ABSTRACT
Agronomic experiments are often replicated over time and space to evaluate how treatments perform across a range of environments. The analysis of experiments conducted for more than one growing season (years) and/or places (locations) is commonly referred to as analysis of combined experiments. Common analyses of these studies treat some effects as fixed, treat others as random, and usually include interactions between fixed and random effects, which we call mixed interactions. Recommendations for how to treat mixed interactions has changed. In the traditional practice, the effects of interactions between fixed and random effects were assumed to sum to zero within each level of a fixed factor. Contemporary practice considers these effects to be mutually independent. This latter assumption is used to construct $F$ tests by many of the statistical analysis programs that are widely used to analyze data from agronomic experiments but is inconsistent with that used in many previously published studies. The assumptions made about mixed interactions in the analysis of variance can result in very different interpretations and can potentially lead to different conclusions. We address the discrepancy between the analyses that were formerly recommended and those that are currently implemented by popular software programs and provide recommendations for analyzing data from combined experiments.

Replication of agronomic experiments over multiple growing seasons and/or geographic locations is commonly done to expand the scope of inference of the experiment over time and space. These analyses are usually performed with both years and locations considered as random factors because the objective is to compute the variability of treatment differences across several environments. Analysis of multi-environment experiments is often referred to as the combined analysis of multiple experiments. McIntosh (1983) described the appropriate analyses of combined experiments based on the constraint that the effects of random interactions, e.g., treatment by year interactions, sum to zero across each level of a fixed factor, e.g., treatment. This assumption reflects a tradition in the analysis of fixed effects that, in most cases, is not appropriate for a random effect (Hocking, 1985, p. 332–334). A more appropriate assumption is that all random effects are mutually independent (Hocking, 1985, p. 332). This change in assumptions reflects a shift in perspective as software and computational capacity has evolved. One consequence of this shift is that conducting the hypothesis tests suggested by McIntosh (1983) can be difficult with modern software, e.g., SAS (Goodnight and Speed, 1978; Littell et al., 2006) or R (R Core Team, 2013). In fact, if variance components estimated by modern software are used to compute certain of the $F$ tests suggested by McIntosh (1983), the results will be incorrect because modern software and McIntosh (1983) define some variance components differently.

The assumption made about mixed interactions affects the inferences that can be made from the analysis (McLean et al., 1991). It is not uncommon for agronomists to make tests of significance based on broad-sense inferences such as described by McIntosh (1983) and then to examine mixed interactions as if they were fixed effects. This is reasonable if the researcher wishes to understand and describe what happened in a given environment but is inconsistent with the broader $F$ tests that have often been made in the analysis of variance. It would be more reasonable to consider years and locations as fixed in an analysis of variance. Making such narrow-sense inferences can be misleading, however, if they are in fact projected to a broader inference space in which the effects of years and locations occur randomly. While it is possible to conceptualize locations used in an experiment as fixed, the effects associated with years almost always occur randomly in dryland production areas because they are completely confounded with the weather that occurred in that year. It would make little sense to draw an inference and make a recommendation related to an event that occurred in the past and is largely unrepeatable.

K.J. Moore, Dep. of Agronomy, Iowa State Univ., Ames, IA 50011; and P.M. Dixon, Dep. of Statistics, Iowa State Univ., Ames, IA 50011. Both authors contributed equally to the paper and may be contacted for further information or questions about the topic. Received 7 Oct. 2013. Accepted 7 Feb. 2014.
*Corresponding authors (kjmoore@iastate.edu; pdixon@mail.iastate.edu).

doi:10.2134/agronj13.0485
Available freely online through the author-supported open access option.
Copyright © 2015 by the American Society of Agronomy, 5585 Guilford Road, Madison, WI 53711. All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Abbreviations: REML, restricted maximum likelihood.
Here we explain the difference between the two assumptions about random effects and describe the consequences of the choice of assumptions about interactions. We also give recommendations about pooling random effects.

**A LINEAR MODEL FOR COMBINED EXPERIMENTS**

One possible additive linear model for an experiment repeated over years and locations is

\[
Y_{ijkl} = \mu + Y_i + L_j + YL_{ij} + B_{(i)k} + T_l + YT_{il} + LT_{jl} + YT_{ijl} + BT_{(i)kl} + e_{ijkl}
\]

where \(Y_{ijkl}\) is the response measured on the \(ijkl\)th experimental unit (plot), \(\mu\) is the overall mean, \(Y_i\) is the effect of the \(i\)th year, \(L_j\) is the effect of the \(j\)th location, \(YL_{ij}\) is the interaction effect of the \(i\)th level of \(Y\) with the \(j\)th level of \(L\), \(B_{(i)k}\) is the effect of the \(k\)th block within the \(i\)th location, \(T_l\) is the effect of the \(i\)th treatment, \(YT_{il}\) is the interaction effect of the \(i\)th level of \(Y\) with the \(l\)th level of \(T\), \(LT_{jl}\) is the interaction effect of the \(j\)th level of \(L\) with the \(l\)th level of \(T\), \(YT_{ijl}\) is the interaction effect of the \(i\)th level of \(Y\) with the \(j\)th level of \(L\) and the \(l\)th level of \(T\), and \(BT_{(i)kl}\) is experimental error, \(NID(0, \sigma_e^2)\).

Inferences based on this model depend on whether each of the factors in the model are considered fixed or random. Fixed factors are those in which the levels included in the experiment are controlled by the researcher and are of specific interest. The interpretation of fixed effects is based on means and is constrained by the levels chosen and included in the experiment. The quantities of interest for a fixed factor are the means and differences (or, more generally, linear combinations) of means. Random factors are assumed to represent a sample from some population. Ideally, levels are chosen randomly from the target population, but often levels are haphazardly chosen and the target population is imagined. The quantity of interest for a random effect is its variance. In ANOVA, the population of random effects is usually considered to be normally distributed, although other distributions can be used (Gbur et al., 2012).

How the model described by Eq. [1] is applied depends on the specific objectives of the experiment and the corresponding inference space that is desired. The most common application considers the treatment to be a fixed effect and all other terms to be random. In particular, the interactions of treatment with year, location, or the combination of year and location are considered to be random effects. Until specifically indicated otherwise, our explanations are for this version of the model.

The issues we discuss are not relevant if all effects in the model, except the residual variance, are considered fixed effects. Tests of all terms in a fixed-effect model use the residual variance as the error term (denominator in an \(F\) test). This has two crucial consequences for the interpretation of the results. The subtle one is the shift in inferential space between narrow-sense inference and broad-sense inference (McLean et al., 1991). An analysis using narrow-sense inference uses the residual variance as the error term and makes conclusions about treatment effects for the specific years and locations in the study. If there are significant interactions involving treatment factor(s), e.g., treatment by location, the usual recommendation is to do and report analyses separately for each location. This approach clearly makes individual conclusions for each location (or perhaps year or combination of years and locations) an analysis using broad-sense inference uses a random treatment interaction or combination of random treatment interactions as the error term. These interactions estimate the consistency of the treatment effect across the study locations (or years). The conclusions are about the treatment effect averaged across the population of locations (or years) sampled in the study (McLean et al., 1991). The test of treatment effects compares the magnitude of the treatment effects to their consistency across locations (or years). A significant interaction involving a treatment factor does not change the structure of a broad-sense analysis. The conclusions are applicable to new locations (or years), while narrow-sense conclusions are not.

Factor effects models for ANOVA are overparameterized, in the sense of having more parameters than group means. In the classic experimental design literature (e.g., Anderson and McLean, 1974; Petersen, 1994; Snedecor and Cochran, 1989; Steel et al., 1997), the extra fixed effect parameters are eliminated by imposing sum-to-zero constraints. Effects for a main effect sum to zero. Effects for interactions sum to zero within each level of the component main effects. When a model included random effects in addition to the residual error, nearly all of the classic texts on experimental design recommended sum-to-zero constraints for fixed effects and some random effects (Anderson and McLean, 1974; Petersen, 1994; Snedecor and Cochran, 1989; Steel et al., 1997). Specifically, the random effects for an interaction between a fixed effect and a random effect were assumed to sum to zero within each level of the fixed effect. Under the sum-to-zero constraint, the treatment by year random effects would sum to zero within each treatment level, as would treatment by location effects and treatment by year by location effects. Mathematically, this constraint imposes a negative correlation among pairs of interaction effects within the same treatment (Hocking, 1985, p. 331). The expected mean squares for the analysis of a combined experiment proposed by McIntosh (1983) are based on the sum-to-zero constraint.

More recent texts (Kuehl, 2000) and, more importantly, many modern statistical software programs use the independence assumption (SAS Institute). Under the independence assumption, all random effects, both main effects (e.g., year) and interactions (e.g., treatment by year) are considered independent. The RANDOM statements of the GLM, MIXED, and GLIMMIX procedures of SAS and the lme() and lmer() functions in R generate \(F\) tests consistent with the independence assumption. When year, location, and block and their interactions are random, some of the tests produced by commonly used software vary significantly from those described by McIntosh (1983), so using McIntosh (1983) as a reference for the analysis is inappropriate.

**ASSUMPTIONS ABOUT EXPECTED MEAN SQUARES**

The choice of assumptions about certain random effects applies only to random effects that are interactions between a fixed effect and a random effect. Therefore, it is only an issue with a mixed model. It is not a concern in either a model with
only fixed effects (and a random error) and or a model with only random effects (and a fixed intercept). The choice of assumption is crucial for the analysis of a combined experiment because the common application of Eq. [1] is a mixed model.

More specifically, it involves variance components associated with mixed interactions. In the traditional approach, the assumption is that the effects of the fixed factor involved in a mixed interaction sum to zero across the levels included in the experiment (Lorenzen and Anderson, 1993). Thus, with respect to testing random main effects and interactions, these mixed variance components are not included in the expected mean squares (Table 1).

The sum-to-zero approach makes intuitive sense from an inferential perspective. Tests for fixed main effects and interactions are made using error terms that include variance components for random factors. That is, the inference space for these tests is across all the random effects to which the fixed treatment factor has been applied. This is the same regardless of the approach used in computing expected mean squares. The difference occurs for random factors and mixed interactions. When not constrained to the sum-to-zero assumption, expected means squares, and thus the F tests determined by them, vary considerably from those where the assumption is made. In this latter case, expected mean squares for any random factor or mixed interaction include variance components for any mixed interaction that includes the factor or interaction of interest

<table>
<thead>
<tr>
<th>Term</th>
<th>MS</th>
<th>Expected MS sum-to-zero constraint</th>
<th>Expected MS independence assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y_t</td>
<td>M1</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>L_t</td>
<td>M2</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>YL_t</td>
<td>M3</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>B(ij)k</td>
<td>M4</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>T_t</td>
<td>M5</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>YLT_t</td>
<td>M6</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>LT_t</td>
<td>M7</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>YLT_t</td>
<td>M8</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
<tr>
<td>BT(ijk)</td>
<td>M9</td>
<td>$\sigma_i^2$</td>
<td>$\sigma_i^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance component</th>
<th>Quantity in independence model</th>
<th>Quantity under sum-to-zero constraint†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>$\sigma_Y^2$</td>
<td>$\frac{\sigma^2}{1 + \tau_Y^2}$</td>
</tr>
<tr>
<td>Location</td>
<td>$\sigma_L^2$</td>
<td>$\frac{\sigma^2}{1 + \tau_L^2}$</td>
</tr>
<tr>
<td>Year × location</td>
<td>$\sigma_{YL}^2$</td>
<td>$\frac{\sigma^2}{1 + \tau_{YL}^2}$</td>
</tr>
<tr>
<td>Block(location)</td>
<td>$\sigma_B^2$</td>
<td>$\tau_B^2 = \sigma_B^2 + \sigma_{B}^2$</td>
</tr>
<tr>
<td>Year × treatment</td>
<td>$\sigma_T^2$</td>
<td>$\tau_T^2 = \sigma_T^2$</td>
</tr>
<tr>
<td>Location × treatment</td>
<td>$\sigma_{LT}^2$</td>
<td>$\tau_{LT}^2 = \sigma_{LT}^2$</td>
</tr>
<tr>
<td>Year × location × treatment</td>
<td>$\sigma_{YLT}^2$</td>
<td>$\tau_{YLT}^2 = \sigma_{YLT}^2$</td>
</tr>
</tbody>
</table>

† t is the number of treatment levels.

| Table 1. Analyses of combined experiments under the sum-to-zero constraint and the independence assumption. In this case, treatments (T) are the only fixed factor. Year (Y), location (L), and block (B) and all interactions involving year, location or block are considered random. We use $\tau^2$ to indicate a variance component in a model with sum-to-zero constraints on the random effects and $\sigma^2$ to indicate a variance component in a model with independent random effects. Subscripts on each variance component indicate the random effect it represents; $\tau[T]$ is the noncentrality parameter for the fixed treatment effects. Coefficients for each term are t (number of treatment levels), b (number of blocks), l (number of locations), and y (number of years). |
|---|---|---|---|
| Term | MS | Expected MS sum-to-zero constraint | Expected MS independence assumption |
| Y_t  | M1 | $\tau_Y^2 = \sigma_Y^2 + \sigma_{Y}^2$ |
| L_t  | M2 | $\tau_L^2 = \sigma_L^2 + \sigma_{L}^2$ |
| YL_t | M3 | $\tau_{YL}^2 = \sigma_{YL}^2 + \sigma_{YL}^2$ |
| B(ij)k | M4 | $\tau_{B}^2 = \sigma_{B}^2 + \sigma_{B}^2$ |
| T_t  | M5 | $\tau_T^2 = \sigma_T^2$ |
| YLT_t | M6 | $\tau_{YLT}^2 = \sigma_{YLT}^2$ |
| LT_t | M7 | $\tau_{LT}^2 = \sigma_{LT}^2$ |
| YLT_t | M8 | $\tau_{YLT}^2 = \sigma_{YLT}^2$ |
| BT(ijk) | M9 | $\tau_{BT}^2 = \sigma_{BT}^2$ |

| Table 2. Relationships among variance components under the sum-to-zero constraint, $\tau^2$, and variance components under the independence assumption, $\sigma^2$. |
|---|---|---|
| Variance component | Quantity in independence model | Quantity under sum-to-zero constraint† |
| Year              | $\sigma_{Y}^2$                      | $\tau_{Y}^2 = \sigma_{Y}^2 + \tau_{Y}^2$ |
| Location          | $\sigma_{L}^2$                      | $\tau_{L}^2 = \sigma_{L}^2 + \tau_{L}^2$ |
| Year × location   | $\sigma_{YL}^2$                     | $\tau_{YL}^2 = \sigma_{YL}^2 + \tau_{YL}^2$ |
| Block(location)   | $\sigma_{B}^2$                      | $\tau_{B}^2 = \sigma_{B}^2 + \tau_{B}^2$ |
| Year × treatment  | $\sigma_{T}^2$                      | $\tau_{T}^2$ |
| Location × treatment | $\sigma_{LT}^2$                  | $\tau_{LT}^2 = \sigma_{LT}^2$ |
| Year × location × treatment | $\sigma_{YLT}^2$                | $\tau_{YLT}^2 = \sigma_{YLT}^2$ |

† t is the number of treatment levels.
locations. Thus, calculations are somewhat more complicated and involved than in the sum-to-zero approach. However, because year, location, and their interaction are all random effects, the focus of their analysis is on their variances, and many statisticians do not compute F tests for them. In many mixed model analyses, the variance components for random factors are computed directly using algorithms such as the restricted maximum likelihood (REML), minimum variance quadratic unbiased estimator (MINQUE), or some other iterative approach (Holland, 2006). Mean squares used for testing fixed effects are then constructed from these variance components. Again, the tests are approximate because the Satterthwaite approximation is used to estimate the degrees of freedom, but the approximations are generally considered to be quite good when all coefficients are positive (Welch, 1956) and less good when some coefficients are negative (Gaylor and Hopper, 1969).

An important consequence of the assumptions made about mixed interactions are potential differences in the variances calculated using the method of moments (McLean et al., 1991). That is, if the variances are estimated from the mean squares and expectations of their variance components, the two assumptions may yield different estimates of the variances. In many agronomic studies, especially those related to plant breeding, variance estimates are of critical importance in calculating other parameters (Holland, 2006). Another concern with using the method of moments to estimate variances is whether or not to include a correction for a finite population for fixed effects in mixed models (McLean et al., 1991). The correction is sometimes applied to variances of mixed interactions to recognize that the sample population is constrained by the fixed effect. The correction factor is not an issue for determining appropriate F tests but can affect the value of a variance estimated using the method of moments.

### ANALYSIS OF COMBINED EXPERIMENTS

Analyses for experiments combined across years and locations are presented in Table 3. This is similar to Table 3 in McIntosh (1983), although the F tests are specified based on the alternative assumption about mixed interactions, i.e., the fixed effects do not sum to zero in a mixed interaction. Analyses are presented for experiments where year is the only random factor, year and location are random, and all factors are random. Other combinations are possible but unlikely to be encountered in agronomic experiments. These F tests are the ones computed by the GLM, MIXED, and GLIMMIX procedures of SAS (SAS Institute). Exactly how these tests are invoked and interpreted by SAS varies slightly among the three procedures, but the assumption with respect to mixed interactions and the resulting tests are identical (see the appendix and compare the syntax for a random statement of each procedure). For analyses of data from designed experiments, using variance components estimated from the expected mean squares (what SAS PROC MIXED calls method = type3) provides better control of the Type I error rate for tests of fixed effects than does REML estimation (Stroup and Littell, 2002; summarized in Littell et al., 2006, p. 150). For the MIXED procedure, the codes for analyzing the three analyses described in Table 3 are:

For Treatment (T) and location (L) fixed, year (Y) random:

```sas
proc mixed method = type3;
class Y L T;
model Yield = T L T*L;
random Y*Y L B(Y*L) T*Y T*L T*Y*L;
run;
```

For Treatment fixed, year and location random:

```sas
proc mixed method = type3;
class Y L B T;
model Yield = T;
random Y L Y*Y L B(Y*L) T*Y T*L T*Y*L;
run;
```

For Treatment, year, and location random:

```sas
proc mixed method = type3;
class Y L B T;
model Yield = ;
random Y L Y*Y L B(Y*L) T*Y T*L T*Y*L;
run;
```

Specifying the method as type3 in PROC MIXED causes the program to compute and display the expected mean squares. Fixed factors are included in the model statement and random factors are specified in the random statement.

To illustrate the three approaches, an example using data from an oat (Avena sativa L.) cultivar trial conducted for 2 yr at
three locations in Iowa is presented. The data set was used only
to obtain expectations for the three cases (Table 3). The trial
included 10 cultivars as treatments, the experiment was blocked
trials in a randomized complete block design for each year
and location, and there were no missing observations.

The expected mean squares generated by SAS for the three cases
are presented in Table 4. Note that the tests specified when year
and location are random and when all factors are random are the
same. This is because, with only one fixed factor (cultivars in this
case), there are no fixed interactions in the model.

Codes for analyzing combined experiments using the
MIXED, GLM, and GLIMMIX procedures of SAS are
included in the appendix. Each procedure uses a different
approach to the analysis but results in very similar tests in the
case of balanced data sets. The default method for calculating
variances in the MIXED and GLIMMIX procedures is REML.
The GLM procedure uses Type III sums of squares and uses
these to compute mean squares. In the case of unbalanced data
sets or situations where distributions are not normal, variances
are heterogeneous, or errors are correlated, use of the MIXED
and GLIMMIX procedures is necessary depending on the
assumption violated. The MIXED procedure is capable of
analyzing data where treatment variances are heterogeneous and
where errors are correlated. Analysis of data with non-normal
distributions requires the use of GLIMMIX or another program
capable of accommodating non-normal data.

Analyzing combined experiments using the sum-to-zero
constraint for the expected mean squares is somewhat difficult
to accomplish in SAS. The TEST statement in GLM can be
used to construct tests based on mean squares for effects where
direct tests are available. It will not calculate $F$ tests based
on linear combinations of mean squares. These must be done
outside the procedure either by hand or by using the SAS
programming language. No analog to the TEST statement of
GLM is available in the MIXED and GLIMMIX procedures.
Mean squares based on the sum-to-zero constraint could be
assembled from the variances calculated by these procedures,
but the process is time consuming and rarely done. The bottom
line is that when using SAS software to analyze combined
experiments, the alternative assumption for expected mean
squares is the most practical approach.

**ALTERNATIVE RULES FOR
EXPECTED MEAN SQUARES**

Snedecor and Cochran (1989) provided a rule-based method
for assembling expected mean squares. Accordingly, the
expected mean square for each model term contains the residual
variance ($\sigma_{BT}^2$), a component for the term itself, and one for
each interaction where all other factors involved are random.
Coefficients for each component are the product of the number
of levels of factors not included in the model term. The error
variance always has a coefficient of one. These rules conform to
the sum-to-zero constraint and produce expected mean squares
consistent with those of McIntosh (1983).

A slightly modified approach can be used to assemble expected
mean squares for the alternative assumption. The rules are
identical, with the exception that all interactions involving
the model term are included in the expectations if any of the
terms involved are random, i.e., a variance component is
added for all mixed interactions containing the model term. A
quick evaluation of the expected mean squares given by SAS
in the example shown above (Table 4) demonstrates that this
modification provides the desired result.

**ALTERNATIVE $F$ TEST**

Snedecor and Cochran (1989) made a point that is
underappreciated today: there are two different $F$ tests for the
main effect of the treatment factor when there are two crossed
random effects, year and location. The tests given in Table
3 use the $MS$ for treatments as the numerator and construct
a composite denominator as a linear combination of mean
squares. That denominator requires a negative coefficient on
the year by treatment interaction (Table 3). The Cochran–Satterthwaite approximation for degrees of freedom
was derived under the assumption that all coefficients in the
linear combination of mean squares are positive (Satterthwaite,
1946). If one or more of the coefficients is negative, the estimated
value of the linear combination may be negative, which is an
inadmissible value for a chi-square random variable. Snedecor
and Cochran (1989) suggested an alternative test that avoids
the negative coefficient but requires a composite numerator and
composite denominator. They suggested the $F$ statistic given by

$$F_A = \frac{MS_T + MS_{ALT}}{MS_{VT} + MS_{LT}}$$

[2]

to test the main effect of treatments. They also implied that
such a test might be less powerful than the one used by SAS and other software that uses $MS_T$ as the numerator and calculates the
appropriate composite denominator $MS_T$. We conducted a small
simulation study to evaluate the performance of both versions of
the test of the main effect of treatment. We simulated data sets
with and without a main effect of treatment and with different
levels of the year by location by treatment variance component.
In one simulation scenario, we changed the variance components
for the year by treatment and year by location interactions. Given
a simulated data set, we calculated the $F$ statistic given in Table 3
and used by SAS and the $F$ statistic given by Eq. [2], used the
Satterthwaite approximation to calculate the denominator df
(and numerator df for Eq. [2]), calculated the $p$ values for the two
tests, and recorded whether or not each was <0.05. We repeated
this 2000 times. When there is no true difference between
the treatments, this provides an estimate of the empirical Type I error
for a nominal 5% test. When there is a true difference between
treatments, this provides an estimate of the power of the test.

The $F$ statistic given in Table 3 and used by SAS rejects the
null hypothesis approximately 5% of the time in scenarios with
no true difference between treatments (Table 5). In contrast,
the test given by Eq. [2] is conservative (empirical rejection rate <5% for a nominal 5% test) when there is a non-zero year
by location by treatment interaction (Table 5). The degree of
conservativeness appears to be associated with the magnitude
of the random year by location by treatment interaction. Larger
values of the three-way interaction variance component are
associated with a smaller empirical rejection rate under the null hypothesis (Table 5).
Table 4. Expected mean squares obtained from the SAS MIXED procedure for variations on the analysis of a combined experiment in which 10 oat cultivars (T) were evaluated in six environments (2 yr [Y] and three locations [L]). Q(T) indicates a quadratic form involving the terms inside the parentheses. The experimental design in each year/location was a randomized complete block (B) design.

<table>
<thead>
<tr>
<th>Source</th>
<th>Expected mean square</th>
<th>Error term</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment (T) and location (L) fixed, year (Y) random</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 90 Var(Y)</td>
<td>MS(Y × L) + MS(Y × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>L</td>
<td>Var(residual) + 3 Var(Y × L × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + Q(1L × T)</td>
<td>MS(Y × L)</td>
</tr>
<tr>
<td>Y × L</td>
<td>Var(residual) + 3 Var(Y × L × T) + 10 Var(B[Y × L]) + 30 Var(Y × L)</td>
<td>MS[10(B[Y × L])] + MS(Y × L × T) − MS(residual)</td>
</tr>
<tr>
<td>B(Y × L)</td>
<td>Var(residual) + 10 Var(B[Y × L])</td>
<td>MS(residual)</td>
</tr>
<tr>
<td>T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T) + Q(T)</td>
<td>MS(Y × T)</td>
</tr>
<tr>
<td>Y × T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T)</td>
<td>MS(Y × L × T)</td>
</tr>
<tr>
<td>L × T</td>
<td>Var(residual) + 3 Var(Y × L × T) + Q(L × T)</td>
<td>MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × L × T</td>
<td>Var(residual) + 3 Var(Y × L × T)</td>
<td>MS(residual)</td>
</tr>
<tr>
<td>Residual</td>
<td>Var(residual)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment fixed, year and location random</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 90 Var(Y)</td>
<td>MS(Y × L) + MS(Y × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>L</td>
<td>Var(residual) + 3 Var(Y × L × T) + 6 Var(L × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 60 Var(L)</td>
<td>MS(Y × L) + MS(L × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × L</td>
<td>Var(residual) + 3 Var(Y × L × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 60 Var(L)</td>
<td>MS[10(B[Y × L])] + MS(Y × L × T) − MS(residual)</td>
</tr>
<tr>
<td>B(Y × L)</td>
<td>Var(residual) + 10 Var(B[Y × L])</td>
<td>MS(residual)</td>
</tr>
<tr>
<td>T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 6 Var(L × T) + 9 Var(Y × T) + Q(T)</td>
<td>MS(Y × T) + MS(L × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T)</td>
<td>MS(Y × L × T)</td>
</tr>
<tr>
<td>L × T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 6 Var(L × T)</td>
<td>MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × L × T</td>
<td>Var(residual) + 3 Var(Y × L × T)</td>
<td>MS(residual)</td>
</tr>
<tr>
<td>Residual</td>
<td>Var(residual)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment, year and location random</td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 90 Var(Y)</td>
<td>MS(Y × L) + MS(Y × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>L</td>
<td>Var(residual) + 3 Var(Y × L × T) + 6 Var(L × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 60 Var(L)</td>
<td>MS(Y × L) + MS(L × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × L</td>
<td>Var(residual) + 3 Var(Y × L × T) + 10 Var(B[Y × L]) + 30 Var(Y × L) + 60 Var(L)</td>
<td>MS[10(B[Y × L])] + MS(Y × L × T) − MS(residual)</td>
</tr>
<tr>
<td>B(Y × L)</td>
<td>Var(residual) + 10 Var(B[Y × L])</td>
<td>MS(residual)</td>
</tr>
<tr>
<td>T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 6 Var(L × T) + 9 Var(Y × T) + 18 Var(T)</td>
<td>MS(Y × T) + MS(L × T) − MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 9 Var(Y × T)</td>
<td>MS(Y × L × T)</td>
</tr>
<tr>
<td>L × T</td>
<td>Var(residual) + 3 Var(Y × L × T) + 6 Var(L × T)</td>
<td>MS(Y × L × T)</td>
</tr>
<tr>
<td>Y × L × T</td>
<td>Var(residual) + 3 Var(Y × L × T)</td>
<td>MS(residual)</td>
</tr>
<tr>
<td>Residual</td>
<td>Var(residual)</td>
<td></td>
</tr>
</tbody>
</table>
When there is a main effect of treatment (and you want the \( F \) test to reject the null hypothesis of no difference), the test given by Eq. [2] is more powerful, i.e., rejects the null hypothesis more frequently than the test used by SAS (Table 5). This is despite the test given by Eq. [2] having a smaller Type I error rate. We believe that the test of Eq. [2] has the larger power because it has a larger Satterthwaite approximate error df (Table 5). The negative denominator coefficient in the test given in Table 3 and used by SAS leads to a smaller denominator df (Dixon, 2014).

In a small proportion of data sets, the test used by SAS led to a negative denominator coefficient in the test given in Table 3 and has a larger Satterthwaite approximate error df (Table 5). The test given by Eq. [2] having a smaller Type I error rate. We have seen proposed analyses of a combined experiment with a 2 \( \times \) 2 by 2 treatment design repeated in two locations and 2 yr that had a total of 22 interactions (e.g., year by \( A \) or location by \( B \)), each with one degree of freedom. This is the ultimate in splitting and leads to tests with very low power because all denominator degrees of freedom are 1.

A final variance issue for combined experiments is heterogeneity or homogeneity of residual variances across the set of experiments. In a typical agricultural experiment at a single site in a single year, treatments are presumed to affect the mean response but not the variance, so using a single pooled residual variance computed from all treatments and replications is standard. Plot-to-plot variation may differ between locations and possibly between years. If experiment-specific residual variances are in fact similar across years and locations, they should be pooled to give better estimates of that common residual variance. In a combined experiment, however, this should not be taken for granted. If experiment-specific residual variances are not similar, they should not be pooled. Modern software (e.g., SAS MIXED) allows you to specify a model with different residual variances for different groups of observations. A classic example demonstrating the principles of pooling is the experiment on early lentil (\textit{Lens culinaris} Medik.) in Syria described by Petersen (1994, p. 207–215). The treatment structure for this study included all combinations of two levels of fertilizer, two levels of weevil control, and two levels of weed control, repeated at eight locations. Peterson's model included eight location interactions, one for each component of the treatment structure. Each location interaction was estimated with 7 df. The location by fertilizer interactions were all small relative to the location by weevil and location by weed interactions (Petersen, 1994, p. 213). Peterson speculated that the eight locations had different

### Table 5. Empirical rejection rates of \( F \) tests at nominal \( p < 0.05 \) using the \( F \) statistic in Table 3 and using the \( F \) statistic in Eq. [2]. In all cases, \( \sigma_i^2 = 0.5 \), \( \sigma_{jl}^2 = 0.25 \), \( \sigma_{jy}^2 = 0.5 \), and \( \sigma_{jyk}^2 = 0.5 \), where \( Y \) indicates years, \( L \) indicates locations, \( B \) indicates blocks, and \( T \) indicates treatments. In all scenarios, there are three treatments. The magnitudes of treatment differences were adjusted to provide adequate power to differentiate among tests, i.e., treatment differences were increased when variances were larger.

<table>
<thead>
<tr>
<th>( \sigma_{y}^2 )</th>
<th>( \sigma_{l}^2 )</th>
<th>( \sigma_{r}^2 )</th>
<th>Treatment main effects</th>
<th>Test statistic</th>
<th>( P(\text{reject}) )</th>
<th>df for numerator</th>
<th>df for error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0, 0, 0</td>
<td>Table 3</td>
<td>0.040</td>
<td>2</td>
<td>5.7</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>0, 0, 0</td>
<td>Table 3</td>
<td>0.029</td>
<td>2</td>
<td>4.2</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>0, 0, 0</td>
<td>Table 3</td>
<td>0.066</td>
<td>2</td>
<td>3.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>0</td>
<td>0.5, 1, 1.5</td>
<td>Table 3</td>
<td>0.39</td>
<td>2</td>
<td>5.6</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>0.5, 1.5</td>
<td>Table 3</td>
<td>0.43</td>
<td>2</td>
<td>8.2</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
<td>2</td>
<td>1, 2, 3</td>
<td>Table 3</td>
<td>0.12</td>
<td>2</td>
<td>4.2</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1</td>
<td>1, 2, 3</td>
<td>Table 3</td>
<td>0.39</td>
<td>2</td>
<td>5.4</td>
</tr>
<tr>
<td>1.0</td>
<td>1.0</td>
<td>1</td>
<td>1, 2, 3</td>
<td>Table 3</td>
<td>0.35</td>
<td>2</td>
<td>5.4</td>
</tr>
</tbody>
</table>

### POOLING

One final issue in the analysis of combined experiments is deciding whether to pool the sources of variation and, if so, what terms to pool. For example, the analysis illustrated in Table 4, with 10 oat cultivars (treatment) evaluated over 2 yr and three locations, includes a treatment by year interaction, a treatment by location interaction, and a treatment by year by location interaction. Is it appropriate to estimate three different interactions, or should the analysis include only a single treatment by environment interaction, where environment is the combination of year and location? Or should it be subdivided even further? If the 10 cultivars represent two breeding groups and five maturity groups, the set of 10 treatments could be divided into a 2 \( \times \) 5 factorial combination of breeding group by maturity group. Should the various interactions be subdivided into breeding group by year, maturity group by year, etc., for a total of nine interaction terms?

In the usual analysis of a combined experiment, these terms are random and serve as the error terms for the matching treatment effects. Even this analysis with nine interactions may inappropriately pool error variances. If the difference between Maturity Group 1 and the other four maturity groups varied considerably across locations but differences among the other four maturity groups were relatively consistent, then the single degree of freedom contrast between Maturity Group 1 and the rest should be separated from the other 3 df among maturity groups. Such an analysis would include two different interactions, one for each group of contrasts. The decision to pool or not should preferably be made in the planning stage of a study, using the experimenter's understanding of the sources and magnitudes of variability. We have seen proposed analyses of a combined experiment with a 2 \( \times \) 2 \( \times \) 2 treatment design repeated in two locations and 2 yr that had a total of 22 interactions (e.g., year by \( A \) or location by \( B \)), each with one degree of freedom. This is the ultimate in splitting and leads to tests with very low power because all denominator degrees of freedom are 1.
levels of infestation by weevils and weeds. The response to the weevil treatment or the weed control treatment would be large at locations with a heavy infestation and small at locations with little infestation. In contrast, the data suggested that the response to fertilization was similar at all locations. Hence, it would be appropriate to keep separate the location by weevil and location by weed interactions, but pool the location by fertilizer interaction with all other location interactions. The result would be a 35 df location by other error term that is a more precise estimate than any of the 7 df estimates.

CONCLUSIONS

The assumptions made about interactions between fixed and random factors in the analysis of variance can result in very different interpretations. Assuming those interactions to sum to zero within each level of a fixed effect in the literature and historically has been the approach used in most agronomic studies. Many software programs, however, take the alternative approach and consider all interactions between fixed and random effects to be independent. Consequently, the tests made and in some cases the variances calculated based on these assumptions vary depending on which are used. Choosing which approach to follow remains the prerogative of the researcher, but it is important to understand that the more traditional tests are difficult to achieve when using many software programs.

APPENDIX

The SAS code for analyzing combined experiments using the GLM, MIXED, and GLIMMIX procedures accepts all default assumptions associated with the procedure and can and should be augmented in some situations. In general, the more a data set departs from the traditional assumptions, the more arguments will need to be included to perform an optimal analysis. We follow Stroup and Littell (2002) where possible. They recommend method = type3 rather than the default method = reml because method = type3 provides better control of Type I error rates for tests of fixed effects. The GLIMMIX procedure provides REML and other estimation methods, but not method = type3. The GLM procedure uses ordinary least squares to calculate the mean squares used in the analysis.
REFERENCES


