

# Statistical Analysis of Mixed Model Factorial Experiments with Missing Factor Combinations: The Case of Asynchronous Cyclic Drought Data

Dawn M. VanLeeuwen

*Agricultural Biometric Service, New Mexico State University, Las Cruces, NM 88003*

Rolston St. Hilaire<sup>1</sup> and Emad Y. Bsoul

*Department of Plant and Environmental Sciences, New Mexico State University, Las Cruces, NM 88003*

ADDITIONAL INDEX WORDS. ANOVA, covariance structure, incomplete data, repeated measures, saturated model, serial correlation, unbalanced data, unsaturated model

**ABSTRACT.** Statistical analysis of data from repeated measures experiments with missing factor combinations encounters multiple complications. Data from asynchronous cyclic drought experiments incorporate unequal numbers of drought cycles for different sources and provide an example of data both with repeated measures and missing factor combinations. Repeated measures data are problematic because typical analyses with PROC GLM do not allow the researcher to compare candidate covariance structures. In contrast, PROC MIXED allows comparison of covariance structures and several options for modeling serial correlation and variance heterogeneity. When there are missing factor combinations, the cross-classified model traditionally used for synchronized trials is inappropriate. For asynchronous data, some least squares means estimates for treatment and source main effects, and treatment by source interaction effects are inestimable. The objectives of this paper were to use an asynchronous drought cycle data set to 1) model an appropriate covariance structure using mixed models, and 2) compare the cross-classified fixed effects model to drought cycle nested within source models. We used a data set of midday water potential measurements taken during a cyclic drought study of 15 half-siblings of bigtooth maples (*Acer grandidentatum* Nutt.) indigenous to Arizona, New Mexico, Texas, and Utah. Data were analyzed using SAS PROC MIXED software. Information criteria lead to the selection of a model incorporating separate compound symmetric covariance structures for the two irrigation treatment groups. When using nested models in the fixed portion of the model, there are no missing factors because drought cycle is not treated as a crossed experimental factor. Nested models provided meaningful F tests and estimated all the least squares means, but the cross-classified model did not. Furthermore, the nested models adequately compared the treatment effect of sources subjected to asynchronous drought events. We conclude that researchers wishing to analyze data from asynchronous drought trials must consider using mixed models with nested fixed effects.

To compare and characterize plant responses to drought, researchers may choose to irrigate plants in cycles, and evaluate changes in water relations, and plant physiological, anatomical, and developmental traits linked to drought tolerance mechanisms (Balok and St. Hilaire, 2002; Feser et al., 2005). Plant taxa, irrigation treatment, and drought cycle are factors that are common to plant cyclic drought experiments. A species, cultivar or seed source could represent plant taxa. Irrigation treatments might consist of a control group that is well irrigated to maintain nonstressed plants and a drought-stressed group that is exposed to deficit irrigation. Different irrigation scheduling criteria are used to impose drought stress. For example, irrigation frequency might be based on plant gravimetric moisture loss due to evapotranspiration (Balok and St. Hilaire, 2002; St. Hilaire et al., 2003), applying water to wet a depth of soil (Tipton, 1994) or on a set interval (St. Hilaire et al., 2004).

In experiments where plant taxa (sources) are subjected to multiple drought events, scientists may apply drought treatments in two ways. First, all plants in each source could be treated equally over the same period if a drought cycle is terminated at a set period and data are taken from plants in both the control and drought treatments. All plants are rewatered and a new cycle starts. Second, drought status could be assessed separately for each source. Since

a drought cycle is terminated at a time particular to this source, each source will have a unique set of drought cycle durations. Also, data collected at the end of a drought cycle will be unique to this source. This produces an asynchronous data set.

Data from synchronous cyclic drought trials might be separately analyzed for each drought cycle (Marron et al., 2003) or all data might be analyzed with a cross-classified model (Feser et al., 2005). The cross-classified model is typically used with factorial treatment structures. Analogous approaches might be considered for asynchronous data. Data for each combination of source and drought cycle from asynchronous cyclic drought trials could be analyzed separately. Such an analysis only allows drought-stressed trees to be compared to control trees and does not allow drought cycles or sources to be compared. Alternatively, attempts to analyze all data using a cross-classified model with the asynchronous data set are problematic because some combinations of drought cycle and source are missing. Truncating later drought cycles for some sources to create a complete data structure that has the same number of drought cycles for all sources resolves problems arising from an incomplete data structure but introduces a new set of problems. This approach compares sources averaged across different amounts of time rather than different numbers of drought cycles. Furthermore, only part of the available data is analyzed; removing later drought cycles for some sources may limit a researcher's ability to determine if plants with shorter drought cycles might be physiologically distinct. Using all available data in a single analysis may be desirable. In addition to allowing

Received for publication 21 May 2005. Accepted for publication 22 Nov. 22 2005. Contribution of the N.M. Agricultural Experiment Station, N.M. State Univ., Las Cruces.

<sup>1</sup>Corresponding author: email rsthilai@nmsu.edu

more hypotheses to be tested, statistical approaches that use all available data may use the data more efficiently and so may be more powerful for some statistical comparisons.

Several difficulties are encountered when attempting to incorporate all data from asynchronous cyclic drought trials into a single analysis. Because multiple observations are taken from each tree, errors might be serially correlated. In addition, it is possible that the drought treatment induces changes to the mean response and greater variability in the data. Both lack of independence and nonconstant variance are violations of typical analysis of variance (ANOVA) model assumptions. Even if a researcher overlooks these violations, the intentionally incomplete data structure causes obvious problems in obtaining fixed effect estimates.

Both violations of ANOVA model assumptions can be addressed using linear mixed model (mixed model) analyses that can be implemented with SAS PROC MIXED software (SAS Institute, Cary, N.C.). Garrett et al. (2004) included mixed models in a discussion of statistical tools that have the potential to improve data analyses in plant pathology. In a paper aimed at encouraging more widespread use of mixed models in agricultural research, Piepho et al. (2003) provide an introduction to "ANOVA-type" mixed models such as those appropriate for analyzing split-plot experiments. The traditional analysis uses the whole-plot error term to model a constant covariance, (and therefore correlation) between errors taken from the same whole plot. Littell et al. (1996) compared PROC GLM to PROC MIXED analyses of split-plot data. For balanced data, PROC GLM and PROC MIXED typically produce the same results. Unlike PROC GLM, PROC MIXED can produce a correct analysis even when data are unbalanced. In a comparison of traditional ANOVA approaches to the likelihood-based methods implemented in PROC MIXED, Littell (2002) concluded that PROC MIXED resolves many of the computational problems associated with applying traditional ANOVA methods and software to mixed models.

Piepho et al. (2004) go beyond the traditional "ANOVA-type" mixed models and discuss applying mixed models to analyzing repeated measurements data. Feser et al. (2005) used mixed model methodology to analyze data from synchronized drought cycles. They incorporated fixed effects for all three main effects and their interactions and used SAS PROC MIXED software to select and model an appropriate repeated measures covariance structure. Padgett-Johnson et al. (2003), however, used a split-in-time (i.e., a split-plot) analysis when analyzing similar data. Because this type of data may exhibit decreasing correlation through time, the split-plot analysis, which models a constant correlation between errors from the same experimental unit may or may not be appropriate. The error structure implied by the split-plot analysis is a special case of the compound symmetric covariance structure. While the compound symmetric covariance structure is one of the most commonly used structures for this type of data (Wolfinger, 1996), other covariance structures should be considered.

A major advantage of mixed model methodology is the capability to appropriately model covariance structures for longitudinal or repeated measures data (Verbeke and Molenberghs, 2000). Analyses based on models with incorrectly specified covariance structure do not provide correct estimates of SES (Diggle et al., 1994; Littell et al., 1998; Schabenberger and Pierce, 2002). Mixed model methodology allows valid inferences based on correct SE estimates (Garrett et al., 2004). Several different parsimonious covariance models including some that model decreasing correlation through time or space are available in SAS PROC MIXED software. Mixed model methodology can also accommodate

modeling heterogeneous covariance structures and multiple random effects. In addition, as long as missing data points occur at random and are not related to treatments, mixed model analyses can accommodate missing data (Verbeke and Molenberghs, 2000). Wolfinger (1996) indicates that mixed models also can accommodate data that are intentionally incomplete.

When there are missing cells in the cross-classified fixed portion of a mixed model, some of the least squares means are not estimable. Because statistical software such as SAS/STAT PROC MIXED will not compute least squares means (LSMs) that are inestimable (SAS version 9.1), obtaining desired estimates and comparisons may become difficult. Additionally, the interpretation of the usual F tests produced using standard software is no longer straightforward (Pendleton et al., 1986).

Analyzing data from asynchronous cyclic drought trials using models for the fixed portion of the model that treat drought cycle as nested within source provides an alternative to the analyses based on cross-classified models. Nested models treat drought cycles as unique events within each source and more readily accommodate different numbers of drought cycles within sources. Furthermore, nested models accommodate comparing sources using all available data. Nested models eliminate some of the difficulties associated with using the cross-classified model to analyze data from asynchronous drought trials.

The objectives of this paper were to apply mixed model methodology to obtain an appropriate model for the repeated measures covariance structure and, once a covariance structure was selected, compare statistical analyses of asynchronous drought cycle data based on the cross-classified fixed effects model to analyses based on the drought cycle nested within source model(s). Our approach was to compare statistical outputs from the cross-classified model with the nested and the reduced nested model. We also discuss contrasts and comparisons that are meaningful to drought cycle nested within source models as well as appropriate interpretations of those comparisons.

## Materials and Methods

The data set used for this study was selected from a study of drought responses of 15 half-sibling sources of bigtooth maples indigenous to Arizona, New Mexico, Texas, and Utah (Table 1). Plants were grown in a greenhouse in New Mexico and irrigated every 2 d (well-irrigated controls) or exposed to drought and irrigated when pots lost 56% of their moisture due to evapotranspiration. Preliminary experiments indicated that a moisture loss of 56% was sufficient to cause severe plant stress but not plant mortality. At the end of every drought cycle, physiological measurements such as midday water potential were taken. Midday water potential was taken from fully expanded leaves with a pressure chamber (PMS Instruments, Corvallis, Ore.). Drought treatment was terminated after 117 d. The number and duration of drought cycles among the plant sources varied because of differences in evapotranspiration (Table 1).

The experimental design began with five complete blocks with a treated and a control tree from each source within each block. Treatment (TRT), source (SOR), and drought cycle (DC) were fixed model factors, and block (BLK) was a random factor. Garrett et al. (2004) discuss fixed vs. random factors; decisions to treat factors as fixed or random were made in a manner consistent with their discussion. For example, because the SOR levels included in the study represented all levels of interest to the researchers, SOR was a fixed factor. Replications were laid out in blocks. Because

Table 1. Seed source, tree designation, latitude, and longitude of 15 half-siblings of bigtooth maples used to test drought responses. Included are the number of drought cycles completed and the mean length of each drought cycle that plants completed at the end of 117 d of drought treatment.

Seed source (state)	Tree designation	Lat. (°N)	Long. (°W)	Drought cycles completed	
				(no.)	[mean length (d)]
Chiricahua Mountains (Ariz.)	AZ	31.592	109.356	5	22
Dripping Springs State Park (N.M.)	DS	32.389	106.813	5	18
Guadalupe Mountains (Texas)	GM1	31.90	104.867	5	17
	GM2	31.90	104.867	5	21
	GM3	31.90	104.867	7	16
	GM4	31.90	104.867	6	17
Lost Maples State Park (Texas)	LMP1	29.49	99.35	6	16
	LMP2	29.49	99.35	7	17
	LMP3	29.49	99.35	5	23
	LMP4	29.49	99.35	5	21
	LMP5	29.49	99.35	5	19
Logan Canyon (Utah)	UN1	41.46	110.490	5	23
	UN2	41.46	110.490	5	20
	UW1	41.46	110.490	5	23
	UW2	41.46	110.490	6	19

the experiment was conducted in a greenhouse, block effects were expected to be small and interactions between the blocking factor and the repeated factor (DC) were not expected. The response variable was midday water potential. Water potential indicates the tension in the plant as it draws water from the growing substrate to satisfy transpiration. Because water potential is measured in negative megapascals, we analyzed magnitudes. Water potential of the plant is usually highest (most negative) around midday. We intended to measure midday water potential for each tree within a source at the end of each drought cycle, however, there was some very slight imbalance in the data. Seven measurements from AZ (tree originating from the Chiricahua Mountains in Arizona), one from GM2 (the tree designated as the second seed source originating from the Guadalupe Mountains of Texas), and one from UW1 (tree designated as the first seed source from a west-facing slope in the Logan Canyon of Utah) were missing because fully expanded leaves were not available at the time of measurement. The total data set consisted of 810 measurements.

We compared the cross-classified model with two other models where drought cycles were nested within source. All analyses were executed using SAS PROC MIXED software (SAS version 9.1) and incorporated appropriate modeling of random effects and an appropriate model for the correlation structure among the residuals within a tree. Our analysis revealed that a model fitting separate compound symmetric (CS) covariance structures to the two treatment groups adequately fit our data.

## Results and Discussion

**THE MODELS.** The SAS code given in Table 2 was used to analyze the data using either a cross-classified model, nested or reduced-nested model for the fixed portion of the model, depending on which of the three model statements is used. The cross-classified model incorporated fixed effects for all main effects and their interactions. The nested model treated DC as nested within TRT. Consequently, this model did not incorporate a drought cycle main effect. To accommodate modeling drought cycle as nested within SOR, DC and the SOR × DC interaction that appear in the cross-classified model were replaced in the nested model with the nested drought cycle effect, DC(SOR), and the TRT\*DC and SOR\*TRT\*DC effects that appear in the

cross-classified model are replaced with the TRT\*DC(SOR) effect. The removal of the DC(SOR) effect from the nested model created the reduced-nested model. This model may be simpler to use in some cases where the researcher chooses to specify contrasts and obtain estimates not automatically provided with an LSMEANS statement.

All models included a block random effect and specified a compound symmetric (CS) covariance structure for observations taken from the same tree. The block random effect was specified using a RANDOM statement while the CS structure was specified using a REPEATED statement. Within the REPEATED statement, DC was defined to be the repeated factor, the SUBJECT=BLK\*SOR\*TRT option uniquely identified each subject (i.e., each tree), and the TYPE=CS option specified the CS covariance structure. The option GROUP=TRT fitted separate covariance parameters for the drought-stress and well-watered plants. Sample variances for the drought-stress plants were larger than those of the well-watered group. The GROUP option incorporated variance heterogeneity into the model. The option DDFM=KR (denominator df = Kenward–Roger) specified that the Kenward–Roger adjusted df and bias-adjusted SEs be used. Statistical simulations show that the Kenward–Roger procedure adequately adjusts df and SE estimates (Schabenberger and Pierce, 2002).

**CHOOSING A MODEL FOR THE COVARIANCE STRUCTURE.** Because repeated measures data such as these data might be expected to be serially correlated with decreasing correlations between observations farther apart in time, several different covariance structures were considered before choosing the CS structure. Competing covariance structures were compared using both the Akaike criterion (AIC) and the Schwarz's Bayesian criterion (SBC). Both are commonly used information criteria (IC). Wolfinger (1996) noted that the most parsimonious covariance structure model that provides a reasonable fit to the data gives the most efficient inferences about the mean model. Schabenberger and Pierce (2002) cautioned modelers to limit the extent to which they model their data. They recommended selecting the most parsimonious model if several covariance structure models with similar IC values emerge. While likelihood ratio tests could be used to compare nested models, interest usually lies in comparing some models that are not nested. Information criteria can be used to compare models that are not nested as long as the fixed

Table 2. SAS code used to analyze asynchronous cyclic drought data using either a cross-classified model, nested or reduced-nested model, depending on which of the three model statements is used. In the code below, “SOR” refers to source of origin of half-sibling plants of bigtooth maples (Table 1), “BLK” to block, “TRT” to irrigation treatment, and “DC” to drought cycles. The output for the cross-classified model and the nested model is given in Tables 3 and 4, respectively. The output for the reduced-nested model is similar to that of the nested model.

```

PROC MIXED DATA = DAT1;
CLASS SOR BLK TRT DC;
MODEL MIDDAY = SOR TRT DC SOR*TRT SOR*DC TRT*DC
SOR*TRT*DC/DDFM=KR; /*cross-classified model*/
MODEL MIDDAY = SOR TRT SOR*TRT DC(SOR) TRT*DC(SOR)/
DDFM=KR; /*nested model*/
MODEL MIDDAY = SOR TRT SOR*TRT TRT*DC(SOR)/DDFM=KR;
/*reduced-nested model*/
RANDOM BLK;
REPEATED DC/SUBJECT=BLK*SOR*TRT TYPE=CS GROUP=TRT;
RUN;

```

effects portion of the model remains constant (Littell et al., 2000; Verbeke and Molenberghs, 2000). When computing IC to select a covariance model, the reduced nested model was used in the fixed portion. As discussed below, all three fixed models are saturated and consequently any of the three would lead to selecting the same covariance model. The reduced nested model requires less computational memory.

Candidate covariance structures should be chosen on the basis of the data structure and the researcher’s understanding of the data being modeled. Descriptive statistics and residual plots may also be useful in identifying candidate structures. Wolfinger (1996) notes that the three most commonly used covariance structures in repeated measures situations are the CS, first-order autoregressive [AR(1)], and unstructured (UN). For the current data, separate covariance structures were fit for control and treated trees. Preliminary descriptive analyses suggested the need for this; both graphical data summaries and sample standard deviations indicated that measurements from drought-treated trees were more variable than those from control trees.

Models considered included one incorporating a CS structure (AIC=1129.8; SBC=1128.7), separate CS structures for the two treatment groups (AIC=1059.4; SBC=1057.5), separate AR(1) structures (AIC=1073.6; SBC=1085.7), and separate AR(1) structures for each combination of drought cycle group and treatment (AIC=1075.2; SBC=1070.1). All models except the first AR(1) model included a random block effect; the block effect was removed from this model because of convergence problems. SAS version 9.1 implements a “smaller is better” version of the IC and, of these models, the model fitting the separate CS structures for the two treatment groups had the smallest AIC (1059.4) as well as the smallest SBC (1057.5). An additional model corresponding to independent errors within a tree was fit (AIC=1153.7; SBC=1152.9). All candidate covariance structures provided an improvement over the independence model.

The CS structure models constant correlation through time and is a simple structure that usually is equivalent to the basic “split-in-time” structure. For these data, the AR(1) structure or a similar structure which models a decreasing correlation through time should be considered among the candidate models. However, neither the AIC nor the SBC suggest that the AR(1) provides a better fit than the CS structure. The AR(1) structure assumes measurements are equally spaced through time. Because a source that underwent seven drought cycles had measurements taken

more frequently than one that had five drought cycles, there are conceptual problems with the first AR(1) model. The second AR(1) model resolves this problem. To create groups with approximately equally-spaced measurements, sources were first grouped into those that underwent five, six, and seven drought cycles, then separate AR(1) structures were fit to each combination of drought cycle group and treatment. This model did not provide better fit than the separate groups CS structure.

The spatial covariance structure SP(POW) provides a generalization of the AR(1) covariance structure that allows for unequally spaced measurements. An attempt to fit this structure using days apart to measure the distance between measurements failed to converge. Piepho et al. (2004) indicate that when this occurs, a simpler correlation model is needed. While model misspecification could cause convergence failure, numerical problems with the data may cause convergence difficulties. SAS documentation or a statistician should be consulted to find the approaches needed to circumvent the numerical problems if the researcher has a particular interest in fitting a model that initially fails to converge.

For repeated measures data, it is often desirable to include an unstructured (UN) covariance model (Littell et al., 1998) in the set of candidate models. For the current data a similar conceptual problem arises as with the AR(1) model. If the variances at each drought cycle are distinct and the correlations between each pair of drought cycles is also distinct, it seems unlikely that the UN covariance structure for trees subjected to five drought cycles would be the same as the covariance structure for trees subjected to seven drought cycles. Consequently, separate UN structures must be fit to each combination of drought cycle group and treatment. The computing memory required to estimate this many covariance parameters is excessive. Consequently, we did not pursue the UN covariance structure. Some additional covariance structures were considered including some allowing heterogeneous variances for DCs (Wolfinger, 1996). The reader is referred to SAS documentation for a complete description of available covariance structures. Several of the references including Littell et al. (2000) and Verbeke and Molenberghs (2000) provide more detail on modeling covariance structure in longitudinal or repeated measures data.

While the parameters in the covariance portion of the model are not typically of primary interest, examining the covariance parameter estimates may provide insight into how the response variable behaves. For the current data, statistical analysis revealed that the block variance component was very small (0.000949) relative to the variance estimates. This suggested that location within the greenhouse had little effect. While the control group residual variance parameter estimate was 0.1263 and the control covariance estimate was -0.00153, the corresponding estimates were substantially larger for the drought treated trees (0.2615 and 0.08179). It is likely that the covariance between control tree measurements is negligible because trees within a source were initially fairly homogeneous. If trees initially had been more heterogeneous, then the covariance parameter estimate might have been larger and positive. In contrast, drought tree measurements were both more variable and more highly correlated through time. Subjecting trees to drought might be expected to induce greater variability since some individual trees respond more severely to the same level of moisture stress than others. For that reason, the higher correlation among drought tree measurements through time is not surprising. Subjecting the trees to drought stress induces a sort of heterogeneity as the individuals that are more susceptible

to that stress tend to have a higher (i.e., more negative) midday water potential than those that are less vulnerable.

**LEAST SQUARES MEANS.** An LSMEANS statement appended to the SAS code in Table 2 gives least squares means (LSMs) based on the estimated generalized least squares estimates (Littell, 2002). For example, the statement “LSMEANS SOR TRT SOR\*TRT SOR\*TRT\*DC;” appended to the SAS code for the cross-classified model, obtains the LSMs corresponding to a subset of the effects listed in the cross-classified model.

Least squares means are obtained by averaging across the levels of other model factors. For the cross-classified model, treatment, most source, and most source  $\times$  treatment LSMs were not estimable (Table 3). This result is expected because, under the crossed-classified model, there are missing cells (due to missing combinations of SOR and DC) (Yandell, 1997). Missing treatment combinations (i.e., missing cells) create difficulties beyond those encountered with unbalanced data where all treatment combinations are represented in the data but in unequal numbers. If all sources had been subjected to the same number of drought cycles, then the few missing observations for AZ, GM2, and UW1 that occurred would not have created this problem because these missing observations did not lead to any missing combinations of source, treatment and drought cycle.

The nested model estimated all LSMs corresponding to model effects (Table 4). For effects included in the reduced-nested model, all LSMs are equal to those computed using the nested model. Under the nested models, there are no missing cells. Modeling drought cycle as nested within source implies that drought cycles are unique to each source and that there may be different numbers of drought cycles within different sources. Under the nested model or the reduced-nested model, the LSM for a given source and treatment combination is the average of the source by treatment by drought cycle LSMs corresponding to that source and treatment combination. Similarly, treatment LSMs are formed to give equal weighting to each of the sources by averaging source by treatment LSMs across the 15 sources.

While the appropriate estimates and comparisons can be obtained with the cross-classified model, individual ESTIMATE statements must be written for each one that is not automatically provided by the LSMEANS statement. Much more statistical programming is required to obtain them than with the nested and reduced-nested models. For this data set, a single ESTIMATE statement can easily require specifying more than a hundred coefficients.

**F TESTS.** Once a model is selected for the covariance structure, inference about fixed effects proceeds almost as if one were using a fixed effects model and PROC GLM (Littell et al., 2000). Piepho et al. (2003) briefly discuss the differences in traditional tools and mixed model methodology but note that Types I, II, and III hypotheses are the same in both cases. Using a cross-classified model, the F tests generated from the analysis of asynchronous drought data are problematic. Type III hypotheses are the default for SAS PROC MIXED software. If the data were merely unbalanced (with all treatment factor combinations represented in the data) then these default tests would be appropriate. However, when there are missing factor combinations, the situation is more complicated. Type I and Type II tests correspond to hypotheses that depend on the parameters of interest and the sample sizes (Pendleton et al., 1986). If unbalanced data are used with the cross-classified model, then Type II tests are misleading (Yandell, 1997). Type III tests are generally recommended for unbalanced data that does not contain missing cells (Pendleton et al., 1986; Yandell, 1997). Type IV hypotheses may depend on arbitrary aspects of the treatment level ordering (Milliken and Johnson, 1992; Pendleton et al., 1986) and are not available in PROC MIXED. Consequently, when there are missing cells, none of the four types of hypotheses can be uniformly recommended and contrasts of interests must be crafted carefully (Pendleton et al., 1986; Urquhart and Weeks, 1978). Under the cross-classified model, F tests automatically produced by SAS software should not be used.

Under the nested model, there are no missing factor combinations and drought cycle is not treated as a “factor” in the usual sense. Although our data were slightly unbalanced, under the nested model there were no missing cells, and concerns about the validity of the F tests are resolved. When there are concerns about the hypotheses being tested, the E3 option can be added to the model statement in the SAS code to show the coefficients that are used to form the F test hypotheses. Type III tests (the SAS PROC MIXED software default) should be used (Yandell, 1997).

The main advantage to using the reduced-nested model is that contrast construction is simplified because there is no DC(SOR) effect. The LSMs and SEs are equal to those obtained using the nested model (Table 4). Additionally, Type III F tests of SOR, TRT, and SOR\*TRT are the same as in the nested model ( $P < 0.0001$  for all three tests). The test of TRT\*DC(SOR) differs. In the nested model, the test of DC(SOR) suggests that drought cycles, averaged across control and treatment, vary within sources ( $P <$

Table 3. Representative output from the LSMEANS and ESTIMATE statement using the statistical analysis system (SAS) code for the cross-classified model. Bigtooth maple sources were subjected to 117 d of cyclic drought.

Effect	Source (SOR)	Treatment (TRT)	Drought cycle (DC)	Estimate	Error	df	<i>t</i>	<i>P</i> > <i>t</i>
SOR	AZ <sup>‡</sup>			Non-est <sup>‡</sup>	.	.	.	.
SOR	DS			Non-est	.	.	.	.
SOR	GM3			2.187	0.8350	62.8	26.2	<0.0001
TRT		Control		Non-est	.	.	.	.
TRT		Drought		Non-est	.	.	.	.
SOR $\times$ TRT	AZ	Control		Non-est	.	.	.	.
SOR $\times$ TRT	AZ	Drought		Non-est	.	.	.	.
SOR $\times$ TRT	GM3	Control		1.957	0.0591	32.2	33.1	<0.0001
SOR $\times$ TRT	GM3	Drought		2.417	0.1550	47.5	15.6	<0.0001
SOR $\times$ TRT $\times$ DC	AZ	Control	1	2.580	0.1586	318	16.3	<0.0001
SOR $\times$ TRT $\times$ DC	AZ	Control	2	1.640	0.1586	318	10.3	<0.0001

<sup>‡</sup>Seed source: AZ = Chiricahua Mountains, Ariz.; DS = Dripping Springs State Park, N.M.; GM3 = the third tree from the Guadalupe Mountains, Texas.

<sup>‡</sup>Non-est = nonestimable.

Table 4. Representative output from the nested model that treated drought cycle as nested within treatment. SAS code is given in Table 2. Bigtooth maple sources were subjected to 117 cyclic drought treatments.

Effect	Source (SOR)	Treatment (TRT)	Drought cycle (DC)	Estimate	Error	df	<i>t</i>	<i>P</i> > <i>t</i>
SOR	AZ <sup>z</sup>			2.102	0.0959	105	21.9	<0.0001
SOR	DS			2.278	0.0898	84.6	25.3	<0.0001
SOR	GM3			2.187	0.8350	62.8	26.2	<0.0001
TRT		Control		1.727	0.0221	4.64	78.2	<0.0001
TRT		Drought		2.427	0.0441	44.2	54.9	<0.0001
SOR × TRT	AZ	Control		1.699	0.0783	82.8	21.7	<0.0001
SOR × TRT	AZ	Drought		2.504	0.1741	73.5	14.4	<0.0001
SOR × TRT	GM3	Control		1.957	0.0591	32.2	33.1	<0.0001
SOR × TRT	GM3	Drought		2.417	0.1550	47.5	15.6	<0.0001
TRT × DC(SOR)	AZ	Control	1	2.580	0.1586	318	16.3	<0.0001
TRT × DC(SOR)	AZ	Control	2	1.640	0.1586	318	10.3	<0.0001

<sup>z</sup>Seed source abbreviations: AZ = Chiricahua Mountains, Arizona; DS = Dripping Springs State Park, N.M.; GM3 = the third tree from the Guadalupe Mountains, Texas.

0.0001). But the test of TRT\*DC(SOR) suggests that, within a given source, the difference between treatments may be somewhat consistent across drought cycles ( $P = 0.2454$ ). The reduced model does not provide separate tests of these two effects. Consequently the test of TRT\*DC(SOR) pools the information in the nested model's DC(SOR) and TRT\*DC(SOR) test. This combined test reflects the variability in drought cycles within sources and yields  $P < 0.0001$ .

**SATURATED AND UNSATURATED MODELS.** The cross-classified, nested and reduced-nested models are saturated models that fit a mean to each cell (that is, to each unique combination of DC, TRT, and SOR). In some cases, researchers fit unsaturated models to overcome the estimability problems encountered with the cross-classified model. For example, researchers may use a main effects only model or a model that excludes drought cycle and all interactions involving drought cycle. Eliminating effects from a model to reduce computational difficulties is not equivalent to choosing an unsaturated model based on applying a model selection procedure to the data. While this strategy may appear to resolve computational problems, an unsaturated model may create other problems.

One problem is that effects have been eliminated from the model even though they may account for a large portion of the response variability. Consequently, estimates of fixed effects remaining in the model are unadjusted for eliminated effects and may be biased, particularly when data are unbalanced. Because the variability associated with the eliminated effects is essentially pooled with the other unexplained variability, variance estimates and SEs of estimates may be larger. Furthermore, estimates of the covariance components are consistent only when the mean is correctly specified (Verbeke and Molenberghs, 2000).

Even though the cross-classified, nested, and reduced-nested models are in some respects equivalent because all three are saturated models, the effects of the nested models more accurately reflect the experimental procedures of asynchronous drought trials. While the nested nature of the drought cycles is not as obvious as when the nested factor is a more familiar concept like genetic lines nested within location or location nested within region, drought cycles are established uniquely within each tree source. Consequently, for asynchronous drought cycle data, drought cycle is nested within source. Since drought cycle is nested within source, the nested models accommodate the different numbers of drought cycles that are expected within different sources. Estimation and inference reflect the more natural fit of the nested models and

these models provide estimates and appropriate tests.

**MEANS COMPARISONS AND TREATMENT EFFECTS.** Researchers should construct hypotheses of interest and test these hypotheses whether or not the tests are automatically provided in an omnibus F test or with the LSMEANS statement and the PDIF option. In the next two sections, we focus on preplanned comparisons. We will follow the convention of Ramsey and Schafer (2002). They suggest that preplanned tests should be conducted without adjusting probability values regardless of the statistical significance of the F test. Probability values for post hoc or unplanned comparisons should be adjusted (Ramsey and Schafer, 2002). We recommend using the reduced-nested model when writing many CONTRAST statements because CONTRAST statements for the reduced-nested model are often shorter than the corresponding CONTRAST statement for the nested model. Researchers should consult a statistician when writing CONTRAST and ESTIMATE statements as these can become very complicated. Contrasts are useful statistical tools because treatment differences or interactions could be confirmed significant with contrasts when an analysis of variance suggests there are no treatment differences or significant interactions (Marini, 2003).

For asynchronous drought cycle trials, specific preplanned comparisons are likely to be of greater interest and importance than any of the omnibus F tests. In particular, within source comparisons of control to drought-stressed groups are likely to be one set of preplanned comparisons. While these experiments are well-designed for purposes of comparing drought treated to control trees within each source, researchers must be cautious when comparing sources to one another. This is particularly true when comparing drought-treated sources to one another. Because drought cycle sequences and durations are unique to each source, there is confounding between sources and the drought treatment sequence, and comparisons between sources must be interpreted carefully.

Asynchronous drought trials are unavoidable even though their analysis and interpretation is complex. Because gravimetric water loss is the basic way to determine water loss from a growing substrate (Handreck and Black, 2002), researchers must craft drought-stress experiments that are clued into gravimetric water loss from the growing substrate. These experimental approaches will produce asynchronous drought cycles if plants are subjected to multiple drought events. While researchers might initially select plants that they anticipate will transpire at equivalent rates, differences in physiology still may create asynchronous trials.

Alternatively, adding moisture to the growing substrate to synchronize drought treatments destroys the concept of gravimetric moisture loss due to transpiration. In this case, some plants would be released temporarily from stress and any pre-stress steady-state rate of growth would be altered. Furthermore, the pattern of resumption of growth after drought stress relief depends on level of stress and the taxa (Hsiao, 1973). So, the addition of moisture to the growing medium could materially affect the outcome of the experiment.

For the nested and reduced-nested models, the LSMEANS statement provides many of the means comparisons likely to be of interest. For example, an LSMEANS statement with a PDIFF option can obtain *t* tests that compare all source  $\times$  treatment estimates to one another. This allows estimates of the average response to be compared across the duration of the experiment and consequently over the same time period. In particular, within source comparisons of drought treatment to control trees are of primary interest and are a subset of the comparisons produced by this LSMEANS statement. For our current data, these comparisons revealed that for all sources, drought-stressed trees had significantly higher (more negative) average midday measurements than control trees.

Another subset of comparisons provided by this same LSMEANS statement is the sources within controls and sources within drought-treatment comparisons. The current dataset reveals that relatively few of these comparisons are significant. However, one of the exceptions is the comparison of the DS (the tree originating from Dripping Springs State Park) droughted trees ( $-2.75$  MPa) to the GM2 droughted trees ( $-3.65$  MPa). The difference is estimated to be  $-0.90$  MPa and is highly significant ( $P = 0.0003$ ). Because drought cycles are nested within sources, it may be desirable to confirm that treatment effects (treatment – control differences) are different for these two sources. The comparison of these treatment effects correspond to assessing whether or not there is a DS vs. GM2  $\times$  treatment interaction. The following estimate statement can be used with either the nested or reduced-nested model and compares the DS to the GM2 treatment effects.

```
ESTIMATE "DS vs. GM2 TREATMENT EFFECTS"  
SOR*TRT 0 0 -1 1 0 0 1 -1;
```

This yields the estimate  $-0.88$  MPa and confirms that the DS and GM2 treatment effects differ ( $P = 0.0008$ ). The estimate of the difference in treatment effects differed little from the estimate of treated source differences. While it is possible for these two estimates to differ substantially, large differences are more likely to occur when comparing sources at a particular drought cycle, for example at the first drought cycle, than when averaging across the duration of the experiment. As previously noted, source differences should be interpreted carefully. In this case, data are averaged across five drought cycles when DS is compared to GM2. However, drought cycle durations averaged 18 d for DS vs. 21 d for GM2. Consequently, it is possible that the difference might be due to the greater stress endured by GM2 rather than to source differences alone. The difference in drought cycle duration is relatively minor, only 3 d. Consequently, the data indicates that DS endured lesser water tension (less stress) even though its drought cycles were shorter.

Treatment effect comparisons correspond to interaction contrasts and control for the influence of environmental conditions on the days when measurements are taken. They are more appropriate than simply comparing treated source means. For example, at the end of the first drought cycle, midday water potential measurements for each source were taken on a single day but the date of

measurement differed among sources. Using treatment effects to compare sources' response to treatment adjusts for the influence of daily fluctuations in environmental conditions.

**ADDITIONAL SUGGESTED CONTRASTS.** In this section, we discuss two additional sets of specific hypotheses and contrasts. The first set of contrasts partitions the drought cycles into those common to all sources and compares averages based on the common drought cycles to one another. In addition, for those sources with more drought cycles, the average of the first set of drought cycles can be compared to the average of the later drought cycles. This comparison can determine whether the response differs between the earlier and later drought cycles or whether there are differences between sources having shorter and longer drought cycle durations. These contrasts may be made within treated or within control plants and also might be used to compare treatment effects.

In our data set, the source AZ (tree originating from the Chiricahua Mountains in Arizona) completed five drought cycles but the source GM3 (the third tree from the Guadalupe Mountains) completed seven. A comparison of AZ to GM3 treatment effects gave  $P = 0.1746$  and suggested no difference (Table 5). This comparison involved averaging across five drought cycles for AZ and seven drought cycles for GM3. A comparison of the AZ to the GM3 early treatment effects yields  $P = 0.0938$ . While the conclusions did not differ in this case, if they had, this or a similar set of comparisons might be very revealing about changes that occur as trees go through more drought cycles. For example, a significant probability value in later drought cycles could indicate that one of the tree sources had acclimated to the drought treatment. For example, the physiology of a plant could have altered so that midday water potential became less negative (lower magnitude) because of improved plant water status.

A comparison of GM3 early controls to GM3 late controls reveals a significant difference with the later controls having a lower mean. This difference is most likely due to seasonal decreases in light and temperature with the concomitant drop in moisture loss from the plant. However, this difference strongly suggests the need to compare treatment effects instead of simply comparing drought treated sources. The comparison of the GM3 early to late treatment effects suggests there may be no difference ( $P = 0.1499$ ) between early and late treatment effects (Table 5).

A second set of contrasts can be obtained using an LSMEANS statement for TRT\*DC(SOR) and PDIFF option. Comparisons of treated to control trees within each level of source and drought cycle are a subset of the TRT\*DC(SOR) LSMS comparisons. These allow researchers to contrast the drought-stressed to the control trees within a source at each drought cycle and to identify break points where the magnitude of the treatment effect changes. These differences can help explore the impact of prolonged drought on the trees. For example, consider the trend of decreasing differences (treatment effects) in midday water potential for the source AZ (Table 6). Additional comparisons can be used to determine which of the individual drought cycle treatment effects differ significantly. Despite the fact that, in the nested model, the F test for TRT\*DC(SOR) was not significant, source AZ comparisons of treatment effects at the different drought cycles reveals that the drought cycle one treatment effect differs significantly from the drought cycle four and the drought cycle five treatment effects (Table 6).

## Recommendations

In both synchronous and asynchronous drought trials, researchers should use mixed model methodology to adequately

Table 5. Representative estimates and comparisons of treatment midday water potential partitioned into early and late drought cycles.

Effect	Estimate value (-MPa)	P
AZ <sup>z</sup> control (ctrl)	1.70	
AZ drought (drt)	2.50	
GM3 <sup>y</sup> ctrl	1.96	
GM3 drt	2.42	
GM3 ctrl early <sup>x</sup>	2.12	
GM3 drt early <sup>x</sup>	2.49	
GM3 ctrl late <sup>w</sup>	1.54	
GM3 drt late	2.24	
AZ vs. GM3 treatment effects	0.34	0.1746
AZ ctrl vs. GM3 ctrl early treatment effects	0.44	0.0938
GM3 ctrl early vs. GM3 ctrl late	0.58	0.0001
GM3 early vs. GM3 late <sup>v</sup> treatment effects	-0.34	0.1499

<sup>z</sup>AZ = Seed source from the Chiricahua Mountains, Ariz.

<sup>y</sup>GM3 = The third seed source from the Guadalupe Mountains, Texas.

<sup>x</sup>The effect "GM3 ctrl early or GM3 drt early" are comparisons based on the first five drought cycles (drought cycles 1–5).

<sup>w</sup>The effect "GM3 ctrl late or GM3 late" are comparisons based on the last two drought cycles (drought cycles 6 and 7).

Table 6. An example of using the Arizona (AZ) source for within source and drought cycle comparisons of drought-stressed to control trees and within source comparisons of drought cycle treatment effects.

Drought cycle	Treatment		Difference (treatment effect)
	Control	Drought	
1	2.58	3.94	1.36* a <sup>z</sup>
2	1.64	2.72	1.08* ab
3	1.55	2.42	0.87* ab
4	1.19	1.64	0.45 b
5	1.55	1.80	0.26 b

\*Indicates significant difference at  $P \leq 0.05$  for within drought cycle treatment effects.

<sup>z</sup>LSMeans within a column (within source drought cycle treatment effects) followed by the same letter are not different at  $P \leq 0.05$ .

model covariance structure features including serial correlation and heterogeneity of variance among the repeated measures as well as heterogeneity of variance between treatment groups. In asynchronous drought trials, the experimental protocol implies that different sources will undergo different numbers of drought cycles and that sets of drought cycles will be unique to each source. A cross-classified model for the fixed portion of the model does not provide the correct analysis of asynchronous drought data because there are missing factor combinations. Nested models reflect the experimental protocol that produced the data and are recommended for asynchronous trials. On the other hand, for synchronized drought trials, a cross-classified model is appropriate. When trials are synchronized, drought cycles are crossed with source rather than nested within sources; a single set of drought cycles is common to all sources and there are no missing combinations of source and drought cycle.

Using the appropriate model does not preclude the need for the researcher to consult with a statistician, and to carefully consider and construct estimates and contrasts to address specific questions of interest. For asynchronous drought trials, researchers might form contrasts to compare sources to each other or treatment effects within a source from one drought cycle to another. Particular care should be taken when interpreting comparisons of drought-treated sources.

## Literature Cited

- Balok, C.A. and R. St. Hilaire. 2002. Drought responses among seven southwestern tree taxa. *J. Amer. Soc. Hort. Sci.* 127:211–218.
- Diggle, P.J., K.-Y. Liang, and S.L. Zeger. 1994. *Analysis of longitudinal data*. Oxford Univ. Press, New York.
- Feser, C., R. St. Hilaire, and D. VanLeeuwen. 2005. Development of in-ground container plants of Mexican elders exposed to drought. *HortScience* 40:446–450.
- Garrett, K.A., L.V. Madden, G. Hughes, and W.F. Pfender. 2004. New applications of statistical tools in plant pathology. *Phytopathology* 94:999–1003.
- Handreck, K. and N. Black. 2002. *Growing media for ornamental plants and turf*. 3<sup>rd</sup> ed. Univ. N.S. Wales Press, Sydney, Australia.
- Hsiao, T.C. 1973. Plant responses to water stress. *Ann. Rev. Plant Physiol.* 24:519–570.
- Littell, R.C. 2002. Analysis of unbalanced mixed model data: A case study comparison of ANOVA versus REML/GLS. *J. Agr. Biol. Environ. Stat.* 7(4):472–490.
- Littell, R.C., J. Pendergast, and R. Natarajan. 2000. Tutorial in biostatistics: Modelling covariance structure in the analysis of repeated measures data. *Stat. Medicine* 19:1793–1819.
- Littell, R.C., P.R. Henry, and C.B. Ammerman. 1998. Statistical analysis of repeated measures data using SAS procedures. *J. Animal Sci.* 76:1216–1231.
- Littell, R.C., G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. 1996. *SAS system for mixed models*. SAS Inst., Cary, N.C.
- Marini, R.P. 2003. Approaches to analyzing experiments with factorial arrangements of treatments plus other treatments. *HortScience* 38:117–120.
- Marron, N., E. Dreyer, E. Boudouresque, D. Delay, J.-M. Petit, F.M. Delmotte, and F. Brignolas. 2003. Impact of successive drought and re-watering cycles on growth and specific leaf area of two *Populus × canadensis* (Moench) clones, 'Dorskamp' and 'Luisa-Avanzo.' *Tree Physiol.* 23:1225–1235.
- Milliken, G.A. and D.E. Johnson. 1992. *Analysis of messy data*. Chapman & Hall, New York.
- Padgett-Johnson, M., L.E. Williams, and M.A. Walker. 2003. Vine water relations, gas exchange, and vegetative growth of seventeen *Vitis* species grown under irrigated and nonirrigated conditions in California. *J. Amer. Soc. Hort. Sci.* 128:269–276.
- Pendleton, O.J., M. Von Tress, and R. Bremer. 1986. Interpretation of the four types of analysis of variance tables in SAS. *Commun. Stat. – Theory Methods* 15:2785–2808.
- Piepho, H.P., A. Büchse, and C. Richter. 2004. A mixed modeling approach for randomized experiments with repeated measures. *J. Agron. Crop Sci.* 190:230–247.
- Piepho, H.P., A. Büchse, and K. Emrich. 2003. A hitchhiker's guide to mixed models for randomized experiments. *J. Agron. Crop Sci.* 189:310–322.
- Ramsey, F.L. and D.W. Schafer. 2002. *The statistical sleuth: A course in methods of data analysis*. 2<sup>nd</sup> ed. Duxbury, Pacific Grove, Calif.
- Schabenberger, O. and F.J. Pierce. 2002. *Contemporary models for the plant and soil sciences*. CRC Press, Boca Raton, Fla.
- St. Hilaire, R., C.F. Feser, T.W. Sammis, and A.S. St. Hilaire. 2003. A system to measure evapotranspiration of in-ground container plants of Mexican elder. *HortTechnology* 13:185–189.
- St. Hilaire, R., A.W. Hockman, and S.M. Chavez. 2004. Propagation and irrigation affect the development of catnip. *Acta Hort.* 629:321–327.
- Tipton, J.L. 1994. Relative drought resistance among selected southwestern landscape plants. *J. Arboricult.* 20:150–155.
- Urquhart, N.S. and D.L. Weeks. 1978. Linear models in messy data: Some problems and alternatives. *Biometrics* 34:696–705.
- Verbeke, G. and G. Molenberghs. 2000. *Linear mixed models for longitudinal data*. Springer-Verlag, New York.
- Wolfinger, R.D. 1996. Heterogeneous variance-covariance structures for repeated measures. *J. Agr. Biol. Environ. Stat.* 1(2):205–230.
- Yandell, B.S. 1997. *Practical data analysis for designed experiments*. Chapman & Hall, London.