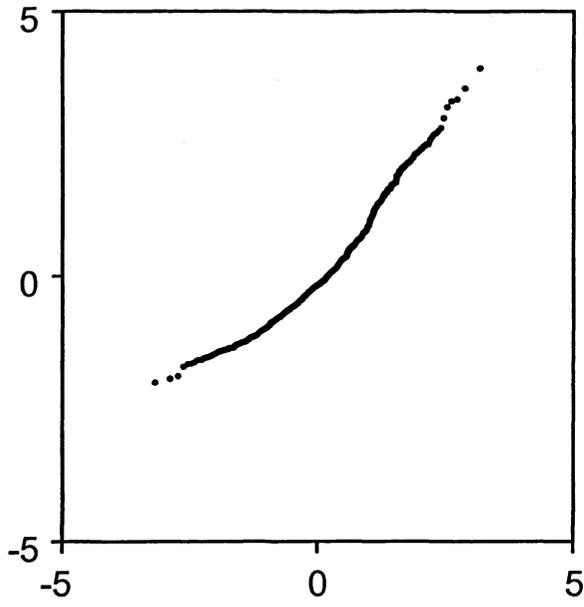


Fig. 1: Normal plot for untransformed data. $r_j = [\text{diag}(\hat{\Sigma}_j)]^{-0.5} (y_j - \hat{\mu}_j)$. Model (2).

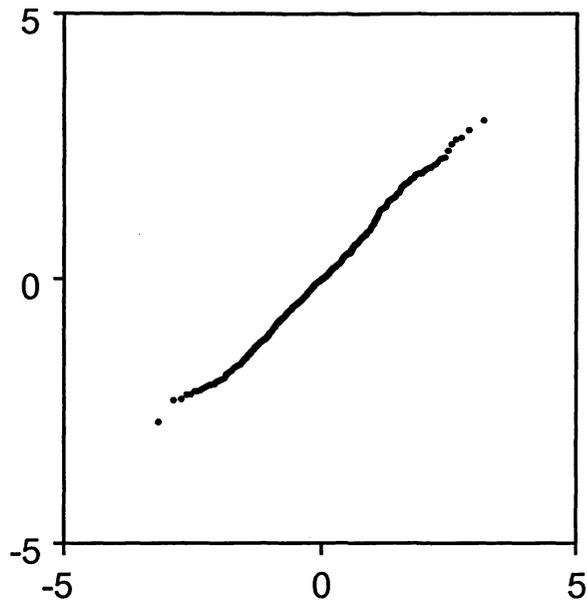


Fig. 2: Normal plot for Box-Cox-transformed data ($z_{ij} = y_{ij}^{0.528}$). $r_j = [\text{diag}(\hat{\Sigma}_j)]^{-0.5} (z_j - \hat{\mu}_j)$. Model (2).

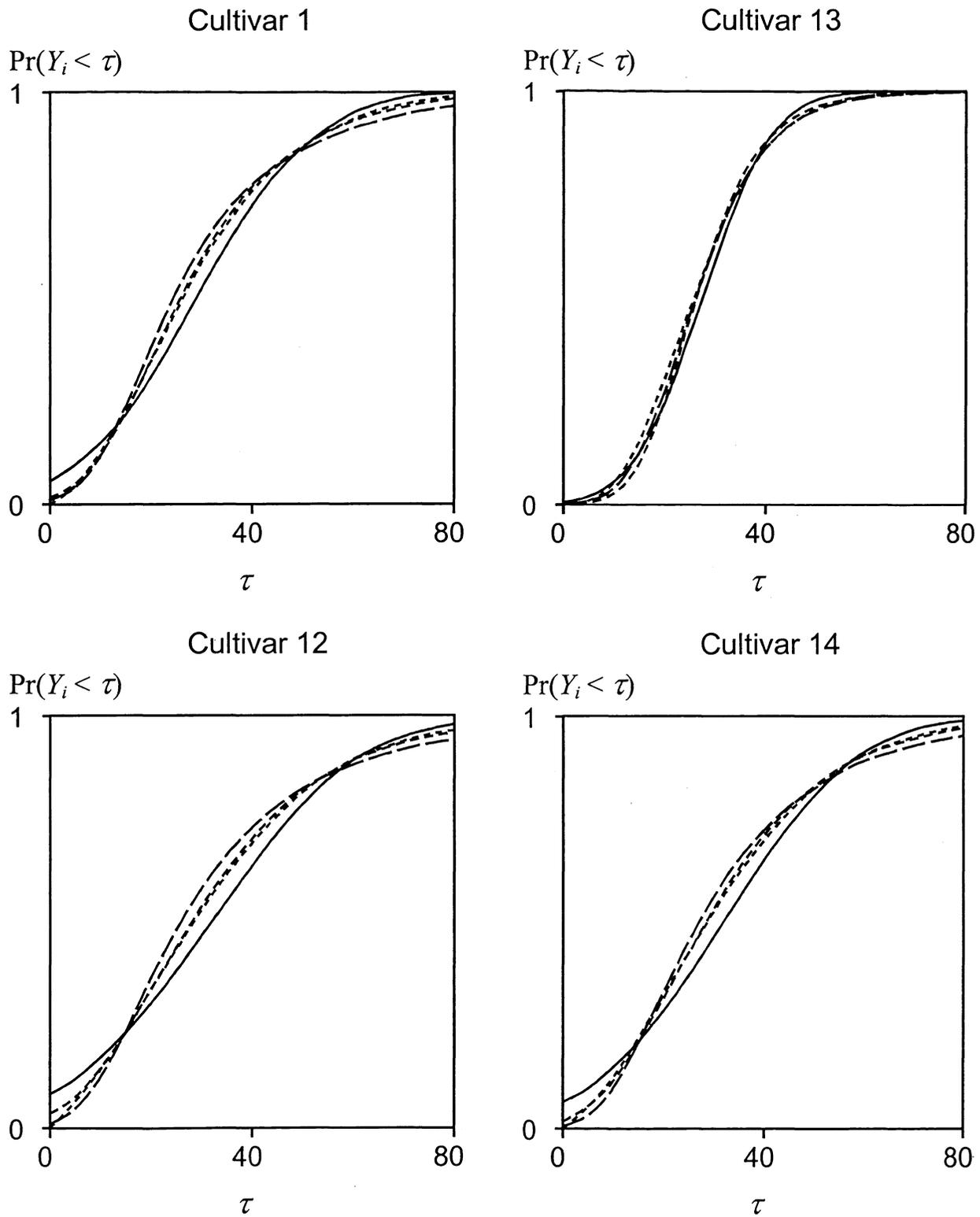


Fig. 3: Plot of risk $\Pr(Y_i < \tau)$ vs. τ under four models. Solid line: untransformed; dotted line: Box-Cox; dashed line: S_L for w_j ; widely dashed line: S_L for $\mu_i + \lambda_i w_j$.

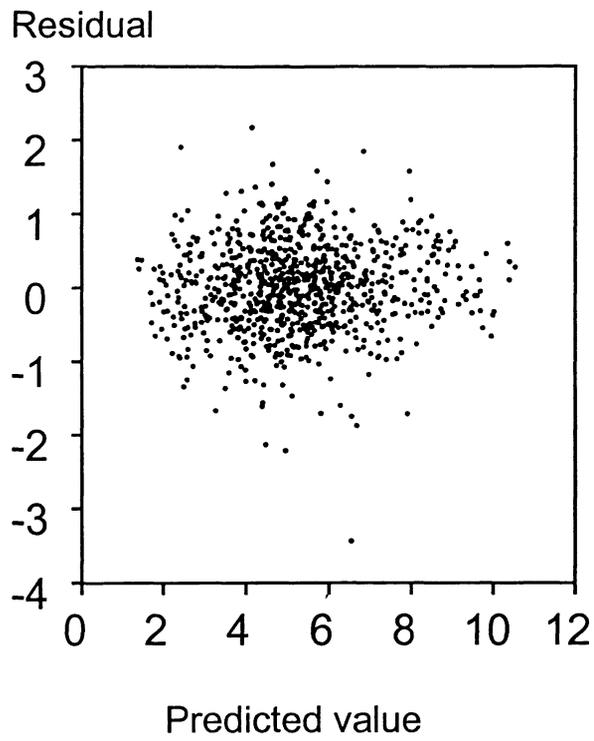


Fig. 4: Plot of residuals $y_{ij}^{0.528} - (\hat{\mu}_i + \hat{\lambda}_i \hat{w}_j)$ vs. predicted values $(\hat{\mu}_i + \hat{\lambda}_i \hat{w}_j)$ under Box-Cox model.