THE UNIVERSITY OF ADELAIDE

Department of Statistics

The University of Adelaide

Department of Statistics

Research Report Series

The analysis of designed experiments and longitudinal data using smoothing splines

by

A.P. Verbyla, B.R. Cullis, M.G. Kenward and S.J. Welham

Research Report 97/4

August 1997

email addresses:
averbyla@stats.adelaide.edu.au
cullisb@agric.nsw.gov.au
M.G. Kenward@ukc.ac.uk
sue.welham@bbsrc.ac.uk

Adelaide, South Australia 5005
Tel: +61 8 8303 5418 Fax: +61 8 8303 3696
averbyla@stats.adelaide.edu.au

The analysis of designed experiments and longitudinal data using smoothing splines

By ARŪNAS P. VERBYLA[†],
The University of Adelaide, South Australia
BRIAN R. CULLIS,
NSW Agriculture, Agricultural Research Institute, Wagga Wagga, Australia
MICHAEL G. KENWARD
University of Kent at Canterbury, U.K.
and SUE J. WELHAM
IACR-Rothamsted, Harpenden, U.K.

SUMMARY

Smoothing splines and other non-parametric smoothing methods are well accepted for exploratory data analysis. These methods have been used in regression, in repeated measures or longitudinal data analysis, and in generalized linear models. However, a major drawback is the lack of a formal inferential framework. An exception which has not been fully exploited is the cubic smoothing spline. The cubic smoothing spline admits a mixed model formulation, which places this non-parametric smoother firmly in a parametric setting. The formulation presented in this paper provides the mechanism for including cubic smoothing splines in models for the analysis of designed experiments and longitudinal data. Thus nonlinear curves can be included with random effects and random coefficients, and this leads to very flexible and informative modelling within the linear mixed model framework. Variance heterogeneity can also be accommodated. The advantage of using the cubic smoothing spline in the case of longitudinal data is particularly pronounced, because covariance modelling is achieved implicitly as for random coefficient models. Several examples are considered to illustrate the ideas.

Keywords: Analysis of variance; Best linear unbiased prediction; Cubic smoothing splines; Longitudinal data; Mixed models; Random coefficient models; Repeated measures; Residual maximum likelihood.

[†]*Address for correspondence*: Department of Statistics, The University of Adelaide, Australia, 5005. E-mail: averbyla@stats.adelaide.edu.au

1 Introduction

There are many situations where the relationship of a response variable as a function of a quantitative variable, for example time, is of importance. In the absence of substantive knowledge which might provide a mechanistic model for the relationship, empirical model building will be required. The presence of other variables or factors may introduce some systematic changes in the relationship. For example, if the relationship can be approximated by a simple linear regression, the intercept and slope may vary between groups or subjects. These coefficients might be fixed or taken as random depending on the context. The extension of this idea to more complex settings in designed experiments is immediate, and the decomposition of effects into one degree of freedom contrasts using orthogonal polynomials is well known in the analysis of variance. If however, the relationship is nonlinear and cannot be approximated by polynomials, difficulties arise.

An important area where modelling is required is the analysis of repeated measures or longitudinal data. Often the impact of treatments is of interest on the relationship between the response and the quantitative variable, time. There is a large literature in this area and no attempt will be made to provide a comprehensive review here. Approaches range from the use of analysis of variance where a split plot analysis is performed with the sub-plot 'treatment' being time (see for example Diggle et al., 1994, section 6.4), the analysis of polynomial contrasts using analysis of variance (Wishart, 1938, Rowell & Walters, 1976), and multivariate methods (Cole & Grizzle, 1966, Grizzle & Allen, 1969). More recently, a joint modelling of mean response profiles and the underlying covariance structure has been adopted (Diggle, 1988, Cullis & McGilchrist, 1990, Verbyla & Cullis, 1992, Diggle et al., 1994), with the mean usually modelled using a linear model. There have been attempts to find flexible low-order parameterisations for the covariance structure (Kenward, 1987).

The joint modelling approach has not been adopted by practitioners in routine analysis of longitudinal data. A major problem is that the approach requires plausible models for both mean response profiles and for the covariance structure and this can be very difficult. As Diggle et al. (1994), p. 102, state in relation to the experiment to be introduced in section 2.1,

> ''Fitting the mean response profiles into the framework of the general linear model proves to be difficult... Low degree polynomials provide poor approximations...''.

In addition, data often exhibit non-stationarity in the covariance structure or possess a complex structure which is very difficult to model parametrically.

An alternative approach is to use linear mixed effects models (Laird & Ware, 1982). Mixed effects models with random coefficients simultaneously provide for modelling of mean response profiles and the covariance structure. However, these models are often too restrictive for modelling mean response profiles, and the implied covariance structure is highly structured and often not flexible enough for longitudinal data.

To overcome some of the problems associated with linear mixed effects models, nonlinear mixed effects models (Lindstrom & Bates, 1990, Davidian & Giltinan, 1995) have been proposed. These models are usually more plausible for the mean response profiles, and there is the implication (not always true) that the nonlinear model reflects the

mechanism driving the process being observed. A major difficulty is that both estimation and inference are approximate, because the nonlinearity precludes exact likelihoods from being found. In addition, the assumptions made for the random effects may be questionable. Methods which avoid these problems are therefore very attractive.

An alternative approach to modelling mean response profiles for designed experiments and longitudinal data is proposed in this paper. The approach is flexible, includes the class of random coefficient models, is easily extended to multi-stratum designs and more complex experiments, and most importantly, for many applications the choice of a parametric model for the covariance structure is obviated. However, the covariance structure is not restricted to that generated by linear random coefficients, but is a consequence of the use of a non-parametric smoother.

Using non-parametric smoothers in place of parametric models introduces flexibility, but these methods are often exploratory and statistical inference is either not available or not well developed. The literature on non-parametric smoothers for a single explanatory variable is extensive; see Hastie & Tibshirani (1990) and Wahba (1990) for an overview. Hastie & Tibshirani (1990) discuss the extension to several explanatory variables using additive models and Wahba (1990) discusses thin-plate splines (see also Green & Silverman, 1994). Additive models provide a simple extension that has computational and practical benefits. Thin plate splines are more flexible but are more difficult computationally.

For the analysis of longitudinal data, the use of non-parametric smoothers for the joint modelling of mean and covariance has received some attention. Zeger & Diggle (1994) use a semi-parametric model and kernel smoothing for modelling the mean CD4 cell numbers in HIV seroconverters, thus introducing flexibility into modelling of the mean response profile. Diggle & Verbyla (1997) consider modelling the covariance structure using local linear smoothing with kernel weights.

Our approach uses the cubic smoothing spline (Green & Silverman, 1994) in conjunction with random coefficients and variance modelling. The need for these aspects is highlighted in the examples presented in section 2 and analysed in section 6. The key result that allows the merging of curve fitting, random coefficients and variance modelling in the context of factorial experiments, is the fact that the cubic smoothing spline can be formulated as a linear mixed effects model (see Verbyla, 1995). Details are given in section 3.1. While this result is implicit in the literature on cubic smoothing splines, see for example Kimeldorf & Wahba (1970), Wahba (1978), and Green & Silverman (1994), p. 52, the approach of this paper places the cubic smoothing spline explicitly in the framework of standard parametric statistical models. As such, practitioners are likely to be able to use the methods in their own work. There are related formulations, for example the state-space representation and the Kalman filter, Wecker & Ansley (1983), and Bayesian forms, Silverman (1985) and Upsdell (1996) (and the associated FLEXI program).

In many designed experiments (and other studies), there is replication at various levels of factor combinations and the possibility of examining the adequacy of the fit of a model. With smoothing splines this is a novel idea. An extended mixed model is proposed for which the smoothing spline model is a (nested) submodel. This extended model allows deviations from the smooth nature of the spline to be modelled and examined for significance. This forms the material of section 3.2.

Many applications involve splines at various levels of a factorial structure, and section 4.1 provides a discussion. In this context, the varying-coefficient models of Hastie

& Tibshirani (1993), their equation (12), relate directly to the models discussed in this paper. The fundamental difference is our use and formulation of smoothing splines and the consequential inferential benefits. To clarify the ideas, section 4.2 provides a symbolic representation (Wilkinson & Rogers, 1973) of the model of section 4.1 and also an analysis of variance table.

Section 4.3 concerns the analysis of longitudinal data. An hierarchy of smoothing spline models is considered, including random coefficient models. Because of the mixed model formulation of the smoothing spline, these components fit naturally into a unified approach to analysis.

These models impose their own implicit covariance structure, so that for longitudinal data analysis the requirement for covariance modelling is removed in most applications. Including smoothing splines for overall mean trends over time and for deviations from that overall trend caused by treatments, both at the main effects and interaction levels, in *lieu* of covariance modelling, is very attractive from a practical point of view. The use of random coefficients when required increases this flexibility.

While random effects provide flexibility in terms of the covariance or correlation structure, there may still be a need to allow for heterogeneity in the variance of the error process. Section 4.5 provides details on one approach to this problem, borrowing the ideas of Verbyla (1993) and Frensham et al. (1997).

Inference is discussed in section 5. Because of the mixed linear model formulation, the natural approach for estimation is based on residual or restricted maximum likelihood (REML) (Patterson & Thompson, 1971). Confidence bands, tests of hypotheses and calibration are also discussed in section 5. Analysis of the examples is presented in section 6 and a discussion of extensions and conclusions are given in section 7.

2 Examples

2.1 *Example 1: Liveweights of cows*

Diggle et al., 1994, pp. 100-101, consider an experiment giving rise to longitudinal data. The aim of the experiment was to study how iron dosing, a factor at two levels (no dosing and a standard amount) and how presence or absence of the organism *M. paratuberculosis*, influence the liveweights of cows. Twenty-eight cows were allocated to treatments in the 2 by 2 factorial design implied by these factors, with 4, 4, 10 and 10 cows for the 4 treatment combinations; in fact one cow died during the study and data on another was removed on advice from the researcher, so that the final sizes were 4, 3, 9 and 10 in the data analysed. The liveweights of the cows were measured at 23 unequally spaced time intervals.

Figure 1 presents a trellis plot (Becker & Cleveland, 1996) of the liveweights under the 2×2 factorial structure; Diggle et al. (1994) consider log-liveweights but the analysis presented in this paper is based on the original scale. The growth pattern over time is reasonably smooth (and appears linear) but there are also consistent non-smooth changes in time; for example note the change between days 179 and 219. Figure 1 also shows that the cows vary in their initial liveweight and in their rate of growth, thereby leading to heterogeneity of variance over time. The individual cow liveweights are plotted against time in Figure 2. The additional information portrayed in this figure is the differing

levels of curvature in the growth pattern of cows over time. All these aspects should be incorporated in the analysis.

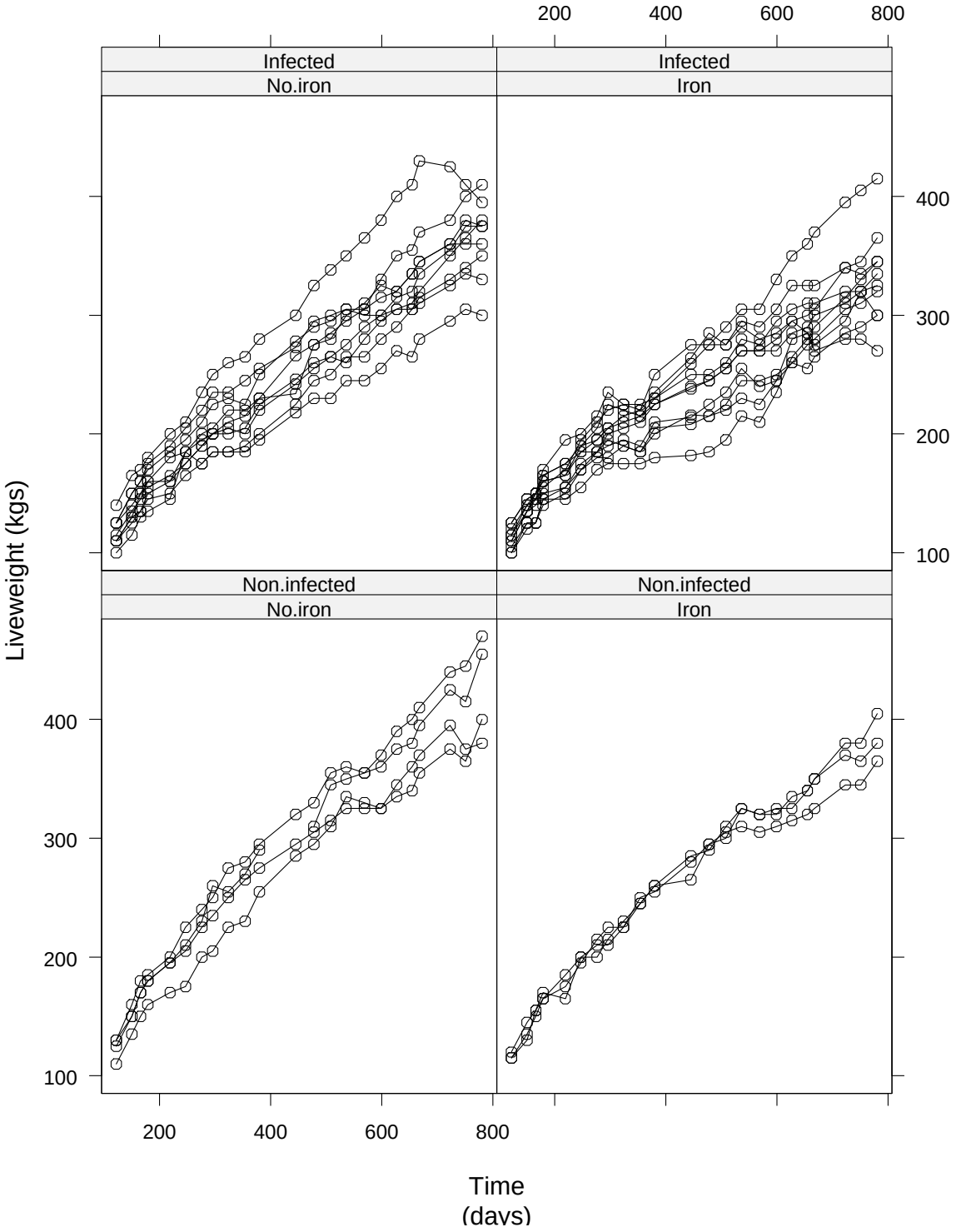

Fig. 1 Cow data: Trellis plot of log-liveweight of cows under the 2x2 factorial treatment structure

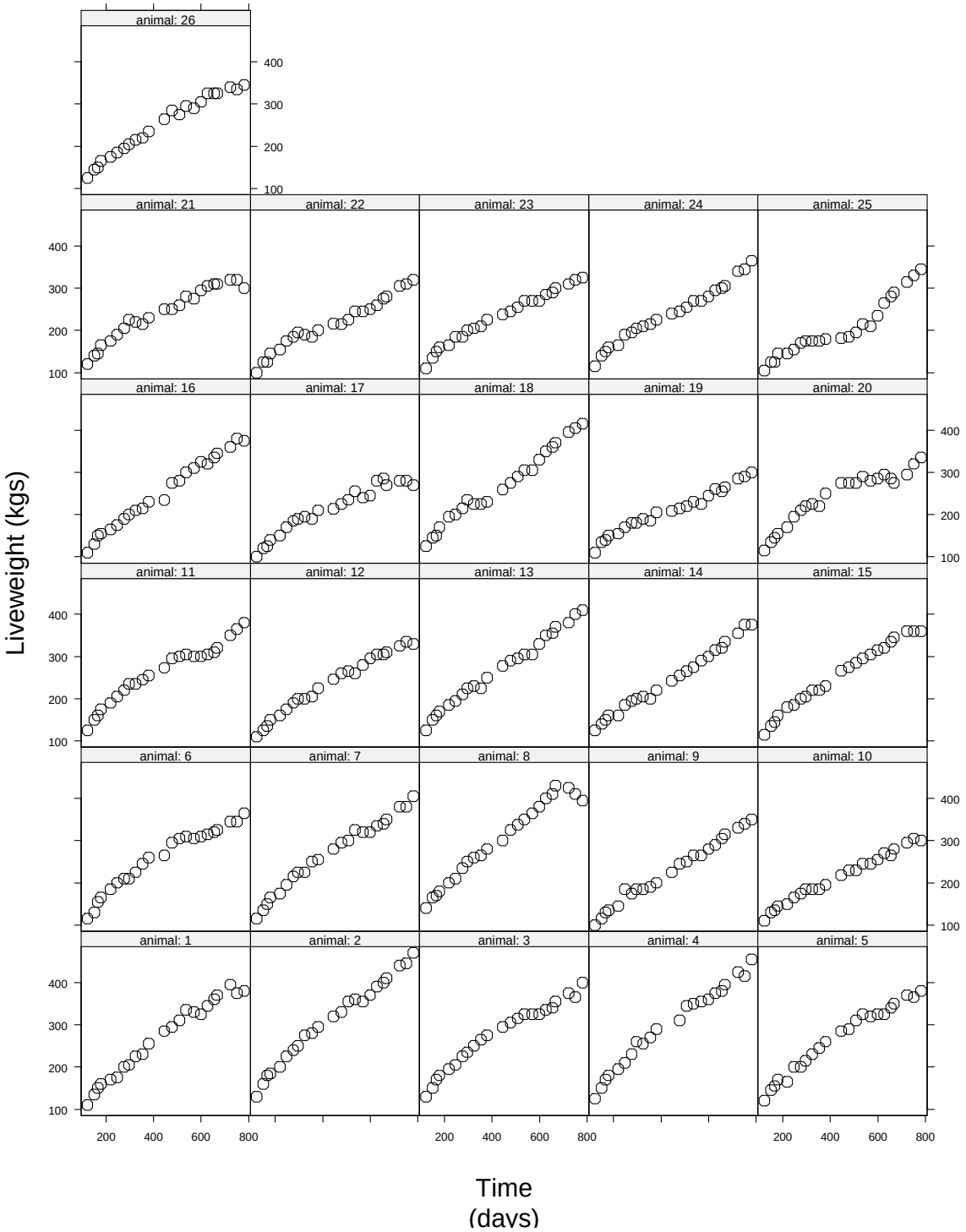

Fig. 2 Cow data: Trellis plot of log-liveweight of individual cows

On the log scale, Diggle et al. (1994) use joint modelling of the mean response profiles and the covariance matrix. The mean response profiles are replaced by the mean profile for the control group and three contrasts. Quadratic polynomials are used for the contrast profiles, while the mean profile for the control treatment is not modelled. The covariance structure, given by equation (5.2.5) in Diggle et al. (1994), is an additive structure consisting of an animal random effect (random intercept), measurement error and a stationary Gaussian process with correlation function, given by

$$\rho(k) = \exp\{-\alpha k^2\}$$

for lag or time separation k.

The main aim of the study was to examine the effect of the treatments on the liveweights of the cows. Tests of hypotheses concerning parameters of the quadratic models for treatment contrasts were carried out by Diggle et al. (1994).

The need to choose the models is clearly difficult as Diggle et al. (1994) state. Some degree of flexibility in model fitting would be advantageous, and if coupled with the ability to test hypotheses of interest could result in an improved analysis.

2.2 *Example 2: ELISA assay for recombinant protein DNase in rat serum*

Davidian & Giltinan (1993) consider analysis of 11 trials, in which duplicate absorbance (optical density) measurements were obtained at each of several known concentrations of the protein DNase. The aim of the experiment was to use a fitted curve to calibrate unknown samples. A trellis plot of the 11 trials is given in Figure 3. The monotonicity of the response profile is evident. The rate of increase in response as concentration increases varies between experiments, and some of the duplicate values show variation at the higher concentrations. These aspects should be incorporated in the model to be used as the calibration curve.

Davidian & Giltinan (1993) consider the four parameter logistic model for the mean response profiles,

$$\mathrm{E}\,(y_{ijk}|\boldsymbol{\beta}_i) = \mu_{ij} = \beta_{i1} + (\beta_{i2} - \beta_{i1})/[1 + \exp\{\beta_{i4}(\log x_j - \beta_{i3}\}]$$

with i indexing the trial ($i = 1, 2, \ldots, 11$), j the dose $j = 1, 2, \ldots, 8$ and $k = 1, 2$ the replicates. The actual dose is denoted by x_j and $\boldsymbol{\beta}_i = [\beta_{i1},\ \beta_{i2},\ \beta_{i3},\ \beta_{i4}]^T$. Zero concentration needs special treatment in this model and Davidian & Giltinan (1993) replace it by the first non-zero concentration divided by 4.

The parameters are allowed vary across trials, with the assumption that $\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta},\ D)$, and because of variance heterogeneity, Davidian & Giltinan (1993) propose a power law for a mean-variance relationship, namely

$$\mathrm{var}\,(y_{ijk}|\boldsymbol{\beta}_i) = \sigma^2 \mu_{ij}^{\theta} \qquad (1)$$

Using a Taylor series expansion, Davidian & Giltinan (1993) obtain an approximate confidence interval in the calibration problem at the individual trial level.

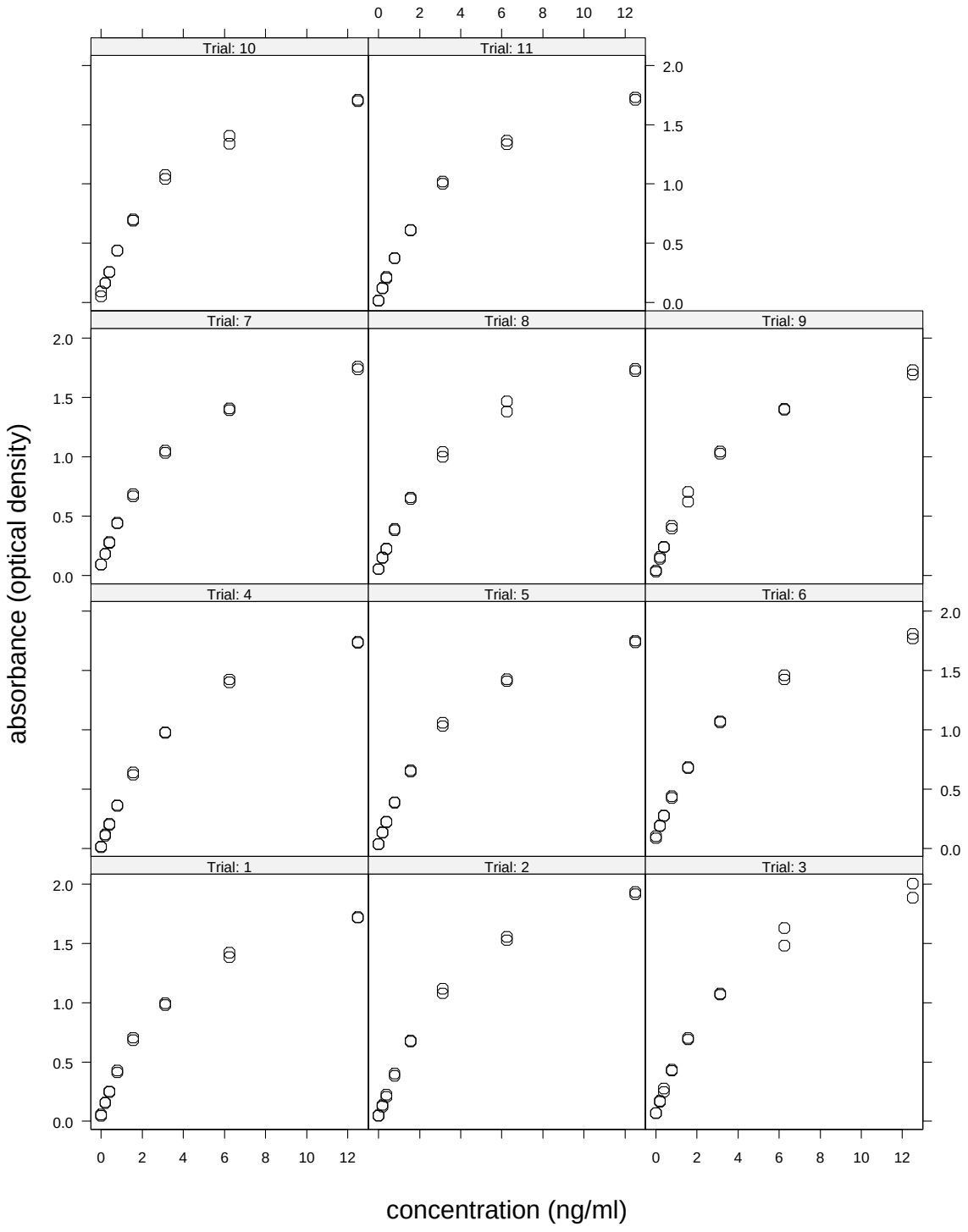

Fig. 3 ELISA assay data: Trellis plot of the eleven experiments (trials) on optimal density

The nonlinear model provides a smooth monotone relationship between absorbance and concentration. However, in some applications there may be a lack of fit problem for the assumed nonlinear model; for example, the data may be insufficient to allow efficient estimation of all the parameters, particularly those associated with the asymptotes. The mean-variance relationship (1) is a strong assumption and introduces additional estimation problems; the plot of studentized residuals given by Davidian & Giltinan (1993) indicates that the assumed mean-variance relationship is not entirely satisfactory. Lastly, the assumption that the random coefficients are jointly normally distributed may be questionable.

An alternative is to use a non-parametric smoother. A model which can be used for the calibration problem and which allows for variance heterogeneity and possible variation between experiments and between concentrations, is discussed in this paper.

2.3 *Example 3: Multi-environment lupin variety trial*

These data were kindly provided by Dr. Wallace Cowling of Agriculture WA. A series of eleven field experiments were conducted in 1991 and 1992 to determine the yield response of a set of narrow leaf lupin varieties to plant density. Nine varieties were used, from the first early flowering variety, Unicrop (released in 1973) to the most recently released variety, 84L:489 or Myallie (released in 1995); see Table 1. The factorial combinations of the varieties with seeding rates (or plant densities) constituted the treatments for each trial. The target densities used in 1991 were 10, 20, . . . , 60 plants per square metre, while in 1992 the target densities were 20, 30, . . . , 70 plants per square metre. The variety 84L:439 was included in 1992 only. All trials were laid out as randomized complete block designs, with either 2 (in 1991) or 3 (in 1992) replicates. Each replicate comprised either 16 rows by 3 columns (in 1991) or 27 rows by 2 columns (in 1992), resulting in trial configurations of 16×6 and 27×6 respectively. The location and year are given in Table 1.

TABLE 1
Lupin data: release dates of varieties and site information (year by location)

Variety	Year of release	site	year	location
Unicrop	1973	1	1991	Badgingerra
Illyarrie	1979	2	1991	Corrigin
Yandee	1980	3	1991	Mt. Barker
Danja	1986	4	1991	Newdegate
Gungurru	1988	5	1991	Wongan Hills
Yorrel	1989	6	1992	Corrigin
Warrah	1989	7	1992	East Chapman
Merrit	1991	8	1992	Mt. Barker
Myallie	1995	9	1992	Mullewa
		10	1992	South Carrabin
		11	1992	Wongan Hills

Figures 4 and 5 provides a plot of yield against seeding rate for each variety and site within each year. The data from columns 5 and 6 for the Mt. Barker site in 1991 was

discarded prior to analysis as these plots were overrun with weeds. There were a total of 28 missing values, caused by various misadventures during sowing and harvesting.

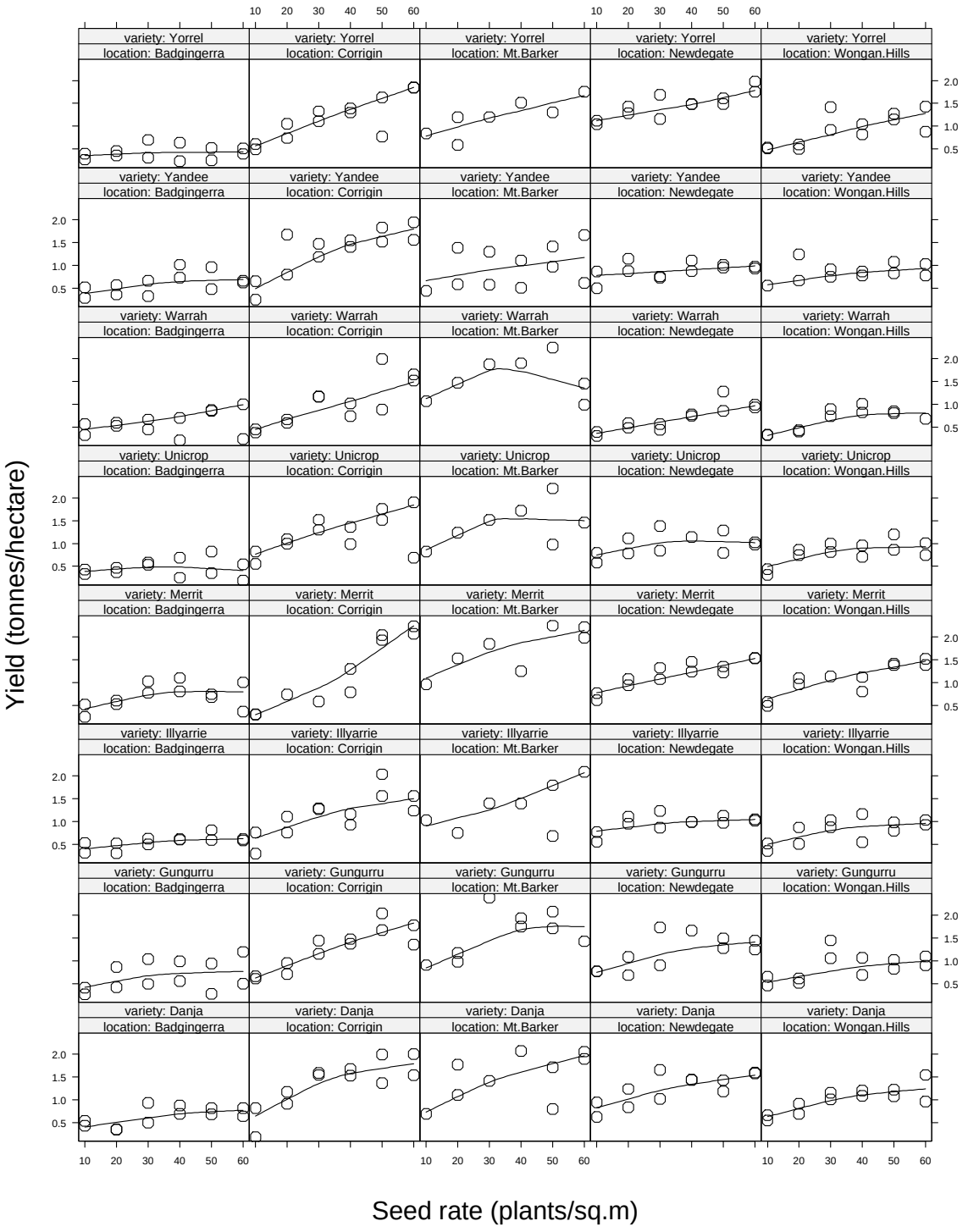

Fig. 4 Lupin data: Trellis plot of yield against seeding rate for each location and variety in 1991

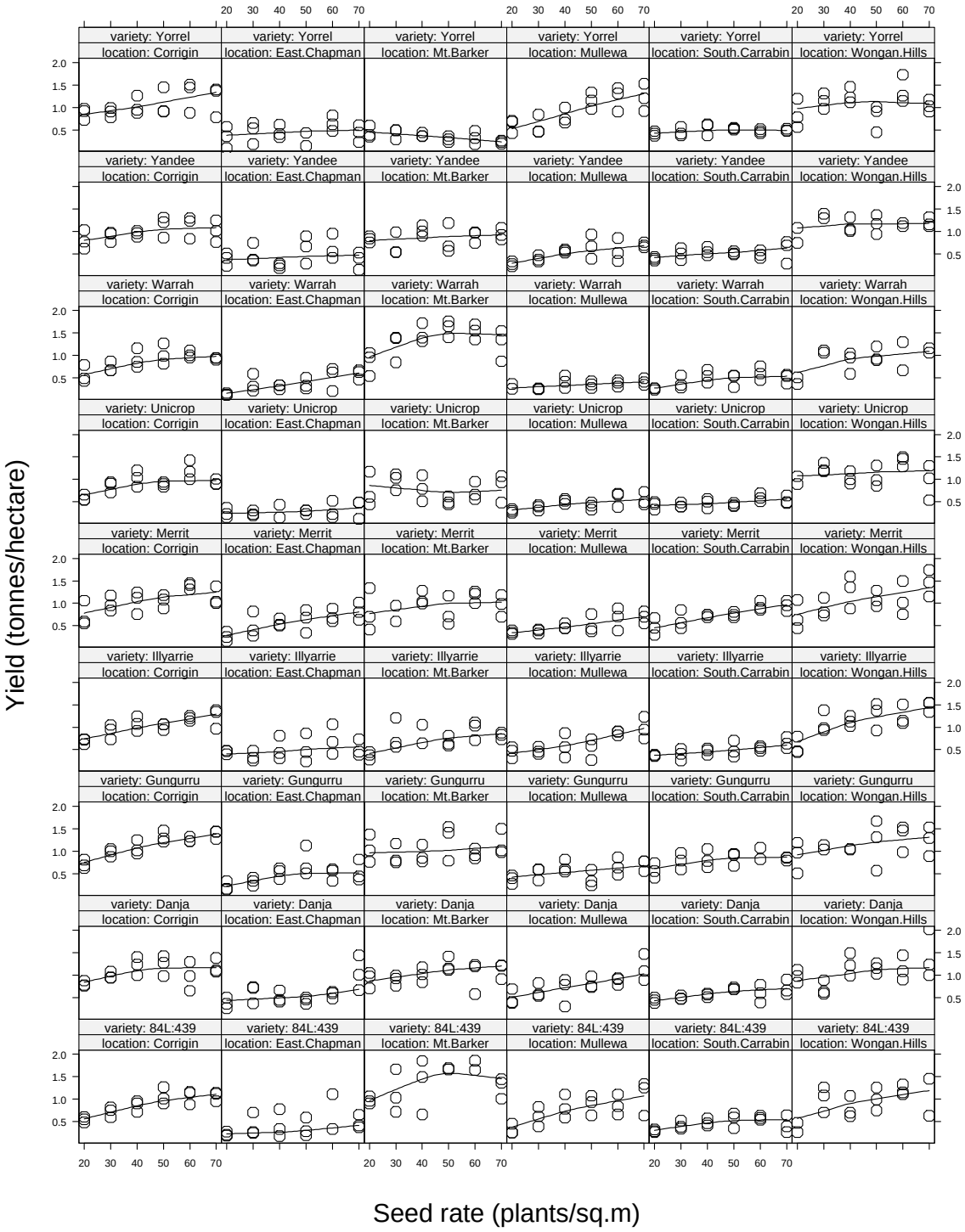

Fig. 5 Lupin data: Trellis plot of yield against seeding rate for each location and variety in 1992

In previous analyses the yield-density relationships were modelled using the equation

$$\mathrm{E}\,(y_{ijkl}|\alpha_{ij},\ \beta_{ij}) = \frac{x_{jk}}{\alpha_{ij} + \beta_{ij} x_{jk}} \tag{2}$$

where i and j index the variety and site (numbered 1 to 11, see Table 1) and k and l index the seeding rate and replicate, y_{ijkl} is the grain yield (in tonnes/ha), x_{jk} is the plant density (in plants/sq. m), $1/\beta_{ij}$ is the asymptotic yield and α_{ij} is the curvature. Thus the coefficients vary according to site and variety. Often, analysis is carried out using a two stage non-linear modelling approach. The first stage consists of fitting equation (2) for each variety by site combination, and in the second stage these coefficients are modelled in an across sites analysis to determine an average curve for each variety. This two-stage approach is inefficient and only approximate, although the non-linear mixed modelling approaches discussed by Davidian & Giltinan (1995), for example, could be adopted. An alternative approach is presented which can easily account for spatial variation and variance heterogeneity in he presence of non-linearity of the yield-density relationships.

3 Smoothing Splines

3.1 *Smoothing splines as mixed models*

Consider a single quantitative explanatory variable T and response variable y, where the observations are (t_i, y_i), $i = 1, 2, \ldots, p$ and $t_1 < t_2 < \ldots < t_p$. A simple model for the relationship of y given $T = t_i$ is

$$y_i = \eta_i + e_i = g(t_i) + e_i \tag{3}$$

where η_i is the mean of y_i, assumed to be a function g of the variable T, and e_i is the random error. If $\boldsymbol{y}$, $\boldsymbol{g}$ and $\boldsymbol{e}$ are the $p \times 1$ vectors of the response variable, the function g at the t_i and the random errors respectively, (3) can be written as

$$\boldsymbol{y} = \boldsymbol{\eta} + \boldsymbol{e} = \boldsymbol{g} + \boldsymbol{e} \tag{4}$$

where $\boldsymbol{e}$ is Gaussian with zero mean vector and covariance matrix $\sigma^2\boldsymbol{R}$, $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{R})$. The matrix $\boldsymbol{R}$ may have unknown parameters; for example $\boldsymbol{R}$ may be diagonal with elements reflecting variance heterogeneity.

Suppose $g(\cdot)$ is a smooth but unknown function to be estimated. Using a likelihood with a specific penalty function to impose the smoothness, leads to the cubic smoothing spline; see Green & Yandell (1985), Green (1987). Green & Silverman (1994) provide a detailed overview but note that the notation used in this paper is consistent with mixed model notation (see Robinson, 1991) and differs from that used by Green & Silverman (1994). In the independent Gaussian case, this penalized log-likelihood approach is equivalent to penalized least squares. The objective is to choose g to maximize the penalized log-likelihood,

$$\begin{aligned} \ell &= -\frac{1}{2}\log(\det\sigma^2\boldsymbol{R}) - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{p}\sum_{j=1}^{p} R^{ij}(y_i - g(t_i))(y_j - g(t_j)) + \lambda_s \int \{g''(t)\}^2 dt\right] \\ &= -\frac{1}{2}\log(\det\sigma^2\boldsymbol{R}) - \frac{1}{2\sigma^2}\left[(\boldsymbol{y} - \boldsymbol{g})^T\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{g}) + \lambda_s \int \{g''(t)\}^2 dt\right] \end{aligned}$$

where λ_s is a parameter governing the amount of smoothing. For given $\boldsymbol{R}$ and λ_s, the solution to this maximization problem is the cubic smoothing spline.

Let $h_j = t_{j+1} - t_j$, $j = 1, 2, \ldots, p-1$. Define $\boldsymbol{\Delta}$ and $\boldsymbol{G}_s$ to be $p \times p-2$ and $p-2 \times p-2$ matrices respectively, with the non-zero elements are (for $i = 1, 2, \ldots, p-2$)

$$\Delta_{ii} = \frac{1}{h_i}, \quad \Delta_{i+1,i} = -\left(\frac{1}{h_i} + \frac{1}{h_{i+1}}\right), \quad \Delta_{i+2,i} = \frac{1}{h_{i+1}} \tag{5}$$

$$G_{s;i,i+1} = G_{s;i+1,i} = \frac{h_{i+1}}{6}, \quad G_{s;ii} = \frac{h_i + h_{i+1}}{3} \tag{6}$$

If $\boldsymbol{K} = \boldsymbol{\Delta}\boldsymbol{G}_s^{-1}\boldsymbol{\Delta}^T$, then at the design points, the cubic smoothing spline is given by

$$\hat{\boldsymbol{g}} = (\boldsymbol{R}^{-1} + \lambda_s \boldsymbol{K})^{-1}\boldsymbol{R}^{-1}\boldsymbol{y} \tag{7}$$

If $\boldsymbol{X}_s$ is a $p \times 2$ matrix whose columns are a vector of ones, $\mathbf{1}_p$ say, and the vector $[t_1\, t_2\, \ldots\, t_p]^T$, $\boldsymbol{Z}_s = \boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{\Delta})^{-1}$, and $\boldsymbol{H} = \sigma^2(\boldsymbol{R} + \lambda_s^{-1}\boldsymbol{Z}_s\boldsymbol{G}_s\boldsymbol{Z}_s^T)$, then it is shown in Appendix A that the solution (7) can be written as

$$\hat{\boldsymbol{g}} = \boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s + \boldsymbol{Z}_s\tilde{\boldsymbol{u}}_s \tag{8}$$

where

$$\begin{aligned} \hat{\boldsymbol{\beta}}_s &= (\boldsymbol{X}_s^T\boldsymbol{H}^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s^T\boldsymbol{H}^{-1}\boldsymbol{y} \\ \tilde{\boldsymbol{u}}_s &= (\boldsymbol{Z}_s^T\boldsymbol{R}^{-1}\boldsymbol{Z}_s + \lambda_s\boldsymbol{G}_s^{-1})^{-1}\boldsymbol{Z}_s^T\boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s) \end{aligned}$$

These results show that the cubic smoothing spline can be written as an estimated straight line ($\boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s$) plus a predicted 'random effect' ($\boldsymbol{Z}_s\tilde{\boldsymbol{u}}_s$), and (8) is therefore a best linear unbiased predictor or BLUP (see Speed, 1991) of a conditional mean vector. If $\boldsymbol{R} = \boldsymbol{I}_p$, the intercept and slope are estimated by ordinary least squares and the two components, linear and nonlinear are orthogonal and hence statistically independent.

The solution (8) implies that at the design points t_i, $g(t)$ can be expressed as a mixed model, the fixed part being a straight line, the random component reflecting the curvature and hence the spline. Thus in (4), $\boldsymbol{g}$ can be expressed as

$$\boldsymbol{g} = \boldsymbol{X}_s\boldsymbol{\beta}_s + \boldsymbol{Z}_s\boldsymbol{u}_s \tag{9}$$

where the $(p-2) \times 1$ random vector $\boldsymbol{u}_s \sim N(\mathbf{0}, \sigma_s^2\boldsymbol{G}_s)$. The smoothing parameter is then a variance ratio, that is $\lambda_s = \sigma^2/\sigma_s^2 = \gamma_s^{-1}$, say, the log-likelihood is a conditional log-likelihood of $\boldsymbol{y}$ given $\boldsymbol{u}_s$, and the penalty function is the log-density function of $\boldsymbol{u}_s$.

An alternative formulation of the spline allows useful results in Green & Silverman (1994) to be used. The vector of second derivatives of the spline at the design points $t_2, \ldots t_{p-1}$ is given by $\boldsymbol{\delta}_s = \boldsymbol{G}_s^{-1}\boldsymbol{u}_s$; at t_1 and t_p the second derivatives of $g(\cdot)$ are zero. Using $\boldsymbol{\delta}_s$, the spline can be calculated at points other than the observed t_i using equation (2.19) of Green & Silverman, 1994, p. 22. Thus if $\boldsymbol{G}_s^*$ is the matrix $\boldsymbol{G}_s$ with zero rows added to the top and bottom, $\boldsymbol{Z}_s^* = \boldsymbol{Z}_s\boldsymbol{G}_s^*$ with ith row $\boldsymbol{z}_{s;i}^{*\,T}$ and $\boldsymbol{\delta}_s^* = [0\ \boldsymbol{\delta}_s^T\ 0]^T$, the estimate of a single $g(t)$ for any $t_i \le t \le t_{i+1}$ is

$$\begin{aligned} \hat{g}(t) &= [1\ t]\hat{\boldsymbol{\beta}}_s + \left[\frac{t - t_i}{h_i}\boldsymbol{z}_{s;i+1}^{*\,T} + \frac{t_{i+1} - t}{h_i}\boldsymbol{z}_{s;i}^{*\,T}\right. \\ &- \left.\frac{(t - t_i)(t_{i+1} - t)}{6h_i}\left\{(t - 2t_i + t_{i+1})\boldsymbol{e}_{i+1}^T - (t - 2t_{i+1} + t_i)\boldsymbol{e}_i^T\right\}\right]\tilde{\boldsymbol{\delta}}_s^* \end{aligned} \tag{10}$$

where $\boldsymbol{e}_j$ denotes the jth standard basis vector of p-dimensional Euclidean space.

The above formulation is important because the cubic smoothing spline can be used in a standard parametric framework. This allows nonlinear trends to be modelled *in a mixed linear model* introducing flexibility and substantial benefits from an inferential point of view.

3.2 *Smoothing splines and replicated data*

Often the values of the explanatory variable are not unique across experimental units. The above results using the mixed model still hold with $\boldsymbol{X}_s$ and $\boldsymbol{Z}_s$ containing duplicate rows as necessary. Formally this means premultiplying both $\boldsymbol{X}_s$ and $\boldsymbol{Z}_s$ by a 'replication' or incidence matrix, $\boldsymbol{N}_g$, say, a binary matrix, in which the ith column is assumed to have n_i (the number of replicate t_i values) unit entries. The total sample size is $n = n_1 + n_2 + \cdots + n_p$. The subscript g is intended to designate grouping. If the unique values, $t_1, t_2, \ldots, t_p$, of the variable T are taken as the levels of a factor, $\boldsymbol{N}_g$ is the corresponding design matrix. If $\boldsymbol{y}$ is the $n \times 1$ vector of observations, equation (4) becomes

$$\boldsymbol{y} = \boldsymbol{N}_g \boldsymbol{\eta} + \boldsymbol{e} \tag{11}$$

with $\boldsymbol{\eta} = \boldsymbol{g}$ remaining as in (9), and the covariance matrix $\sigma^2 \boldsymbol{R}$ is now an $n \times n$ positive-definite symmetric matrix. The smoothing spline is fitted to the means at the replicated values, with subsequent weighting to account for possibly different levels of replication.

With replication at the t_i, some assessment of the adequacy or lack of fit of smoothing splines should be possible. A method similar to the standard methods for fixed effects in analysis of variance is developed and involves nesting the smoothing spline model in an extended model.

Because the smoothing spline is a random trend model, that is a fixed component plus a random component, a natural approach is to extend the model by including a non-smooth random component. Thus in place of (4) consider

$$\boldsymbol{\eta} = \boldsymbol{g} + \boldsymbol{u}_g = \boldsymbol{X}_s \boldsymbol{\beta}_s + \boldsymbol{Z}_s \boldsymbol{u}_s + \boldsymbol{u}_g \tag{12}$$

so that

$$\boldsymbol{y} = \boldsymbol{N}_g (\boldsymbol{X}_s \boldsymbol{\beta}_s + \boldsymbol{Z}_s \boldsymbol{u}_s + \boldsymbol{u}_g) + \boldsymbol{e} \tag{13}$$

The extra component of variation in the model for $\boldsymbol{y}$, namely $\boldsymbol{u}_g$, is assumed to be $N(\boldsymbol{0}, \sigma_g^2 \boldsymbol{I}_p)$ and hence is a standard random effect for T taken as a factor; this type of term is labelled RLOF in following sections. This is a natural way to allow for departures from a smooth random trend, allowing for lack of fit of the smooth model which manifests itself as structure in the underlying covariance structure. There are many ways the covariance structure may depart from the structure implied by the spline mixed model and the random errors and $\boldsymbol{u}_g$ is aone possible way to allow for such departures. In using $\boldsymbol{u}_g$ there may also be a lack of power in detecting various forms of departure. Tests concerning σ_g^2 are discussed in section 5.3.

4 Modelling using Smoothing Splines

4.1 *Designed experiments*

To illustrate the ideas, consider the following simple situation. Suppose we have two crossed factors, a qualitative treatment factor A, and a quantitative factor T. Let y_{ijk} denote the response variable for which the factor T has value t_i ($i = 1, 2, \ldots, p$), the factor A is at level j ($j = 1, 2, \ldots, a$) and k indexes the replicate within the combination (i, j) ($k = 1, 2, \ldots, n_{ij}$); $n = \sum_{i=1}^{p} \sum_{j=1}^{a} n_{ij}$. Consider the saturated model

$$y_{ijk} = \eta_{ijk} + e_{ijk}$$

where η_{ijk} is the mean of y_{ijk} and e_{ijk} is the random error. If μ, τ_i, α_j and ζ_{ij} are the overall mean, factor T, factor A, and factor T by factor A interaction effects, consider the two-way analysis of variance additive model

$$\begin{aligned} \eta_{ijk} &= \mu + \tau_i + \alpha_j + \zeta_{ij} \\ &= \mu + \tau(t_i) + \alpha_j + \zeta_j(t_i) \end{aligned} \tag{14}$$

where in (14) the main effect due to T ($\tau(t_i)$), and the A by T interaction effects ($\zeta_j(t_i)$), are written explicitly as functions of T. Collecting over T, (14) becomes

$$\boldsymbol{\eta}_{jk} = \mathbf{1}_p \mu + \boldsymbol{\tau} + \mathbf{1}_p \alpha_j + \boldsymbol{\zeta}_j \tag{15}$$

where $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}_j$ are $p \times 1$ vectors which are functions of the factor T.

The terms α_j, $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}_j$ in (15) may be fixed or random depending on the context. The latter two terms are modelled by using (12), so that each is taken to be a smooth function of T, given by the cubic smoothing spline, plus a term for possible departures from smoothness. Thus $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}$ have both fixed and random components. Explicitly

$$\boldsymbol{\tau} = \boldsymbol{X}_{s,0}\boldsymbol{\beta}_{s,0} + \boldsymbol{Z}_{s,0}\boldsymbol{u}_{s,0} + \boldsymbol{u}_{g,0} \tag{16}$$

$$\boldsymbol{\zeta}_j = \boldsymbol{X}_{s,j}\boldsymbol{\beta}_{s,j} + \boldsymbol{Z}_{s,j}\boldsymbol{u}_{s,j} + \boldsymbol{u}_{g,j} \tag{17}$$

where in (16) and (17), the $\boldsymbol{X}$ and $\boldsymbol{Z}$ matrices with subscript s refer to smoothing spline linear and random component design matrices given by (9) in section 3.1.

In each of the fixed components of equations (16), and (17) there are intercept and slope parameters. With μ in the model, not all the intercepts are identifiable; there is also one redundant slope parameter. These are removed in the analysis.

The spline random effects, $\boldsymbol{u}_{s,j}$, $j = 0, 1, \ldots, a$, and the lack of fit random effects, $\boldsymbol{u}_{g,j}$, $j = 0, 1, \ldots, a$, contribute to the covariance structure through their associated variance components. It is usually assumed that $\boldsymbol{u}_{s,0} \sim N(\mathbf{0}, \sigma^2_{s,0}\boldsymbol{G}_{s,0})$ and $\boldsymbol{u}_{s,j} \sim N(\mathbf{0}, \sigma^2_s \boldsymbol{G}_{s,j})$, $j = 1, 2, \ldots, a$ and similarly $\boldsymbol{u}_{g,0} \sim N(\mathbf{0}, \sigma^2_{g,0}\boldsymbol{I})$ and $\boldsymbol{u}_{g,j} \sim N(\mathbf{0}, \sigma^2_g \boldsymbol{I})$, $j = 1, 2, \ldots, a$. The spline (and also RLOF) variance component is assumed the same for all interaction terms $\boldsymbol{\zeta}_j$. However, more complicated covariance structures are possible if required. More complex factorial structures give rise to similar considerations.

The full conditional mean vector is

$$\begin{aligned} \boldsymbol{\eta} &= \mathbf{1}_n \mu + \boldsymbol{N}_\tau \boldsymbol{\tau} + \boldsymbol{N}_\alpha \boldsymbol{\alpha} + \sum_{j=1}^{a} \boldsymbol{N}_{\zeta,j} \boldsymbol{\zeta}_j \\ &= \mathbf{1}_n \mu + \boldsymbol{N}_\tau \boldsymbol{\tau} + \boldsymbol{N}_\alpha \boldsymbol{\alpha} + \boldsymbol{N}_\zeta \boldsymbol{\zeta} \end{aligned} \tag{18}$$

where the matrices in (18) are replication matrices in the sense of section 3.2 and $\boldsymbol{\tau}$ and $\boldsymbol{\zeta}_j$ are replaced by (16) and (17). Equation (18) is a mixed linear model.

4.2 *Symbolic and analysis of variance representation*

To clarify the modelling approach, it is useful to consider the Wilkinson & Rogers (1973) symbolic representation (and extensions of the representation) of the above model, and the analysis of variance representation to be used in the examples.

Let y denote the observation vector, 1, a and t denote the constant (or intercept) and factors A and T, respectively, and following Wilkinson & Rogers (1973) let $a.t$ denote their interaction. Let lin(t) denote t as a quantitative variable which will contribute a single regression parameter (slope). Let spl(t) denote the spline random component based on the quantitative t and ran(t) denote the random lack of fit or random effect with t as a factor. The model (14) is then

$$y \sim 1 + a + t + a.t$$

while the mixed model (18) is

$$y \sim 1 + a + \text{lin}(t) + \text{spl}(t) + \text{ran}(t) + a.\text{lin}(t) + a.\text{spl}(t) + a.\text{ran}(t)$$

This model can be represented in an analysis of variance table as given in Table 2. The table is divided into constant (1), main effects for A (a) and T (t), and the A by T interaction ($a.t$). The latter two are sub-divided into the linear, spline and random t components; fixed effects are labelled by F and have associated degrees of freedom (df), whereas spline and random lack of fit components are labelled by R and RLOF respectively, and have associated number of effects in $\boldsymbol{u}_s$ and $\boldsymbol{u}_g$, respectively. The table provides a convenient summary of (18).

TABLE 2
Two-way design: AOV Decomposition

Term	Decomposition	Type	df or ne
Constant	1	F	1
a			
	a	F	$a-1$
t			
	lin(t)	F	1
	spl(t)	R	$p-2$
	ran(t)	RLOF	p
$a.t$	a.lin(t)	F	$a-1$
	a.spl(t)	R	$a(p-2)$
	a.ran(t)	RLOF	ap
Error		R	

4.3 *Repeated Measures or Longitudinal Data*

In the longitudinal setting, the quantitative variable T is time. Let $y_{jk}(t_{ijk})$ be the response variable at the ith time t_{ijk} for unit k at level j of factor A. Consider as in section 4.1 the saturated model

$$y_{jk}(t_{ijk}) = \eta_{ijk} + e_{ijk}$$

where η_{ijk} represent the unit by time (as a factor) means and e_{ijk} is the random error. Decomposing η_{ijk} into additive components for an overall effect, a factor A effect and a unit effect, all at time t_{ijk},

$$\eta_{ijk} = \mu(t_{ijk}) + \alpha_j(t_{ijk}) + \zeta_{jk}(t_{ijk})$$

the vector of means for unit k in treatment j then equals

$$\boldsymbol{\eta}_{jk} = \boldsymbol{\mu} + \boldsymbol{\alpha}_j + \boldsymbol{\zeta}_{jk} \tag{19}$$

where each vector is thus a function of time.

In a similar manner to (16) and (17), (12) is applied to each term in (19), so that

$$\boldsymbol{\mu} = \boldsymbol{X}_s\boldsymbol{\beta}_s + \boldsymbol{Z}_s\boldsymbol{u}_s + \boldsymbol{u}_g \tag{20}$$

$$\boldsymbol{\alpha}_j = \boldsymbol{X}_{s,j}\boldsymbol{\beta}_{s,j} + \boldsymbol{Z}_{s,j}\boldsymbol{u}_{s,j} + \boldsymbol{u}_{g,j} \tag{21}$$

$$\boldsymbol{\zeta}_{jk} = \boldsymbol{X}_{s,jk}\boldsymbol{\beta}_{s,jk} + \boldsymbol{Z}_{s,jk}\boldsymbol{u}_{s,jk} \tag{22}$$

The subscripts again indicate spline and random group (time in this context) terms. The full vector of means becomes

$$\boldsymbol{\eta} = \boldsymbol{N}_\mu\boldsymbol{\mu} + \boldsymbol{N}_\tau\boldsymbol{\tau} + \boldsymbol{\zeta} \tag{23}$$

and the $\boldsymbol{N}$ matrices are replication matrices.

Thus splines are incorporated at the 'overall mean' level, at the treatment A level and at the unit level. The random time effect for examining the adequacy of the spline fit appears at all but the unit level. There is usually no replication at the unit level, and hence this type of term cannot be included over and beyond the error term.

As in section 4.1, some fixed effects are aliased and the redundancies are removed. Additional considerations discussed in section 4.4 where a symbolic representation of the model (23) is given; the equivalent analysis of variance table follows as in section 4.2.

4.4 *Random coefficients*

There may be situations where heterogeneity is apparent at the treatment or factor A level of the experiment. In this case, the intercept and slope parameters $\boldsymbol{\beta}_{s,j}$ may be taken as random coefficients with zero mean vector and a 2×2 covariance matrix to allow for the heterogeneity, as in Laird & Ware (1982). Furthermore, the model for $\boldsymbol{\zeta}_{jk}$, (22), allows for different intercepts and slopes for the individual experimental units in the longitudinal setting. Again these are not all identifiable and furthermore are usually not of direct interest. A common approach is to allow a random intercept and slope at the unit level. The BLUPs of the intercept and slope random effects then provide the adjustment to the intercept and slope determined by other terms in the model.

Random coefficients will be denoted by RC in analysis of variance tables given in the analyses of section 6. A superscript may be attached to link the coefficients appearing in different parts of the table. Usually there are three terms in an RC structure, a random intercept, a random slope and a correlation.

In symbolic terms, the situation in section 4.3 given by equations (20), (21), (22) and (23) can be expressed as

$$y \sim 1 + \text{lin}(t) + \text{spl}(t) + \text{ran}(t) + a + a.\text{lin}(t) + a.\text{spl}(t) + a.\text{ran}(t) + \text{unit} + \text{unit.lin}(t) + \text{unit.spl}(t)$$

where unit and unit.lin(t) are correlated random effects, random intercept and slope terms for each experimental unit, and unit.spl(t) is a unit random splines component. These terms increase the flexibility of the implicit covariance structure generated.

4.5 *Variance Heterogeneity*

Variance heterogeneity can be a problem in designed experiments and longitudinal data. A transformation may be used to overcome the problem as in the analysis by Diggle et al. (1994) of the data on cows given in section 2.1, or a random coefficients model as in section 4.4 may be used. It may still be the case that the error process has non-constant variance. Davidian & Giltinan (1995) propose mean-variance power laws in several examples. Another possibility is to use a log-linear model for the variances; see for example Harvey (1976), Aitkin (1987), Verbyla (1993) and Frensham et al. (1997). This approach has the advantage of maintaining parameter independence between the mean and covariance structure, allowing easier estimation. It is used in analysing the data of section 2.2.

If $\sigma^2 \boldsymbol{R} = \text{diag}\left(\sigma^2_{ijk}\right)$ so that σ^2_{ijk} denotes the variance of y_{ijk} or $y_{jk}(t_{ijk})$, the log-linear model is of the form

$$\log \sigma^2_{ijk} = \boldsymbol{w}^T_{ijk}\boldsymbol{\phi} \tag{24}$$

where $\boldsymbol{w}_{ijk}$ is a vector of explanatory variables and $\boldsymbol{\phi}$ ($q \times 1$ say) is a vector of unknown parameters. The first element in $\boldsymbol{w}_{ijk}$ is 1, so that if $\phi_2 = \ldots = \phi_q = 0$ the variance is constant, namely $\sigma^2_{ijk} = \exp(\phi_1)$.

5 Estimation and Inference

5.1 *Estimation*

The models considered in this paper are mixed linear models, and estimation is based on residual or restricted maximum likelihood (Patterson & Thompson, 1971), with best linear unbiased prediction (Robinson, 1991) for the random effects.

Cross-validation is the standard data-driven method for determining the smoothing parameters for the splines which may be in the fitted model. Speed (1991) points out that an alternative method, called generalized maximum likelihood (GML) by Wahba (1985), is REML; see also Thompson (1985) in a discussion of Silverman (1985). Wahba (1985) shows the behaviour of GML (or REML) and generalized cross-validation is similar for small to medium sample sizes; see also Kohn et al. (1991). For the models used in this paper, REML provides a straightforward method for determining the amount of smoothing required for a particular component, in the presence of other variance components or parameters and other smoothing parameters.

Appendix B provides some details regarding estimation and software implementations.

5.2 *Confidence or probability bands*

A fundamental requirement is the provision of confidence or probability bands. The usual approach in the smoothing literature is to take the estimated mean profile at a

particular point plus or minus twice the estimated standard error. With the mixed model formulation, this type of operation can be seen to be equivalent to a simple prediction interval approach used in regression.

Consider the simple case of a single spline component with additional fixed and random components. The mixed model equations will be of the form

$$\begin{bmatrix} \boldsymbol{X}_s^T\boldsymbol{R}^{-1}\boldsymbol{X}_s & \boldsymbol{X}_s^T\boldsymbol{R}^{-1}\boldsymbol{Z}_s & \boldsymbol{X}_s^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}_s^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}_s^T\boldsymbol{R}^{-1}\boldsymbol{X}_s & \boldsymbol{Z}_s^T\boldsymbol{R}^{-1}\boldsymbol{Z}_s + \gamma_s^{-1}\boldsymbol{G}_s^{-1} & \boldsymbol{Z}_s^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}_s^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X}_s & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z}_s & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X}_s & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z}_s & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_s \\ \tilde{\boldsymbol{u}}_s \\ \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{u}} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{X}_s^T \\ \boldsymbol{Z}_s^T \\ \boldsymbol{X}^T \\ \boldsymbol{Z}^T \end{bmatrix} \boldsymbol{R}^{-1}\boldsymbol{y} \qquad (25)$$

where the subscript s refers to the spline and $\boldsymbol{\beta}$ and $\boldsymbol{u}$ represent additional fixed and random effects that may be present. If $\boldsymbol{C}$ denotes the matrix on the left hand side of (25), $\sigma^2\boldsymbol{C}^{-1}$ is the variance-covariance matrix of prediction errors, Robinson (1991), p. 16, a result that allows approximate prediction or probability intervals to be constructed.

A prediction interval for the spline at one of design points $T = t_i$ say, can be found by using the results

$$\mathrm{E}\left\{\boldsymbol{x}_{s,i}^T\hat{\boldsymbol{\beta}}_s + \boldsymbol{z}_{s,i}^T\tilde{\boldsymbol{u}}_s - (\boldsymbol{x}_{s,i}^T\boldsymbol{\beta}_s + \boldsymbol{z}_{s,i}^T\boldsymbol{u}_s)\right\} = \boldsymbol{0} \qquad (26)$$

$$\mathrm{var}\left\{\boldsymbol{x}_{s,i}^T\hat{\boldsymbol{\beta}}_s + \boldsymbol{z}_{s,i}^T\tilde{\boldsymbol{u}}_s - (\boldsymbol{x}_{s,i}^T\boldsymbol{\beta}_s + \boldsymbol{z}_{s,i}^T\boldsymbol{u}_s)\right\} = \sigma^2\boldsymbol{a}^T\boldsymbol{C}^{-1}\boldsymbol{a} \qquad (27)$$

where $\boldsymbol{x}_{s,i}^T = [1\ t_i]$, $\boldsymbol{z}_{s,i}^T$ is the row of $\boldsymbol{Z}_s$ corresponding to t_i, and $\boldsymbol{a} = [\boldsymbol{x}_{s,i}^T\ \boldsymbol{z}_{s,i}^T]$. Thus an interval can be constructed in the manner of regression. This is equivalent to the Bayesian intervals of Silverman (1985) which has pointwise coverage, but not equivalent to the intervals of Wahba (1983) which have the required coverage across the design space. Because an explicit expression is available for the estimate of a spline at any $T = t$, see equation (10), pointwise confidence intervals can be found for any t.

The intervals constructed using (26) and (27) assume the variance parameters are known. Kenward & Roger (1997) provide methods for adjusting the standard errors of fixed effects in a general linear model, allowing for the estimation of variance components; these methods are however computationally intensive. Work to provide efficient computation of the adjustments and to incorporate these ideas for BLUPs is in progress.

5.3 *Tests of hypotheses*

While estimation of effects is usually of prime importance, tests of hypotheses will inevitably be required to assess the significance of effects. For example, tests may be required in the model selection process.

Tests of significance of random effects are carried out using residual maximum likelihood ratio tests (REMLRT). Most tests of interest involve the significance of individual variance components. For example, the lack of fit of a spline and the significance of the spline-random-component, involve testing if the corresponding variance component is

significantly different from zero. In these cases, the null hypothesis specifies that the variance component is zero, and hence on the boundary of the parameter space.

Non-regular likelihood ratio theory is required when testing hypotheses where the parameter is on the boundary of the parameter space; see for example Chernoff (1954), Moran (1971) and Self & Liang (1987) who discuss this general problem, while Stram & Lee (1994) (see also the correction, *Biometrics* **51**, p. 1196) consider the specific issue of tests concerning variance components and random coefficients. For a single variance component, Stram & Lee (1994) state that the asymptotic distribution of the likelihood ratio test is a 50:50 mixture of chi-square variates on zero degrees of freedom (which has probability 1 at the value of zero) and 1 degree of freedom. Thus the approximate P-value for the statistic $D = -2\log\Lambda$ is easily calculated as $1 - \{0.5 + 0.5\Pr(\chi^2_1 \leq d)\} = 0.5\{1 - \Pr(\chi^2_1 \leq d)\}$ where d is the observed value of D; thus the P-value is actually half the standard asymptotic P-value.

For testing that k variance components are zero, the mixture involves chi-square variates from 0 to k degrees of freedom. The mixing probabilities depend on the geometry of the situation through the expected information matrix for the outer model evaluated at the null hypothesis; see for example case 7 of Self & Liang (1987).

Fixed effects are examined using Wald statistics (again the approach of Kenward & Roger, 1997 has not yet been implemented) and an asymptotic chi-square for the null distribution.

Tests involving both fixed and random effects, for example testing the equality of two or more splines, are likely to be of interest in many applications. If the spline estimates are available at the same values of the variable T, Wald statistics based on (26) and (27) could be used. If T differs, it may be necessary to examine the equality sequentially, firstly by testing for a common random-spline-component using a REMLRT, and then using a Wald test for the fixed effects. In fact this sequential approach may highlight interesting patterns.

5.4 *Tests of hypotheses: simulations*

A small simulation study was carried out to examine the mixture approximation in small samples. 5000 simulations were run for each of two configurations (10 groups of 4 observations, 20 groups of 2 observations) and examination of the test of no spline component when data were generated by a straight line model and the test for departures from smoothness when the underlying curve was firstly a spline and secondly a straight line.

The results of this small study are presented graphically in Figure 6. The labels indicate the configuration and the test (RLOF, spline or RLOF in the presence of a spline). The mixture representation gives a conservative test in all cases, in that the nominal theoretical levels are greater than the empirical levels. This is at least comforting. It is likely that in large samples the approximation will become rather more accurate. Notice that the proportion of zero values is larger that 0.5 in all cases. This is in part because the underlying quadratic approximation may not very good for small sample sizes, estimation on the boundary can be difficult, and the actual level of convergence of the iteration process will vary across simulations.

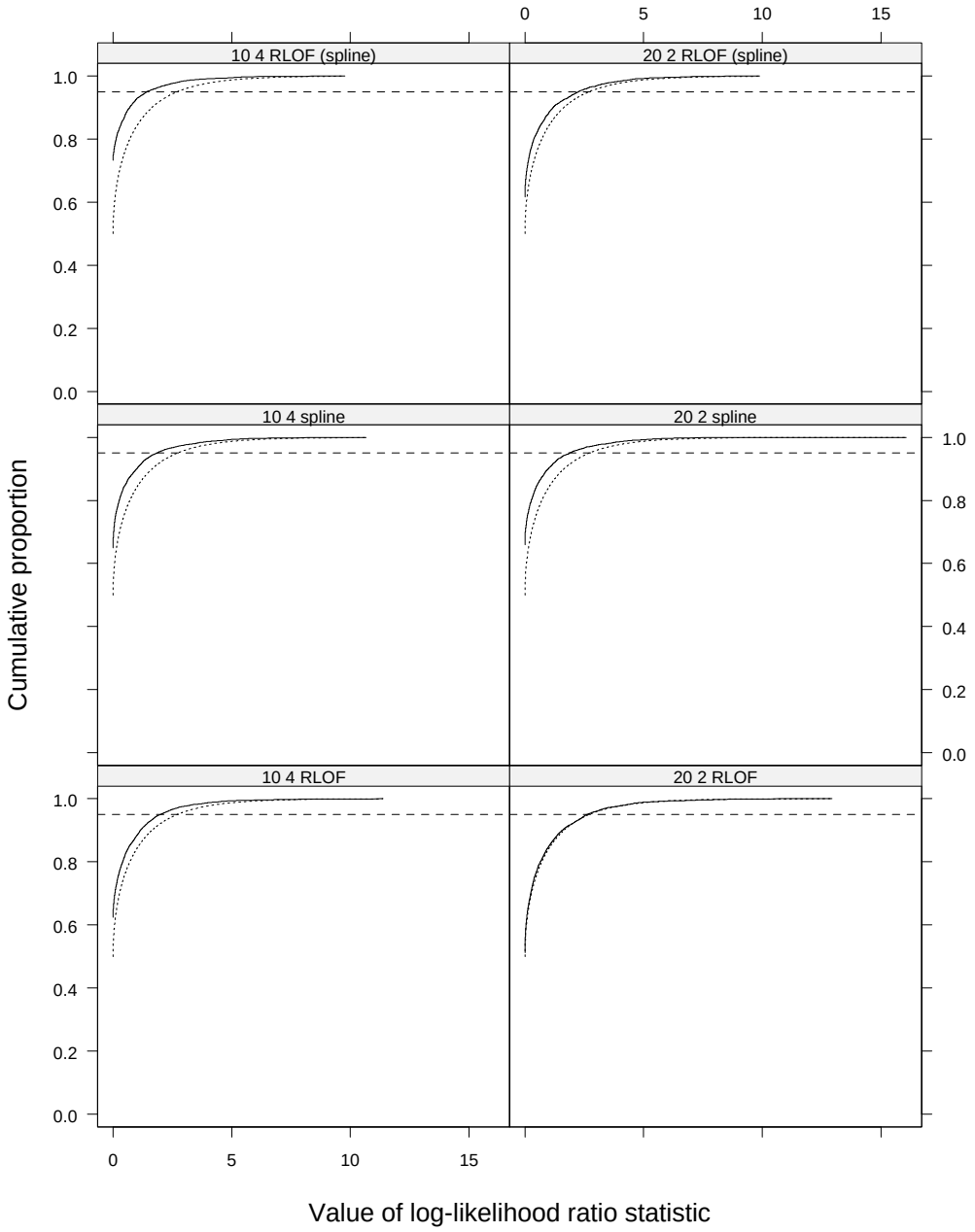

Fig. 6 REML likelihood ratio test: A simulation study. The bold line represents the empirical cumulative distribution function from the simulations, while the curved broken line is the asymptotic cumulative distribution function and the horizontal broken line is the 0.95 probability line. The simulations are, test of spline component: 10 groups of 4 observations; test of spline component: 20 groups of 2; test of random component: 10 groups of 4 observations; test of random component: 20 groups of 2 observations; 10 groups of 4 with a spline; and 20 groups 2 with spline.

5.5 *Calibration*

For the ELISA data of section 2.2, suppose we wish to estimate the concentration for a given the value of the absorbance. If y_0 denotes the given absorbance and t_0 denotes the concentration, then given a spline fit, a simple approach to determine t_0 is to solve

$$\hat{g}(t_0) = y_0 \tag{28}$$

Because the estimated cubic spline $\hat{g}$ is such a simple function which is monotone in the context of calibration (if it is not monotone the whole procedure of calibration must be placed in some doubt), numerical solution of (28) is quite easy. The determination of an interval estimate of t_0 remains.

For nonlinear models there are basically two approaches to this problem. The first is to use a Taylor series expansion as in Davidian & Giltinan (1993) to provide an approximate interval estimate; in simulation studies for nonlinear mixed effects models, Davidian & Giltinan (1995), p. 294, found that the linearization approach was very poor for coverage and near the boundaries. The second approach involves inversion of a prediction interval to determine the corresponding interval for t_0; Davidian & Giltinan (1995) found this to be the best in their simulation study. This second approach is considered for the cubic smoothing spline.

For a single $g(t)$, if $\boldsymbol{a}_0^T = [\boldsymbol{x}_{s,0}^T \; \boldsymbol{z}_{s,0}^T]$ are the fixed and random component design vectors at $t = t_0$ and $\sigma^2\omega_0$ is the variance of y_0 (ω_0 allows for dependence on t_0 and other variables), and

$$\sigma_0^2 = \text{var}\,(y_0 - \hat{g}(t_0)) = \sigma^2 \left(\omega_0 + \boldsymbol{a}^T \boldsymbol{C}^{-1} \boldsymbol{a}\right)$$

is the prediction variance, with $\boldsymbol{C}$ as in section 5.2, then for given variance components and under Gaussian assumptions, unconditionally

$$y_0 - \hat{g}(t_0) \sim N(\boldsymbol{0}, \sigma_0^2)$$

and an approximate $100(1-\alpha)\%$ interval for t_0 can be found by solving

$$\frac{(y_0 - \hat{g}(t_0))^2}{\hat{\sigma}_0^2} < \chi^2_{1-\alpha}(1) \tag{29}$$

where $\chi^2_{1-\alpha}(1)$ is the $1-\alpha$ probability point of the chi-square distribution on 1 degree of freedom. A numerical procedure is required to solve (29).

6 Examples

6.1 *Example 1: Liveweights of cows*

The model to be fitted to the liveweights of cows is motivated by the features discussed in section 2.1. Table 3 provides and analysis of variance table which is sub-divided into Time, Animal and Time by Animal effects, with each of these effects subsequently sub-divided into appropriate components.

The overall smooth pattern of growth is modelled using a spline at the Time level. Thus the terms lin(time) and spl(time) are listed and the RLOF term ran(time) is also included. The Animal effects are made up of the factorial treatment structure and the

random intercept term. The Time by Animal effects are subdivided into smooth patterns over time in the factorial treatments, as given by lin(time) by main effects and interactions, and spl(time) by main effects and interactions. A treatment by ran(time) is included for departures from smoothness at the treatment level, together with the random slope (and correlation with the random intercept term) for animals. Finally the splines are allowed at the animal level to pick up the curvature of individual animals see in Figure 2.

In the model specified by Table 3, there are 8 estimable fixed effect parameters, and each random term, R, RLOF or RC, provides a single variance (or covariance) component. Thus there are 11 variance/covariance components in the full model whereas a fully parameterised covariance matrix would have 276 parameters. The random-spline components for the factorial effects are generated for all levels of the main effects and interactions as are RLOF components. Splines (and RLOF effects) for each level within a main effect or within an interaction are taken to have the same variance component and hence the same smoothing parameter. If necessary, it is possible to provide for a more highly parameterized underlying covariance structure by introducing further variance components.

TABLE 3
Cow data: AOV Decomposition

Term	Decomposition	Type	df or ne
Constant	1	F	1
Time			
	lin(time)	F	1
	spl(time)	R	21
	ran(time)	RLOF	23
Animal			
	treatments (trt)		
	infection	F	1
	iron	F	1
	infection.iron	F	1
	animal	RC	26
Animal.Time			
	trt.lin(time)		
	infection.lin(time)	F	1
	iron.lin(time)	F	1
	infection.iron.lin(time)	F	1
	trt.spl(time)		
	infection.spl(time)	R	42
	iron.spl(time)	R	42
	infection.iron.spl(time)	R	84
	trt.ran(time)	RLOF	92
	animal.lin(time)	RC	26
	(correlation)	RC	
	animal.spl(time)	R	546
Error		R	

Various models were fitted to the data and are summarized in Table 4. The significance of effects was examined by testing the significance of the corresponding (single) variance component using the REMLRT of section 5.3. Maximized residual log-likelihoods, REMLRTs and P-values based on the mixture distribution of section 5.3 are also given.

TABLE 4
Cow data: models fitted and tests of significance

Model	Terms	log-likelihood	$-2\log\Lambda$	P-value
1	full	-1575.75		
2	1 -animal.spl(time)	-1774.22	396.94	0
3	1 -trt.ran(time)	-1586.85	22.20	0
4	1 -ran(time)	-1580.45	9.40	0.001
5	1 -ran(time)-trt.ran(time)	-1593.36	35.22	0
6	1 -infection.iron.spl(time)	-1575.75	0	1
7	1 -iron.spl(time)	-1575.75	0	1
8	1 -infection.spl(time)	-1578.92	6.34	0.028
9	1 -iron.spl(time)-infection.iron.spl(time)	-1575.75	0	1
10	1 -iron.spl(time)-infection.iron.spl(time)†	-1583.30		

†Model 10 has reduced fixed effects. The other models have the fixed effects of Table 3.

The test of a spline or curvature effect at the animal level is highly significant (model 2 against model 1). The animal spline effect is illustrated by examining plots of studentized residuals. Figure 7 is a trellis plot of the studentized residuals after fitting model 1. There is no evidence of systematic behaviour. Figure 8 is the corresponding plot when splines at the animal level are omitted. Several animals exhibit regular patterns in the residuals over time highlighting the difference in curvature in animal growth. Figure 9 gives the estimated semi-variograms (Diggle et al., 1994) of the two models, the left panel for the model including the animal spline term and the right panel omitting animal splines. The first plot suggests a constant estimated variance and little apparent remaining correlation, while the second shows remaining correlation structure in the residuals from the reduced model.

Figure 10 is a plot of studentized residuals against fitted values. The variance heterogeneity evident in Figure 1 has been accounted for by the animal random coefficients and animal splines and has been achieved without a transformation.

There are significant departures from smooth behaviour, as indicated by the REMLRTs for model 3 and model 4 against model 1 (Table 4). These departures may be due to effects not measured or out of control to the experimenter. Furthermore, while they occur, they may not of prime interest and their inclusion allows estimation of the smooth treatment effects over time.

The focus of the experiment was the effect of infection and iron dosing. Firstly, the splines for iron by infection interactions and the iron main effect are not significant (models 6 and 7 against model 1); the variance components for these effects converged to zero. The splines for the infection factor are however significant (model 8 compared to model 1). Therefore curvature effects due to treatments are restricted to the infection factor. The final model (model 10) is model 9 with reduced fixed effects as discussed

below. The residual log-likelihood for model 10 is not comparable to models 1 to 9 because the fixed effects model differs.

In standard linear mixed models, the variance components pre-multiply known matrices with unit diagonal elements. However, both spline-random-effects and random coefficients result in non-unit diagonal elements in the contributing matrices. Variance component estimates for several models fitted are given in Table 5 and are adjusted by multiplying by the mean of the diagonal elements to indicate the relative contribution of the components. Removal of terms can dramatically alter the estimates of the variance components that remain. For example, the overall spline variance component for model 1 converged to zero. This is in line with the visual impression that the growth is predominantly linear over the age span of the cows. However, removal of the lack of fit terms ran(time) and trt.ran(time) (model 5) produces a large component for the overall spline and the splines for the infection factor. These large components imply a rough rather than smooth fit and can therefore be seen to be doing the work of the lack of fit terms in this reduced model. Including lack of fit terms is clearly important in this application.

TABLE 5
Cow data: adjusted variance component estimates

Stratum	Term	Model 1	Model 2	Model 5	Model 9	Model 10
Time						
	spl(time)	0.00	44.29	2398.4	0.00	0.00
	ran(time)	9.59	9.02	-	10.07	10.08
Animal						
	animal	117.5	111.6	117.5	118.2	129.7
Time.Animal						
	infection.spl(time)	201.2	92.39	1526.2	172.0	202.2
	iron.spl(time)	0.00	0.00	0.00	-	
	infection.iron.spl(time)	0.00	10.84	-	-	-
	trt.ran(time)	6.93	0.00	-	6.58	6.67
	animal.lin(time)	1463.2	1424.2	1464.9	1471.9	1320.7
	(correlation)	(-0.107)	(-0.109)	(-0.107)	(-0.107)	(-0.160)
	animal.spl(time)	178.5	-	160.1	172.0	172.1
Error	Residual	20.63	89.30	23.99	20.91	20.92

Estimates of fixed effects and their standard errors are given in Table 6 for models 9 and 10. For model 9, the only significant effects are the overall linear trend and the infection factor intercept and slope effect. If all interactions involving iron and the linear iron effect are removed (model 10), the main effect for iron becomes significant. The approximate Wald tests used should however be viewed with some caution because they do not include adjustment of the standard errors as in Kenward & Roger (1997). Furthermore, model selection is likely to mean that the stated standard errors are too small.

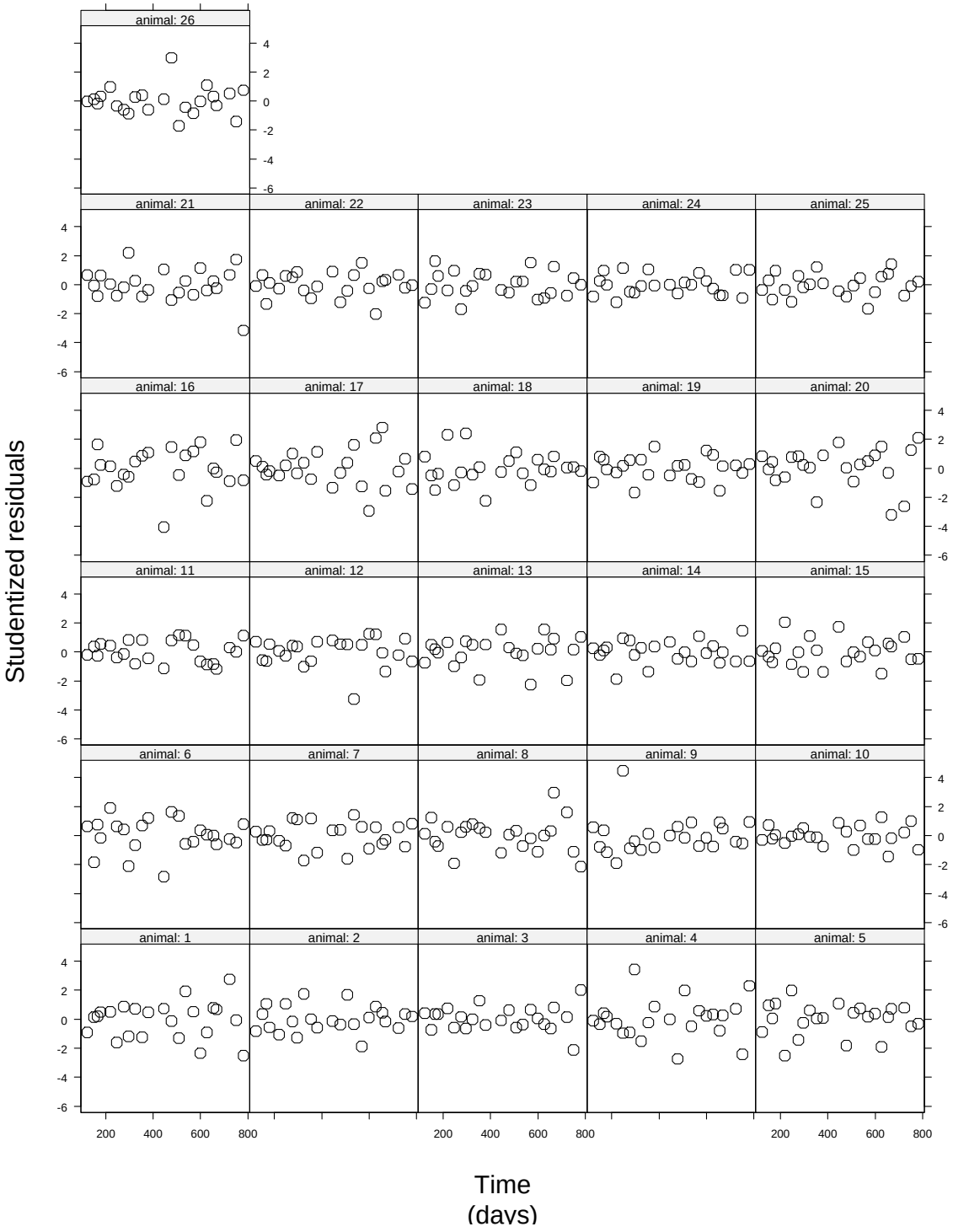

Fig. 7 Cow data: studentized residuals against time for the full model

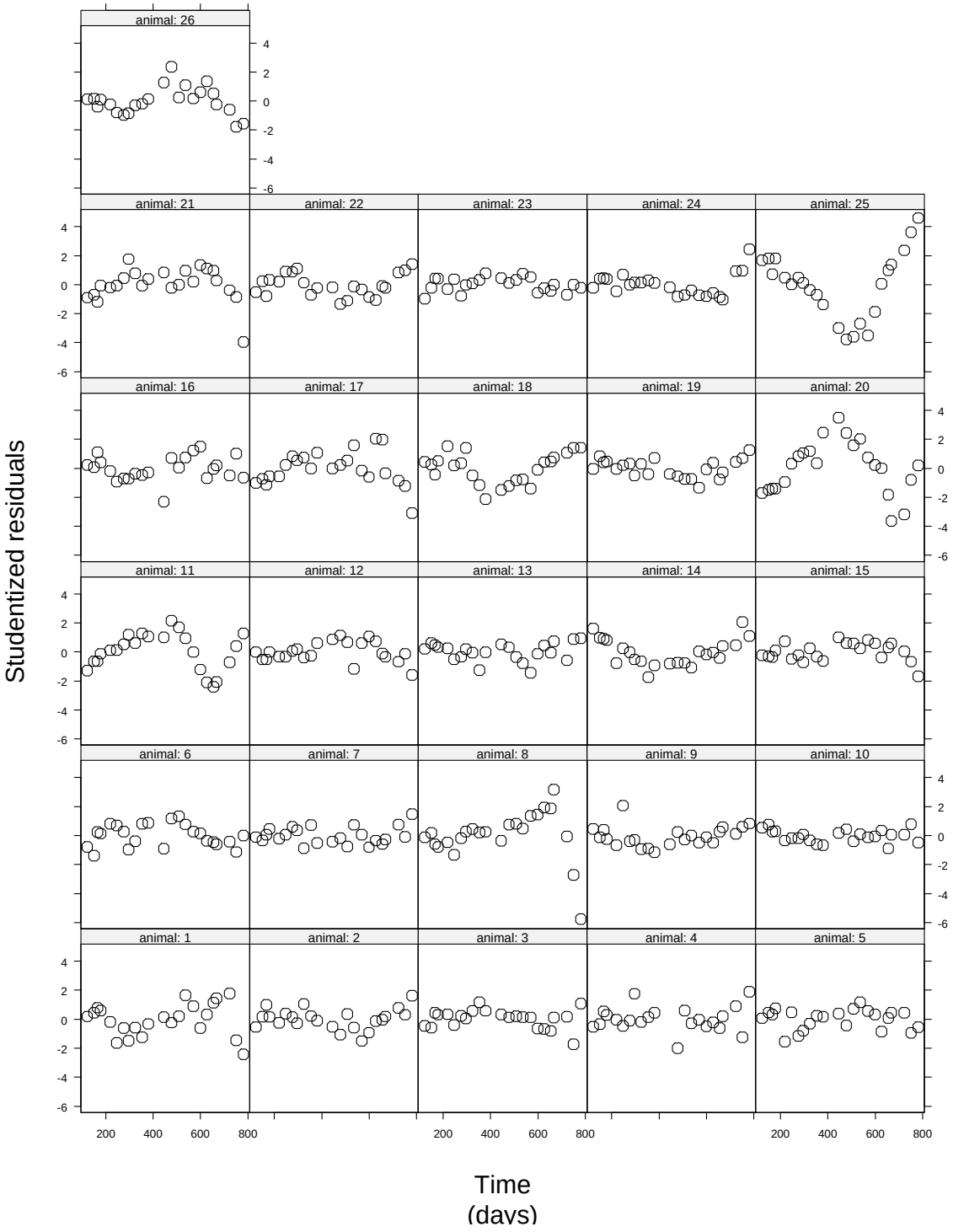

Fig. 8 Cow data: studentized residuals against time for the model without animal splines

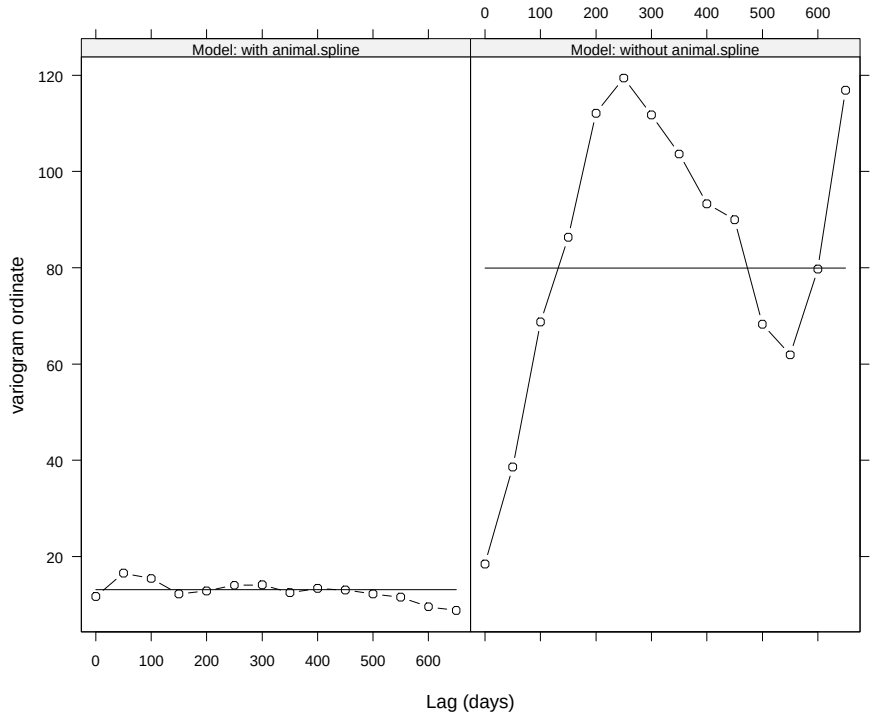

Fig. 9 Cow data: variograms for the model with animal splines, and without animal splines.

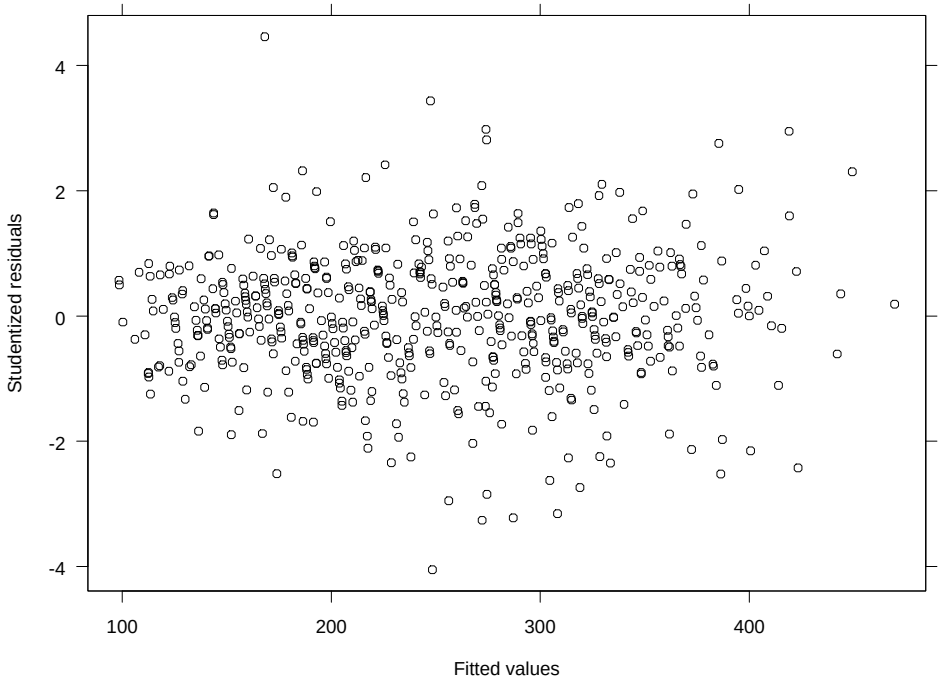

Fig. 10 Cow data: studentized residuals against fitted values for the full model

TABLE 6
Cow data: estimates of fixed effects for models 9 and 10 together with Wald statistics

	Model 9			Model 10		
Term	Est	SE	Est/SE	Est	SE	Est/SE
Constant	279.31	6.39	43.77	283.62	5.21	54.47
lin(time)	40.98	6.34	6.48	39.99	4.66	8.58
infection	-28.68	7.71	-3.72	-34.75	5.57	-6.24
iron	-13.82	9.41	-1.47	-21.32	5.03	-4.25
iron.infection	-11.48	11.10	-1.04	-	-	-
infection.lin(time)	-4.58	7.68	-0.60	-7.27	5.51	-1.32
iron.lin(time)	-2.66	9.36	-0.28	-	-	-
iron.infection.lin(time)	-4.32	11.04	-0.39	-	-	-

Figure 11 is the smooth estimate of the infection contrast, with approximate 95% confidence bands. Infection produces lower log-liveweights at essentially an increasing rate over time. There is an upturn in the curve around 530 days that does not have a physical explanation.

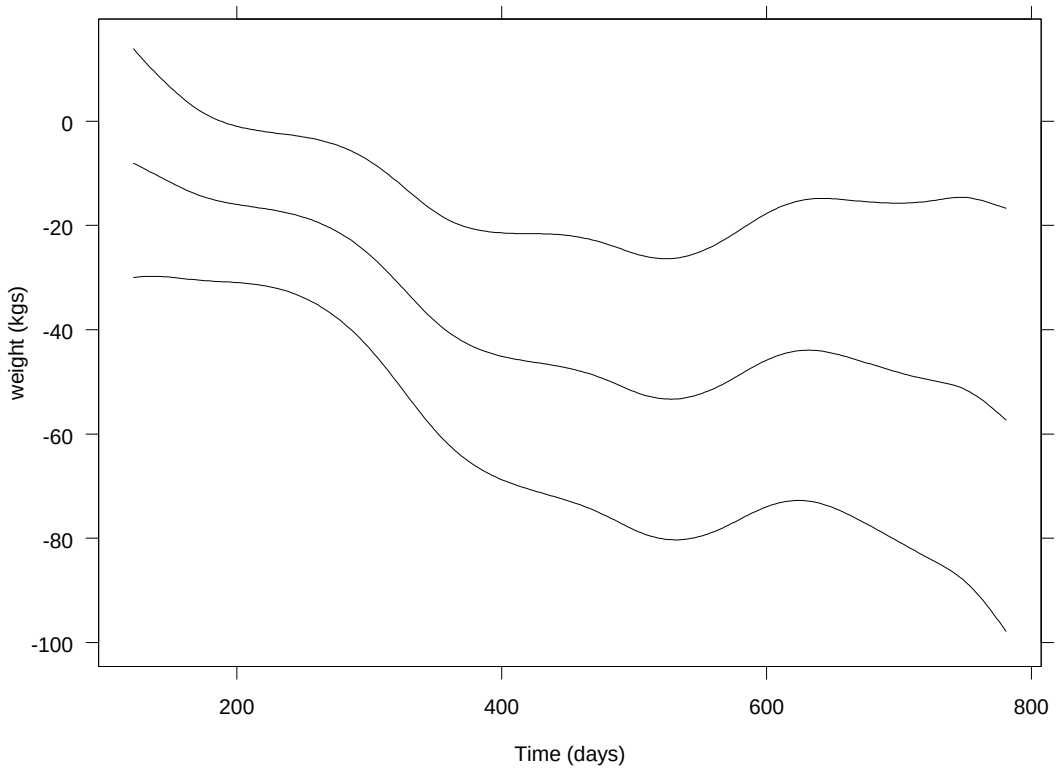

Fig. 11 Cow data: infection contrast with approximate pointwise 95% confidence bands

Figure 12 is a plot of fitted values for the smooth and smooth plus random time effects for the four factorial treatment combinations, both with approximate 95% confidence

bands. The time and time by treatment random components can now be clearly appreciated.

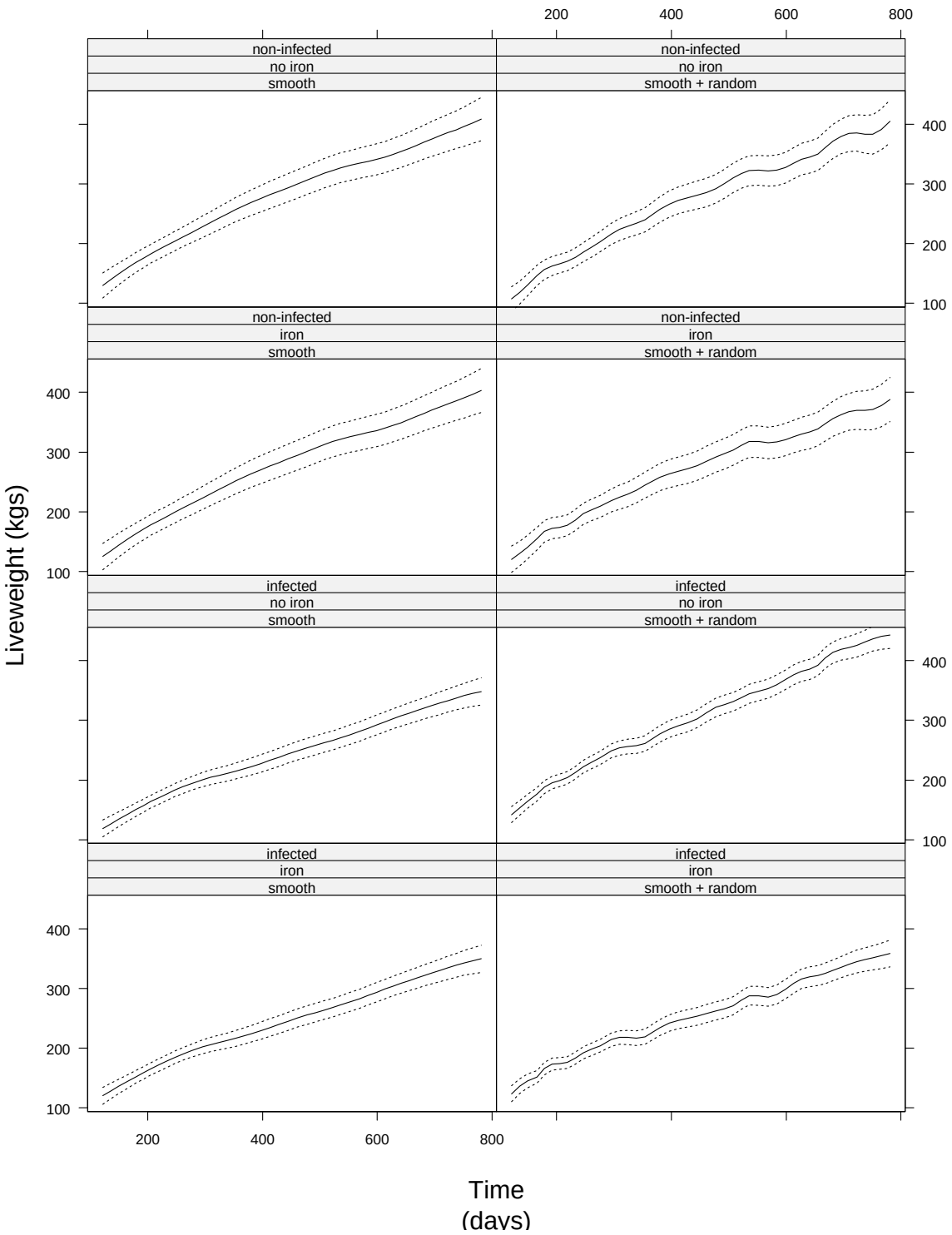

Fig. 12 Cow data: fitted values for the factorial treatment combinations with smooth only and smooth plus random time terms, with approximate pointwise 95% confidence bands

In summary, the overall trend is linear, but there are smooth departures as captured by splines at the animal level, and non-smooth departures as captured by the random time and treatment by time effects. Iron dosing provides a constant negative effect on growth over time, while infection inhibits growth in a nonlinear and increasing manner over time.

The general conclusions that have been drawn with respect to the treatments, qualitatively match those of Diggle et al. (1994) if their results are mapped to the factorial structure used here. The results of our analysis apply to the original scale, do not require explicit covariance modelling, and have the added advantage of allowing for flexible nonlinear smooth trends and departures from that smooth behaviour.

6.2 *Example 2: ELISA assay for recombinant protein DNase in rat serum*

For the assay data of Example 2.2, an analysis of variance decomposition is given in Table 7.

TABLE 7
ELISA assay data: AOV Decomposition

Term	Decomposition	Type	df or ne
Constant	1	F	1
Trial			
	trial	RC	11
Concentration			
	lin(conc)	F	1
	spl(conc)	R	6
	ran(conc)	R	8
Trial.Concentration			
	trial.lin(conc)	RC	11
	(correlation)	RC	
	trial.spl(conc)	R	66
	trial.ran(conc)	R	88
Error	heterogeneous	R	

The conditional or BLUP residuals from a fit of the model given in Table 7 shows clear variance heterogeneity, in line with the analysis by Davidian & Giltinan (1993); see Figure 13. In contrast to these authors, a log-linear model for the variances was fitted, namely

$$\log \sigma^2_{ij} = \theta + \omega_i + \psi_j \tag{30}$$

where i and j index the trials and concentrations respectively. The ω_i and ψ_j require constraints ($\sum \omega_i = 0 = \sum \psi_j$) while in terms of previous development $\sigma^2 = \exp(\theta)$. In (24), $\boldsymbol{\phi} = [\theta \ \boldsymbol{\omega}^T \ \boldsymbol{\psi}^T]^T$ where the non-redundant elements are included in the vectors of the trial and concentration effects, $\boldsymbol{\omega}$ and $\boldsymbol{\psi}$. This variance model has the effect of stabilizing the spread of the resultant residuals, as seen in Figure 13.

Model reduction was carried out by dropping terms from the model of Table 7; details are given in Table 8. If the full model as specified in Table 7 is fitted, the terms for non-smooth random fluctuations are seen to be unimportant (models 2, 3, and 4); this

is not surprising given the data depicted in Figure 3. All remaining terms are required including those in the log-linear model for the error variances (30); the variance effects are particularly strong.

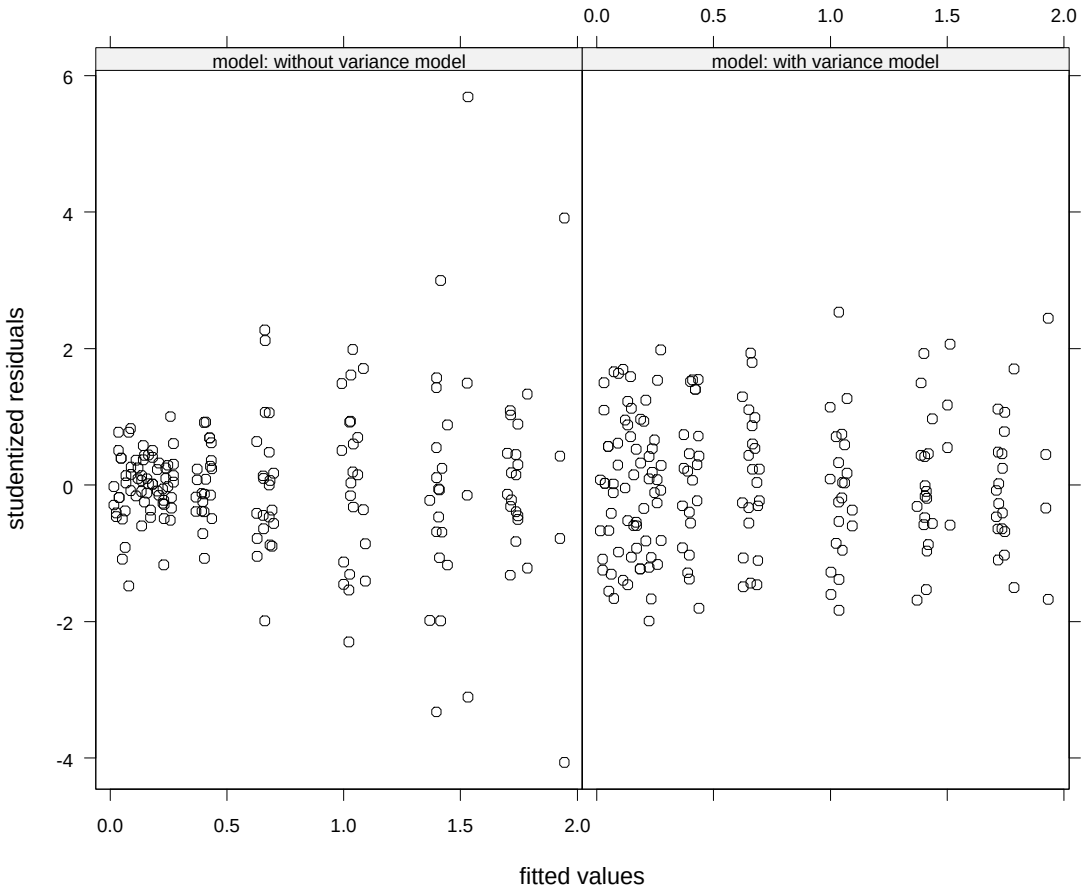

Fig. 13 ELISA assay data: Trellis plot of studentized residuals from the two models, without the variance model and with the variance model.

Figures 14 and 15 are plots studentized residuals against fitted values for the individual trials, both with and without the variance model. The scale is identical in both plots. For the model without variance modelling, there are some very large studentized residuals, particularly for trial three at the two largest concentrations. The variance model certainly repairs this aspect in the subsequent plot but also provides a more uniform spread over concentration. It is clear that a reduced variance model may achieve the same ends. It is also clear that the experimental procedure will occasionally throw up such outlying values, and there is a need to allow for these 'poor' trials. An approach which includes random effects in the variance model may well be the answer. This is not pursued here.

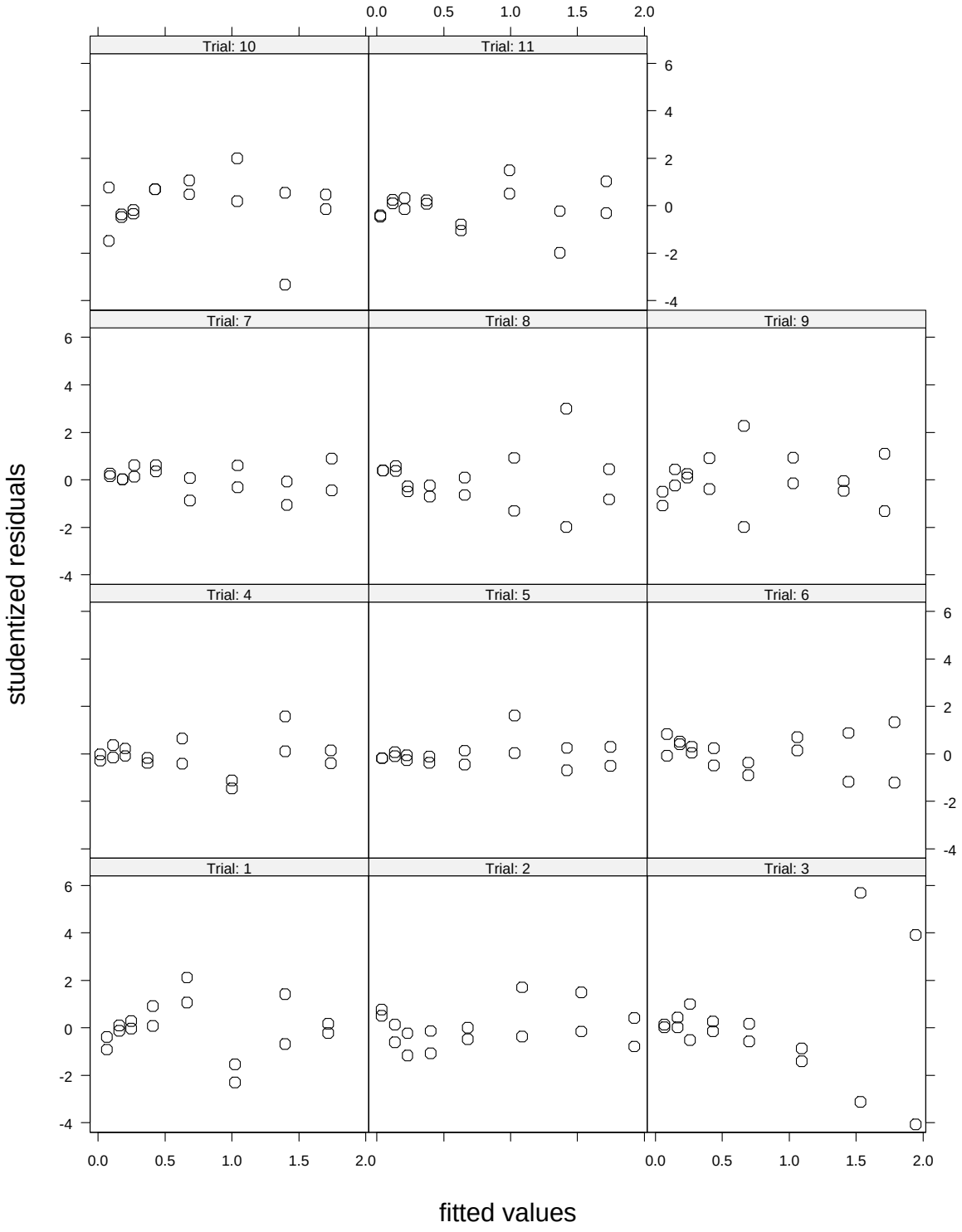

Fig. 14 ELISA assay data: studentized residuals against time for each trial without variance model.

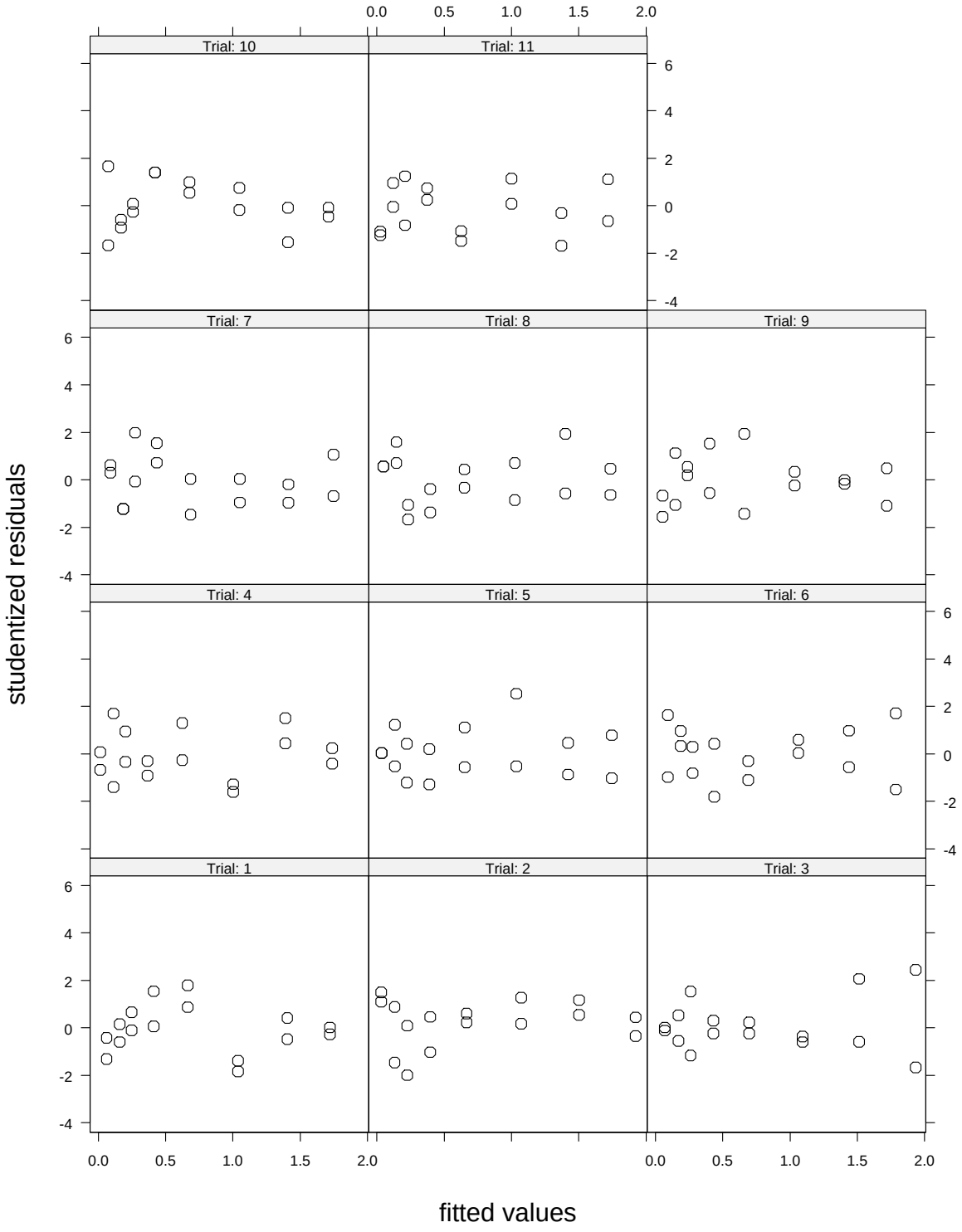

Fig. 15 ELISA assay data: studentized residuals against time for each trial with variance model.

TABLE 8
ELISA assay data: models fitted and REML likelihood ratio tests (a subscript in the P-value denotes the number of zeros)

Model	Experimental	log Variance	log-likelihood	Models	$-2\log\Lambda$	df	P-value
1	full AOV	$\theta+\omega_i+\psi_j$	599.92				
2	1-ran(conc)	$\theta+\omega_i+\psi_j$	599.92	2 v 1	0	1	
3	1-trial.ran(conc)	$\theta+\omega_i+\psi_j$	599.30	3 v 1	1.24	1	0.13
4	1-ran(conc)-trial.ran(conc)	$\theta+\omega_i+\psi_j$	599.30				
5	4	$\theta+\omega_i$	558.39	5 v 4	81.82	7	0
6	4	$\theta+\psi_j$	581.88	6 v 4	34.84	10	0.0_31
7	4	θ	531.71	7 v 4	135.18	17	0
8	4 -trial.spl(conc)	$\theta+\omega_i+\psi_j$	596.26	8 v 4	6.08	1	0.007

Variance component estimates are given in Table 9. The size of the components reflects the strength of the effects present in the data. The difference between models 4 and 7 is most pronounced in the error variance component.

TABLE 9
ELISA assay data: adjusted variance component estimates and variance heterogeneity estimates

		Experimental component $\times 10^4$	
Stratum	Term	Model 4	Model 7
Trial			
	trial	6.22	5.82
Concentration			
	spl(conc)	418.4	416.2
Trial.Concentration			
	trial.lin(conc)	97.84	112.2
	(correlation)	(-0.261)	(-0.088)
	trial.spl(conc)	0.392	0.585
Error		1.701	4.111
	Heterogeneity: variance parameter estimates		
Factor	Model 4: Estimates on log scale (zero sum constraint)		
Trial (i,ω_i)	(1,0.50), (2,0.71), (3,1.14), (4,-0.49), (5,-1.68), (6,-0.55), (7,-0.62), (8,0.17), (9,0.70), (10,0.74), (11,-0.62)		
Concentration (x_j,ψ_j)	(0,-0.70), (0.195,-1.98), (0.39,-1.39), (0.78,-0.76), (1.56,0.59), (3.125,1.35), (6.25,1.89), (12.5,0.99)		

Figure 16 gives the fitted splines and their approximate pointwise 95% confidence bands which highlight the heterogeneity due trials and concentrations. This heterogeneity has implications for calibration.

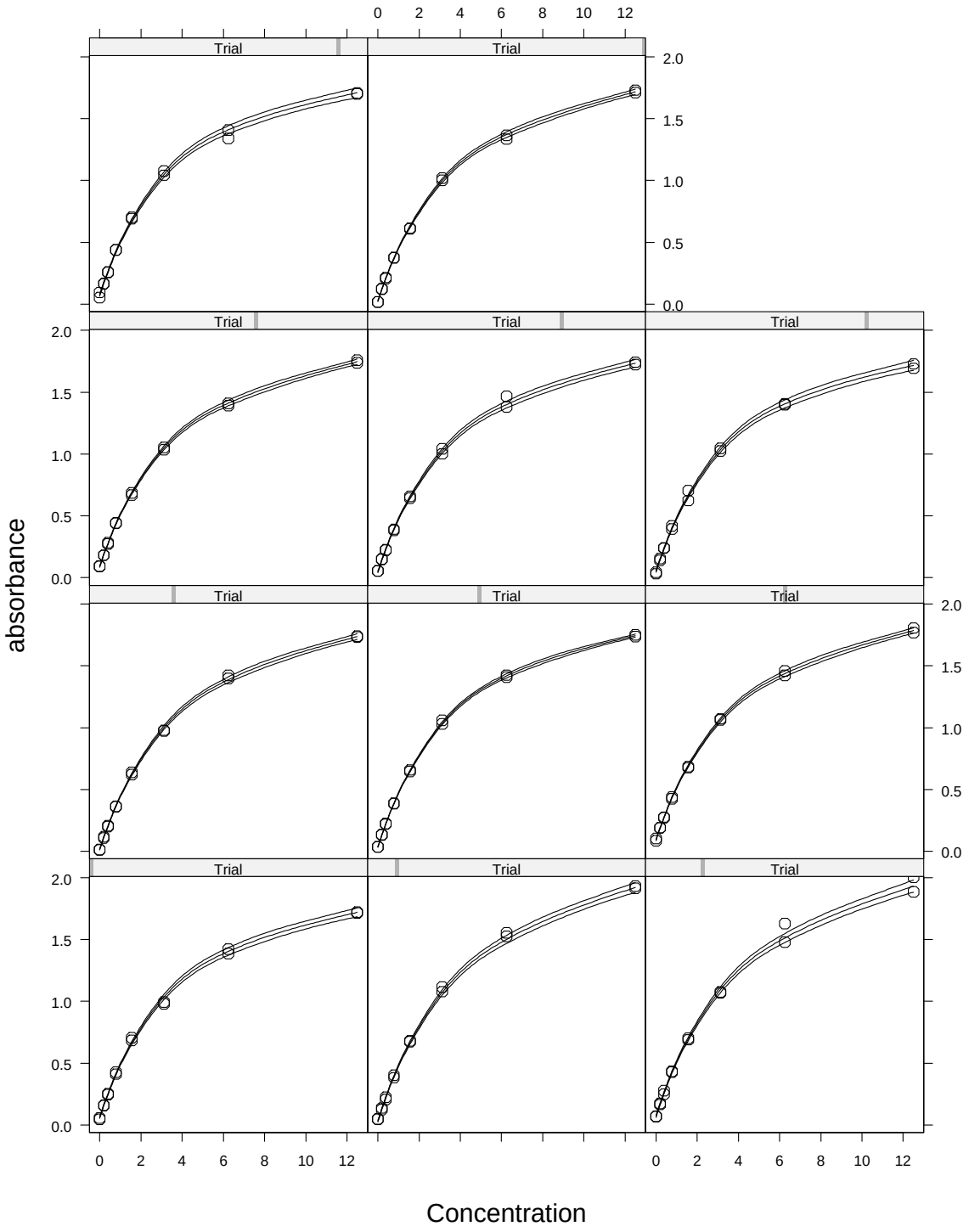

Fig. 16 ELISA assay data: Trellis plot of fitted splines for the eleven trials on optimal density with approximate pointwise 95% confidence bands

The aim of the study was calibration. A difficulty is that the variability depends on both the concentration and the particular trial. Thus an interval estimate for any calibration will depend on both the trial and concentration (the aim of the calibration). In order to find an interval estimate, the estimated variance at each concentration was calculated by averaging over trials on the log scale. The estimated variances at concentrations other than the observed concentrations are found by simple linear interpolation. The estimates and calibrations using this approach are given in Table 10. The second set is based on a model with constant variance. The estimates and the calibration intervals are similar, though the constant variance model produces intervals that are too narrow for large concentrations and too wide at low concentrations.

TABLE 10
ELISA assay data: calibration point and interval estimates

	Model 4			Model 7		
		Interval			Interval	
y_0	Estimate	Lower	Upper	Estimate	Lower	Upper
0.10	0.091	0.052	0.12	0.093	0.0043	0.17
0.25	0.41	0.37	0.46	0.41	0.32	0.51
0.50	1.04	0.94	1.14	1.04	0.92	1.16
0.75	1.88	1.69	2.12	1.86	1.70	2.07
1.00	2.92	2.70	3.13	2.93	2.70	3.13
1.25	4.45	3.45	5.08	4.48	3.84	4.97
1.50	7.30	6.25	7.84	7.18	6.25	7.27
1.75	12.05	10.40	14.12	12.01	10.52	13.79

6.3 *Example 3*

As for the previous examples, an analysis of variance decomposition is presented; see Table 11. The first part of the table, labelled Experimental, includes terms which account for the aims and structure of the experiment. Thus site (year by location combinations, see Table 1), variety and seeding rate and two and three factor interactions of these factors appear in this section of the table. Seeding rate is subdivided into lin(sr), spl(sr) and ran(sr). Site and site.lin(sr) are taken as fixed effects; these are likely to be large sources of variation as seen in Figures 4 and 5. Variety, variety.lin(sr), site.variety and site.variety.lin(sr) are taken as random coefficients (RC), with an associated correlation between pairs of intercepts and slopes.

The second part of Table 11, labelled Spatial, gives the additional terms that have been included to account for extraneous and global spatial variation. Unlike the first part of the table which is largely determined by the experimental design, this requires analysis of the particular set of trials.

The modelling process begins by determining the appropriate spatial models for each trial. Initially the spatial variation within each trial is modelled by a separable autoregressive process for columns and rows, with different variance and correlation coefficients for each site. Spline terms were not included during this initial model

building phase.

TABLE 11
Lupin data: AOV Decomposition. The Experimental section refers to the design aspects of the experiment. The spatial model terms are all specific to sites. Thus those labelled as fixed (F) are nested, while the terms labelled as random (R) are heterogeneous.

	Experimental		
Term	Decomposition	Type	df or ne
Constant	1	F	1
Site		F	10
Seedrate			
	lin(sr)	F	1
	spl(sr)	R	5
	ran(sr)	RLOF	7
Variety		RC[1]	9
Site.Seedrate			
	site.lin(sr)	F	10
	site.spl(sr)	R	55
	site.ran(sr)	RLOF	77
Variety.Seedrate			
	variety.lin(sr)	RC[1]	9
	(correlation)	RC[1]	
	variety.spl(sr)	R	45
	variety.ran(sr)	RLOF	63
Site.Variety		RC[2]	99
Site.Variety.Seedrate			
	site.variety.lin(sr)	RC[2]	99
	(correlation)	RC[2]	
	site.variety.spl(sr)	R	495
	site.variety.ran(sr)	RLOF	693
	Spatial		
Term	Site	Type	df or ne
site.lin(col)	2, 5, 6, 8, 9, 10	F, nested	6
site.lin(row)	3, 5, 7, 8	F, nested	4
site.fixed row	2, 3, 5, 7, 8, 9, 11	F, nested	21
site.ran(row)	3, 6, 7, 9	R heterogeneous	97
site.ran(col)	1, 2, 3, 7, 9, 10	R heterogeneous	34

Figure 17 presents the sample variograms (scaled by the site trend error variance) of the residuals (see Gilmour et al., 1997) in a wireframe trellis plot for each site. In general the sample variograms are not consistent with the assumed model for spatial variation. Strong cyclic patterns in the rows are evident for several sites, with the most obvious alternating behaviour at site 8 (1992, Mt. Barker). We included a row factor coded as 1,2,3,4,1,2,3,4. . . at those sites where there was this cyclic effect (see Table 11).

The plant breeder suggested that this cyclic pattern probably resulted from a differential germination, dependent on the seeding direction, which was performed in a serpentine manner. Global trend was also modelled by the inclusion of linear row and column covariates whenever the variogram displayed a noticeable drift along the first row or column of the wireframe plot (Table 11 and Figure 17). Additional extraneous effects were required for several sites, and were accounted for by including random row and column effects. These effects are often a result of variations in plot length (column) or serpentine harvesting (row). The reduction in semi-variance for $row = 0$ is indicative of the presence of column effects, for example; see Gilmour et al. (1997) for a more complete discussion. Measurement error variance has been included whenever the component was estimable.

Table 12 presents the sequence of models fitted to these data, beginning with the baseline model, which includes all extra spatial components, such as fixed row, linear row and column and random row and column effects. This baseline model has no spline terms but does have RLOF terms used in the preliminary analysis to determine the spatial models. All REML log likelihoods are comparable as the fixed effects do not change from this base model. Inclusion of spl(sr) and its interactions with site, variety and site.variety reduced all four RLOF variance components to zero and prevented convergence of the estimation procedure because all variance components are constrained to be positive (which can be relaxed for some components but not for splines). Dropping these lack of fit terms, as well as columns for site 7 (1992, East Chapman) and variety.site.spl(sr) gave a REML log-likelihood of 1557.81, with all variance components remaining positive.

TABLE 12
Lupin data: models fitted and REML likelihood ratio tests of significance (subscript in P-value indicates the number of zeros)

Model	Terms	Variance Parameters Spatial, Experimental	log-likelihood	$-2\log\Lambda$	P-Value
1	Base	47, 8	1548.20		
2	1 + spl(sr) - site.variety.ran(sr)	47, 11	1553.2†		
3	2 -site.ran(col)(7)-ran(sr)-site.ran(sr) -variety.ran(sr)-site.variety.spl(sr)	46, 7	1557.81		
4	3 + RC^1, RC^2 correlation	46, 9	1568.43	23.26	0
5	4 - meas. error (10)	45, 9	1568.43	0	1
5(a)	5 - variety.spl(sr)	45, 8	1563.14	10.58	0.0_35
5(b)	5 - site.spl(sr)	45, 8	1566.14	5.58	0.0_29
5(c)	5 -spl(sr)	45, 8	1559.72	19.42	0.0_55

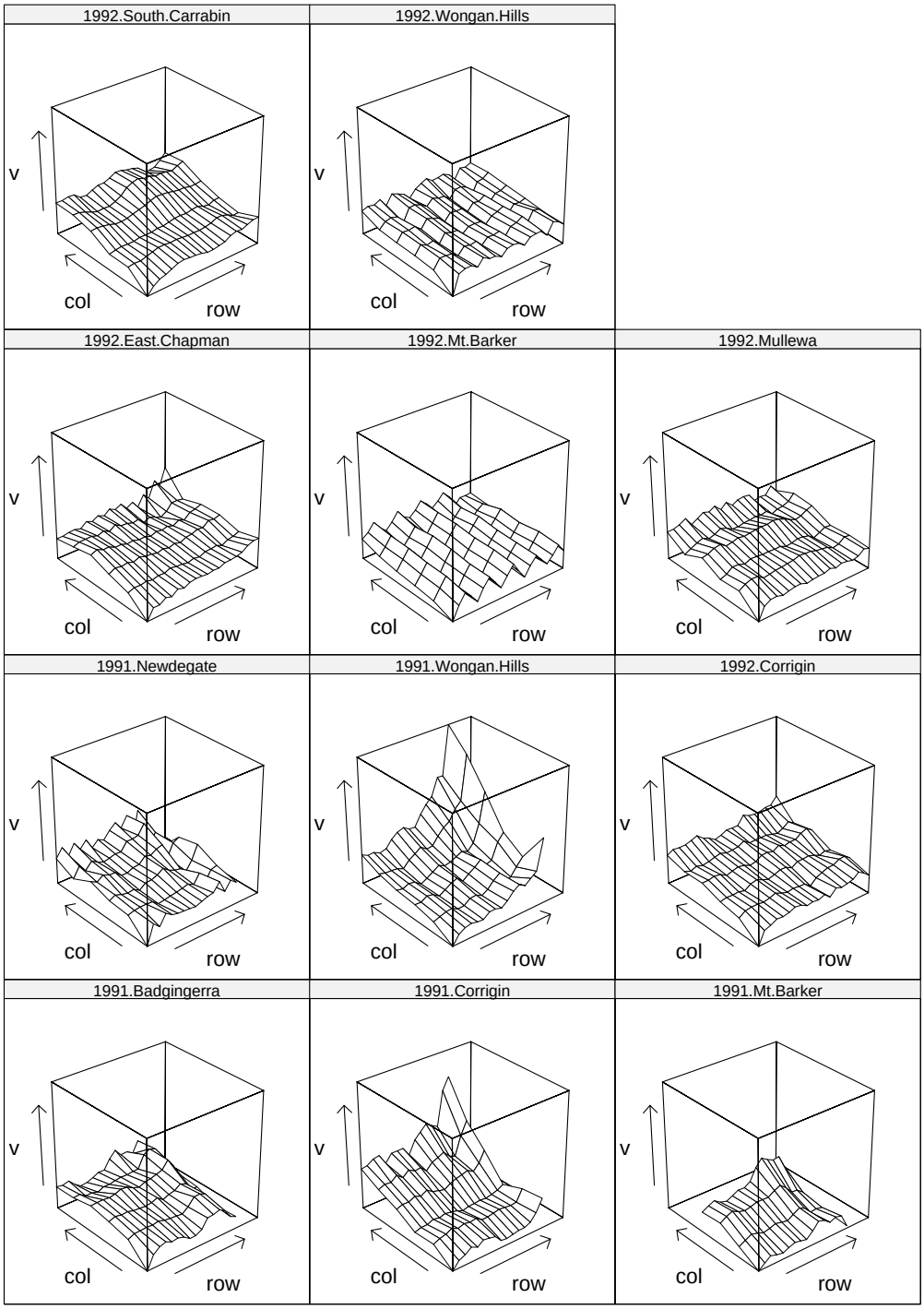

Fig. 17 Lupin data: Wireframe of sample variograms of the residuals from model 1

Inclusion of correlation coefficients for the variety and variety.lin(sr) and site.variety and site.variety.lin(sr), increased the REML log-likelihood by over 10 units, but reduced the measurement error variance for site 10 (1992, South Carrabin) to near zero, and so this was subsequently deleted. Three submodels were then examined by deleting each of the spl(sr) terms in succession (Table 12). No further reduction was possible.

Asymptotic Wald tests of the fixed effects in the model are presented in Tables 13 and 14. All tests are highly significant, with the exception of the linear row effect for site 7 (1992, East Chapman), which could be dropped from the model.

TABLE 13
Lupin data: Wald tests for spatial fixed effects

Site	Year	Location	Term linear row (df=1)	Term linear column (df=1)	Term Fixed row (df=3)
1	1991	Badgingerra	-	-	-
2	1991	Corrigin	-	3.92	23.7
3	1991	Mt. Barker	8.41	-	35.5
4	1991	Newdegate	-	-	-
5	1991	Wongan Hills	3.13	10.96	25.4
6	1992	Corrigin	-	13.99	-
7	1992	East Chapman	1.39	-	20.0
8	1992	Mt. Barker	4.84	8.35	187.2
9	1992	Mullewa	-	5.57	5.84
10	1992	South Carrabin	-	7.24	-
11	1992	Wongan Hills	-	-	52.9

TABLE 14
Lupin data: Wald tests for site and seed rate fixed effects

Term	df	Wald statistic
site.lin(sr)	10	154.8
site	10	193.7
lin(sr)	1	182.2

Tables 15 and 16 presents the REML estimates of the variance parameters. Some further simplification could be achieved by setting some of the autoregressive parameters to zero, particularly for site 4 (1991, Newdegate). This is of little consequence in the analysis and so these have been retained. There was a very strong association between

the intercept and slope for each variety and site by variety combination.

TABLE 15
Lupin data: spatial variance parameter estimates

Site	ran(row)	ran(col)	ρ_c	ρ_r	meas error	error
1	-	0.0104	0.140	0.185	-	0.0327
2	-	0.0250	0.167	0.310	-	0.0400
3	0.00259	0.0817	-0.018	0.471	-	0.0505
4	-	-	0.109	-0.028	-	0.0329
5	-	-	0.084	0.562	0.00439	0.0177
6	0.00113	-	0.221	0.322	-	0.0241
7	0.000189	-	0.570	0.902	0.0132	0.0366
8	-	-	0.177	0.740	0.0129	0.0181
9	0.00453	0.00408	-0.034	0.251	-	0.0173
10	-	0.000458	0.170	0.522	-	0.0103
11	-	-	0.103	0.194	-	0.0540

TABLE 16
Lupin data: adjusted variance component estimates for experimental effects

Term	estimate
spl(sr)	0.00330
variety.spl(sr)	0.000524
site.spl(sr)	0.000396
variety	0.00550
variety.lin(sr)	0.00114
correlation	0.712
site.variety	0.0189
site.variety.lin(sr)	0.000521
correlation	0.986

Figure 18 presents the sample variograms of the residuals from model 5 (again scaled by site trend error variance). Each variogram is more consistent with the assumed error process than for model 1. The spatial correlation in most cases is only weak and short range, except when a measurement error term has been included.

Figure 19 presents a trellis plot of the fitted curves for each variety and approximate 95% confidence bands. There is some divergence of the curves as the plant density increases, with the more recent varieties having higher yields at higher densities. An exception to this behaviour is the variety Warrah, which is not tolerant to dry conditions and performs badly in the arid regions of Australia where some of the trials were conducted. The varieties Merrit and Danja have a pronounced increase in yield as density increases, whereas the other varieties tend to stabilize in terms of yield. This behaviour is consistent with the the hypotheses of the plant breeder.

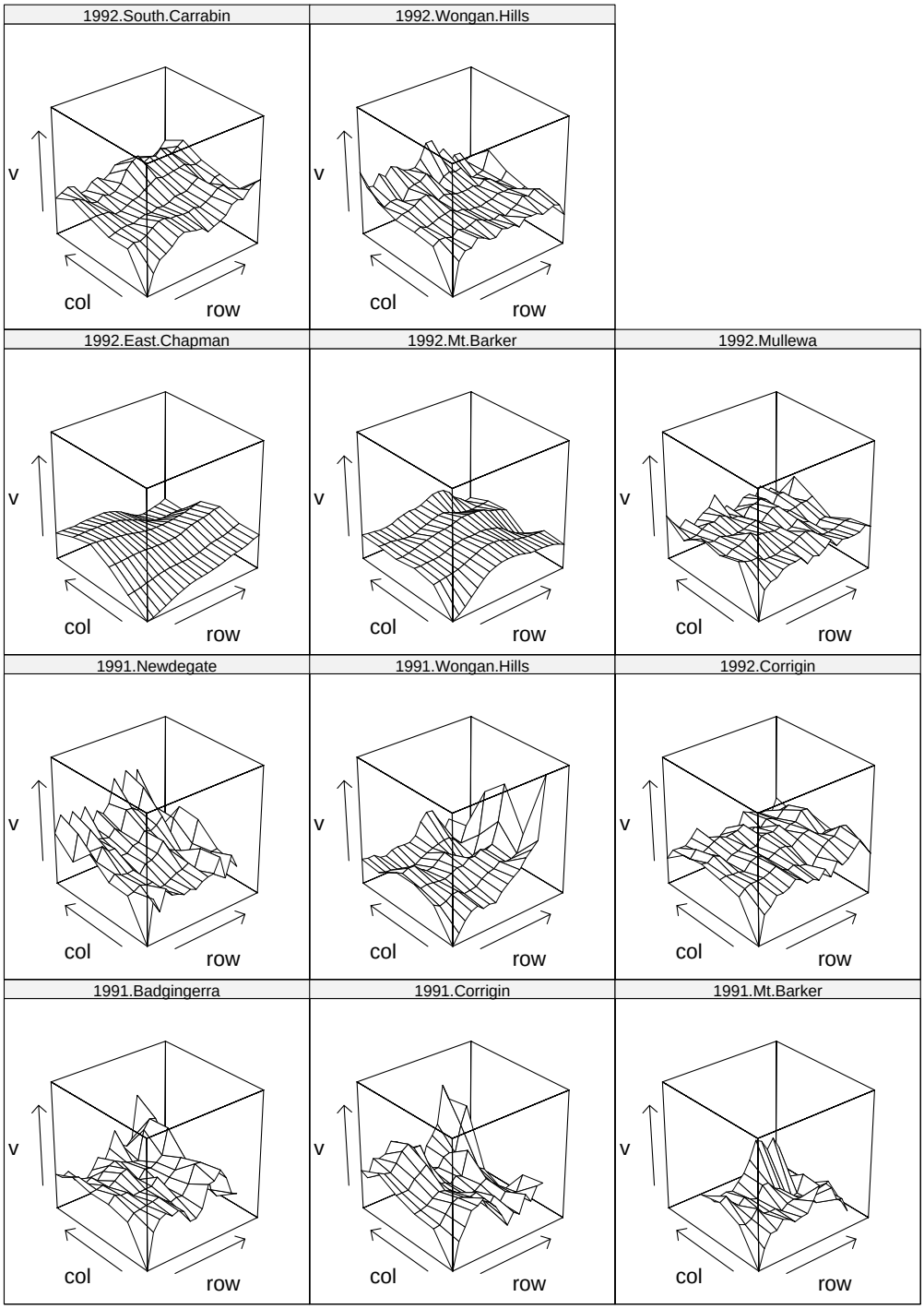

Fig. 18 Lupin data: Wireframe of sample variograms of the residuals from model 5

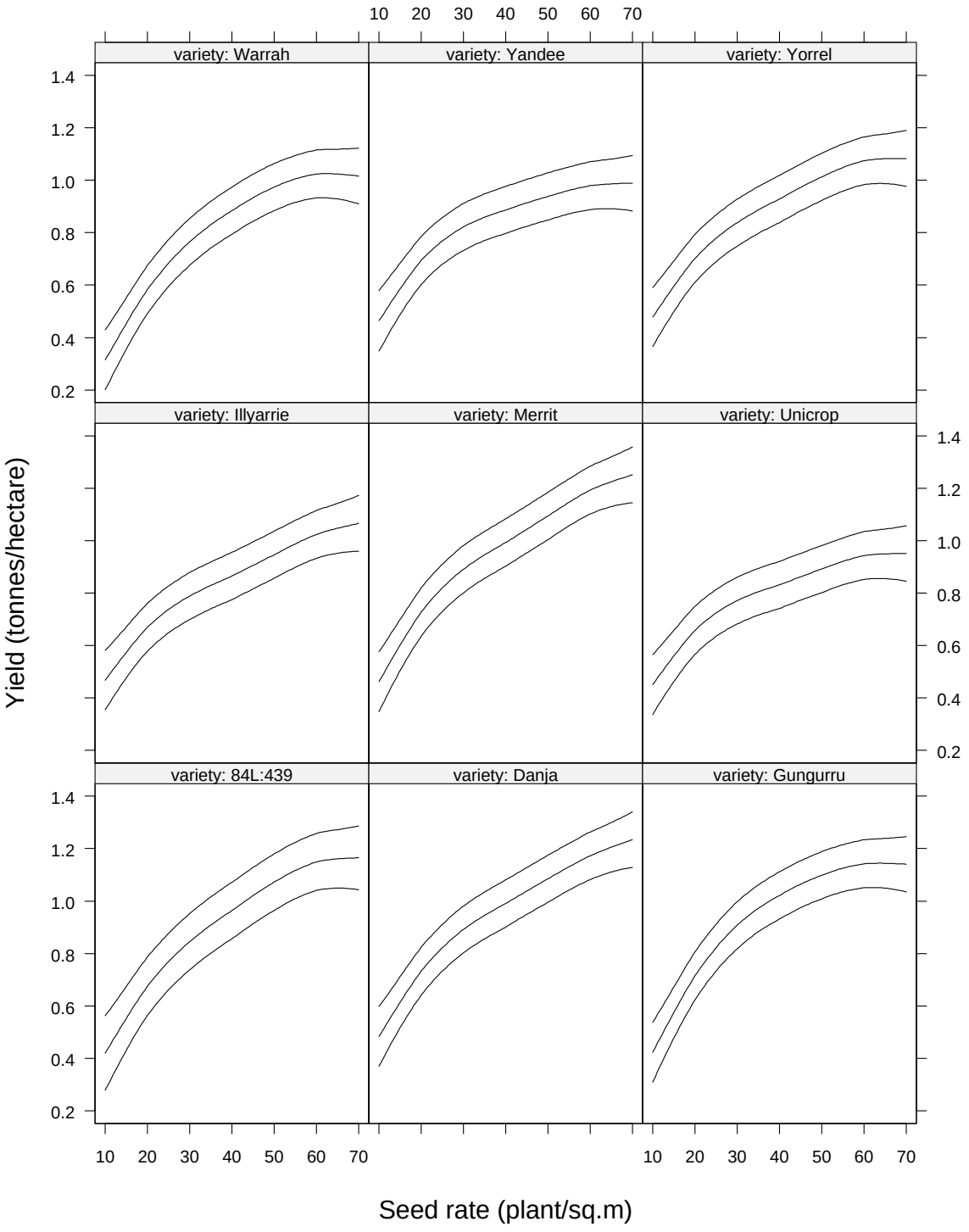

Fig. 19 Lupin data: fitted yield-density curves for each variety from model 5

7 Extensions and Discussion

The examples highlight the flexibility and the wide applicability of the proposed approach. In contrast to nonlinear mixed effects models which have difficulties in both estimation and inference, the linear mixed model approach of this paper provides straightforward estimation and inference for smooth nonlinear relationships in complex experiments and longitudinal studies. In the latter case the need to model temporal correlations is largely overcome. The analysis of variance decomposition provides a further link to standard methods and allows formulation of appropriate models.

The ideas extend to the case of several quantitative explanatory variables. Tensor-splines and thin-plate splines also admit a mixed model formulation. The tensor-splines involve Kronecker products of single structures and as such fit very neatly with one approach to the spatial analysis of field experiments; see Gilmour et al. (1997). Thin plate splines are formulated more simply via the variogram (see Chapter 7 of Green & Silverman, 1994), but in principle can be incorporated into the approach of this paper. Wahba (1990) discusses analysis of variance type spline decompositions, which incorporate higher dimensional splines and those ideas could be extended to allow for additional factorial structure in the manner of this paper.

These ideas can be used in generalized linear models. If splines are incorporated in the linear predictor in the manner of this paper, the resulting model is a generalized linear mixed model; these models are enjoying extensive study, see Schall (1991), Breslow & Clayton (1993), McGilchrist (1994). In particular, the penalized likelihood approach fits in with the hierarchical generalized linear model formulation of Lee & Nelder (1996). The splines introduce correlation structure and so may be useful for non-Gaussian longitudinal data.

Lastly, the availability of software for fitting these models is crucial. Development continues with more efficient algorithms being incorporated (see Appendix B) to provide rapid estimation in very large studies, such as multi-site variety trials.

Acknowledgements

We thank Wallace Cowling for providing the Lupin data and Robin Thompson for help in developing the sparse matrix approach given in Appendix B. We also thank Arthur Gilmour who has developed ASREML, a program for analysing data using most of the ideas of this paper. This program is an outstanding contribution to the area of mixed model analysis. ASREML is available by anonymous ftp at `ftp.res.bbsrc.ac.uk` in `pub/uploads/aar`. We gratefully acknowledge the financial support of the Australian Research Council, The University of Adelaide Research Grant scheme, and the Grains Research and Development Corporation.

Appendix A: Derivation of BLUP form of the cubic smoothing spline

For a mixed linear model,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{e}$$

with $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2\gamma\boldsymbol{G})$ and $\boldsymbol{e} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{R})$, the mixed model equations (see for example Searle et al., 1992, p. 276) are

$$\begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{Z} + \gamma^{-1}\boldsymbol{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}^T\boldsymbol{R}^{-1}\boldsymbol{y} \end{bmatrix} \tag{31}$$

Let $\boldsymbol{C}$ denote the matrix on the left-hand-side of (31), and let the partition of $\boldsymbol{C}$ and hence of $\boldsymbol{C}^{-1}$ implied by (31) be denoted by

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{bmatrix} \quad \text{and} \quad \boldsymbol{C}^{-1} = \begin{bmatrix} \boldsymbol{C}^{11} & \boldsymbol{C}^{12} \\ \boldsymbol{C}^{21} & \boldsymbol{C}^{22} \end{bmatrix}$$

For given γ, $\boldsymbol{G}$ and $\boldsymbol{R}$, the estimate of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ is given by

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{u}} \end{bmatrix} = \boldsymbol{C}^{-1} \begin{bmatrix} \boldsymbol{X}^T \\ \boldsymbol{Z}^T \end{bmatrix} \boldsymbol{R}^{-1}\boldsymbol{y} \tag{32}$$

The estimate of $\boldsymbol{g}$ given by (7) can be written as (recalling $\lambda_s = \gamma_s^{-1}$),

$$\begin{aligned} \hat{\boldsymbol{g}} &= \left\{\boldsymbol{R} - \boldsymbol{R}\boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{R}\boldsymbol{\Delta} + \gamma_s\boldsymbol{G}_s)^{-1}\boldsymbol{\Delta}^T\boldsymbol{R}\right\}\boldsymbol{R}^{-1}\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{R}\boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{R}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^T)\boldsymbol{y} \\ &\quad + \boldsymbol{R}\boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{R}\boldsymbol{\Delta})^{-1}\left\{(\boldsymbol{\Delta}^T\boldsymbol{R}\boldsymbol{\Delta})^{-1} + \lambda_s\boldsymbol{G}_s^{-1}\right\}^{-1}(\boldsymbol{\Delta}^T\boldsymbol{R}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^T\boldsymbol{y} \end{aligned}$$

Using the identity

$$\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{X}_s(\boldsymbol{X}_s^T\boldsymbol{R}^{-1}\boldsymbol{X}_s)^{-1}\boldsymbol{X}_s^T\boldsymbol{R}^{-1} = \boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{R}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}^T$$

and defining $\boldsymbol{Z}_s = \boldsymbol{\Delta}(\boldsymbol{\Delta}^T\boldsymbol{\Delta})^{-1}$, $\hat{\boldsymbol{g}}$ can be written as

$$\begin{aligned} \hat{\boldsymbol{g}} &= \boldsymbol{X}_s\boldsymbol{C}_{11}^{-1}\boldsymbol{X}_s^T\boldsymbol{R}^{-1}\boldsymbol{y} \\ &\quad + (\boldsymbol{Z}_s - \boldsymbol{X}_s\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12})(\boldsymbol{C}_{22}^{-1} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12})^{-1}(\boldsymbol{Z}_s^T\boldsymbol{R}^{-1} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{X}_s^T\boldsymbol{R}^{-1})\boldsymbol{y} \end{aligned}$$

Using standard results on the inverse of partitioned matrices, the expression can be written as

$$\begin{aligned} \hat{\boldsymbol{g}} &= (\boldsymbol{X}_s\boldsymbol{C}^{11}\boldsymbol{X}_s^T + \boldsymbol{X}_s\boldsymbol{C}^{12}\boldsymbol{Z}_s^T + \boldsymbol{Z}_s\boldsymbol{C}^{21}\boldsymbol{X}_s^T + \boldsymbol{Z}_s\boldsymbol{C}^{22}\boldsymbol{Z}_s^T)\boldsymbol{R}^{-1}\boldsymbol{y} \\ &= [\,\boldsymbol{X}_s \quad \boldsymbol{Z}_s\,]\boldsymbol{C}^{-1}\begin{bmatrix} \boldsymbol{X}_s^T \\ \boldsymbol{Z}_s^T \end{bmatrix}\boldsymbol{R}^{-1}\boldsymbol{y} \end{aligned}$$

so that $\hat{\boldsymbol{g}}$ is using (32)

$$\hat{\boldsymbol{g}} = [\,\boldsymbol{X}_s \quad \boldsymbol{Z}_s\,]\begin{bmatrix} \hat{\boldsymbol{\beta}}_s \\ \tilde{\boldsymbol{u}}_s \end{bmatrix} = \boldsymbol{X}_s\hat{\boldsymbol{\beta}}_s + \boldsymbol{Z}_s\tilde{\boldsymbol{u}}_s$$

as claimed in section 3.1.

Appendix B: Computation

The Average Information (AI) algorithm (Gilmour et al., 1995) is used for REML estimation, and the methods of this paper are available as S-PLUS functions (Chambers & Hastie, 1992), in GENSTAT 5.4 (Payne et al., 1995) and as a stand alone FORTRAN program ASREML (Gilmour et al., 1996). Details are available from the authors.

While the mixed model formulation provides the unifying link, using that formulation is computationally inefficient. The software listed above uses sparse matrix manipulations and an absorption/backsubstitution based on the mixed model equations (see Gilmour et al., 1995 and the references therein). This is very efficient for standard mixed model analysis. The following modification of the cubic smoothing spline utilises the tridiagonal nature of matrices involved to permit sparse matrix methods to be used when splines are included in the model. The idea follows Green & Silverman (1994), p. 121. The additional aspect presented here is the separation of linear and random components, similar to the mixed model formulation.

For simplicity of notation, a single spline term in the unreplicated case is considered. If $\boldsymbol{N}_g$ is the replication matrix, the results below are modified in the replicated case by replacing $\boldsymbol{R}^{-1}$ by $\boldsymbol{N}_g^T \boldsymbol{R}^{-1} \boldsymbol{N}_g$.

Write

$$\begin{aligned} \boldsymbol{\eta} &= X_s \boldsymbol{\beta}_s + (\boldsymbol{g} - X_s \boldsymbol{\beta}_s) \\ &= X_s \boldsymbol{\beta}_s + \boldsymbol{g}^*, \quad \text{say} \end{aligned}$$

Then

$$\boldsymbol{\Delta}^T \boldsymbol{g} = \boldsymbol{\Delta}^T \boldsymbol{g}^* = \boldsymbol{G}_s \boldsymbol{\delta}_s \tag{33}$$

and the mixed model solution for $\boldsymbol{\beta}_s$ and $\boldsymbol{g}^*$ is (recall $\boldsymbol{K} = \boldsymbol{\Delta} \boldsymbol{G}_s^{-1} \boldsymbol{\Delta}^T$)

$$\begin{bmatrix} \boldsymbol{X}_s^T \boldsymbol{R}^{-1} \boldsymbol{X}_s & \boldsymbol{X}_s^T \boldsymbol{R}^{-1} \\ \boldsymbol{R}^{-1} \boldsymbol{X}_s & \boldsymbol{R}^{-1} + \gamma_s \boldsymbol{K} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_s \\ \tilde{\boldsymbol{g}}^* \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_s^T \boldsymbol{R}^{-1} \boldsymbol{y} \\ \boldsymbol{R}^{-1} \boldsymbol{y} \end{bmatrix} \tag{34}$$

The equivalent sparse formulation includes $\boldsymbol{\delta}_s$ in the vector of unknowns in (34) and incorporates the constraint (33) to give

$$\begin{bmatrix} \boldsymbol{X}_s^T \boldsymbol{R}^{-1} \boldsymbol{X}_s & \boldsymbol{X}_s^T \boldsymbol{R}^{-1} & \boldsymbol{0} \\ \boldsymbol{R}^{-1} \boldsymbol{X}_s & \boldsymbol{R}^{-1} & \gamma_s^{-1} \boldsymbol{\Delta} \\ \boldsymbol{0} & \gamma_s^{-1} \boldsymbol{\Delta}^T & -\gamma_s^{-1} \boldsymbol{G}_s \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}_s \\ \tilde{\boldsymbol{g}}^* \\ \hat{\boldsymbol{\delta}}_s \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_s^T \boldsymbol{R}^{-1} \boldsymbol{y} \\ \boldsymbol{R}^{-1} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}$$

The matrices $\boldsymbol{\Delta}$ and $\boldsymbol{G}_s$ being banded, allow sparse methods to be used in the calculations.

References

Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using GLIM. *Appl. Statist.* **36**, 332--339.

Becker, R. A. & Cleveland, W. S. (1996). *S-PLUS Trellis Graphics User's Manual*. Seattle: MathsSoft Inc, Murray Hill: Bell Labs.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9--25.

Chambers, J. M. & Hastie, T. J. (1992). *Statistical Models in* S. Pacific Grove: Wadsworth & Brooks/Cole.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573--578.

Cole, J. W. L. & Grizzle, J. E. (1966). Applications of multivariate analysis of variance to repeated measurements experiments. *Biometrics* **22**, 810--828.

Cullis, B. R. & McGilchrist, C. A. (1990). A model for the analysis of growth data from designed experiments. *BIometrics* **46**, 131--142.

Davidian, M. & Giltinan, D. M. (1993). Some simple methods for estimating intra-individual variability in nonlinear mixed effects models. *Biometrics* **49**, 59--73.

Davidian, M. & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. New York: Chapman Hall.

Diggle, P. J. (1988). An approach to the analysis of repeated measures. *Biometrics* **44**, 959--971.

Diggle, P. J., Liang, K.-Y., & Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Diggle, P. J. & Verbyla, A. P. (1997). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics* under revision.

Frensham, A., Cullis, B. R., & Verbyla, A. P. (1997). Genotype by environment variance heterogeneity in a two stage analsyis. *Biometrics* accepted.

Gilmour, A. R., Cullis, B. R., & Verbyla, A. P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agric., Biol. and Environ. Statist.* accepted.

Gilmour, A. R., Thompson, R., & Cullis, B. R. (1995). Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440--1450.

Gilmour, A. R., Thompson, R., Cullis, B. R., & Welham, S. J. (1996). *ASREML*. Biometric Bulletin: New South Wales Agriculture.

Green, P. J. (1987). Penalized likelihood for general semi-parametric regression models. *Int. Statist. Rev.* **55**, 245--259.

Green, P. J. & Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

Green, P. J. & Yandell, B. (1985). Semi-parametric generalized linear models. In Gilchrist, R., Francis, B. J., & Whittaker, J., editors, *Generalized linear models*, volume 32 of *Lecture Notes in Statistics*, pages 44--55. Berlin: Springer.

Grizzle, J. E. & Allen, D. M. (1969). Analysis of growth and dose response curves. *Biometrics* **25**, 357--381.

Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica* **44**, 460--465.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models (with discussion). *J. R. Statist. Soc.* B **55**, 757--796.

Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Appl. Statist.* **36**, 296--308.

Kenward, M. G. & Roger, J. H. (1997). Small sample inference for fixed effects from

restricted maximum likelihood. *Biometrics* accepted.

Kimeldorf, G. S. & Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat.* **41**, 495--502.

Kohn, R., Ansley, C. F., & Tharm, D. (1991). The performance of Cross-validation and maximum likelihood estimators of spline smoothing parameters. *J. Am. Statist. Assoc.* **86**, 1042--1050.

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

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *J. R. Statist. Soc.* B **58**, 619--678.

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

McGilchrist, C. A. (1994). Estimation in generalized mixed models. *J. R. Statist. Soc.* B **56**, 61--69.

Moran, P. A. P. (1971). Maximum likelihood estimation in non-standard conditions. *Proc. Camb. Phil. Soc.* **70**, 441--450.

Patterson, H. D. & Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58**, 545--554.

Payne, R. W., Lane, P. W., Digby, P. G. N., Harding, S. A., Leech, P. K., Morgan, G. W., Todd, A. D., Thompson, R., Tunnicliffe Wilson, G., Welham, S. J., & White, R. P. (1995). *Genstat 5 Release 4 Reference Manual*. Oxford: Oxford University Press.

Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statist. Sci.* **6**, 15--51.

Rowell, J. G. & Walters, R. E. (1976). Analysing data with repeated observations on each experimental unit. *J. Agric. Sci., Camb.* **87**, 423--432.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719--727.

Searle, S. R., Casella, G., & McCullogh, C. E. (1992). *Variance components*. New York: Wiley.

Self, S. G. & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82**, 605--610.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to nonparametric regression curve fitting (with discussion). *J. R. Statist. Soc.* B **47**, 1--52.

Speed, T. P. (1991). Comment on That BLUP is a good thing: The estimation of random effects, by G. K. Robinson. *Statist. Sci.* **6**, 44.

Stram, D. O. & Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171--1177.

Thompson, R. (1985). Comment on Silverman, Some aspects of the spline smoothing approach to non-parameteric regression curve fitting. *J. R. Statist. Soc.* B **47**, 43.

Upsdell, M. P. (1996). Choosing an appropriate covariance function in Bayesian smoothing. In Bernardo, J. M., Berger, J. O., Dawid, A. P., & Smith, A. F. M., editors, *Bayesian Statistics 5, Proceedings of the Fifth Valencia International Meeting, June 5-9, 1994*, pages 747--756. Oxford: Clarendon Press.

Verbyla, A. P. (1993). Modelling variance heterogeneity: residual maximum likelihood and diagnostics. *J. R. Statist. Soc.* B **55**, 493--508.

Verbyla, A. P. (1995). A mixed model formulation of smoothing splines and testing linearity in generalized linear models. Research Report 95/5, Department of Statistics, The University of Adelaide.

Verbyla, A. P. & Cullis, B. R. (1992). The analysis of multistratum and spatially correlated repeated measures data. *Biometrics* **48**, 1015--1032.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Statist. Soc.* B **40**, 364--372.

Wahba, G. (1983). Bayesian ''confidence intervals'' for the cross-validated smoothing spline. *J. R. Statist. Soc.* B **45**, 133--150.

Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.* **4**, 1378--1402.

Wahba, G. (1990). *Spline models for observational data*. Philadelphia: SIAM.

Wecker, W. P. & Ansley, C. F. (1983). The signal extraction approach to nonlinear regression and spline smoothing. *J. Am. Statist. Assoc.* **78**, 81--89.

Wilkinson, G. N. & Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Appl. Statist.* **22**, 392--399.

Wishart, J. (1938). Growth-rate determinations in nutrition studies with the bacon pig, and their analysis. *Biometrika* **30**, 16--28.

Zeger, S. L. & Diggle, P. J. (1994). Semi-parametric models for longitudinal data with application to cd4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689--699.