Kansas State University Libraries New Prairie Press

Conference on Applied Statistics in Agriculture

2001 - 13th Annual Conference Proceedings

USING PROC NLMIXED TO ANALYZE A TIME OF WEED REMOVAL STUDY

Erin E. Blankenship

Sean Evans

Walter W. Stroup

Stevan Z. Knezevic

See next page for additional authors

Follow this and additional works at: https://newprairiepress.org/agstatconference

🔮 Part of the Agriculture Commons, and the Applied Statistics Commons



This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

Blankenship, Erin E.; Evans, Sean; Stroup, Walter W.; and Knezevic, Stevan Z. (2001). "USING PROC NLMIXED TO ANALYZE A TIME OF WEED REMOVAL STUDY," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1225

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

Author Information Erin E. Blankenship, Sean Evans, Walter W. Stroup, and Stevan Z. Knezevic

USING PROC NLMIXED TO ANALYZE A TIME OF WEED REMOVAL STUDY

Erin E. Blankenship^a, Sean Evans^b, Walter W. Stroup^a, and Stevan Z. Knezevic^c

^aDepartment of Biometry, University of Nebraska–Lincoln, Lincoln, NE 68583-0712 ^bDepartment of Agronomy and Horticulture, University of Nebraska–Lincoln, Lincoln, NE 68583-0915 ^cHaskell Agricultural Laboratory, University of Nebraska, Concord, NE 68728-2828

ABSTRACT

Many studies in weed science involve fitting a nonlinear model to experimental data. Examples of such studies include dose-response experiments and studies to determine the critical period of weed control. The experiments typically use block designs and often have additional complexity such as split-plot features. However, nonlinear models are typically fit using software such as SAS PROC NLIN that are limited to a single error term and whose ability to account for blocking is either awkward or lacking entirely. For example, Seefeldt et al. (1995) only proceeded in fitting the nonlinear model after establishing that the block effect was negligible. Issues such as multiple error terms in split-plot designs are simply not dealt with at all. In this paper, we examine a weed removal study carried out as a split-plot design with blocks and illustrate the use of SAS PROC NLMIXED to account for blocks and the two-level error structure.

KEY WORDS: split-plot design; nonlinear regression; logistic model; Gompertz model; critical period of weed control.

1 Introduction

Weed control is an essential part of crop management. However, efficient weed control is also important. Reducing the amount of herbicide is not only cost effective, but may also reduce possible environmental contamination and impact on non-target organisms (Hall et al., 1992). One way to reduce herbicide use is to determine the times during growing season in which weed control is most effective in increasing yield. This idea is best illustrated graphically. The plot in Figure 1 demonstrates the effect of increasing duration of weed presence on crop yield. As weeds are left in the field for longer periods of time (measured in growing degree days, gdd, in this example), the crop yield decreases. However, allowing weeds to remain in the crop for a short period of time at the beginning of the growing season does not appear to significantly decrease yield. For approximately the first 200 gdd, the yield is fairly constant. Figure 2 shows the effect of increasing duration of weed control on crop yield. As weeds are controlled for longer periods of time, the crop yield increases. However, as with increasing durations of weed presence, it does not appear that weeds must be controlled for the entire growing season. After approximately 700 gdd the yield is fairly constant, indicating that suspending weed control at this time will not adversely affect crop yield. The plots in Figures 1 and 2, therefore, indicate that there may be some time at both the beginning and end of the growing season when weed presence has little impact on yield. The period of time in which weed interference does significantly impact yield is commonly known as the critical period of weed control (CPWC). The CPWC is defined as the point at which the grower must intervene in the development of the weed, and the length of time the intervention should continue (Weaver and Tan, 1983).

Time of weed removal studies focus on determining the critical period of weed control. They are typically carried out over one or two years and involve a single crop, as each crop has its own CPWC. There are generally two types of treatment, corresponding to the two plots shown in Figures 1 and 2. One type of treatment involves allowing weeds to remain in the crop for increasing durations, after which the plots are kept weed-free for the remainder of the growing season. The other type of treatment requires the crop be kept weed-free for increasing durations, after which the weeds are allowed to grow freely. A season long weed-free treatment is also usually maintained so that response data can be expressed as the percentage of the average yield from the weed-free plots (Oliver, 1988). The timing of the weed removal has been based on days after planting (Weaver and Tan, 1983; Weaver et al., 1992), days after emergence (Baziramakenga and Leroux, 1992; Ferrero et al., 1996; Singh et al., 1996) and growth stage of the crop (Hall et al., 1992; Van Acker et al., 1993). Time of weed removal studies are typically carried out as randomized complete block designs. Occasionally, treatments other than the time of weed removal are also applied, typically in a split-plot fashion. For example, different densities of weeds may be inflicted upon the crop (Baziramakenga and Leroux, 1994) or different row spacings may be used (Mulugeta and Boerboom, 2000).

The critical period of weed control is estimated by fitting non-linear regressions to the sets of treatments leading to Figures 1 and 2. Methods in use to date often lead to inappropriate or insufficient analyses because they ignore or avoid complications resulting from blocking or multiple error terms. The purpose of this paper is 1) to review traditional methods; 2) to point out their shortcomings; and 3) to show how nonlinear mixed models, implemented using PROC NLMIXED can address these problems. An example demonstrates the use of PROC NLMIXED to account for blocks and the two-level error structure necessitated by additional treatment application in split-plots. We also, briefly, discuss estimating the critical period of weed control.

2 Traditional Analysis

The data are usually first analyzed with an analysis of variance to determine if the time of weed removal, or possibly an additional treatment, has a significant effect on the response. The various times of weed removal are then used as levels of an independent variable to fit a nonlinear model to the yield response. Typically, a logistic model is used to fit the increasing duration of weed interference data, and a Gompertz model is used to fit the increasing duration of weed-control data (Hall et al., 1992; Van Acker et al., 1993). The parameterization of the logistic model commonly fit is

$$Y = \left(\frac{1}{D * \exp(K * (T - x)) + F} + \frac{F - 1}{F}\right) * 100\%,\tag{1}$$

where Y is the yield response, expressed as the percentage of the weed-free crop yield, T is the removal time (i.e., the independent variable), the parameter x is the point of inflection and the parameters D, K, and F are simply termed "constants" (Hall et al., 1992; Van Acker et al., 1993). The parameterization of the Gompertz model that is often used is

$$Y = A * \exp(-B * \exp(-k * T)), \tag{2}$$

where Y and T are again the response and independent variables, respectively, A is the yield asymptote, and B and k are constants (Hall et al., 1992; Van Acker et al., 1993). The parameter A could, if expressing the response as a percentage of the weed-free yield, be held constant at 100.

The scientist uses the fit of these nonlinear models to determine the critical period of weed control. The beginning of the critical period is the time (day) of weed removal which corresponds to some pre-specified level of yield loss in the logistic curve. The end of the critical period is the time of removal which corresponds to the same level of yield loss in the Gompertz curve. For example, the weed scientist may decide that a 5% loss is a significant yield loss. (S)he would then find the day of weed removal which gives 95% yield response in both curves, and these days would define the beginning and end of the critical period of weed control for that crop. An example of a CPWC is shown in Figure 3. The solid lines show the fitted logistic and Gompertz curves. The horizontal dashed line shows the 95% yield response, and the vertical dashed lines indicate the time of weed removal that corresponds to 95% yield response in the two curves. The logistic curve crosses the 95% yield line at about 200 growing degree days. Therefore, one concludes that the presence of weeds prior to about 200 growing degree days or after about 600 growing degree days not jeopardize crop yields.

While this approach appears to be appropriate, often the way the methods are implemented is inappropriate or insufficient. For example, while blocks are usually accounted for in the initial analysis of variance, they are typically ignored at the nonlinear modeling stage. Some authors recognize this peculiarity. Seefeldt et al. (1995), for example, only fit the nonlinear models after establishing that the block effect was negligible. This is not, however, a realistic approach because in a carefully blocked experiment we expect to see a significant block effect. A robust method for fitting these nonlinear models needs to account for the possibility of significant differences between the blocks.

Experimental treatments other than time of weed removal are also often applied, typically in a split-plot fashion. The nonlinear models are then usually fit separately for each level of the additional treatment, rather than dealing with the multiple error terms imposed by the design. For example, Baziramakenga and Leroux (1994) used three different weed densities in addition to different times of weed removal. The logistic and Gompertz models were then fit separately for each weed density.

Some data analysts account for blocking and split-plot effects by determing the CPWC using multiple comparison procedures such as least significant differences or Duncan (Barrentine, 1974; Friesen, 1978; Weaver and Tan, 1983; Harris and Ritter, 1987) within the framework of the analysis of variance. However, the problems with this approach have been discussed at great length (Cousens, 1988), and this discussion has motivated the use of non-linear curves to determine the CPWC (Hall et al., 1992; Weaver et al., 1992). Statistical inference for the estimates of the endpoints of the CPWC has not been fully explored. Singh et al. (1996) propose methods for constructing confidence intervals for the endpoints of the critical period, but they only demonstrate their methods for linear and logistic models. Also, their methods do not seem to be easily adapted to mixed effects models, which most time of weed removal studies require. The appropriate analysis requires non-linear modeling of the treatment design and the ability to model blocking and split-plot variability–i.e., nonlinear mixed models. This method is demonstrated in Section 4.

3 Motivating Experiment

We use the following example, presented during a consulting session, to demonstrate the use of nonlinear mixed models to fit the regressions required to estimate the CPWC. The study, designed and carried out by co-authors Evans and Knezevic, is typical of time of weed removal studies with an additional treatment. The objective of the study was to evaluate the effects of nitrogen supply on the critical period for weed control for dryland corn in competition with a naturally occurring weed population.

The experiment was laid out as a split-plot design in four randomized complete blocks,

and was carried out at two different locations in Nebraska (Mead and Concord) in both 1999 and 2000. The whole plot factor is applied nitrogen, which was administered at three levels: 0, 60 and 120 hg N ha⁻¹. The subplot factor is time of weed removal. As with most such studies, two sets of subplot treatments were applied: increasing durations of weed interference and increasing durations of weed control. There were five times of removal in each type of treatment (weedy or weed-free up to third leaf stage, sixth leaf stage, ninth leaf stage, fifteenth leaf stage, and anthesis), and season-long weedy and weed-free treatments were also maintained. The removal times, although determined by leaf stage, where expressed as growing degree days.

The response variable used in this study is the percentage of the yield from the weed-free plots by block, rather than the percentage of the yield from the weed-free plots averaged over block as has been used in previous studies. This is because previous studies have not incorporated blocks in the nonlinear fitting stage of the analysis. The analysis we propose will include block effects, and so there is no need to average over blocks to obtain the response yields.

When this experiment was brought to us, PROC NLMIXED (SAS Institute, Inc., 1999) had recently been released, and we saw this as a good opportunity to see if NLMIXED could be used to analyze time of weed removal studies and estimate the critical period of weed control. We found that it could be used successfully, and the method we used is presented in the next section.

4 Analysis Using PROC NLMIXED

The general framework for any statistical model can be expressed as

$$y = \text{mean model} + \text{variability model}.$$
 (3)

The mean model describes how the response, y, changes based on the experimental treatments. The variability model expresses how random error is introduced into the response. In time of weed removal studies, the mean model is a non-linear regression model, and the variability model encompasses the block effect, whole-plot error effect and split-plot error effect. Each of these effects must be specified in PROC NLMIXED. We will start by examining the general syntax for PROC NLMIXED, and considering how it might be used for fitting the nonlinear models used in time of weed removal studies. The general syntax for PROC NLMIXED is (SAS Institute, Inc., 1999):

```
proc nlmixed;
```

```
parms parm1=start_val1 parm2=start_val2 . . . parmp=start_valp;
model dependent_variable ~ distribution;
random random_effects ~ distribution subject=variable;
run;
```

There are three main statements in PROC NLMIXED: PARMS, MODEL and RANDOM. The PARMS statement lists the parameters used in the model and their starting values. The MODEL statement gives the conditional distribution of the response given the random effects. The RANDOM statement gives the distribution of the random effects.

We begin by examining the Gompertz function for modeling increasing durations of weed control. For the Gompertz function, the form of the mean model follows from Equation (2). We also need to incorporate the blocks and split-plot elements. Thus, the model

$$\operatorname{relyield}_{iik} = A_i * \exp(-B_i * \exp(-k_i * T_j)) + R_k + w_{ik} + \epsilon_{ijk}$$
(4)

where i = 1, 2, 3 nitrogen levels, j = 1, 2, 3, 4, 5, 6, 7 weed removal times, k = 1, 2, 3, 4 blocks, R_k is the block effect, w_{ik} is the whole plot error effect, assumed to be distributed iid $N(0,\sigma_w^2)$, and ϵ_{ijk} is the split plot error effect, assumed to be distributed iid $N(0,\sigma^2)$, can be used to model increasing durations of weed control. This is the model familiar from use with traditional split-plot experiments. The only difference is that the mean effect is a nonlinear function.

Because PROC NLMIXED does not have a CLASS statement, the qualitative variables must be defined explicitly. For example, to set up classification variables for levels of the whole plot factor we could define, in the DATA step: in1=(n=1); in2=(n=2); in3=(n=3);. The mean effect could then be expressed as

$$\mu_{ij} = (A1 * in1 + A2 * in2 + A3 * in3) * \exp(-(B1 * in1 + B2 * in2 + B3 * in3)) * \exp(-(k1 * in1 + k2 * in2 + k3 * in3) * T)).$$

Also as a consequence of no CLASS statement, we must identify the whole plot experimental unit. Here, we used the variable whole=10*blk + n. This variable identifies the block by nitrogen level interaction, and therefore uniquely specifies the whole plot experimental unit. We also need to define classification variables for the block. We will treat the blocks as fixed effects, and use the $\sum_{k=1}^{4} R_k = 0$ constraint to remove singularities. The block indicator variables could then be defined as: ib1=(blk=1)-(blk=4); ib2=(blk=2)-(blk=4); ib3=(blk=3)-(blk=4);. The block effect is then expressed as bloc=blk1*ib1 + blk2*ib2 + blk3*ib3.

There are several parameters that require starting values. The default value in SAS is 1, and PROC NLMIXED can be sensitive to choice of starting values, especially for non-linear regression or variance component parameters. For the block effects (blk1, blk2, blk3), starting values were all set to 0. For the Gompertz parameters (A1, A2, A3, B1, B2, B3, k1, k2, k3) starting values were obtained by fitting non-linear models using PROC NLIN to treatment LSMEANS obtained from PROC MIXED. The parameter estimates from PROC NLIN were used as starting values in PROC NLMIXED. The PROC MIXED step also yielded covariance parameter estimates of nitrogen level by block and the residual error. These covariance parameter estimates were used as starting values for whole and split plot error effects, respectively.

To demonstrate the use of PROC NLMIXED with our time of weed removal study, we will use the data collected at Concord, Nebraska in 1999. The SAS code is:

Note that the temp variable gives the expected value of the response conditional on the random effects. The parameter estimates obtained are shown in Table 1 and the fitted Gompertz curve for nitrogen level one is shown in Figure 4.

We can also use PROC NLMIXED to estimate the differences in the parameters for different nitrogen levels, and thus test to see if nitrogen level has an effect on relative yield. For this example, we could add the following:

estimate 'a1 vs a2' a1-a2 df=71;

Note that the ESTIMATE statement uses syntax similar, but not identical, to other SAS procedures. The degrees of freedom must be specified. If the df= option is not specified SAS will use the total number of observations minus one as the degrees of freedom. PROC NLMIXED also has a CONTRAST statement.

The Gompertz function is often used to model increasing durations of weed control, and therefore the end of the critical period of weed control. The treatments involving increasing durations of weed interference are usually modeled using the logistic function. The parameterization of the logistic function commonly used was displayed in Equation (1), and is

$$Y = \left(\frac{1}{D * \exp(K * (T - x)) + F} + \frac{F - 1}{F}\right) * 100\%.$$

Note that the model in this form is overparameterized. Consider the model term $D * \exp(K * (T - x))$. This can be rewritten as

$$D * \exp(K * (T - x)) = D * \exp(KT) * \exp(-Kx)$$

= $\exp(\ln D) * \exp(KT) * \exp(-Kx)$
= $\exp(KT) * \exp(\ln D - Kx)$
= $\exp(KT) * \exp(K(\ln D/K - x))$
= $\exp(KT) * \exp(Kx^*).$

The parameter D is therefore redundant, and can be dropped from the model. The logistic function we used to model the increasing durations of weed interference is the threeparameter logistic model

$$Y = \left(\frac{1}{\exp(K * (T - x)) + F} + \frac{F - 1}{F}\right) * 100\%.$$
 (5)

We incorporated split-plot elements and blocks and found starting values as we did with the Gompertz model. The SAS code for the Concord, 1999 data is:

The fitted logistic model for nitrogen level is one is shown in Figure 5.

5 Estimating the Critical Period of Weed Control

The goal of the experiment was to estimate the critical period of weed control. This estimation is not a trivial problem, as it involves essentially inverse predicition, or calibration, with a nonlinear mixed effects model. Much of the discussion of the estimation of and inference for the CPWC endpoints is deferred to a forthcoming paper, but we will outline the general estimation procedure. First, a percent loss level is chosen arbitrarily by the weed scientist. For example, one researcher may consider 10% to be a significant yield loss while another may use 5%. The beginning and end of the critical period are then estimated by substituting the yield level corresponding to the chosen loss into the logistic and Gompertz equations, respectively, and finding the day in the growing season that corresponds to that yield.

For our time of weed removal experiment we considered 5% to be a significant loss level, which implies that a relative yield of 95% was substituted into both the fitted logistic and Gompertz equations to estimate the endpoints of the CPWC. Substituting 95% into the logistic model for nitrogen level one at Concord in 1999 estimates of the beginning of the CPWC as 211.5 growing degree days. The fitted Gompertz equation at 95% yield gives an

estimate of the end of the CPWC at 580.37 growing degree days. We were also interested in constructing confidence intervals for these estimates, as well as testing for significant differences between endpoints with different nitrogen levels.

We adapted the calibration method presented in Schwenke and Milliken (1991) to nonlinear mixed models, and used that adapted method to construct confidence intervals and run hypothesis tests. For example, a 95% confidence interval for the beginning of the CPWC for nitrogen level one at Concord in 1999 is (208.06, 214.93) and a 95% confidence interval for the end of the CPWC is (578.46, 582.29). This critical period is shown in Figure 6. A more thorough investigation of these calibration techniques will be presented in a forthcoming paper.

6 Conclusion

Analysis of time of weed removal studies using PROC NLMIXED is a great improvement over the current analysis techniques. PROC NLMIXED can incorporate random effects, as well as multi-level error structure necessitated by split-plot experimental designs that are impossible or awkward to implement in PROC NLIN or other nonlinear model fitting software packages currently used by weed scientists. We have used an experiment from weed science to illustrate the use of NLMIXED, but there are numerous disciplines in which both a complex experimental design and a nonlinear model are required. Until recently, software has not been available to implement both the experimental design features and the nonlinear model. With the advent of PROC NLMIXED, these features can coexist and the benefits of this procedure must be disseminated among both subject-matter scientists and statistical consultants.

However, much work remains to be done on this problem. In this paper, the blocks were treated as fixed effects, and often they should be considered random. Multiple random effects can be implemented in PROC NLMIXED. Also, from the plots of the fitted nonlinear models, it appears that we may have a problem with nonconstant variance. This may also need to be addressed. Finally, PROC NLMIXED only carries out maximum likelihood estimation. We know from PROC MIXED that type I error can be inflated when carrying out inference with maximum likelihood estimates of variance components, as opposed to restricted maximum likelihood estimates. Whether this type I error inflation occurs in NLMIXED as well needs to be investigated.

References

- Barrentine, W. L. (1974) "Common Cocklebur Competition in Soybeans," Weed Science, 22, 600–603.
- [2] Baziramakenga, R. and Leroux, G. D. (1994) "Critical Period of Quackgrass (*Elytrigia repens*) Removal in Potatoes (*Solanum tuberosum*)," Weed Science, 42, 528–533.
- [3] Cousens, R. (1988) "Misinterpretations of Results in Weed Research Through Inappropriate Use of Statistics," Weed Research, 28, 281–289.
- [4] Ferrero, A., Scanzio, M., and Acutis, M. (1996) "Critical Period of Weed Interference in Maize," Proceedings Second International Weed Control Congress, Copenhagen, 171–176.
- [5] Friesen, G. H. (1978) "Weed Interference in Pickling Cucumbers (*Cucumis sativus*)," Weed Science, 26, 626–628.
- [6] Hall, M. R., Swanton, C. J., and Anderson, G. W. (1992) "The Critical Period of Weed Control in Grain Corn (Zea mays)," Weed Science, 40, 441–447.
- [7] Harris, T. C. and Ritter, R. L. (1987) "Giant Green Foxtail (Setaria viridis var. major) and Fall Panicum (Panicum dichotomiflorum) Competition in Soybeans (Glycine max)," Weed Science, 35, 663–668.
- [8] Mulugeta, D. and Boerboom, C. (2000) "Critical Time of Weed Removal in Glyphosate-Resistant *Glycine max*," Weed Science, 48, 35–42.
- [9] Oliver, L. R. (1988) "Principles of Weed Threshold Research," Weed Technology, 2, 398–403.
- [10] SAS Institute, Inc. (1999) SAS On-Line Doc, Version 8. Cary NC: SAS Institute.
- [11] Schwenke, J. R. and Milliken, G. A. (1991) "On the Calibration Problem Extended to Nonlinear Models," *Biometrics*, 47, 563–574.

- [12] Seefeldt, S. S., Jensen, J. E., and Fuerst, E. P. (1995) "Log-Logistic Analysis of Herbicide Dose-Response Relationships," Weed Technology, 9, 218–227.
- [13] Singh, M., Saxena, M. C., Abu-Irmaileh, B. E., Al-Thahabi, S. A., and Haddad, N. I.
 (1996) "Estimation of Critical Period of Weed Control," Weed Science, 44, 273–283.
- [14] Van Acker, R. C., Swanton, C. J., and Weise, S. F. (1993) "The Critical Period of Weed Control in Soybean [*Glycine max* (L.) Merr.]," *Weed Science*, 41, 194–200.
- [15] Weaver, S. E., Kropff, M. J., and Groeneveld, R. M. W. (1992) "Use of Ecophysiological Models for Crop-Weed Interference: The Critical Period of Weed Interference," Weed Science, 40, 302–307.
- [16] Weaver, S. E. and Tan, C. S. (1983) "Critical Period of Weed Interference in Transplanted Tomatoes (*Lycopersicon esculentum*): Growth Analysis," Weed Science, 31, 476–481.

| Parameter | Estimate | Standard Error |
|-----------|----------|----------------|
| blk1 | -0.3359 | 2.3471 |
| blk2 | 4.8686 | 2.3471 |
| blk3 | -4.0276 | 2.3471 |
| al | 104.12 | 4.3169 |
| a2 | 105.12 | 4.1796 |
| a3 | 102.99 | 5.7657 |
| b1 | 1.1499 | 0.1830 |
| b2 | 0.9878 | 0.1537 |
| b3 | 0.5229 | 0.1026 |
| k1 | 0.004358 | 0.000787 |
| k2 | 0.004776 | 0.000969 |
| k3 | 0.003079 | 0.001143 |
| s2w | 36.4028 | 11.1761 |
| s2e | 117.83 | 11.1761 |

Table 1: Parameter estimates from the Gompertz model applied to the Concord, 1999 data.



Figure 1: Increasing duration of weed presence. The time that weeds are allowed to remain in the field is measured in growing degree days.



Figure 2: Increasing duration of weed control. The time that weeds are controlled is measured in growing degree days.



Figure 3: The critical period of weed control, with 5% yield loss. The CPWC begins at about 200 growing degree days and ends at about 600 growing degree days.



Figure 4: The fitted Gompertz curve for the Concord, 1999 data at nitrogen level one.



Figure 5: The fitted logistic curve for the Concord, 1999 data at nitrogen level one.



Figure 6: The critical period of weed control for nitrogen level one at Concord in 1999. The estimated CPWC is from 211.5 to 580.37 growing degree days.