SLU

Linear mixed models for series of variety trials

Johannes Forkman
Swedish University of Agricultural Sciences
Department of Crop Production Ecology
Box 7043, SE-750 07 Uppsala, Sweden
johannes.forkman@vpe.slu.se

Summary

When using linear mixed models for analysis of series of variety trials, the statistical inference space is dependent on the design of the series. In this paper, four inference spaces are recognized for one-year crop variety trials: single-year local, long-term local, single-year regional and long-term regional. In addition, different linear mixed models are appropriate for different designs. Five models are defined that are useful for 13 different types of series of variety trials. The standard statistical analysis includes estimation of variance components using the REML method and estimation of variety means by generalised least squares. Although this standard method gives the best linear estimates, provided correct variance components, the estimated differences between the test varieties and the control variety occasionally do not agree with yearly or local results. An alternative method is outlined, termed the control method, which does not have this problem.

Key words and phrases: Inference spaces, Linear mixed models, Series of variety trials, VCU testing.

1. Introduction

The usefulness of linear mixed models for analysis of series of variety trials is well recognized (Smith, Cullis and Thompson, 2005; Spilke, Piepho and Hu, 2005). In the statistical analysis of series of one-year crop variety trials, effects of years, locations and trials may, with advantage, be modelled as random (i.e. as if they were random samples from normally distributed populations). Variety-by-year and variety-by-location interactions may also be modelled as random, enabling all relevant variance components to be taken into account when testing the varieties.

Series of variety trials may be analysed with one-stage analysis or two-stage analysis. One-stage analysis uses a complete dataset comprising all observations of the response variable from the trials. In this case it is necessary to consider the experimental designs of the trials, for example blocking, in the analysis of the series. Two-stage analyses are performed in two steps. In the first step, averages are estimated in all separate trials, and in the second step, the analysis of the series is carried out on the estimated averages from the first step. Using two-stage analysis, the experimental designs are not included in the analysis of the series (i.e. in the second step). This is convenient, especially when different experimental designs have been used in the trials of the series. This paper discusses two-stage analyses of series of variety trials.

Some series of variety trials extend over many years, others comprises trials only from a single year. Also, some series include many locations, but others only trials from a single location. When using linear mixed models for analysis of series, the statistical inference space

1

is dependent on the design of the series with regard to years and locations. Section 2 defines four inference spaces for series of one-year crop variety trials.

Field researchers may find it hard to specify a mixed model appropriate for the analysis of a given series. As a guide, Section 3 lists 13 different types of series, and suggests five different linear mixed models for the analyses. Some of the models were advocated by Patterson (1997). Piepho, Büchse and Emrich (2003) provided a useful general guideline for setting up linear mixed models for agricultural experiments.

Section 4 discusses statistical analysis of series of varieties using linear mixed models. The standard method uses maximum likelihood. The variance components are usually estimated with the residual maximum likelihood method (REML). The fixed effects, i.e. the variety means, are estimated through weighted least squares, given the estimates of the variance components. Variety trials that comprise several years are often highly unbalanced, because new varieties enter the trials every year, and older, less performing, varieties are phased out. Sometimes in analyses of unbalanced series including many years, the estimated difference between a specific test variety and the control variety does not agree with yearly estimates. Section 4 outlines an alternative method, the control method, which does not share this problem. Section 5 uses a small series of pea trials as an example.

2. Inference spaces

If the series include trials from a single year inference can be made for that specific year only. It is not possible to extend the results of the study to other years, because varieties may be sensitive to weather conditions, which vary randomly between years. A variety that is top performing one year may not perform well in the following year. Similarly, inference cannot be made about the performance of the varieties at other locations if the series only include trials from one single location (i.e. from one experimental station).

The statistical inference space may be termed *single-year local* if all trials of the series are from a single year and from a single location. If the series include more than one year, but only include trials from a single location, the inference space may be called *long-term local*. Such series provide information about variety-by-year interactions, but not about variety-by-location interactions. In contrast, the series may include trials from many locations, all carried out during the same year. The statistical inference space may in this case be termed *single-year regional*, because inference can be made about the performance of the varieties in the region that the locations represent, but not about the performance in other years than the investigated. Series comprising many locations and years may finally be called *long-term regional*. Such series examines both variety-by-year interactions and variety-by-location interactions.

Researchers planning series of variety trials should be aware that the statistical inference space, when using mixed models, is implied by the design of the series, as illustrated in Figure 1. For example, if the series comprise trials from two years, the inference space is either long-term local or long-term regional, and the objective is to make inference about long-term differences between varieties. However, based on only two years, lasting differences can hardly be established with any precision. Clearly, more years are needed for investigation of long-term differences. The researcher may in this case prefer analysing the two years of the series separately, or with a fixed effects model, but with such analyses long-term differences are not estimated.
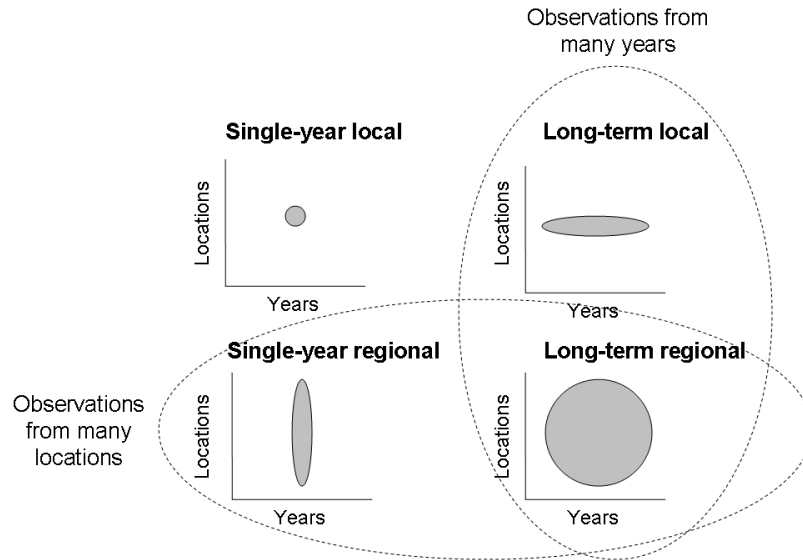
Figure 1. Inference spaces for series of variety trials. The inference space is single-year local when the series include observations from a single year and from a single location. The inference space is long-term local or long-term regional when the series include observations from more than one year. The inference space is single-year regional or long-term regional when the series includes observations from more than one location.

3. Statistical models

This section provides five useful models for analysis of series of variety trials. The models may be used for 13 common types of series, enumerated from *i* to *xiii*.

A short-hand model notation, previously used by Patterson (1997) and Piepho, Büchse and Emrich (2003), is utilised. In this notation, the linear model is simply specified by giving the names of the explanatory factors of the model, separated by plus signs. Interactions, for example between the factors A and B, are written as A·B. Fixed effects are separated from random effects by a colon. On the left-hand side of the colon, the fixed factors are specified, and on the right-hand side the random effects. The response variable, as well as the residual error term, is implicit. All random effects and the random error term are assumed to be independent and normally distributed with expected value 0.

Model 1 is

$$\text{Variety : Trial .}$$

Model 1 is useful when

- *i.* There are trials from one single year and one single location,
- *ii.* There are trials from several years but only from one location, and there is only one trial per year,
- *iii.* There are trials from several locations but only from one single year, and there is only one trial per location,
- *iv.* There are trials from several years and several locations, but there is only one trial per year and one trial per location.

3

The factors Location and Year are not included in Model 1, because in series of type *i*, there is only one year and one location; in series of type *ii*, there is only one location and years are confounded with trials; in series of type *iii*, there is only one year, and locations are confounded with trials, and in series of type *iv* years and locations are confounded with trials.

When locations are confounded with trials (i.e. in series of types *iii* and *iv*), Model 1 is identical to the first basic REML model recommended by Patterson (1997). When years are confounded with trials (i.e. in series of types *ii* and *iv*), Model 1 is identical to the second basic REML model proposed by Patterson (1997).

Model 2 is

$$\text{Variety : Year + Variety·Year + Trial .} \qquad (3.1)$$

Model 2 is useful when

v. There are trials from several years, and several trials per year. The trials are from one single location.
vi. There are trials from several years and several locations. There is only one trial per location. There are several trials per year.
vii. There are trials from several locations and several years. There are several trials per location and year. All trials from a specific location are from the same year. All trials from a specific year are from the same location.

The factor Location is not included in Model 2, because in series of type *v* there is only one location; in series of type *vi*, locations are confounded with trials, and in series of type *vii*, locations are confounded with years.

Model 2 is identical to the fourth basic REML model proposed by Patterson (1997) when the interaction between years and locations is confounded with trials (i.e. in series of type *vi*).

Model 3 is

$$\text{Variety : Location + Variety·Location + Trial .}$$

Model 3 is useful when

viii. There are trials from several locations, and several trials per location. The trials are from one single year.
ix. There are trials from several locations and several years. There is only one trial per year. There are several trials per location.
vii. There are trials from several locations and several years. There are several trials per location and year. All trials from a specific location are from the same year. All trials from a specific year are from the same location.

The mathematical structure in Model 3 is the same as in Model 2, but with Location substituted for Year. The factor Year is not included in Model 3, because in series of type *viii* there is only one year; in series of type *ix*, years are confounded with trials, and in series of type *vii*, years are confounded with locations. For series of type *vii*, either Model 2 or Model 3 may be used.

4

Model 4 is

Variety : Year + Location + Variety·Year + Variety·Location + Trial .

Model 4 is useful when

x.   There are trials from several years and several locations. There are several locations per year and several trials per location. All trials from a specific location are from the same year.

xi.   There are trials from several locations and several years. There are several years per location and several trials per year. All trials from a specific year are from the same location.

xii.   There are trials from several years and several locations. There are several trials per location and per year, but only one trial per year and location.

The interaction Year·Location is not included in Model 4, because in series of type *x*, this interaction is confounded with locations, in series of type *xi* with years, and in series of type *xii* with trials.

Model 4 is identical to the third basic REML method proposed by Patterson (1997), when the interaction between years and locations are confounded with trials (i.e. in series of type *xii*)

Model 5 is

Variety : Year + Location + Variety·Year + Variety·Location + Year·Location + Trial .

Model 5 is useful when

xiii.   There are several years per location and several locations per year. There are several trials per year and location.

In series of type *xiii*, no factors or interactions are confounded with other factors or interactions. Consequently, Model 5, which is the full model, may be used.

4. Methods for statistical analysis

The variance components of the model may be estimated with the REML method, which gives the same estimates as the traditional ANOVA method when the series is balanced (Robinson, 1987). Given the estimates of the variance components, the fixed effects may be estimated by the method of generalised least squares. It is well known that the generalised least squares estimates of the fixed effects are the best linear unbiased estimates with regard to standard error, if the variance components are known (Goldberger, 1962).

The variances in the estimates of the variety means in Models 1–5, and the covariances between the variety means, are dependent on the variance components of the model. The standard procedure involves estimating the variances and the covariances for the means by exchanging the unknown variance components for their estimates, and carrying out an approximate F-test of the hypothesis of no difference between the varieties, $H_0: \alpha_i = \alpha_1$, $i = 2$, ..., I, as well as approximate t-tests for pairwise comparisons (Littell et al., 2006).

Results of analyses of series of variety trials are often presented in tables with estimated variety means, i.e. estimates of α_i, $i = 1, 2, ..., I$. Usually one of the varieties, which is included in all trials of the series, is regarded as a control variety, and the ratios of the estimates of the expected values of the other varieties to the estimate of the expected value of the control variety are shown in the table. This common interest in ratios, rather than in differences, suggests analysis on the logarithmic scale. When calculated on the logarithmic scale, the estimates of the parameters α_i, $i = 1, 2, ..., I$, may easily be back-transformed to the original scale by the exponential function. Similarly, calculated confidence intervals for differences on the logarithmic scale may be back-transformed to confidence intervals for ratios.

In short, the standard method for analysis of series of variety trials using mixed models involves the REML method for estimation of variance components and the generalised least squares method for estimation of variety means. The estimated variance components are taken into account through the use of approximate F- and t-tests. The analysis may be performed on the logarithmic scale.

Occasionally, in highly unbalanced series, the standard method produces estimates of variety means that, when presented as ratios to the mean of the control variety, appear inconsistent with the ratios obtained in the trials. An example of this phenomenon is given in Section 5. For such cases, an alternative method could be preferred, based on direct comparisons with the control. In the following such a method is outlined, which will be termed the control method.

Using the control method, the variance components are estimated through the REML method, exactly as with the standard method. However, instead of estimating the variety means by the method of generalised least squares, their differences to the mean of the control variety are estimated by weighted average differences. For example, using Model 2 and eq. (3.2), let d_{pj} denote the difference between the means of test variety p and the control variety in year j, $j = 1, 2, ..., J$. The expected difference δ_p is estimated by

$$\delta_p^* = \frac{\sum_{j=1}^{J} w_{pj} d_{pj}}{\sum_{j=1}^{J} w_{pj}} \ , \tag{4.1}$$

where

$$w_{pj} = \left(\hat{\sigma}_g^2 + \frac{\hat{\sigma}_e^2}{K_{pj}} \right)^{-1} . \tag{4.2}$$

In (4.2), $\hat{\sigma}_g^2$ and $\hat{\sigma}_e^2$ are the REML estimates of σ_g^2 and σ_e^2, respectively, and K_{pj} is the number of trials including the test variety p and the control variety, in the j:th year. The variance in (4.1) is estimated by

$$\text{var}(\delta_p^*) = \frac{1}{\sum_{j=1}^{J} w_{pj}} \ ,$$

and

$$\delta_p^* \pm t \frac{1}{\sqrt{\sum_{j=1}^{J} w_{pj}}} \quad , \tag{4.3}$$

is an approximate $(1 - \alpha)$ % confidence interval for δ_p. In (4.3), t is the $(1 - \alpha/2)$:th percentile of a t distribution with $M - I - J + 1$ degrees of freedom, where M is the total number of variety-by-year combinations.

Let y_{cjk} denote the observation of the control variety in the k:th trial of the j:th year. The expected value α_c of the control variety is estimated by a weighed average of means, i.e. by

$$\alpha_c^* = \frac{1}{\sum_{j=1}^{J} u_j} \sum_{j=1}^{J} \frac{u_j}{K_j} \sum_{k=1}^{K_j} y_{cjk} \quad ,$$

where K_j is the number of trials including the control variety in the j:th year, and

$$u_j = \left(\hat{\sigma}_a^2 + \hat{\sigma}_g^2 + \frac{\hat{\sigma}_t^2}{K_j} + \frac{\hat{\sigma}_e^2}{K_j} \right)^{-1} \quad . \tag{4.4}$$

In (4.4), $\hat{\sigma}_a^2$, $\hat{\sigma}_g^2$, $\hat{\sigma}_t^2$ and $\hat{\sigma}_e^2$ are the REML estimates of σ_a^2 σ_g^2 σ_t^2 and σ_e^2, respectively The expected value of α_i is estimated by $\alpha_c^* + \delta_i^*$, $i = 1, 2, ..., I$.

5. Example

A small series of in total 22 pea trials were carried out during the five years period 2001–2005 on varying locations. The statistical inference space was consequently long-term regional. The series comprised eight test varieties and one control variety: Celine. The series was analysed, on the logarithmic scale, using Model 2 (eq. 3.1), with the standard method as well as the control method. The variance components were estimated, and the standard analysis performed, using the mixed procedure in SAS 9.2 (Littell et al., 2006).

The standard deviations σ_a, σ_g, σ_t and σ_e were estimated, using the REML method, to 0,242; 0,030; 0,121 and 0,071 respectively. The probability value of the approximate F-test was 0,001, indicating differences between the varieties.

The standard method produced the results of Table 1. Using this method, the test variety Exclusiv was estimated to produce 2 % larger yield than the control variety Celine. This result is surprising given yearly information about the performance of Exclusive in comparison with Celine. Within the series, Exclusiv was only included in six trials: four in 2004, and two in 2005. The average difference on the logarithmic scale between Exclusiv and Celine was 0,040 in 2004 and 0,074 in 2005. With calculations on the original scale, the average was 4,3 % larger with Exclusiv than with Celine in 2004 and 5,7 % larger with Exclusiv than with Celine in 2005.

Clearly, the standard method does not provide average differences to the control variety. Some researchers could prefer the control method, which does calculate weighted average differences. The results of the control method are presented in Table 2. Using the control method, the ratio between Exclusiv and Celine is estimated to 105 %.

Table 1. Results of an analysis of a series of 22 pea trials, using the standard method performed on logarithmic scale. The table includes the numbers of trials, the numbers of years and estimated averages and ratios to the control variety Celine, with 95 % confidence intervals, after back-transformation to the original scale.

Variety	N Trials	N Years	Mean (kg/ha)	Rel. (%)	95 % Conf. int.
Clara	19	5	4952	95	(90 , 101)
Jackpot	20	4	4995	96	(90 , 102)
Faust	22	5	5129	99	(93, 105)
Stilo	6	2	5233	101	(92 , 110)
Exclusiv	6	2	5294	102	(93 , 112)
Tudor	6	2	4310	83	(76 , 91)
Brutus	20	5	4843	93	(88 , 99)
Pinochio	16	3	5015	96	(90 , 103)
Celine	*22*	*5*	*5198*	*100*	

Table 2. Result of an analysis of a series of pea trials, performed on logarithmic scale, using the control method. The table includes the numbers of trials, the numbers of years and estimated averages and ratios to the control variety Celine, with 95 % confidence intervals, after back-transformation to the original scale.

Variety	N Trials	N Years	Mean (kg/ha)	Rel. (%)	95 % Conf. int.
Clara	19	5	4933	95	(89 , 101)
Jackpot	20	4	4970	96	(90 , 102)
Faust	22	5	5122	99	(93, 105)
Stilo	6	2	5412	104	(94 , 116)
Exclusiv	6	2	5475	105	(95 , 117)
Tudor	6	2	4457	86	(77 , 96)
Brutus	20	5	4843	93	(88 , 99)
Pinochio	16	3	4928	95	(88 , 102)
Celine	*22*	*5*	*5191*	*100*	

Given the estimates of the variance components, the ratio 102 % in Table 1, between Exclusiv and Celine, was calculated based on all 22 trials of the series. The ratio 105 % in Table 2 was calculated based on only the six trials including Exclusiv. Trials not including Exclusiv did not directly affect the estimate of the performance of Exclusiv. Trials without Exclusiv did only indirectly have an influence on the estimate of the performance of Exclusiv, because in the calculation of the ratio 105 %, a function of the estimated variance components were used as weights, and the variance components were estimated based on all 22 trials.

Many researchers do not accept that conclusions about the performance of Exclusiv be based on trials others than the ones including Exclusiv. These researchers may prefer the control method to the standard method. On the other hand, the control method is theoretically less efficient than the standard method. This is easily seen by comparing the widths of the confidence intervals of Tables 1 and 2.

6. Discussion

Linear mixed models are very useful for analysis of series of variety trials. Using mixed models, all relevant variance components can be accounted for when testing differences between varieties. The design of the series implies different inference spaces. If the inference space is long-term, the series should be run over many years; otherwise long-term differences cannot usually be established with adequate precision. Similarly, if the inference space is regional, the series should comprise many locations within the region, because otherwise systematic differences between varieties cannot usually be estimated sufficiently well.

Occasionally, in unbalanced series, the standard method mixed model estimates of the differences between the test varieties and the control variety are not consistent with yearly results, as shown in the example of Section 5. This inconsistency is not due to the logarithmic transformation or mixed modelling; it is an effect of the series being unbalanced with regard to years and varieties. Exactly the same phenomenon occurs in traditional analysis of variance, which includes a single error term, carried out on original scale (Forkman, 2007).

Given correct estimates of the variance components, the standard method, using generalised least squares, is the best linear unbiased method for comparison of varieties. The standard method is for this reason recommended. However, the control method, sketched in this paper, may serve as an alternative when it is considered very important that the results of the analysis of the series harmonise with yearly or local results, and when the result for a specific variety is required to be based on only the experiments comprising that variety.

References

Forkman J., 2007, Analysing and reporting series of variety trials. *In* H.P. Piepho and H. Bleiholder (eds.), *Agricultural field trials – today and tomorrow*. Proceedings of the international symposium 8-10 Oct 2007, Stuttgart-Hohenheim, Germany. Verlag Grauer.

Goldberger A.S.,1962, Best linear unbiased prediction in the generalized linear regression model. Journal of the American Statistical Association 57, 369-375.

Littell R.C., Milliken G.A., Stroup W.W., Wolfinger R.D., Shabenberger O., 2006. *SAS for mixed models*. 2nd ed. SAS Institute Inc.

Patterson H.D., 1997, Analysis of series of variety trials. *In* R.A. Kempton, and P.N. Fox (eds.) *Statistical methods for plant variety evaluation*. Chapman & Hall.

Piepho H.P., Büchse A., Emrich K., 2003, A hitchhiker's guide to mixed models for randomized experiments. Journal of Agronomy and Crop Science 189, 310-322.

Robinson D.L., 1987, Estimation and use of variance components. The Statistician 36, 3-14.

Smith A.B., Cullis B.R., Thompson R., 2005, The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143, 449-462.

Spilke J., Piepho H.P., Hu X., 2005, Analysis of unbalanced data by mixed linear models using the mixed procedure of the SAS system. Journal of Agronomy and Crop Science 191, 47-54.