LINEAR AND NONLINEAR MIXED-EFFECTS MODELS

Douglas M. Bates

Jose C. Pinheiro

Recommended Citation

LINEAR AND NONLINEAR MIXED-EFFECTS MODELS

Douglas M. Bates

Department of Statistics
University of Wisconsin – Madison

José C. Pinheiro

Bell Laboratories
Lucent Technologies

Abstract

Recent developments in computational methods for maximum likelihood (ML) or restricted maximum likelihood (REML) estimation of parameters in general linear mixed-effects models have made the analysis of data in typical agricultural settings much easier. With software such as SAS PROC MIXED we are able to handle data from random-effects one-way classifications, from blocked designs including incomplete blocked designs, from hierarchical designs such as split-plot designs, and other types of data that may be described as repeated measures or longitudinal data or growth-curve data. It is especially helpful that the new computational methods do not depend on balance in the data so we are able to deal more easily with observational studies or with randomly missing data in a designed experiment.

We describe some of the new computational approaches and how they are implemented in the nlme3.0 library for the S-PLUS language. One of the most powerful features of this language is the graphics capabilities, especially the trellis graphics facilities developed by Bill Cleveland and his coworkers at Bell Labs. Although most participants in this conference may be more familiar with SAS, and most of the models described here can be fit with PROC MIXED or the NLINMIX macro or new PROC NLMIXED in SAS version 7, some exposure to the combination of graphical display and model-fitting approaches from S-PLUS may be informative.

We show how data exploration with trellis graphics, followed by fitting and comparing mixed-effects models, followed by graphical assessment of the fitted model can be used in a variety of situations. On some occasions, such as modeling growth curves, a linear trend or polynomial trend or other types of linear statistical models for the within-subject time dependence are just not going to do an adequate job of representing the data. In those cases, a nonlinear model is more appropriate. We show how the concept of a random coefficient model can be extended to nonlinear models so as to fit nonlinear mixed-effects models.

1   Introduction

The mixed-effects model has been one of the mainstays of applied statistics in agriculture. Indeed, much of the theory and practice of mixed-effects modeling was developed directly for agricultural applications.

Early use of mixed-effects models would often rely on balanced experimental designs to allow estimation of the variance components through sums-of-squares decompositions. It was realized that maximum likelihood (ML) or restricted maximum likelihood (REML) estimators of the variance components and the fixed effects in the models could be defined but computing these estimates was just too difficult when the design was unbalanced. Now, thanks to some developments in computing techniques and to the remarkable increase in computing power available to the average researcher or consultant, we are able to use REML or ML estimation routinely. The MIXED procedure in SAS is one of the most flexible ways of defining and fitting linear mixed-effects models.

It can be surprising to see the range of statistical models or statistical analysis techniques that can be expressed as mixed-effects models. Some of the examples in Littell, Milliken, Stroup and Wolfinger (1996) include: *one-way* classification with random-effects, *blocked* designs (complete or incomplete), *split-plot* or *strip-plot* designs, *repeated-measures* data or *longitudinal* data,
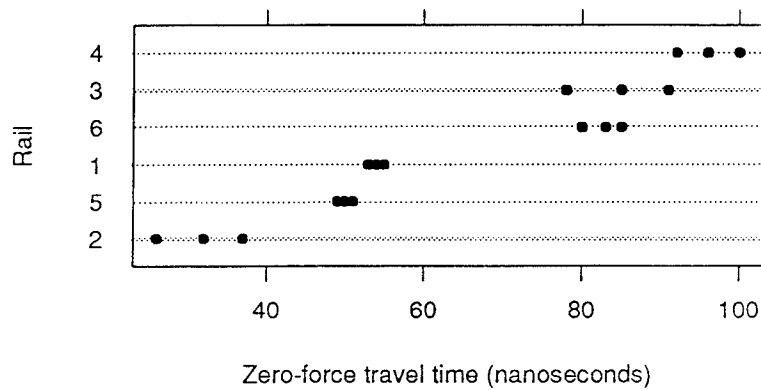
Figure 1: Travel times of ultrasonic waves in railway rails - three replications in each of six rails.

growth-curve data, *panel* data, *analysis of covariance*, *multilevel* models, and *hierarchical linear models*. That is just a sample of the models and analysis methods that can be expressed with mixed-effects models.

In §2 we illustrate some of the graphical presentation methods that can complement the analytic methods for grouped or clustered data. These graphical methods are based on the trellis graphics system developed for the S language. In §3 we examine the Laird-Ware formulation of the linear mixed-effects model and some computational methods for determining the MLE's or REML estimates. We describe some extensions to the case of nonlinear mixed-effects models in §4.

2 Graphical presentation of grouped data

Mixed-effects models are applied to data where the responses are grouped according to one or more classification factors. The simplest structure for such grouped data is a one-way classification. An example from Devore (1995, §10.3) involves non-destructive testing of railway rails for internal flaws. Six rails were selected and the travel time of a type of ultrasonic wave through the rail was recorded three times for each rail. The data are shown as a dotplot in Figure 1. Because the rails represent a sample from the population of rails to which this technique could be applied,

the deviations in the mean travel time for a particular rail are represented as a random variable $A_i, i = 1, \ldots, 6$ rather than a fixed parameter $\alpha_i, i = 1, \ldots, 6$ in the model

$$y_{ij} = \mu + A_i + \epsilon_{ij}, \quad A_i \sim \mathcal{N}(0, \sigma_A^2), \quad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \quad i = 1, \ldots, 6; \; j = 1, \ldots, 3 \quad (1)$$

The grouping of these observations is a simple structure. Each of the 18 observations has been made on one of the six rails hence the data are classified according to one level of classification. There are no other covariates associated with the response. The dotplot of Figure 1 would be a common way of graphing such data. In keeping with some general principles of trellis graphics, we have ordered the rails according to increasing mean travel time as you read from bottom to top. This enhances the comparison of between-rail to within-rail variation.

The parameters of the model (1) are μ, the overall mean travel time, and the two variance components σ^2 and σ_A^2.

The next level of complication we consider is repeated measures data with a continuous covariate. A common type of repeated measures data is longitudinal data where we observe the same subject (or, more generally, "experimental unit") over time. Hand and Crowder (1996) provide an example of the weights of 50 baby chicks followed for 20 days after hatching. The chicks are grouped into 4 treatment groups that received different dietary supplements. These data are plotted in a trellis plot in Figure 2.

Figure 2 illustrates the origin of the term "trellis plots". The plot is formed from panels laid out in a regular grid, like a garden trellis, according to some factor in the data. Furthermore, the scales on each of the panels are the same, making it easier to compare between panels as well as to look for patterns within panels. Within each treatment group the panels are ordered by increasing maximum recorded weight for the chick.

By scanning through the panels we can see that there was some mortality in the control group and that the overall growth pattern appears to be sigmoidal, although there are some exceptions.
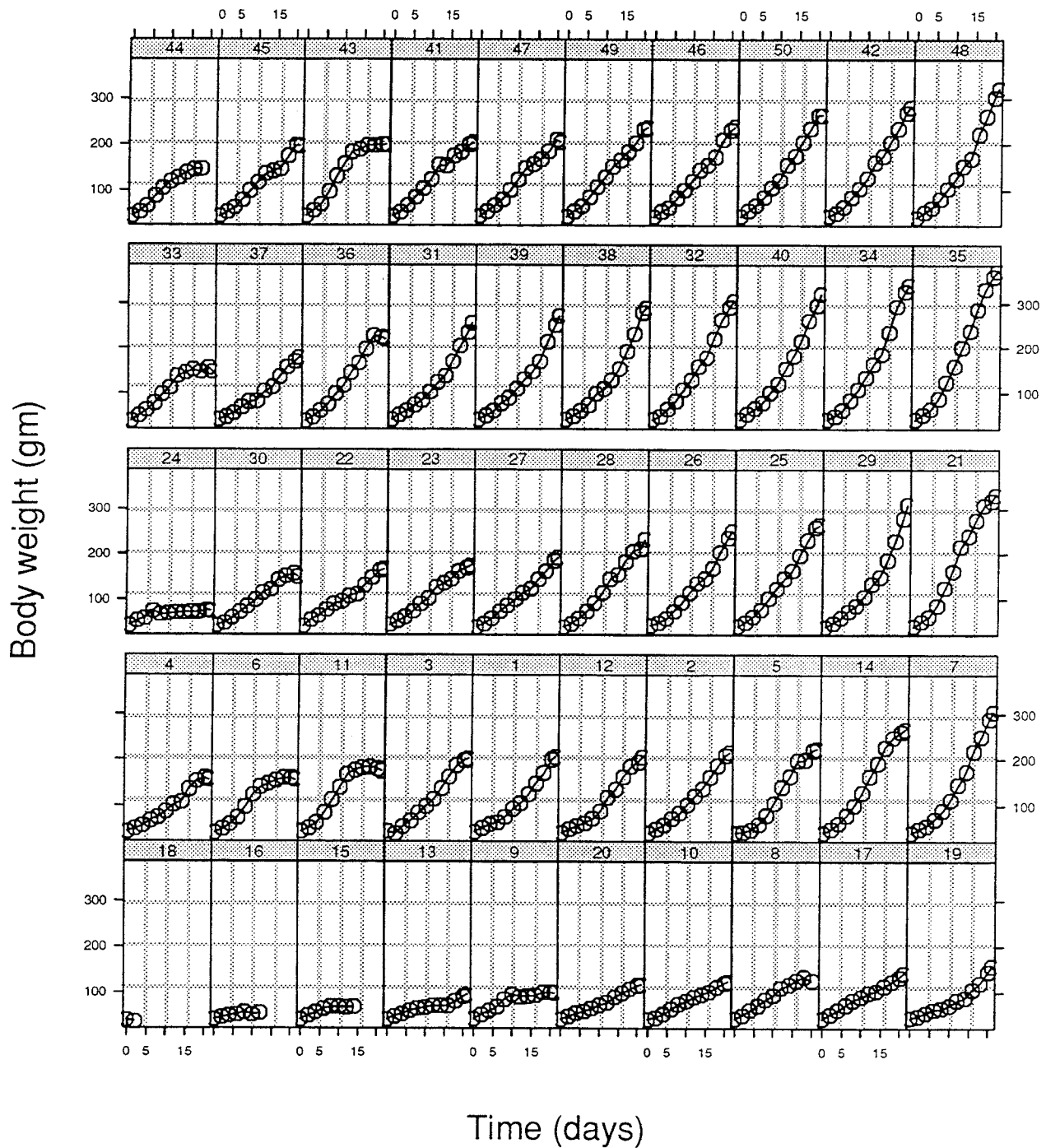
Figure 2: Growth curves for the weights of baby chicks from time of hatching. The 50 chicks are grouped into a control group of 20 chicks and three treatment groups of 10 chicks each. Different groups received different amounts of a dietary supplement.
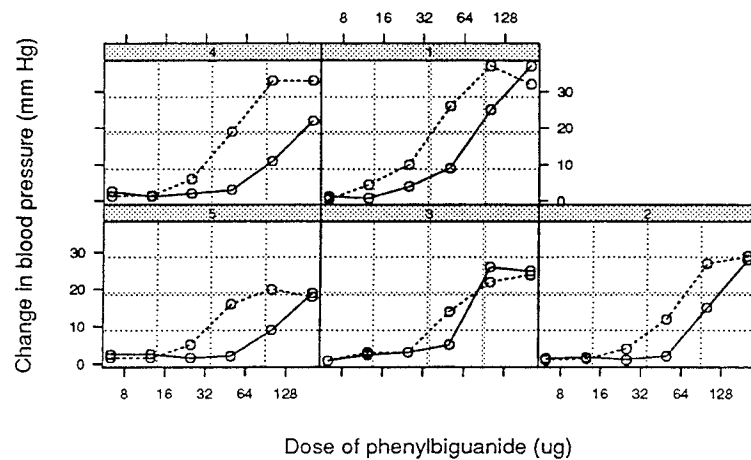
Figure 3: Change in blood pressure versus dose of phenybiguanide for five rabbits. There were two runs, one with a placebo and one with treatment with MDL 72222, for each rabbit. The responses with the placebo are joined by a dotted line; those with MDL 72222 by a solid line.

In other designs we have factors that change within a subject. Ludbrook (1994) describes an experiment where each of 5 rabbits is subjected to increasing doses of phenylbiguanide while the change in blood pressure is measured. The experiment was performed after treatment with MDL 72222 and after treatment with a placebo. The data are shown in Figure 3.

As a final example of the graphical methods, we consider the soybean growth data described in Davidian and Giltinan (1995). Two varieties, "Forrest" and a experimental variety, were planted in eight plots each during three consecutive growing seasons: 1988, 1989, and 1990. On several occasions throughout the growing season plants were selected randomly for harvest and the average leaf weight per plant was measured. This is a rather complicated experimental design but the patterns within the plots and the patterns between varieties and years can be neatly summarized in a plot such as Figure 4.

With the trellis graphics library we can easily overlay the lines on the plots to compare the years for each variety or to compare the varieties for each year (Figure 5). These trellis displays are easily produced with the trellis and nlme libraries in S-PLUS.
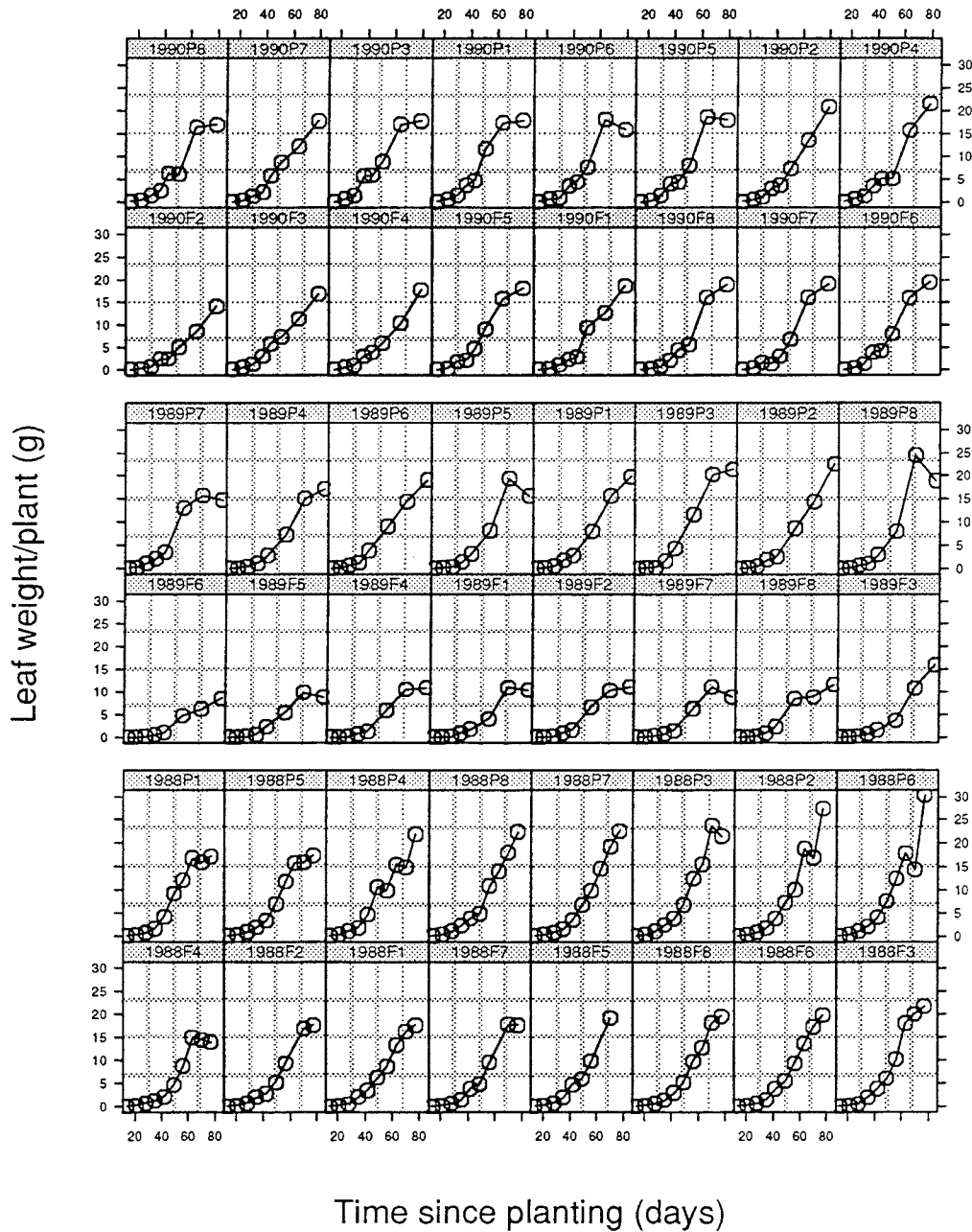
Applied Statistics in Agriculture

Figure 4: Leaf weight per plant versus time since planting for eight plots each of two varieties of soybeans in three consecutive growing seasons. The two varieties are "Forrest" (F) and "Plant Introduction #416937" (F).
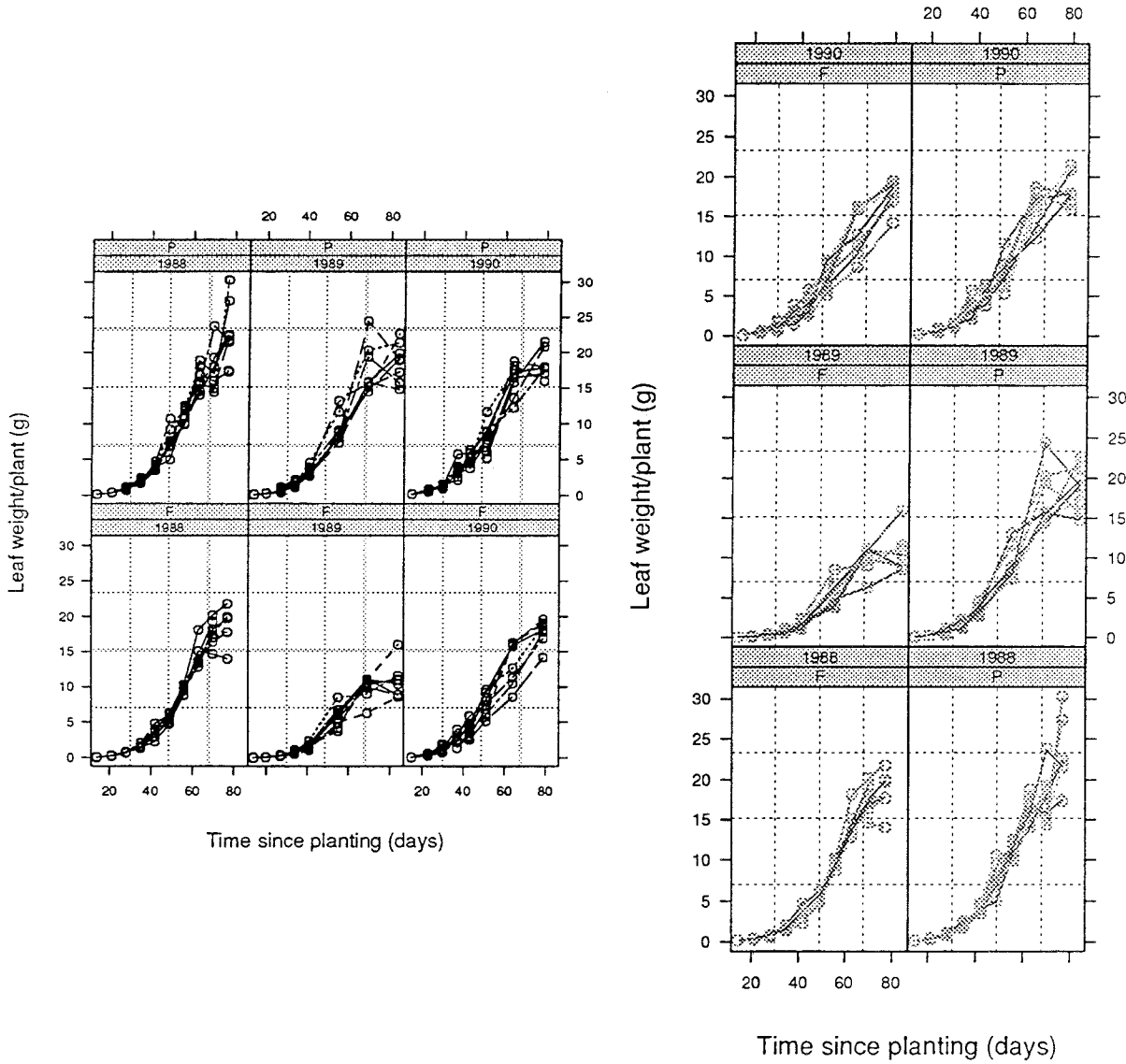
Figure 5: Leaf weight per plant versus time since planting for eight plots for the soybean data. The panels on the left are arranged to facilitate comparisons between years for each variety. The panels on the right are arranged to facilitate comparisons between varieties for each year.

Applied Statistics in Agriculture

3   Computing methods for ML and REML estimation

As described in Laird and Ware (1982) the models for the rail and the chick weight examples can be written as

$$y_i = X_i\beta + Z_ib_i + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I), \quad b_i \sim \mathcal{N}(0, \sigma^2 D) \tag{2}$$

where X_i and Z_i are design matrices for the ith group. In the rails example these design matrices are particularly simple

$$X_i = Z_i = [1\ 1\ 1]' \quad i = 1, \ldots, 6 \tag{3}$$

The assumption of $\epsilon_i \sim \mathcal{N}(0, \sigma^2 I)$ can be relaxed, as is done with the REPEATED statement in PROC MIXED. We show an example in §4. The Laird-Ware formulation can be extended to multiple nested levels of random effects as shown in Bates and Pinheiro (1998).

We have altered the Laird-Ware formulation for the linear mixed-effects model by expressing the variance-covariance matrix for the random effects as a *relative* variance $\sigma^2 D$ where σ^2 is the variance of the within-group noise ϵ_{ij} and D is a general positive-definite matrix. Just as a positive number can be expressed as the exponential of another number (its logarithm), we can express

$$D = e^A \tag{4}$$

for some general symmetric matrix A. A is called the matrix logarithm of D. If we let θ be the

non-redundant elements of A we can write the likelihood in the Laird-Ware model as

$$
\begin{aligned}
L(\beta, \theta, \sigma^2 | y) &= \prod_{i=1}^{M} p(y_i | \beta, \theta, \sigma^2) \\
&= \prod_{i=1}^{M} \int p(y_i | b_i, \beta, \sigma^2) \, p(b_i | \theta, \sigma^2) \, db_i \\
&= \prod_{i=1}^{M} \frac{1}{\sqrt{(2\pi\sigma^2)^{n_i} |D|}} \int \frac{\exp\left[\frac{-1}{2\sigma^2}\left(\|y_i - X_i\beta - Z_i b_i)\|^2 + b_i' D^{-1} b_i\right)\right]}{(2\pi\sigma^2)^{q/2}} \, db_i
\end{aligned}
\tag{5}
$$

The inverse of a variance-covariance matrix is sometimes called the precision matrix. If we take the "square-root" of the inverse of D, which we call the *relative precision factor* Δ,

$$
\Delta(\theta) = e^{-A(\theta)/2} \quad \Rightarrow \quad \Delta'\Delta = D^{-1}
\tag{6}
$$

the log-integrand in (5) becomes

$$
\|y_i - X_i\beta - Z_i b_i)\|^2 + b_i' \Delta' \Delta b_i
\tag{7}
$$

which is the residual sum-of-squares for a regression of $\tilde{y}_i$ on $\tilde{X}_i$ and $\tilde{Z}_i$ where

$$
\tilde{y}_i = \begin{bmatrix} y_i \\ 0 \end{bmatrix} \quad \tilde{X}_i = \begin{bmatrix} X_i \\ 0 \end{bmatrix} \quad \tilde{Z}_i = \begin{bmatrix} Z_i \\ \Delta \end{bmatrix}
\tag{8}
$$

This is sometimes called the "pseudo-data" representation of a penalized regression problem.

The important point of these transformations and re-expressions is that given θ, we can efficiently evaluate the log-likelihood or the restricted-log-likelihood for the profiled model. That is, for a value of θ we can easily calculate the log-likelihood at the conditionally optimal values of β and σ^2, given θ. If desired, $\widehat{\beta}(\theta)$, $\widehat{\sigma}^2$, and the BLUP's of the b_i can also be calculated as regression estimates but they do not have to be evaluated explicitly just to get the log-likelihood.

Applied Statistics in Agriculture

For a given θ the profiled log-likelihood is

$$\ell(\widehat{\beta}(\theta), \theta, \widehat{\sigma}^2(\theta)) = k - N \log \|r\| + \sum_{i=1}^{M} \log \text{abs} \frac{|\Delta|}{\sqrt{|\tilde{Z}_i' \tilde{Z}_i|}} \qquad (9)$$

where r is the residual for the penalized least-squares problem.

The log-restricted-likelihood is

$$\ell_R(\widehat{\beta}(\theta), \theta, \widehat{\sigma}^2(\theta)) = k^* - (N - p) \log \|r\| - \log \text{abs} |R| + \sum_{i=1}^{M} \log \text{abs} \frac{|\Delta|}{\sqrt{|\tilde{Z}_i' \tilde{Z}_i|}} \qquad (10)$$

where R is from the decomposition for the calculation of $\widehat{\beta}$.

It is an advantage that these methods for ML or REML estimation of mixed-models are not restricted to balanced data — we simply need to express the formulas for X_i and Z_i and the grouping factor. The method can be extended to multiple, nested grouping factors (but not to crossed grouping factors). Another point to note is that, once the Z_i are known, we can establish reasonable starting estimates $\Delta^{(0)}$ since the Δ must be comparable in size to the Z_i. This makes the process of fitting the models more automatic. Either EM or Newton-Raphson iterations can be used to optimize ℓ or ℓ_R w.r.t. θ. During the optimization it is very useful to have θ unconstrained.

The lme function in version 3.0 of nlme library for S can be used to compute these estimates. For the rails example it would look like

```
S> fm1.Rail <- lme( travel ~ 1, data = Rail, random = ~ 1 | Rail )
S> summary( fm1.Rail )
Linear mixed-effects model fit by REML
 Data: Rail
      AIC      BIC   logLik
   128.18  130.68  -61.089
Random effects: ~ 1 | Rail
         (Intercept) Residual
StdDev:       24.805   4.0208
Fixed effects: Value Std.Error DF t-value p-value
(Intercept)      66.5    10.171 12   6.538       0
S> intervals( fm1.Rail )   # Approximate 95% confidence intervals
                        lower    est.   upper
```

```
Fixed Effects (Intercept) 44.339 66.500 88.661
Random Effects (Rail) sd  10.246 24.805 60.056
Within-groups sd           2.283  4.021  7.080
```

for REML estimates. Maximum likelihood estimates can be obtained in a similar manner.

Note that the definitions of the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), also called Schwartz's Bayesian Criterion (SBC), used here are different from those used in SAS PROC MIXED. In PROC MIXED the criteria are defined as the log-likelihood minus a penalty on the number of parameters in the model. An algebraically larger value indicates a superior model. The definition employed here, from Sakamoto, Ishiguro and Kitagawa (1986) and from Schwartz (1978), uses $-2\times$log-likelihood plus a penalty so the algebraically smaller values indicates superior models. Some draft documentation for SAS PROC NLMIXED indicates that both definitions will be used and presented there.

4   Extensions to nonlinear models

The soybean growth data from Davidian and Giltinan (1995), the blood pressure data from Ludbrook (1994), and perhaps also the chick weight data from Hand and Crowder (1996), all show a sigmoidal shape with respect to time (or log(dose) in the case of the blood pressure data). It is difficult to fit such sigmoidal patterns with a model that is linear in the parameters.

In other situations the data may indicate an asymptote for the within-subject profile or another pattern that is not easily modeled with linear models. We may also have a mechanistic model for the response, such as one of the compartment models used with pharmacokinetic data. Such mechanistic models are usually nonlinear in the parameters.

For all these reasons we would seek to extend the Laird-Ware formulation to handle nonlinear model forms. We can think of this either as extending linear mixed-effects models to nonlinear forms or as extending nonlinear fixed-effects to a hierarchical model.

A hierarchical model for nonlinear mixed-effects can be written

$$\beta_i = A_i\beta + B_i b_i \qquad b_i \sim \mathcal{N}(0, \sigma^2 D)$$

$$y_{ij} = f(\beta_i, x_{ij}) + \epsilon_{ij} \qquad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \tag{11}$$

for observation j, $j = 1, \ldots, n_i$ in group i, $i = 1, \ldots, M$. In model (11) β_i includes both fixed effects β and random effects b_i. A_i and B_i are design matrices so, for example, $A_1 = [I\ 0]$ and $A_2 = [0\ I]$ would give different means for different groups.

As with linear mixed-effects models, our focus in parameter estimation is on β, σ^2 and D. However, the process also produces more precise "estimates" of b_i (and hence β_i) by "borrowing strength" from the rest of the sample from the population. The likelihood for the nonlinear mixed-effects model can be rewritten in a form similar to (5) as

$$L(\beta, \theta, \sigma^2 | y) = \prod_{i=1}^{M} \frac{1}{\sqrt{(2\pi\sigma^2)^{n_i} |D|}} \int \frac{\exp\left[\frac{-1}{2\sigma^2}\left(\|y_i - f_i(\beta, b_i)\|^2 + b_i' D^{-1} b_i\right)\right]}{(2\pi\sigma^2)^{q/2}} db_i \tag{12}$$

If all the components of β that have random effects associated with them occur linearly in $f(\beta_i, x_{ij})$ then f_i is linear in b_i and the integral in (12) can be evaluated explicitly. Otherwise we can use various techniques for evaluating the integral such those as described in Pinheiro and Bates (1995). These include: a linear mixed-effects approximation where we find the modes of β and b_i, conditional on θ, and evaluate the (log)likelihood for a linear approximation to f_i there; Laplacian integration where we find the conditional modes $\hat{b}_i(\beta, \theta, \sigma^2)$ and evaluate the log integrand and its Hessian; (adaptive) Gaussian integration where we find the conditional modes and use the Hessian of the log integrand to lay out a grid for further evaluation; and Monte Carlo approximations such as Markov-Chain Monte Carlo or importance sampling. In what follows we will be using the linear mixed-effects approximation.

To model the soybean leaf weights over time we might use a logistic growth function

$$f(\beta_i, t_{ij}) = \beta_{i1} / \left\{ 1 + \exp\left[(\beta_{i2} - t_{ij}) / \beta_{i3} \right] \right\} \tag{13}$$

where β_{i1} is the asymptotic leaf weight per plant for plot i (written Asym in what follows), β_{i2} is the time post planting when the weight reaches half of this asymptote (xmid), and β_{i3} is the time to go from (approximately) one-quarter the asymptote to one-half the asymptote (scal).

There are two *outer* factors in the data Year and Variety. These do not change within a plot so they are outer to this level of random effects. We also notice from Figure 4 that there is a greater level of variability in the responses when the responses are larger so we may need to model this heteroscedasticity.

We begin by fitting a logistic growth model to the data from each plot. To fit a nonlinear model to a single set of data in S we use nls; to fit multiple sets, we use nlsList. A *self-starting* nonlinear regression model called SSlogis has been defined for the logistic growth model to make this task easier.

```
> Soybean.lis <- nlsList( SSlogis, Soybean ) # get separate fits
> coef( Soybean.lis )          # display fitted parameters by plot
            Asym      xmid     scal
1988F4   15.1513    52.834   5.1766
1988F2   19.7455    56.575   8.4067
...
1988P2   36.6539    66.555  11.9146
1988P6  162.6227   104.818  17.9205
1989F6    8.5098    55.276   8.8573
...
1989P8        NA        NA       NA
1990P4   26.1327    61.203  10.9738
```

The algorithm did not converge for one plot (1989P8) and gave unusual estimates for another (1988P6). We can compare the parameter estimates using separate confidence intervals on these estimates by plot, as in Figure 6. Figure 6 shows systematic differences between years and between varieties in the asymptote, Asym, and perhaps systematic differences in xmid and scal. The Asym
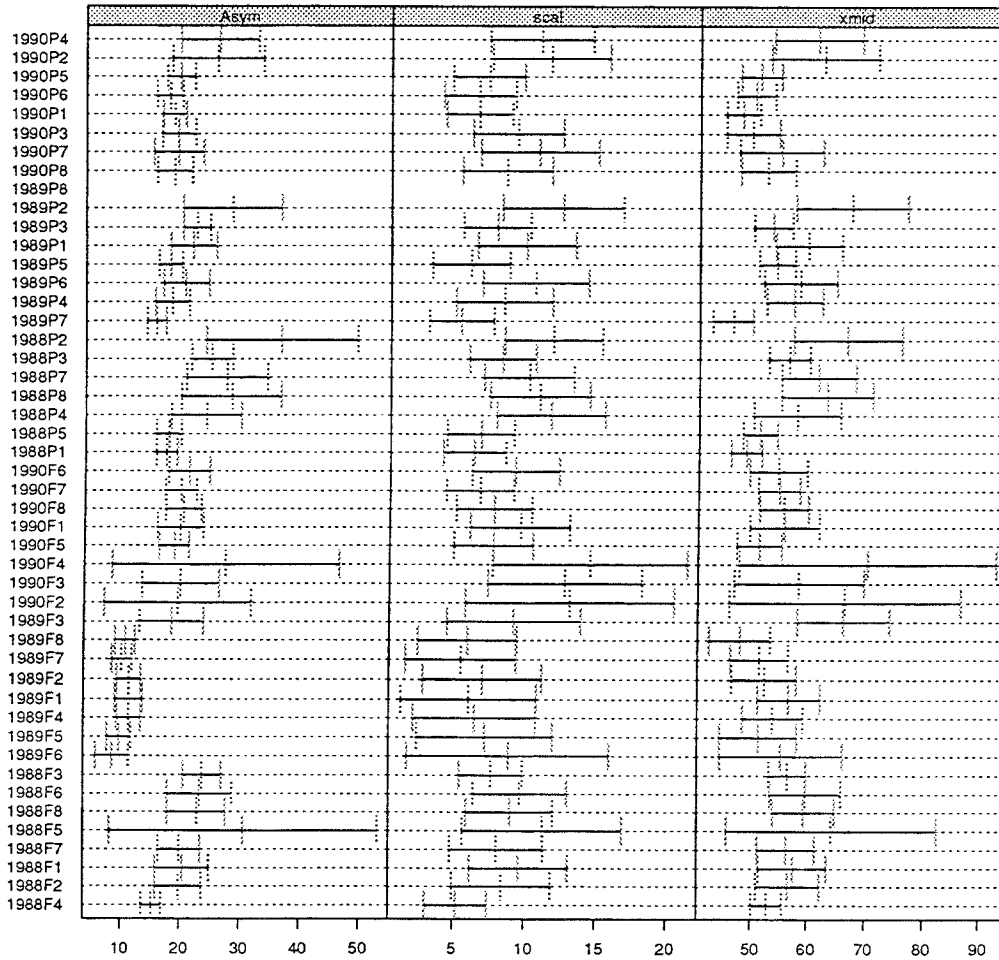
Figure 6: Individual 95% confidence intervals on the parameters of a logistic growth function fit to each soybean plot. The unusual estimates from the plot labelled 1988P6 have been omitted.
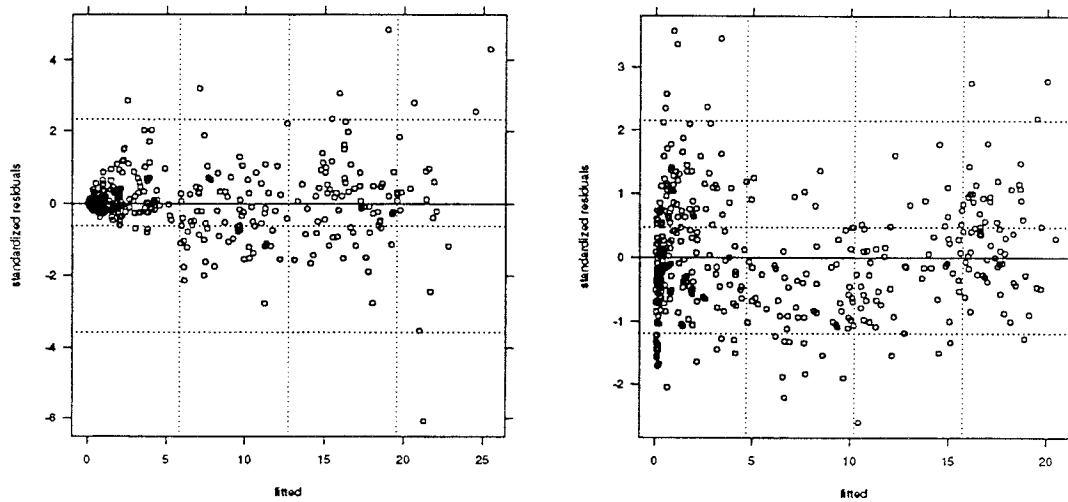
Figure 7: Standardized residuals versus fitted values for nonlinear mixed-effects fits of the logistic growth model to the soybean weight data. The panel on the left shows the residuals from the initial model fit. The panel on the right shows the weighted residuals from a weighted fit.

values seem to have substantial variability with the same year and variety. We model this with a random effect. We may also need random effects for the other parameters

An initial nonlinear mixed-effects fit to the data could ignore the systematic differences in Variety and Year and simply include random effects β_{i1}, β_{i2}, and β_{i3}. The purpose here is to obtain the conditional modes of the random effects. (These are like the BLUP's for random effects in linear mixed-effects models except they are neither linear nor unbiased and it is not certain in which they sense are "best".) We then relate the conditional modes to the covariates and modify the fixed-effects part of the model, if indicated. Something else shows up first, however. A plot of the residuals versus the fitted values (Figure 7) clearly indicates heteroscedasticity.

```
> Soybean.nlme <- nlme(Soybean.lis)  # fit nonlinear mixed-effects
> plot(Soybean.nlme)                 # produces Figure 7 (left panel)
```

We re-fit the model assuming the variance σ^2 of the within-group noise term, ϵ_{ij}, is a power of the fitted response and estimating that power. This is the type of extension to the Laird-Ware formulation we mentioned in §3

Applied Statistics in Agriculture

```
> Soybean.nlme2 <- update( Soybean.nlme, weights = varPower() )
> plot( Soybean.nlme2 )              # produces Figure 7 (right panel)
```

The residual plots show much greater stability of variance in the weighted analysis. We can

also use a likelihood ratio test to check if the extra parameter is significant.

```
> anova( Soybean.nlme, Soybean.nlme2 )
              Model df    AIC    BIC  logLik      Test Lik.Ratio p-value
Soybean.nlme      1 10 1499.2 1539.4 -739.62
Soybean.nlme2     2 11  740.0  784.2 -359.02 1 vs. 2    761.2         0
```

It is interesting to do a graphical comparison of, say, the parameter estimates for each plot in

the different fits. By comparing the original parameter estimates done separately for each plot with

those from the mixed-effects model

```
> plot(compareFits(coef(Soybean.lis), coef(Soybean.nlme2)), # Figure 8
+       subset = -16, layout = c(3,1))
```

we can see that the nonlinear mixed-effects fit produces much more homogeneous parameter esti-

mates than do the individual fits. This is in keeping with the idea of "borrowing strength" between

plots to refine the parameter estimates. However, close examination of Figure 8 will also reveal

that, especially for Asym, there is a strong effect for Year and for Variety. We also notice this

in the original data plots. This dependence should be incorporated into the fixed-effects for the

model. First we produce another plot (Figure 9) of the conditional modes of the random-effects

for the different levels of the outer factors to emphasize this dependence.

```
> plot(random.effects(Soybean.nlme2, aug = T), outer = ~ Year*Variety)
```

We can see that there are Year and Variety effects in these estimates. We proceeded to fit

several other models, using fixed effects in the expressions for Asym, xmid, and scal to account

for dependencies on year and variety. After each model fit we examine the conditional modes of the

random-effects to look for systematic dependencies on covariates. The final model includes terms

for Year and Variety and their interaction for Asym, Year and Variety but no interaction for

xmid, and Year only for scal. A plot of the original data and the predictions from this model is
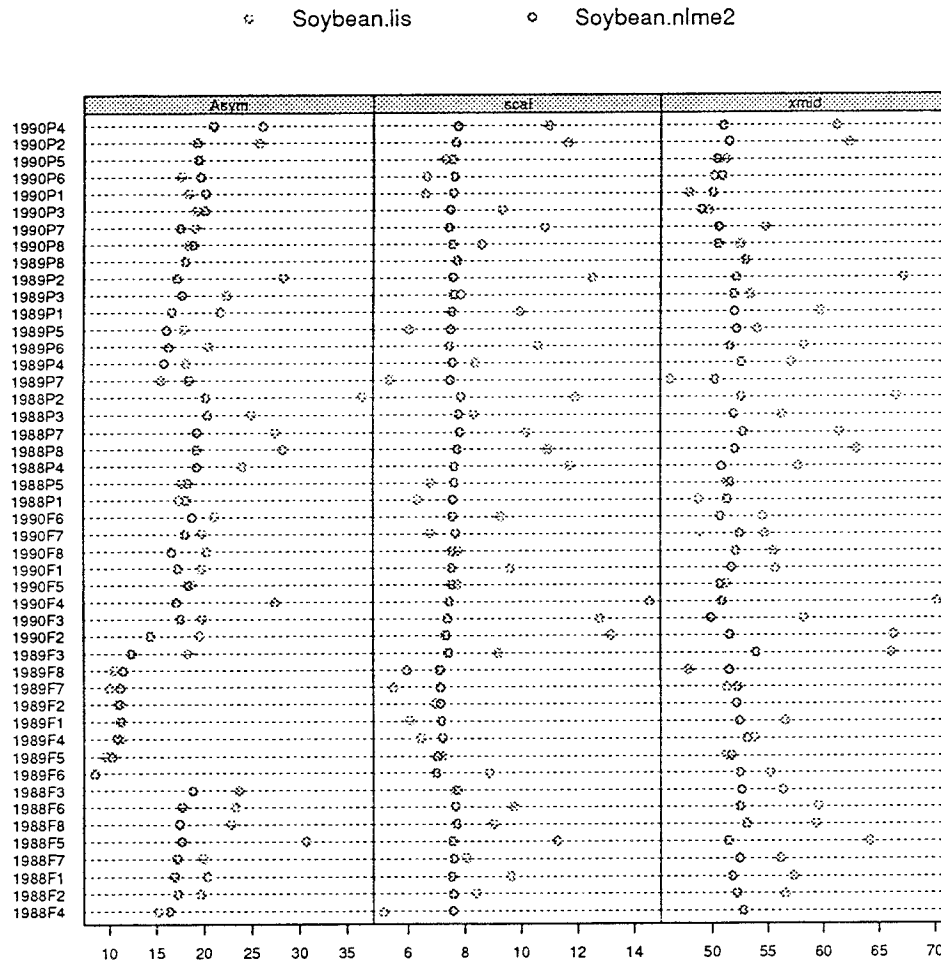
shown in Figure 10.

Figure 8: Comparison of the fitted parameters by plot from separate fits to each plot and from a weighted nonlinear mixed-effects fit of the logistic growth model to the soybean leaf weight data
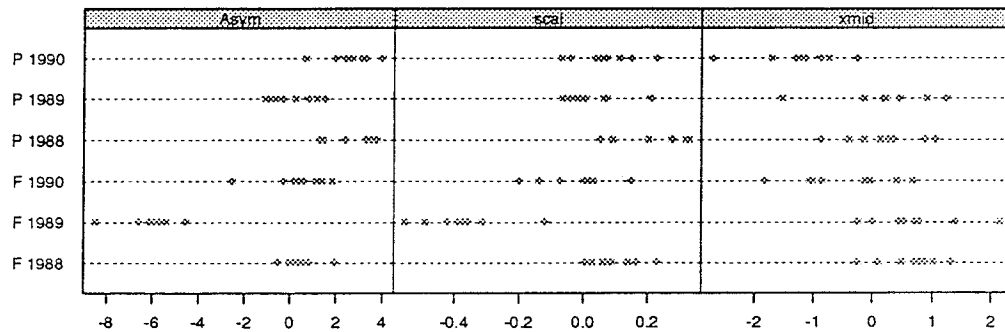
Figure 9: Conditional modes of the random effects for the parameters in a nonlinear mixed-effects fit of the logistic growth model to the soybean leaf weight data

These model fits indicate that by providing a more realistic model for the fixed-effects, incorporating the effects of Year and Variety and their interaction, we can eliminate the need for different random-effects terms. In this case the only random effect needed in the final model is that for Asym. Further exploration of residual plots did not indicate deficiencies in the model.

5   Summary

It is well-known that linear mixed-effects models provide a flexible, versatile formulation of many common statistical models. We see here that extensions to generalized linear mixed-models and nonlinear mixed-effects can provide even more versatility for modeling when required to describe the within-group behaviour of the data. It helps that efficient, reliable computational methods are now available for fitting such models by no amount of computation can replace plotting and examining the data, the parameters, the fitted models, ...
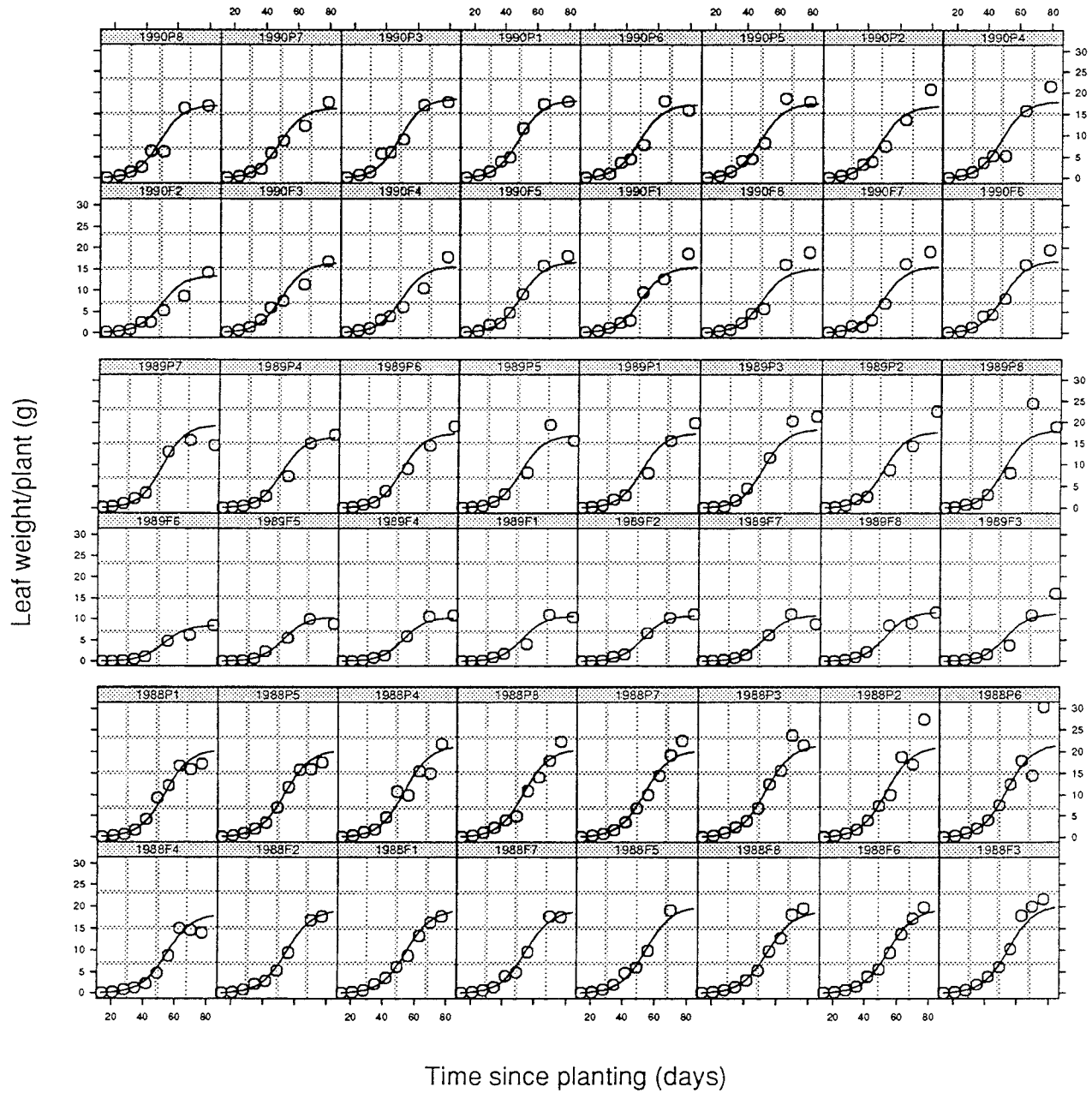
Time since planting (days)

Figure 10: Soybean leaf weight data and the predicted response curves from a nonlinear mixed-effects model.

References

Bates, D. M. and Pinheiro, J. C. (1998). Computational methods for multilevel models, *Technical Report 01-98*, Statistics and Information Analysis, Bell Laboratories, Lucent Technologies.

Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*, 1st edn, Chapman & Hall, London.

Devore, J. L. (1995). *Probability and Statistics for Engineering and the Sciences*, 4th edn, Duxbury.

Hand, D. and Crowder, M. (1996). *Practical Longitudinal Data Analysis*, Texts in Statistical Science, Chapman and Hall.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics* **38**: 963–974.

Littell, R. C., Milliken, G. A., Stroup, W. W. and Wolfinger, R. D. (1996). *SAS System for Mixed Models*, SAS Institute Inc., Cary, NC.

Ludbrook, J. (1994). Repeated measurements and multiple comparisons in cardiovascular research, *Cardiovascular Research* **28**: 303–311.

Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model, *Journal of Computational and Graphical Statistics* **4**(1): 12–35.

Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*, D. Reidel Publishing Co.

Schwartz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**: 461–464.