DEPARTMENT OF STATISTICS
University of Wisconsin
1210 West Dayton St.
Madison, WI 53706

TECHNICAL REPORT NO. 906

October 1993

NONLINEAR MIXED EFFECTS CLASSES AND METHODS FOR S

by

José Pinheiro
Douglas M. Bates
Department of Statistics
and
Mary Lindstrom
Department of Biostatistics

Nonlinear Mixed Effects Classes And Methods For S

José C. Pinheiro, Douglas M. Bates, and Mary J. Lindstrom
Department of Statistics **Dept. of Biostatistics**
University of Wisconsin — Madison

ABSTRACT

We describe extensions to the nonlinear modeling facilities in release 3 of S and S-plus. These extensions provide classes and methods for fitting and analyzing nonlinear mixed effects models with the two-stage estimation method described by Lindstrom and Bates (1990). They are implemented in a combination of S and C code and complement the classes and methods for the linear mixed effects model that we previously contributed to the *StatLib* collection. These methods include an implementation of the "loose-coupling" approach to the pseudo-data step which reduces the overall computational burden in problems with many clusters.

Keywords: Longitudinal data, overdispersion

1. INTRODUCTION

We describe a set of S classes and methods for the analysis of nonlinear mixed effects models. The source code, written in S and C, and the documentation are available in the S collection at StatLib. Details on how to obtain this and other software from StatLib can be found in Newton (1993).

In this paper we will focus on repeated measures data, i.e. data generated by observing a number of *clusters* repeatedly under different experimental conditions. Usually observations on the same cluster are made at different times, as in longitudinal studies. An example of such data is the pine tree growth data given in Kung (1986). A total of 14 sources (seeds) of Loblolly pine were planted in the southern United States and the tree heights (in ft.) were measured at 3, 5, 10, 15, 20, and 25 years of age. Figure 1 shows a plot of these data. Kung (1986) used a logistic curve to model the trees' growth, but an asymptotic regression model seems to explain the observed growth pattern better[1]. This model can be expressed as

$$f(t, \boldsymbol{\phi}) = \phi_1 - \phi_2 e^{-\phi_3 t} \tag{1.1}$$

where t denotes the tree's age, ϕ_1 the asymptotic height, ϕ_2 the difference between ϕ_1 and the height at age zero, and ϕ_3 the growth rate.

2. THE NONLINEAR MIXED EFFECTS MODEL

Repeated measures data are often analyzed using a two-stage model. The nonlinear mixed effects model described by Lindstrom and Bates (1990) is such a two-stage model that in some ways generalizes the linear mixed effects model described in Laird and Ware (1982). In the first stage the *j*th observation on the *i*th cluster is modeled as

$$y_{ij} = f(\boldsymbol{\phi}_i, \boldsymbol{x}_{ij}) + e_{ij} \tag{2.1}$$

where f is a nonlinear function of a cluster specific parameter vector $\boldsymbol{\phi}_i$ and the covariate vector $\boldsymbol{x}_{ij}$, and e_{ij} is a normally distributed noise term. In the second stage the cluster specific parameter vector is modeled as

$$\boldsymbol{\phi}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim \mathcal{N}(0, \sigma^2\boldsymbol{D}), \tag{2.2}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of fixed population parameters, $\boldsymbol{b}_i$ is a q-dimensional random effects vector associated with the *i*th cluster, $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are design matrices for the fixed and random effects respectively[2], and $\sigma^2\boldsymbol{D}$ is a (general) covariance matrix. It is further assumed that observations made on different clusters are independent and that the e_{ij} are *i.i.d.* $\mathcal{N}(0,\sigma^2)$ and independent of the $\boldsymbol{b}_i$.

Assuming that all three parameters ϕ_1, ϕ_2, and ϕ_3 in equation (1.1) have both a fixed and a random component, the nonlinear mixed effects version of this model for the pine trees is

$$y_{ij} = (\beta_1 + b_{i1}) - (\beta_2 + b_{i2})e^{-(\beta_3+b_{i3})t_j} + e_{ij}, \tag{2.3}$$

---

[1] We also tried the logistic model, the Gompertz model, the Morgan, Mercer, and Flodin model, and the Weibull type model (Ratkowsky, 1990). The asymptotic regression gave the best overall fit.

[2] In most applications $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are incidence matrices, i.e. matrices whose entries are either zero or one.

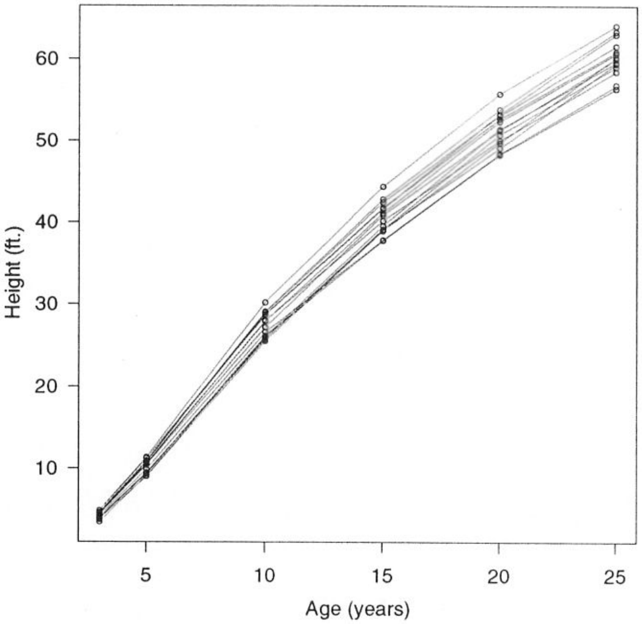

Figure 1: Loblolly pine heights at different ages.

In this case $i = 1, \ldots, 14$, $j = 1, \ldots, 6$, $\boldsymbol{A}_i = \boldsymbol{B}_i = \boldsymbol{I}$ and hence $\boldsymbol{\phi}_i = \boldsymbol{\beta} + \boldsymbol{b}_i$.

3. PARAMETER ESTIMATION

The parameters to be estimated in a nonlinear mixed effects model are the fixed effects $\boldsymbol{\beta}$, the cluster specific variance σ^2 and the scaled variance-covariance matrix of the random effects, **D**. Since **D** is a symmetric matrix, we only need $q(q+1)/2$ parameters to represent it. We will denote by $\boldsymbol{\theta}$ any such minimal set of parameters that determine **D**. In most applications the cluster specific parameters $\boldsymbol{b}_i$, estimated by the conditional modes, are also of interest.

The estimation procedure proposed by Lindstrom and Bates (1990) is an alternating algorithm. Based on the current estimate of $\boldsymbol{\theta}$, the conditional modes of the random effects, and the conditional estimates of $\boldsymbol{\beta}$ are obtained. Using these and a first order Taylor approximation to the likelihood, new estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ are determined. The estimation procedure iterates between these two estimation stages until some pre-established convergence criterion is met.

It can be shown that minimizing

$$g(\boldsymbol{\beta}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_M \mid \boldsymbol{\theta}, \boldsymbol{y}) = \sum_{i=1}^{M} \left(\|\boldsymbol{y}_i - \boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\|^2 + \boldsymbol{b}_i^T \boldsymbol{D}^{-1} \boldsymbol{b}_i \right) \tag{3.1}$$

in $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$, for given $\boldsymbol{\theta}$, produces the conditional modes of the random effects. Here $\boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)$ is the n_i-dimensional vector of predictions with

$$\{\boldsymbol{f}_i(\boldsymbol{\beta}, \boldsymbol{b}_i)\}_j = f(\boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \boldsymbol{x}_{ij}) \tag{3.2}$$

By appropriately augmenting the response vector $\boldsymbol{y}_i$ with zeroes, this minimization can be converted to a nonlinear least squares problem. This estimation step is called the *pseudo-data nonlinear least squares (PNLS)* step. Details may be found in Lindstrom and Bates (1990).

In order to estimate $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and σ^2 we should ideally use the likelihood

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 \mid \boldsymbol{y}) = \int \ell(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{b})\, p(\boldsymbol{b})\, d\boldsymbol{b} \tag{3.3}$$

where $p(\boldsymbol{b})$ is the multivariate normal density of $\boldsymbol{b}$. In general this integral does not have a closed-form expression when the expectation function f is nonlinear in $\boldsymbol{b}$. To avoid this difficult integral Lindstrom and Bates (1990) considered a first order Taylor series expansion of f around the current estimate of $\boldsymbol{\beta}$ and the conditional modes $\boldsymbol{b}_i$, obtaining an approximate log-likelihood of

$$\begin{aligned} \tilde{\ell}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 \mid \boldsymbol{y}) = {} & \\ -\frac{1}{2} \sum_{i=1}^{M} \Bigg(& \log \left|\sigma^2 (\boldsymbol{I} + \widehat{\boldsymbol{Z}}_i^{(w)} \boldsymbol{D} \widehat{\boldsymbol{Z}}_i^{(w)T})\right| \\ & + \sigma^{-2} [\widehat{\boldsymbol{w}}_i^{(w)} - \widehat{\boldsymbol{X}}_i^{(w)} \boldsymbol{\beta}]^T (\boldsymbol{I} + \widehat{\boldsymbol{Z}}_i^{(w)} \boldsymbol{D} \widehat{\boldsymbol{Z}}_i^{(w)T})^{-1} \\ & [\widehat{\boldsymbol{w}}_i^{(w)} - \widehat{\boldsymbol{X}}_i^{(w)} \boldsymbol{\beta}] \Bigg), \end{aligned} \tag{3.4}$$

with the *(w)* superscript indicating quantities based on the current estimates of $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$, $||$ indicating the determinant of a matrix,

$$\begin{aligned} \widehat{\boldsymbol{Z}}_i &= \left.\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{b}_i^T}\right|_{\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}}, \quad \widehat{\boldsymbol{X}}_i = \left.\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{\beta}^T}\right|_{\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}}, \quad \text{and} \\ \widehat{\boldsymbol{w}}_i^{(w)} &= \boldsymbol{y}_i - \widehat{\boldsymbol{f}}_i^{(w)} + \widehat{\boldsymbol{X}}_i^{(w)} \widehat{\boldsymbol{\beta}}^{(w)} + \widehat{\boldsymbol{Z}}_i^{(w)} \widehat{\boldsymbol{b}}_i^{(w)} \end{aligned}$$

This approximate log-likelihood is the log-likelihood of a linear mixed effects model where $\widehat{\boldsymbol{w}}^{(w)}$ is the vector of observations, $\boldsymbol{\beta}$ represents the fixed effects, and $\widehat{\boldsymbol{X}}^{(w)}$ and $\widehat{\boldsymbol{Z}}^{(w)}$ are the fixed and random effects design matrices as described in Lindstrom and Bates (1988). The second estimation step is therefore called the *linear mixed effects (LME)* step.

Alternatively the approximate residual log-likelihood (Harville, 1974) can replace the log-likelihood in the *LME* step. This residual log-likelihood is given by

$$\begin{aligned} \tilde{\ell}_R(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}) = {} & \\ & \tilde{\ell}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2 | \boldsymbol{y}) \\ & -\frac{1}{2} \sum_{i=1}^{M} \log \left|\widehat{\boldsymbol{X}}_i^{(w)} (\boldsymbol{I} + \widehat{\boldsymbol{Z}}_i^{(w)} \boldsymbol{D} \widehat{\boldsymbol{Z}}_i^{(w)T})^{-1} \widehat{\boldsymbol{X}}_i^{(w)T}\right| \end{aligned} \tag{3.5}$$

4. THE NLME CLASS IN S AND RELATED METHODS

Based on the estimation approach described in the previous section an integrated set of S (Becker, Chambers and Wilks, 1988) functions was developed to fit and analyze nonlinear mixed effects models. These use the object-oriented methodology introduced in Chambers and Hastie (1992). In this section we illustrate the use of these functions by analyzing the Loblolly pine tree data. Details on all the available options and the syntax of the arguments are in the *help* files.

The main function is `nlme()`. It fits a general nonlinear mixed effects model, as defined in (1.1), using the two-stage algorithm described in section 3. In the *PNLS* step we make use of the loosely coupled structure of the nonlinear least squares minimization problem, as described in Soo and Bates (1992). This allows the complexity of the calculations involved to increase linearly with the number of clusters M, instead of growing with M^2 as would be the case if standard nonlinear least squares techniques were used.

The `nlme()` function generates an object to which is assigned the class `nlme`. Methods for standard generic functions such as `print()`, `summary()`, and `plot()` are defined for this class.

Let us consider the nonlinear mixed effects model for the Loblolly pine data in which all the parameters are assumed to have both fixed and random effects, as given in (2.1). As in most nonlinear methods, we require starting estimates for some of the parameters. The description of the parameters following equation (1.1) and the data in Figure 1 indicate that, for these trees, $\phi_1 \simeq 95$, $\phi_2 \simeq 100$, and $\phi_3 \simeq 0.044$. Because the patterns for different trees are similar, we use these starting estimates for $\boldsymbol{\beta}$ and set all $\boldsymbol{b}_i = 0$ initially. The data are stored in a data frame called `pinetree` with columns `height`, `age`, and `seed`. A simple call to `nlme()` to fit model (2.1) would be

```
> pine.fit <-
+  nlme(model = height ~ A - B*exp(-C * age),
+    fixed = list(A ~ ., B ~ ., C ~ .),
+    random = list(A ~ ., B ~ ., C ~ .),
+    cluster =  ~ seed, data = pinetree,
+    start = list(fixed = c(95, 100, 0.0443)))
```

The first argument is a nonlinear model formula describing the response and the nonlinear model, *f*.

The second and third arguments, `fixed` and `random`, are lists of formulas that define the fixed and random effects, respectively. In these formulas a "." on the right hand side of a formula indicates that a single parameter is associated with the effect, but any linear formula in S could be used instead[3]. That gives considerable flexibility to the model, as time-dependent parameters can be easily incorporated (e.g. when a formula in the `fixed` list involves a covariate that changes with time). Usually every parameter in the model will have an associated fixed effect, but it may, or may not, have an associated random effect. Since we assumed that all random effects have mean zero, the inclusion of a random effect without a corresponding fixed effect would be unusual.

The `cluster` argument defines the cluster of each observation. An S expression or a formula with no left hand side can be used here. `Data` names a data frame and `start` provides a list of starting values for the iterative algorithm. Only the fixed effects starting estimates are required. The default value for the random effects is zero and starting estimates for $\boldsymbol{\theta}$ and σ^2 are automatically generated using a formula given in Laird, Lange and Stram (1987) if they are not supplied.

Further information on the arguments of `nlme()` is available in the help file.

A brief description of the estimation results can be obtained through the `print` method of the `nlme` class. This method only gives the estimates for the variance/covariance matrix of the random effects ($\sigma^2\boldsymbol{D}$), the cluster specific variance (σ^2), and the fixed effects ($\boldsymbol{\beta}$).

```
> pine.fit
Call:
  Model: height ~ A - B * exp( - C * age)
  Fixed: list(A ~ ., B ~ ., C ~ .)
 Random: list(A ~ ., B ~ ., C ~ .)
Cluster:  ~ seed
   Data: pinetree

Variance Components Estimates:
  Variance/Covariance/Corr. of Random Effects
            A            B             C
A   52.5813674   51.6526300 -2.310920e-02
B    1.0000000   50.7402967 -2.270114e-02
C   -0.8842338   -0.8842381  1.298984e-05

  Cluster-specific Residual Variance: 0.3965168

Fixed Effects Estimates:
       A         B          C
  102.258  110.845 0.03916749

Number of Observations: 84
Number of Clusters: 14
```

The most interesting results at this point in the analysis are the very strong positive correlation between b_1 and b_2 and the small variance of b_3. Also of interest is the negative correlation between b_3 and the other two random effects.

A more complete description of the estimation results is obtained with `summary`

```
> summary(pine.fit)
  . . .
```

---

[3] Note that the formulas could be directly incorporated in the model declaration. The approach used in `nlme()` allows for more efficient calculation of derivatives and will be useful for `update` methods that will be incorporated in the code in the future.

```
Convergence at iteration: 7
Loglikelihood: -34.61809

Variance Components Estimates:
  Variance/Covariance/Corr. of Random Effects
             A           B             C
A   52.5813674   51.6526300 -2.310920e-02
B    1.0000000   50.7402967 -2.270114e-02
C   -0.8842338   -0.8842381  1.298984e-05

  Cluster-specific Residual Variance: 0.3965168

Fixed Effects Estimates:
          Value Std.Error(C) z ratio(C)
A 102.25796518  2.711028378   37.71925
B 110.84499565  2.556649810   43.35556
C   0.03916749  0.001464481   26.74496

  Conditional Correlations of Fixed Effects
          A          B
B  0.9965924
C -0.9360038 -0.9213363
  .  .  .
```

We can see that the fixed effects are being estimated reasonably precisely (the approximate standard errors are only of the order of 2% of the estimates' values), but their estimators are highly correlated, especially $\widehat{\beta_1}$ and $\widehat{\beta_2}$.

As mentioned above b_1 and b_2 show a very high positive correlation and b_3 does not vary substantially among seeds. This gives some indication that perhaps some of the parameters (or linear combinations of them) could be considered as purely fixed, i.e. without an associated random effect. We will return to this issue below.

Plots of random effects, residuals, and fitted values can be obtained using the `plot` method of class `nlme`. The following call will produce a scatter plot matrix of the random effects, as shown in Figure 2.

```
> plot(pine.fit)
```

This plots shows clearly the almost perfect linear relationship between the random effects for A and B. It also indicates the negative correlation between the last random effect and the first two.

The same `plot` method can be used to generate diagnostic plots of the residuals and the fitted values which can be used to assess the quality of the fit. For example

```
> plot(pine.fit, option = 'r')
```

produces the plots in Figure 3.

The first plot, observed versus fitted values, indicates that the asymptotic regression model does a reasonable job of explaining the Loblolly pine trees growth. The points fall very close to the displayed $y = x$ line, indicating a strong agreement between the observed and fitted values. The second plot shows the major flaw of the model: all the residuals at the earliest age are positive, while all those at the second measurement are negative. Nevertheless the residuals are all relatively small in magnitude and the underfitting/overfitting problem is not present for later ages, which are the ones of most interest in this type of study. The final plot, featuring the boxplots of the residuals by cluster, suggests that the dispersion is homogeneous over clusters.

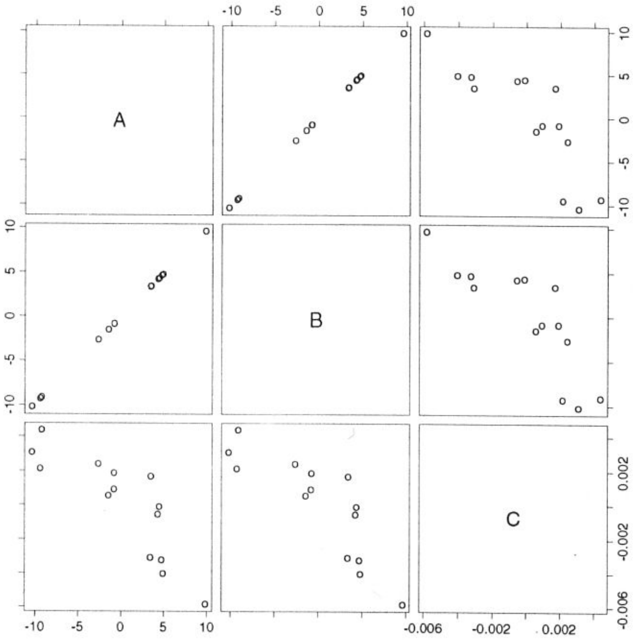

Figure 2: Scatter Plot Matrix of Random Effects.

One of the major questions of interest for these data is predicting the tree's height for later ages. The `predict` method for the `nlme` class can be used for this. As an example we will use it to predict the heights for the first seed up to 50 years of age. First we create a new data frame containing information on the seed number and the ages for which the predictions are desired as

```
> new.pinetree <- data.frame(seed=rep(301,6),
+   age = seq(25, 50, len = 6))
```

The predictions are then obtained with

```
> predict(pine.fit, ~ seed, new.pinetree)
  cluster population.average cluster.specific
1     301           60.62278         61.30597
2     301           68.02777         68.48960
3     301           74.11575         74.32499
4     301           79.12096         79.06519
5     301           83.23597         82.91573
6     301           86.61911         86.04361
```

where the `population.average` column refers to the pre-

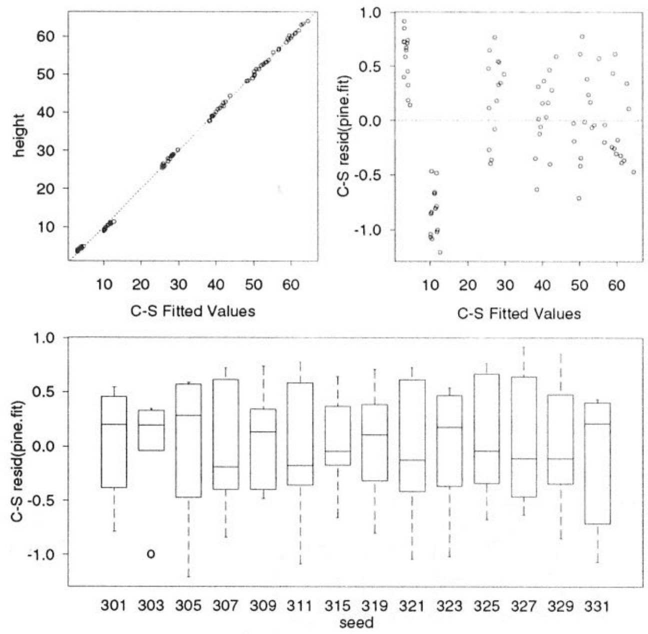

Figure 3: Residuals and Fitted Values Plots

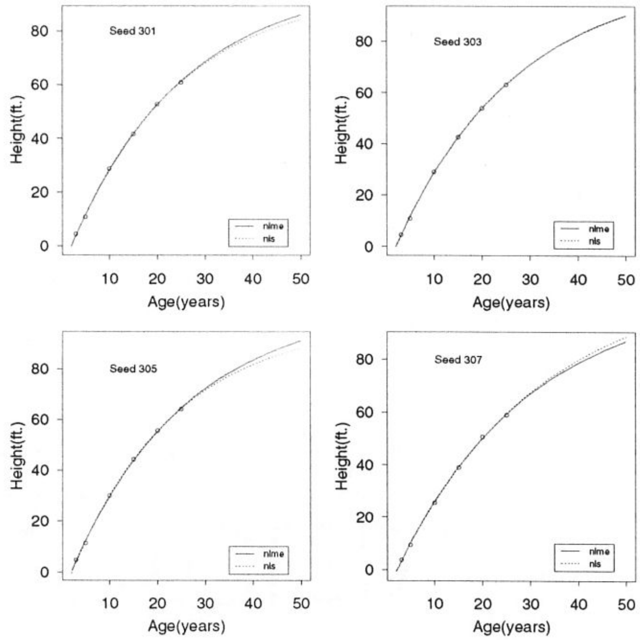

Figure 4: Fitted Curves for First Four Seeds

dictions when the random effects are set to zero and the `cluster.specific` uses the conditional modes for seed 301. It is interesting to note that these predictions are consistent with what is observed in practice since, as mentioned in Kung (1986), Loblolly pine trees on a good site can achieve heights of 90 feet or more by 50 years of age.

The `predict` method can also be used for plotting smooth fitted curves by calculating fitted values at closely spaced times . Figure 4 presents the fitted curves for the first four seeds using a total of 200 time points between 2 and 50 years. For the sake of comparison we also include the fitted curves corresponding to the individual nonlinear least squares fits[4]. The two fits almost coincide, especially for the earlier ages.

One interesting question in the model building process is how to decide whether a parameter should be fixed for all clusters, or should include a random component. The literature does not give a definite answer, especially for nonlinear models. One possibility is to analyze the eigenstructure of the estimated variance-covariance matrix of the random effects ($\boldsymbol{D}$), checking for eigenvalues close to zero and use the approximate log-likelihood of section 2 to create likelihood ratio tests for evaluating whether particular combinations of the random effects could be eliminated from the model. The validity of such analysis will, of course, depend on the accuracy of the approximation of the log-likelihood.

For the pine tree data the eigenvalues of $\widehat{\boldsymbol{D}}$ are 262.21, 7.82×10^{-6}, and 7.13×10^{-6}, suggesting that the model is overparametrized and possibly just one random effect is needed.

The corresponding orthonormalized eigenvectors are

$$\begin{aligned} \boldsymbol{u}_1^T &= (-0.7134, -0.7007, 0.0003) \\ \boldsymbol{u}_2^T &= (-0.6925, 0.7050, -0.1533) \\ \boldsymbol{u}_3^T &= (-0.1071, 0.1096, 0.9882) \end{aligned} \tag{4.1}$$

The vectors $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ indicate that $b_1 - b_2$ and/or b_3 could be eliminated from the model. We decided to go one step at a time and first eliminate only b_3. The corresponding call to `nlme()` and (pieces of) the associated `summary` are given below.

```
> pine.fit.AB <- nlme(height ~ A-B*exp(-C*age),
+   list(A ~ ., B ~ ., C ~ .),list(A ~ ., B~ .),
+   ~seed,pinetree,list(fixed=c(95,100,0.0443)))
> summary(pine.fit.AB)
  . . .
Convergence at iteration: 3
Loglikelihood: -37.53919

Variance Components Estimates:
  Variance/Covariance/Corr. of Random Effects
          A        B
A 13.1269677 13.03844
B  0.9999386 12.95210


  Cluster-specific Residual Variance: 0.5161905
```

---

[4] The `nls()` function was used to fit the model and the fitted values were obtained through the `predict` method of the `nls` class.

```
Fixed Effects Estimates:
          Value Std.Error(C) z ratio(C)
A 101.47369856  2.414937496   42.01918
B 110.09821064  2.203230618   49.97126
C   0.03939487  0.001325846   29.71300

  Conditional Correlations of Fixed Effects
          A          B
B  0.9950441
C -0.9089410 -0.8791107
 . . .
```

The resulting approximate log-likelihood is -37.54, which, when compared with the value of -34.62 for the full model, gives a likelihood ratio statistic of 5.84. Since the removal of b_3 from the full model implies three fewer parameters in the reduced model factorization of **D**, we compare the likelihood ratio statistic to a χ^2 with 3 degrees of freedom. The corresponding p-value is 0.12, indicating that the random effect could be dropped from the model. We observe a considerable decrease in the variances of the remaining two random effects, when compared to the values obtained for the full model. The correlation between the two random effects continues to be extremely high. The fixed effects estimates do not change much, nor do their approximate standard errors.

The natural question now is whether we can drop other random effects from the model. We tried removing each of the random effects separately, but in both cases the log-likelihood ratio statistic turned out to be highly significant. Table 1 summarizes the findings.

Table 1: Likelihood Ratio Statistics for Random Effects

Effect Removed	Approximate Log-likelihood	Likelihood Ratio	p-value
b_1	-76.53	86.72	< 0.0001
b_2	-61.01	52.78	< 0.0001

The eigenvalues of the estimated $\boldsymbol{D}$ matrix in the reduced model are 26.08 and 0.0008, giving a clear indication that the model is still overparametrized. The eigenvector corresponding to the smallest eigenvalue is $\boldsymbol{u_2}^T = (0.705, -0.709)$ suggesting that $b_1 - b_2$ should be eliminated. That led us to consider the following reparametrization of the original asymptotic regression model

$$f(t, \phi) = \phi_1 + (\phi_2 - \phi_1)\, e^{-\phi_3 t} \qquad (4.2)$$

where ϕ_1 and ϕ_3 continue to have the same interpretation as before, but ϕ_2 now represents the height at age zero. Using this reformulation of the model we get the values shown in table 2 for the log-likelihood where the likelihood ratio statistics are calculated with respect to the full model (i.e. with all three effects as random). We conclude that the reparametrized model with just one random effect for the asymptotic height is adequate to fit the data. In order to find out about the cluster specific parameter estimates for this final fit the `cs.coef` method can be used. Note that as height at age zero and growth rate do not have an associated random effect, their respective estimates do not vary with cluster. Therefore we extract only the cluster specific asymptotic heights.

Table 2: Likelihood Ratio Statistics for Random Effects in the Reparametrized model

Random Effects	Approximate Log-likelihood	Likelihood Ratio	p-value
none	-85.65	102.08	<0.0001
b_1	-37.55	5.88	0.318
b_1 and b_2	-37.54	5.86	0.119
b_1, b_2, and b_3	-34.61		

```
> pine.fit.noBC <-
+   nlme(height ~ A + (B - A) * exp(-C * age),
+   list(A ~ ., B ~ ., C ~ .),list(A ~ .),~seed,
+   pinetree,list(fixed=c(A=100,B=-10,C=0.039)))
> cs.coef(pine.fit.noBC)[, 1]
      301      303      305      307      309
 103.4589 106.0564 108.5432 99.08856 105.036

      311      315      319      321      323
 100.0454 101.3371 102.6419 99.08609 104.4332

      325      327      329      331
 99.75504 96.4305 95.8819 98.46592
```

These values are in agreement with the expected behavior of Loblolly pine trees at older ages, as reported in Kung (1986). An interesting aspect of this example is that neither the random effect for ϕ_1 alone nor that for ϕ_2 alone can be dropped from the model, but a linear combination of them can. The example also illustrates how the eigenstructure of the estimated $\boldsymbol{D}$ matrix can help in the model building process.

5. CONCLUSIONS

The classes and methods described here provide tools for analyzing nonlinear mixed effects models. As they are defined within the S environment, all the powerful analytical and graphical machinery present in S is simultaneously available. The analysis of the Loblolly pine tree illustrates some of the available features, but many other features are available.

Some questions remain open to future investigation. The first one concerns the validity of the approximation to the marginal log-likelihood described in section 3. All estimation results depend heavily on the accuracy of that approximation, and further study is necessary to evaluate its properties. Numerical techniques , such as Gaussian quadrature (Smith, Skene, Shaw, Naylor and Dransfield, 1985; Davis and Rabinowitz, 1984), or Monte Carlo methods, such as the Metropolis algorithm (Metropolis, Rosenbluth, Rosenbluth and Teller, 1953) or importance sampling (Rubinstein, 1981), could be used to evaluate the marginal log-likelihood and compare the results with the approximation. We will consider replacing the approximation by more exact methods, such as these, in future implementations of the code. The estimation method described in section 3 would be used to obtain *good* initial estimates for the final iterative process, in which the approximation would be replaced by the *exact* log-likelihood. The software design allows for that sort of upgrading without major modifications in the code.

Another issue that deserves a more thorough treatment is that of choosing which parameters should be random effects and which purely fixed effects. Comparison of nested models can certainly be done using a likelihood ratio approach, as illustrated in the Loblolly pine tree example. For that matter, the use of an *exact* log-likelihood would certainly give more accurate results. One of the problems with this approach is deciding which way to construct the nesting; from smaller to larger models, or the other way around. Starting with a model where all parameters are random effects and then removing unnecessary terms sounds tempting, but has the major disadvantage that convergence may not be achieved if the model is overparametrized. In these cases the **D** matrix may become seriously ill-conditioned, making it difficult or impossible to converge. The smaller to larger approach seems preferable, but has the disadvantage of the large number of models that may have to be fitted before the desired one is found. There is yet another important aspect that is overlooked by the model nesting approach. Sometimes, as in the Loblolly pine tree example, a linear combination of random effects can be considered fixed, but none of the individual terms can. A good strategy to detect random effects overparametrization is to examine the eigenvalues of the estimated **D** matrix and see when one, or more, are close enough to zero. The associated eigenvector(s) would then give an estimate of the linear combination of the random effects that could be taken as fixed. If more than one eigenvalue approach zero, rotations of the corresponding eigenvectors could be used to give more meaningful random effects combinations, similarly to what is done in factor analysis. Of course, the question that remains open is what *close enough* means. More research is needed on this.

REFERENCES

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics*, Wadsworth, Belmont, CA.

Chambers, J. M. and Hastie, T. J. (eds) (1992). *Statistical Models in S*, Wadsworth, Belmont, CA.

Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*, second edn, Academic Press, New York.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts, *Biometrika* **61**: 383–385.

Kung, F. H. (1986). Fitting logistic growth curve with predetermined carrying capacity, *ASA Proceedings of the Statistical Computing Sect.* pp. 340–343.

Laird, N., Lange, N. and Stram, D. (1987). Maximum likelihood computations with repeated measures: Application of the EM algorithm, *Journal of the American Statistical Association* **82**: 97–105.

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

Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *Journal of the American Statistical Association* **83**: 1014–1022.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data, *Biometrics* **46**: 673–687.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, E. (1953). Equation of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1092.

Newton, H. J. (1993). New developments in statistical computing, *American Statistician* **47**: 146–147.

Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models*, Marcel Dekker, New York.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*, Wiley, New York.

Smith, A. F. M., Skene, A. M., Shaw, J. E. H., Naylor, J. C. and Dransfield, M. (1985). The implementation of the Bayesian paradigm, *Communications in Stat., Part A–Th. and Meth.* **14**: 1079–1102.

Soo, Y.-W. and Bates, D. M. (1992). Loosely coupled nonlinear least squares, *Computational Statistics and Data Analysis* **14**: 249–259.