Perspectives on Fixed and Random Effects

Kevin Wright

2017-08-28

This document provides a collection of perspectives on fixed and random effects from published papers and other sources.

[This document had a maddening error such that the citation key "robinson1991that" was printed as "1". Tried everything to fix it.]

Robinson (1991) 1

I agree with Tukey's remark... "our focus must be on questions, not models". The choice of whether a class of effects is to be treated as fixed or random may vary with the question which we are trying to answer.

Stroup and Mulitze (1991)

In general, practicing statisticians have tended to treat the distinction between fixed and random effects as an either-or affair, even while acknowledging that in many instances, the line between the two can be rather subtle. However, in mixed linear models, a third form of inference is available, although it is not as widely known as the two traditional forms. *Predictions* of the form $\mathbf{K'B + M'U}$ may be obtained from the solutions of the *mixed model equations* described by Henderson (1975) and Harville (1976):

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + G^{-1}} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X'R^{-1}Y} \\ \mathbf{Z'R^{-1}Y} \end{bmatrix} \tag{1}$$

Henderson has shown that $\mathbf{b}$ in the solution of the mixed model equation is the same $\mathbf{b}$ as in the Generalized Least Squares equation. The solution $\mathbf{u}$ for $\mathbf{U}$ has various names in the literature: Henderson referred to it as the "best linear unbiased predictor" or BLUP; Harville alternately called it a "shrinkage estimator" or a "realized value of a random variable". They apply the same nomenclature to $\mathbf{K'B + M'U}$. Henderson called this a "predictable function", which is defined to be "predictable" if $\mathbf{K'B}$ is estimable. Thus $\mathbf{K'b + M'u}$ is the BLUP of the predictable function $\mathbf{K'B + M'U}$. This is the terminology that will be used in this article.

What is the practical value of BLUP relative to BLUE? consider a factor with a total of T levels. Suppose that T is *large* but not so large that it is impractical to observe all T levels in a single experiment. For example, in many experiments, such as agronomic variety yield tests, T may be between 20 and 100. Suppose further that the objective is to assess treatment means or differences. Under conventional methods, such an effect would be defined as fixed. Yet if the distribution of the effects is reasonably symmetric, BLUP's are typically more efficient than BLUE's. In other words, when T is large and all levels are observed, modeling the effect as random is often preferable despite the fact that it would be classified as fixed using traditional definitions. This appears to be particularly true with unbalanced data. [Text omitted.]

The point is that, in many applications, the data analyst has a dilemma: Should an effect be classified as an element of $\mathbf{B}$ (fixed) and a BLUE obtained, or as an element of $\mathbf{U}$ (random) and a BLUP obtained? The traditional distinction between fixed and random effects is not helpful; it may, in fact, lead the data analyst to choose the less efficient alternative.

Shayle R Searle (1991)

There are two other properties of $\mathbf{u}$ that are so useful to animal breeders: one is the good ranking property of $\mathbf{u}$, that in ranking sires by their corresponding values in $\mathbf{u}$ one is maximizing the probability of correctly ranking the sires, as shown by Portnoy (1982) and the other useful feature of $\mathbf{b}$ and $\mathbf{u}$ is that provided $\mathbf{K}'\mathbf{b}$ is estimable (i.e. $\mathbf{K}' = \mathbf{t}'\mathbf{X}$ for some $\mathbf{t}$, then the best estimator of $\mathbf{K}'\mathbf{B} + \mathbf{M}'\mathbf{U}$ is $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$.

Williams and Talbot (1994)

For alpha designs it is normal to specify the replicates as fixed effects and the [incomplete] blocks within replicates as random effects so that treatment information can be recovered from between blocks. For row-column designs, replicates again should be fixed with the rows and columns as random effects. For latinized designs, the long columns are specified as fixed effects if not much treatment information can be recovered from between long columns.

van Eeuwijk, Denis, and Kang (1996)

Another important question is the choice of terms as fixed or random. Two main types of arguments can be distinguished. A first type of argument is based on sampling considerations. Do the genotypes and or environments in the experiment constitute a sampled from a population to which the inference is directed? The second kind of arguments is more pragmatic, and involves the desirability of shrinkage and recovery of information, and the convenience of choosing a model term random when many parameters are associated with the term. With regard to shrinkage we may question whether it is reasonable to shrink estimates deviating from the mean of the sample back towards that mean. Or, should relatively good genotypes pay for being an element of a relatively bad sample, while relatively bad genotypes benefit from being an element of a relatively good sample? Considerations concerning recovery of information play a role when data are unbalanced. At all times it must be possible to assess whether the random effects indeed could have come from the assumed distribution. For example, the estimation of a variance component, at least 10 degrees of freedom should be available, otherwise it is preferable to take the term fixed. The same remark applies to Shukla's approach; many environments are needed for accurate estimates of individual genotypic variances.

Besag and Higdon (1999)

In a discussion of Besag and Higdon (1999), Arthur Gilmour et al. [Page 731] state:

To avoid overfitting, we insist on collaborative modelling and only include fixed effects if there is an identifiable cause. Fitting random row, column and spline effects by using restricted maximum likelihood (REML) protects against overfitting

A. Smith, Cullis, and Gilmour (2001)

One possibly contentious issue is the choice between fixed and random effects in the analysis. We have considered this in great depths and lament the lack of direction in the statistical literature. The standard textbook notion of effects being random if they have been sampled from a population and fixed if attention is confined only to those effects in the model (S. R. Searle 1971) is unhelpful and can lead to a circular argument. In our opinion, the choice depends on the aim of the analysis. In terms of variety effects, our aim is to *predict* future performance. This is best achieved by assuming the effects to be random. Initially, plant breeders and evaluators were skeptical about the use of blups. They now accept the method because the predictions have been more realistic. It is no longer true that yield gains observed by farmers are substantially lower than those predicted by CVEPs. We do not wish to predict environmental effects. The effects could be assumed random in order to recover information on varieties, but the variance component for environments is usually

so large that very little information is recovered. The magnitude of the component also means there is very little shrinkage of environmental effects with the result that the BLUPs and BLUEs are almost identical. We therefore assume that environmental effects are fixed.

Crawley (2002)

Fixed effects influence only the *mean* of [the response] y.

Random effects influence only the *variance* of y.

[Page 669] When one or more of the explanatory variables group observations (either temporally or spatially), observations within groups are usually correlated and not independent. This contravenes one of the fundamental assumptions of standard statistical models: *independence of errors*. Mixed effects models take care of this non-independence of errors by modelling the covariance structure introduced by the grouping of the data.

A major benefit of a random effects model is that it economises on the number of degrees of freedom that are used up by factor levels.

Are there enough levels in a factor to estimate the variance of the effects? No means an effect should be fixed.

Schabenberger and Pierce (2002)

[Page 627] Before proceeding further with random field linear models we need to remind the reader of the adage that *one modeler's random effect is another modeler's fixed effect*.

Piepho, Bøuchse, and Emrich (2003)

[Page 312] A factor is random when the observed levels can be regarded as randomly sampled from a population (e.g. environments and sampling units). Alternatively, a factor is random if it represents a randomization unit (e.g. plots). Otherwise the factor is usually taken as fixed (e.g. non-randomized blocks and treatments). If comparisons are to be made among the levels of a factor (e.g. treatments), the factor is considered as fixed, regardless of whether or not it is random by design. If a factor is random, then all effects containing that factor are random.

[Page 316] A random effect which does not contain (is not aliased with) the fixed effect to be tested may be taken as fixed in the analysis. This is advantageous mainly when the random effect has less than five or 10 levels, in which case the variance component estimate will be very unreliable.

[Page 319] A treatment factor may be considered random due to the sampling design, and yet there is an interest in the specific treatment levels tested in an experiment. If the number of levels is large, it is advantageous to consider the factor as random and obtain estimates of random effects under that model (S R Searle, Casella, and Mcculloch 1992, 18), e.g. in a plant breeding trial evaluating a large set of lines derived from a single cross. A popular estimation method for this purpose is known as best linear unbiased prediction (BLUP) and it is often used in plant and animal breeding S R Searle, Casella, and Mcculloch (1992). BLUPs may be more efficient than estimators assuming fixed effects.

Welham et al. (2004)

As a sample of environments in which varieties are evaluated, environment main effects might be fitted as random so that variety comparisons are combined across environments. However, previous experience indicates that, in South Australian conditions, environment differences are so large that there is little practical difference between fitting these main effects as fixed or random terms. As environment variability is not in itself of interest in this analysis, environment effects are fitted as fixed. Variety and variety $\times$ environment

effects are partitioned and fitted as random terms. This gives better predictions of future variety performance assessed over the whole set of varieties than a model with a fixed variety term, due to the minimum mean square error property of best linear unbiased predictions (BLUPs), as discussed by Patterson and Silvey (1980). The selection strategy using a random variety term is thus more efficient. In addition, more complex variance models can then be used to investigate variety variability across environments (see e.g. (A. Smith, Cullis, and Gilmour 2001). This use of random terms to get more appropriate predictions contrasts with that in the split-plot design of example 1, where random terms were used solely to define error strata.

Gelman (2005)

A persistent point of conflict in the ANOVA literature is the appropriate use of fixed or random effects, an issue which we must address since we advocate treating *all* batches of effects as sets of random variables. Eisenhart (1947) distinguishes between fixed and random effects in estimating variance components, and this approach is standard in current textbooks (Kirk 1995). However, there has been a stream of dissenters over the years; for example, Yates (1967):

> ... whether the factor levels are a random selection from some defined set (as might be the case with, say, varieties), or are deliberately chosen by the experimenter, does not affect the logical basis of the formal analysis of variance or the derivation of variance components.

Before discussing the technical issues, we briefly review what is meant by fixed and random effects. It turns out that different–in fact, incompatible–definitions are used in different contexts. (See also Kreft and de Leeuw 1998, sec. 1.3.3, for a discussion of the multiplicity of definitions of fixed and random effects and coefficients, and Robinson (1998) for a historical overview.) Here we outline five definitions that we have seen:

1. Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts α_i and fixed slope β corresponds to parallel lines for different individuals i, or the model $y_{it} = \alpha_i + \beta t$.
   (Kreft and de Leeuw 1998, 12) thus distinguish between fixed and random coefficients.
2. Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. S R Searle, Casella, and Mcculloch (1992), (section 1.4) explore this distinction in depth.
3. "When a sample exhausts the population, the corresponding variable is *fixed*; when the sample is a small (i.e., negligible) part of the population the corresponding variable is *random*" (Green and Tukey 1960).
4. "If an effect is assumed to be a realized value of a random variable, it is called a random effect" (LaMotte 1983).
5. Fixed effects are estimated using least squares (or, more generally, maximum likelihood) and random effects are estimated with shrinkage ["linear unbiased prediction" in the terminology of

   1. This definition is standard in the multilevel modeling literature (Snijders and Bosker 1999) and in econometrics.

   In the Bayesian framework, this definition implies that fixed effects $\beta_j^{(m)}$ are estimated conditional on $\sigma_m = \infty$ and random effects $\beta_j^{(m)}$ are estimated conditional on σ_m from the posterior distribution.

Of these definitions, the first clearly stands apart, but the other four definitions differ also. Under the second definition, an effect can change from fixed to random with a change in the goals of inference, even if the data and design are unchanged. The third definition differs from the others in defining a finite population (while leaving open the question of what to do with a large but not exhaustive sample), while the fourth definition makes no reference to an actual (rather than mathematical) population at all. The second definition allows fixed effects to come from a distribution, as long as that distribution is not of interest, whereas the fourth and fifth do not use any distribution for inference about fixed effects. The fifth definition has the virtue of mathematical precision but leaves unclear when a given set of effects should be considered fixed or random. In summary, it is easily possible for a factor to be "fixed" according to some of the definitions above and

"random" for others. Because of these conflicting definitions, it is no surprise that "clear answers to the question 'fixed or random?' are not necessarily the norm" (S R Searle, Casella, and Mcculloch 1992, 15).

One way to focus a discussion of fixed and random effects is to ask how inferences change when a set of effects is changed from fixed to random, with no change in the data. For example, suppose a factor has four degrees of freedom corresponding to five different medical treatments, and these are the only existing treatments and are thus considered fixed (according to definitions 2 and 3 above). Suppose it is then discovered that these are part of a larger family of many possible treatments, and so it is desired to model them as "random". In the framework of this paper, the inference about these five parameters $\beta_m^{(m)}$ and their finite-population and super-population standard deviations, s_m and σ_m, will not change with the news that they actually are viewed as a random sample from a distribution of possible treatment effects. But the super-population variance now has an important new role in characterizing this distribution. The difference between fixed and random effects is thus not a difference in inference or computation but in the ways that these inferences will be used. Thus, we strongly disagree with the claim of (Montgomery 1986, 45) that in the random effects model, "knowledge about particular [regression coefficients] is relatively useless." We prefer to sidestep the overloaded terms "fixed" and "random" with a cleaner distinction by simply renaming the terms in definition 1 above. We define effects (or coefficients) in a multilevel model as *constant* if they are identical for all groups in a population and *varying* if they are allowed to differ from group to group. For example, the model $y_{ij} = \alpha_j + \beta x_{ij}$ (of units i in groups j) has a constant slope and varying intercepts, and $y_{ij} = \alpha_j + \beta_j x_{ij}$ has varying slopes and intercepts. In this terminology (which we would apply at any level of the hierarchy in a multilevel model), varying effects occur in batches, whether or not the effects are interesting in themselves (definition 2), and whether or not they are a sample from a larger set (definition 3). Definitions 4 and 5 do not arise for us since we estimate all batches of effects hierarchically, with the variance components σ_m estimated from data.

A. B. Smith, Cullis, and Thompson (2005)

With the widespread adoption of mixed model analyses for multi-environment trial (MET) data there has been a dichotomy of thought as to the classification of variety effects as fixed or random. This is evident from the examples presented in the previous sections. The present authors believe the choice depends on the aim of the analysis and consideration of the properties of the two types of estimation procedures, namely empirical best linear unbiased prediction (E-BLUP) for random effects and empirical best linear unbiased estimation (E-BLUE) for fixed effects.

If the aim of the analysis is selection (that is, to identify the best varieties of those under consideration) then the rankings of the estimated variety effects are required to be as close as possible to the rankings of the true variety effects. In more exact terms, a set of estimates of variety effects is required that best predict the true effects. By definition, this implies the use of BLUP so that variety effects should be regarded as random. The optimality properties of BLUP are based on the assumption that the variance parameters in the model are known. In general, this is not the case and the parameters are estimated from the data. The only question that remains, therefore, is whether the estimates of the variance parameters are sufficiently precise to ensure that the optimality of BLUP is maintained with E-BLUP.

If the aim of the analysis is to determine the difference between specific pairs of varieties, then the use of BLUP as an estimation method is inappropriate since the BLUP of a specific difference is biased. Thus, in this case variety effects should be regarded as fixed.

The key issue, therefore, is a clear definition of the aim of the analysis. In order to pursue this, common practice is followed with differentiating between breeding and evaluation programmes, although the distinction is sometimes hazy. Breeding programmes are concerned with the early stages of varietal evaluation. Finney (1980) refers to this as the 'cradle to kindergarten' phase) in which large numbers (often greater than 1000) of new breeding lines are grown in small numbers (usually less than 3) of field trials. The 'best' lines are selected to continue to the next stage of testing, in which fewer lines are evaluated in more locations. The process culminates in the testing of a small number (usually less than 40) of elite breeding lines, together with commercial standard varieties, in a large number of trials that span a wide range of geographic locations

and several growing seasons. On the basis of these trials, a new breeding line may be recommended for commercial use and thence make the transition to a commercial variety. [text ommitted]

It is clear that the aim of the analysis of breeding data is selection so that the use of random variety effects is appropriate. Some statisticians advocate the use of random effects in this setting because they regard that the varieties themselves are a random sample from a population. After some unspecified number of stages of selection, this ceases to be a reasonable assumption so that at this point variety effects are regarded as fixed. The present authors do not adhere to this line of reasoning.

Most of the literature on methods for analysing MET data appears to be focused on evaluation data (or at least the example data-sets used are of this nature). It is in this setting that the fixed versus random variety effects issue is most heatedly debated. We believe that the aim of analysis of data from evaluation programmes such as those in Australia and the UK is still selection but it is now the farmer making the selection decisions rather than the plant breeder. The farmer wishes to know which varieties are best for his/her environment. These views are shared by Patterson and Silvey (1980) in their landmark paper describing the analysis of data from the UK evaluation system in which it was stated that 'The main objective of a series of NL or RL trials is to identify, with minimum selection error, the best varieties for cultivation and use.' Thus, once again the rankings of estimated variety effects (possibly within specific environments) are required to correlate well with the true rankings. In contrast, a seed company may wish to know the difference between their potential new variety and other commercial varieties, an aim that would require the use of fixed variety effects. The present authors believe, however, that the analysis of evaluation data is conducted 'for the common good', that is, to allow farmers to identify and thus adopt the best varieties for their environment. The assumption of random variety effects for both breeding and evaluation data is therefore made.

Of course, with balanced data and orthogonal analyses, the rankings of varieties would be the same in both the fixed and random variety settings. Even so, the present authors still prefer the use of random variety effects since the resultant predictions of genetic gain are more realistic than those based on fixed variety effects. The latter are generally over-optimistic due to selection bias (Patterson and Silvey 1980). An additional key advantage with the use of random variety effects is that it allows a valid analysis of data combined across stages of selection (often corresponding to a sequence of years). The analysis of such data is crucial for plant breeders since it provides more reliable estimates of variety main effects (being based on all relevant data, not merely the data for the current year) and since years are synonymous with seasons the analysis provides information on variety by season interactions.

[Concluding remarks] Despite the clear benefits of the general mixed model approach, adoption within plant breeding and crop variety evaluation programs has been very slow. In particular, the use of the more complex (and informative) models and the assumption of random rather than fixed variety effects is not widespread. This is in stark contrast to animal breeding programmes, in which REML and BLUP have been used for many years as the basis for selection and estimation of breeding values and genetic parameters. The reasons for the difference between disciplines are unclear but may have historical foundations. Plant breeding data are derived from field trials that were originally analysed (as far back as the 1930s) using an ANOVA framework where treatment (variety) effects were regarded as fixed and block effects as random. The approach was extended to MET data by regarding environments as blocks. This doctrine remained unchallenged until relatively recently when statisticians began to advocate the use of more general mixed models for MET data. It has therefore required a major culture change for plant breeding programmes to adopt the more complex models and only a small number have done so. The challenge therefore remains to improve adoption worldwide.

A historical argument against the use of mixed models for plant breeding data was the lack of suitable software. As discussed in [this paper], this is no longer an issue as the tools to fit complex mixed models to large MET data sets are now available.

A further challenge is to encourage the use of random rather than fixed variety effects. This is not an easy task, particularly as this is still a controversial topic among statisticians. As discussed earlier, the present authors believe that variety effects should be assumed to be random since this minimizes selection errors when identifying the best varieties, it provides more realistic predictions of genetic gain and allows a valid analysis of data combined across stages of selection.

S R Searle, Casella, and Mcculloch (2006)

[Page 16] In some situations the decision as to whether certain effects are fixed or random is not immediately obvious. Take the case of year effects, for example, in studying wheat yields: are the effects of years on yield to be considered fixed or random? The years themselves are unlikely to be random, for they will probably be a group of consecutive years over which the data have been gathered or the experiments run. But the effects on yield may reasonably be considered random, subject, perhaps, to correlation between yields in successive years. Of course, if one was interested in comparing specific years for some purposes, then treating years as random would not be appropriate.

In endeavoring to decide whether a set of effects is fixed or random, the context of the data, the manner in which they were gathered and the environment from which they came are the determining factors. In considering these points the important question is that of inference: are the levels of the factor going to be considered a random sample from a population of values? "Yes"–then the effects are to be considered as random effects. "No"– then, presumably, inferences will be made just about the levels occurring in the data and the effects are considered as fixed effects. Thus when inferences will be made about a population of effects from which those in the data are considered to be a random sample, the effects are considered as random; and when inferences are going to be confined to the effects in the model, the effects are considered fixed.

Another way of putting it is to ask the questions "DO the levels of a factor come from a probability distribution"? and "Is there enough information about a factor to decide that the levels of it in the data are like a random sample"? Negative answers to these questions mean that one treats the factor as a fixed effects factor and estimates the effects of the levels. Affirmative answers mean treating the factor as a random effects factor and estimating the variance component due to that factor. In that case, if one is also interested in the realized values of those random effects that occur in the data, then one also uses a prediction procedure for those values (see Section 3.4).

Schabenberger (2006)

Comments by Oliver Schabenberger at the Applied Statistics in Agriculture conference.

Often overheard: One of the greatest advantages of mixed models is that you can treated effects as fixed or random depending on the kind of analysis tha tyou are interested in. This is not true.

What they mean: Feel free to move effects from fixed to random as the analysis requires.

What they do not mean (but what is appropriate interpretation): Choose different inference spaces depending on whether you want to fix a random effect at is realization or whether you want to perform inference with respect to the random effects distribution.

Yang et al. (2009)

The determination of whether an effect is fixed or random in GE studies is not always easy and has been debated in the literature. Some statisticians have made a pragmatic suggestion that there should be enough information in the data to estimate variance and covariance parameters of random effects with sufficient precision. For example, Stroup and Mulitze (1991) suggested that a factor (genotype or environment) should have more than 10 levels before it is considered random. A. B. Smith, Cullis, and Thompson (2005) argued that genotypic effects should be random because selection of best lines or cultivars through rankings rather than comparisons is the main goal either in the early 'breeding' phase or in advanced 'evaluation' phase. Plant breeders would usually consider that years and their interactions with genotypes are random but debate considerably about how locations should be viewed. Part of the location effect would be 'fixed' because it represents known physical properties (e.g., soil type of a location) or long-term average (e.g., precipitation or other agro-climatic patterns) of the same location at some future time. However, the goal of most crop

improvement programs is to infer about future performance at many untested locations. Thus, it is our opinion that location effects and their interactions with genotypes should be random as well.

In the early phase of the breeding programs, hundreds or thousands of breeding lines need to be evaluated. The large number of breeding lines is considered as a random sample from a breeding population. It is thus reasonable to assume that genotypic and GE effects are random. After several cycles of selection, however, the number of lines is considerably reduced and these lines are now ready for a comparison with standard 'check' cultivars. At this stage, the breeding lines or cultivars may be reasonably considered as fixed (but as discussed earlier, A. B. Smith, Cullis, and Thompson 2005 argued against this consideration). Thus, FA(2) biplot is preferred at the early stage of breeding programs whereas the SREG2 or AMMI2 biplots may be useful at the late stage of breeding program.

Doug Bates

From Doug Bates, R mailing list, 6 Apr 2011

The acronym BLUP (Best Linear Unbiased Predictor) is not appropriate in the case of generalized linear or nonlinear mixed models. Alan James once told me, in reference to the Lindstrom and Bates (1990) article about nonlinear mixed-effects models that he "liked the idea of finding the random effects values that would be the BLUP's - except that they are not linear and not unbiased and there is no clear sense in which they are 'best' ".

I prefer to think of these values as the conditional modes. They are the values of the random effects that maximize the conditional density of the random effects, given the observed data (and for a fixed, "known" values of the parameters).

References

Besag, Julian, and David Higdon. 1999. "Bayesian Analysis of Agricultural Field Experiments." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61: 691–746. doi:10.1111/1467-9868.00201.

Crawley, Michael J. 2002. *Statistical Computing*. John Wiley.

Eisenhart, C. 1947. "The Assumptions Underlying the Analysis of Variance." *Biometrics* 3: 1–21.

Finney, D J. 1980. "Contribution to the Discussion of: Statutory Recommended List Trials of Crop Varieties in the United Kingdom." *Journal of the Royal Statistical Society Series A* 143: 245–46.

Gelman, Andrew. 2005. "Analysis of Variance–Why It Is More Important Than Ever." *The Annals of Statistics* 33: 1–53.

Green, B F, and J W Tukey. 1960. "Complex Analyses of Variance: General Problems." *Psychometrika* 25: 127–52.

Harville, David. 1976. "Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects." *The Annals of Statistics* 4: 384–95.

Henderson, C. R. 1975. "Best Linear Unbiased Estimation and Prediction Under a Selection Model." *Biometrics* 31: 423–48.

Kirk, R E. 1995. *Experimental Design: Procedures for the Behavioral Sciences*. Brooks/Cole, Belmon, MA.

Kreft, I, and J de Leeuw. 1998. *Introducing Multilevel Modeling*. Sage, London.

LaMotte, L R. 1983. "Encylopedia of Statistical Sciences." In, 3:137–41. Wiley, New York.

Montgomery, D C. 1986. *Design and Analysis of Experiments*. Wiley, New York.

Patterson, H D, and V Silvey. 1980. "Statutory and Recommended List Trials of Crop Varieties in the United

Kingdom." *J. Roy. Statist. Soc. Ser. A* 143: 219–52.

Piepho, Hans-Pieter, A. Bøuchse, and K. Emrich. 2003. "A Hitchhiker's Guide to Mixed Models for Randomized Experiments." *Journal of Agronomy and Crop Science* 189: 310–22. doi:10.1046/j.1439-037X.2003.00049.x.

Portnoy, S. 1982. "Maximizing the Probability of Correctly Ordering Random Variables Using Linear Predictors." *J Multivariate Anal* 12: 256.

Robinson, G K. 1998. "Encyclopedia of Biostatistics." In. Wiley, Chichester.

Schabenberger, Oliver, and Francis J Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press.

Searle, S R, G Casella, and C E Mcculloch. 1992. *Variance Components*. Wiley, New York.

———. 2006. *Variance Components*. Wiley, New York.

Searle, S. R. 1971. *Linear Models*. John Wiley.

Searle, Shayle R. 1991. "C R Henderson, the Statistician; and His Contributions to Variance Components Estimation." *J Dairy Sci* 74: 4035–44.

Smith, A B, B R Cullis, and R Thompson. 2005. "The Analysis of Crop Cultivar Breeding and Evaluation Trials: An Overview of Current Mixed Model Approaches." *Journal of Agricultural Science* 143: 1–14.

Smith, Alison, Brian Cullis, and Arthur Gilmour. 2001. "The Analysis of Crop Variety Evaluation Data in Australia." *Aust. N. Z. J. Stat.* 43: 129–45.

Snijders, T. A. B., and R J Bosker. 1999. *Multilevel Analysis*. Sage, London.

Stroup, W W, and D K Mulitze. 1991. "Nearest Neighbour Adjusted Best Linear Unbiased Prediction." *The American Statistcian* 45: 194–200.

van Eeuwijk, FA, JB Denis, and MS Kang. 1996. "Genotype-by-Environment Interaction." In, 15–50. CRC.

Welham, Sue, Brian Cullis, Beverley Gogel, Arthur Gilmour, and Robin Thompson. 2004. "Predictions in Linear Mixed Models." *Aust. N. Z. J. Stat.*, 325–47.

Williams, A, and M Talbot. 1994. *Alpha+. Experimental Designs for Variety Trials and Many-Treatment Experiments*.

Yang, Rong-Cai, José Crossa, Paul L. Cornelius, and Juan Burgueno. 2009. "Biplot Analysis of Genotype X Environment Interaction: Proceed with Caution." *Crop Science* 49: 1564–76. doi:10.2135/cropsci2008.11.0665.

Yates, F. 1967. "A Fresh Look at the Basic Principles of the Design and Analysis of Experiments." In *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, 4:777–90. Univ. California Press, Berkeley.