Analysis of Agricultural Field Trials in the Presence of Outliers and Fertility Jumps

Ross H. Taplin; Adrian E. Raftery

Biometrics, Vol. 50, No. 3 (Sep., 1994), 764-781.

Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28199409%2950%3A3%3C764%3AAOAFTI%3E2.0.CO%3B2-Q

BIOMETRICS 50, 764–781
September 1994

Analysis of Agricultural Field Trials in the Presence of Outliers and Fertility Jumps

Ross H. Taplin

Mathematics MPS, Murdoch University,
Murdoch WA 6150, Australia

and

Adrian E. Raftery

Department of Statistics, GN-22,
University of Washington,
Seattle, Washington 98195, U.S.A.

SUMMARY

We show how a Bayesian analysis of a fertility model incorporating many of the previously suggested models can account for uncertainty about which fertility model provides the best approximation in any given trial. We also show how uncertainty about anomalies such as outliers and fertility jumps can be accounted for. We argue that this is preferable to conditioning on an "appropriate" model, and show by examples how accounting for such possible anomalies can both influence support for a particular fertility model and reduce the dependence of treatment estimates on the choice of fertility model.

1. Introduction

An objective of most agricultural field trials is to assess the effect of treatment on yields. This must be done in the presence of unknown fertility trends through the field. Many stochastic models for these fertility trends have been proposed and shown to be efficient competitors to classical blocking designs (see Section 2). In this paper we shall restrict attention to trials where plot fertilities are correlated in only one direction. In the case of a rectangular lattice of plots this is often assumed when the long, thin nature of the plots makes those with common long edges spatially close compared to those with common short edges.

When using a stochastic model for field fertility in the analysis of a particular trial, two questions must be addressed. The first is the choice of fertility model, and the second is the treatment of anomalies such as outliers and fertility jumps, i.e., sudden level shifts in fertility. Even when several such possibilities are entertained, it is not uncommon to condition on a single model and a single set of anomalies, which are determined by model diagnostics; see, for example, Cullis, McGilchrist, and Gleeson (1991) and Martin (1990). Here we propose to account for uncertainty about the fertility model and possible anomalies rather than to condition on an "appropriate" model. This will be particularly important when several choices are almost equally appropriate (see Section 4.3).

We first define a general but simple state-space model for agricultural field trials, which turns out to include many of the models in current use (Section 2). We then show how this framework allows us to model outliers and fertility jumps quite easily (Section 3). The importance of doing this and of accounting fully for the associated uncertainty is illustrated in several examples (Section 4).

2. A Fertility Model

2.1 *Model Definition*

We suppose that the trial consists of one or more *blocks* of plots. Here a block refers to a row of adjacent plots, with plots in different blocks assumed to bear little relationship to each other. The

Key words: Bayesian analysis; Random effects; Residual maximum likelihood; State-space models.

plots are numbered 1 through n along the row within each block, and block by block. Plot t is therefore at the beginning of a new block if $t = 1$ or if plots t and $t - 1$ are in different blocks.

We assume that the data can be decomposed additively:

$$Y = D\tau + F + \varepsilon, \tag{1}$$

where Y is the n-vector of yields, τ is the v-vector of treatment effects (which may include other covariates), D is the corresponding design matrix, F is the n-vector of fertility levels, and ε is the n-vector of measurement errors.

We model F recursively to correspond to walking through the plots from plot 1 to plot n, with t therefore indexing time as well as the plots. Starting from the assumption that the fertility is approximately locally linear, we suppose that the fertility level at plot t given the previous plots is approximately the fertility on plot $t - 1$ plus some measure of the rate of increase of the fertility at this plot. Given that the fertility level is fluctuating around some local mean value, we may expect the fertility level on plot t to be closer to this mean value (as otherwise it may become unbounded). We define G_{t-1} to be the fertility gradient between the fertility level at plot t and plot $t - 1$, and note that the more slowly G changes with time the smoother the corresponding fertility level F becomes.

More precisely, we shall model the plot fertilities conditional on the parameters λ_1, λ_2, and σ^2_{grad}, where $0 \leqslant \lambda_2 \leqslant \lambda_1 < 1$ and $\sigma^2_{\text{grad}} \geqslant 0$, recursively as follows:

Case 1: Same block

$$\begin{pmatrix} F \\ G \end{pmatrix}_t = \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} F \\ G \end{pmatrix}_{t-1} + \begin{pmatrix} 0 \\ \xi_t \end{pmatrix}, \quad \xi_t \overset{\text{iid}}{\sim} N(0, \sigma^2_{\text{grad}}) \tag{2}$$

Case 2: New block

$$\begin{pmatrix} F \\ G \end{pmatrix}_{t|(t-1)} \sim N(0, \boldsymbol{\Sigma}_{\text{uncond}}) \tag{3}$$

$$\Sigma_{\text{uncond}} = \begin{pmatrix} \dfrac{1 + \lambda_1\lambda_2}{(1 - \lambda_1^2)(1 - \lambda_1\lambda_2)} & \dfrac{\lambda_2}{(1 - \lambda_1\lambda_2)} \\ \dfrac{\lambda_2}{(1 - \lambda_1\lambda_2)} & 1 \end{pmatrix} \frac{\sigma^2_{\text{grad}}}{1 - \lambda_2^2}$$

where we use the notation $(t - 1)$ to indicate the set $1, 2, \ldots, t - 1$.

The assumption that $|\lambda_1|, |\lambda_2| < 1$ results in the distribution of $\binom{F}{G}$ being stationary and it is the distribution unconditional on other plots that we use when at the beginning of a block. We are assuming here that while the plot at the beginning of a new block is not adjacent in the field to the previous plot, the two plots are relatively close. Thus while the two plots have similar characteristics (i.e., the same marginal distribution), they can be assumed independent.

The intuition of a smoothly varying fertility suggests that both λ_1 and λ_2 should be nonnegative. The restriction $\lambda_1 \geqslant \lambda_2$ is enforced to aid identifiability as the marginal distribution of Y_t remains unchanged if we interchange λ_1 and λ_2. Indeed, the marginal distribution of the fertility level F_t also remains unchanged, with this interchange affecting only the distribution of the fertility gradient G_t. The choice of λ_1 being the larger coincides with our intuition governing the definition of F and G.

In line with ε representing measurement error, we shall assume that the ε_t $(t = 1, \ldots, n)$ are independent (and independent of F), and further make the distributional assumption,

$$\varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2_{\text{obs}}), \quad t = 1, \ldots, n, \quad \sigma^2_{\text{obs}} \geqslant 0.$$

While ε could be absorbed into F, we make the distinction as the term ε is directly interpretable as measurement error and observed data often support its presence (Whittle, 1954). The distribution of Y is therefore dependent on the parameter $\theta^* = (\sigma^2_{\text{grad}}, \sigma^2_{\text{obs}}, \lambda_1, \lambda_2)$, or $\theta = (\sigma^2, \%fert, \lambda_1, \lambda_2)$, where

$$\sigma^2 = \frac{1 + \lambda_1\lambda_2}{(1 - \lambda_1^2)(1 - \lambda_1\lambda_2)(1 - \lambda_2^2)} \sigma^2_{\text{grad}} + \sigma^2_{\text{obs}}$$

is the variance of any observation Y_t given D and τ only, and

$$\%fert = 100\left(1 - \frac{\sigma^2_{\text{obs}}}{\sigma^2}\right)$$

is the percentage of this variance explained by the fertility.

The condition $0 \leq \lambda_2 \leq \lambda_1 < 1$ forms a triangle in (λ_1, λ_2) parameter space, with an open side (being the bound to nonstationarity) and two closed sides. This suggests two natural special cases for modelling field fertility, with the more general parameter values lying between these special cases. (We ignore the nonstationary side; see Section 6.) The first special case is obtained by forcing $\lambda_2 = 0$. This model, while assuming that the fertility levels at consecutive plots are related, assumes that the fertility gradients are independent. This typically results in jagged fertility trends. The second case results by forcing $\lambda_1 = \lambda_2$ and potentially produces smoother fertility trends than the first.

Figure 1 displays these two special cases in relation to our (λ_1, λ_2) parameter space, together with various fertility models previously proposed for field fertility. These include the simplest conditional autoregression CAR (Besag, 1974), the simplest simultaneous autoregression SAR (Whittle, 1954), the first-difference model I(1) (Besag and Kempton, 1986), the second-difference model I(2) (Green, Jennison, and Seheult, 1985), and the ARIMA(1, 1, 0) (Gleeson and Cullis, 1987), AR(1) (Patterson, 1983), and white noise models.

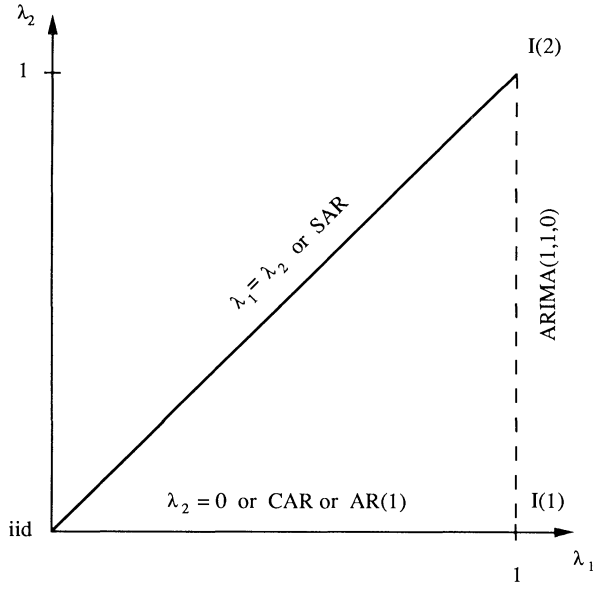

Figure 1. Comparison of our fertility model with the simplest simultaneous autoregression (SAR), conditional autoregression (CAR), first- and second-order integrated models (I(1) and I(2)), ARIMA(1, 1, 0), and the model of independence (iid).

Our fertility model is a special case of the AR(2) model, with λ_1 and λ_2 being the roots of the characteristic equation. Forcing these roots to be real and positive eliminates models with oscillating correlograms that can take negative values. We argue that these phenomena do not conform with the notion of a fertility trend, where the correlations should decrease with increasing distance and remain nonnegative. If present, it is assumed that they can be more directly modelled (such as with direction of harvester as a covariate), and not included with the fertility. Williams (1952) considered experimental designs and maximum likelihood under the general AR(2) fertility model without measurement error.

The model with $\lambda_1 = \lambda_2$ is the AR(2) model with equal roots of the characteristic equation. The resulting autocorrelation function has the form

$$\rho_k = \left(1 + \frac{1 - \lambda_1^2}{1 + \lambda_1^2} k\right)\lambda_1^k.$$

This autocorrelation function is unlike those for any other AR(2) model in that while it is monotonically decreasing away from zero, it is flat at the origin rather than spiked. This special case of the

AR(2) model was not studied by Box and Jenkins (1976), for example. The same autocorrelation structure also arises from what may be considered the simplest truly bilateral simultaneous autoregression, namely $F_t = a(F_{t-1} + F_{t+1}) + \xi_t$, where $a = \lambda_1/(1 + \lambda_1^2)$. Whittle (1954) considered the two-dimensional version of this process and described data that supported this phenomenon.

2.2 *Bayesian Estimation*

To calculate the posterior distribution of τ we must calculate the integral

$$p(\tau|Y) = \int p(\tau, \theta|Y)\, d\theta \propto \int p(Y|\tau, \theta)p(\tau, \theta)\, d\theta,$$

where $p(Y|\tau, \theta)$ is the likelihood and $p(\tau, \theta) = p(\tau|\theta)p(\theta)$ is our prior for (τ, θ). In general, this integral cannot be evaluated analytically, and so a numerical technique must be used. The computation can be simplified by using a multivariate normal prior for $(\tau|\theta)$ (or with simple modifications a mixture of multivariate normals). It then follows that $p(\tau|Y, \theta) \propto p(Y|\tau, \theta)p(\tau|\theta)$ is also a multivariate normal, and $p(\tau|Y) = \int p(\tau|Y, \theta)p(\theta|Y)\, d\theta$ is a mixture of these normal distributions.

The mean and variance of $(\tau|Y, \theta)$ and the likelihood given θ, $p(Y|\theta)$, can be calculated directly using the Kalman filter with state $X_t = (F_t, G_t, \tau_1, \ldots, \tau_v)$ even in the presence of missing values (Kalman, 1960). This suggests the use of importance sampling to estimate the posterior for τ by first simulating θ_i $(i = 1, \ldots, N)$ independently from the sampling importance density $f(\theta)$, and then forming the estimate

$$\hat{p}(\tau|Y) = \frac{1}{N}\sum_{i=1}^{N} p(\tau|\theta_i, Y)\frac{p(\theta_i|Y)}{f(\theta_i)}. \tag{4}$$

The distribution $f(\theta)$ is chosen such that it can be easily simulated from, and is as close to the unknown $p(\theta|Y)$ as possible. As N approaches infinity, $\hat{p}(\tau|Y)$ approaches $p(\tau|Y)$.

Another approximation is

$$p(\tau|Y) = p(\tau|\hat{\theta}, Y), \tag{5}$$

where $\hat{\theta}$ is the value of θ that maximises the posterior distribution of θ, $p(\theta|Y)$. A numerical maximisation of $p(\theta|Y)$ is therefore required to obtain $\hat{\theta}$. We shall employ this approximation (5) in Section 3.3.

Figure 2 displays the posterior of a treatment contrast (early spraying − no spraying) in the mildew control trial (Draper and Guttman, 1980; Jenkyn et al., 1979) estimated by equations (4) and (5) under the prior $p(\tau, \theta) \propto \sigma^{-2}$. Similar results were obtained with other vague priors. While conditioning on $\hat{\theta}$ resulted in a good approximation with considerably less computation, as expected it underestimates the uncertainty about the value of this contrast. It obtains the same mean but its variance is underestimated by about 19%.

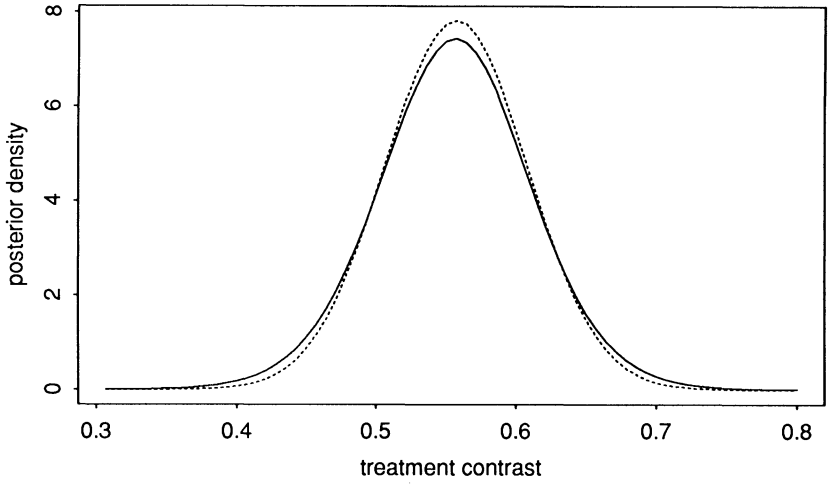

Figure 2. Posterior distribution for early spraying − no spraying in the mildew trial; solid line: true posterior; dotted line: approximation (5).

The accuracy of the approximation (5) is not surprising when one notes its similarities with restricted maximum likelihood (REML) estimation (Patterson and Thompson, 1971; Laird and Ware, 1982). When the prior $p(\tau, \theta) \propto 1$ is used, $\hat{\theta}$ equals the REML estimate of θ and the $(1 - \alpha)$ highest posterior density region of the approximate posterior distribution of τ equals the $(1 - \alpha)$ confidence interval from conditioning on the REML estimate of θ. To correct the underestimation of uncertainty that occurs in REML, McGilchrist (1989) suggested inflating the REML estimate of σ^2 by n/df, where n is the number of observations and $df = n - \dim(\theta)$. However, here this increases the variance by only about 9% compared with the 19% that would be required to make it exact.

Note that a better choice of prior than $p(\tau, \theta) \propto 1$ is likely to be available. Even ignoring the fact that θ contains variance terms for which a non-uniform prior may be more appropriate, for many trials prior information about τ will be available. This is particularly likely to be the case when τ represents the effect of different varieties. In early generation variety trials the prior for τ should reflect the genetic relationship between the varieties [see, for example, Cullis et al. (1990)]. For later trials, the information derived from earlier trials with these varieties can be utilised.

An alternative to using a prespecified prior is to use a random effects model where the variety effects have an exchangeable joint distribution. The prior distribution of the variety effects can be estimated from the data, leading to an empirical Bayes approach. If that distribution has a parametric form, we have a parametric empirical Bayes model (Morris, 1983). A simple parameterisation of the distribution for this population is to assume τ multivariate normal with mean $\mu_\tau \cdot 1$ and covariance matrix $\sigma_\tau^2 C_\tau$. Here C_τ is a prespecified covariance matrix for τ, 1 is the v-vector of ones, and μ_τ and σ_τ^2 are unknown scalars that could, for example, be estimated by the mode of their posterior density. This is similar to the approach employed by Cullis et al. (1990) for unreplicated treatments.

Figure 3 illustrates the effect of using a random effects model instead of a fixed effect model on the ARC 2 trial (a trial with 38 treatments replicated 3 times and nothing unusual in the way of outliers or fertility jumps). Here we have used the approximation just described with C_τ equal to the identity and the model with $\lambda_1 = \lambda_2$. Note that the random effects model results in estimated treatment effects (posterior means) with a smaller range. The standard deviations of the treatment difference posteriors were similarly reduced. This shrinkage agrees with our intuition that the treatment with the largest estimated effect from the fixed effect model is likely to be overestimated. The reason for this treatment's favourable result is likely to be due at least in part to random deviations (as it would be completely if the true variety effects were equal).

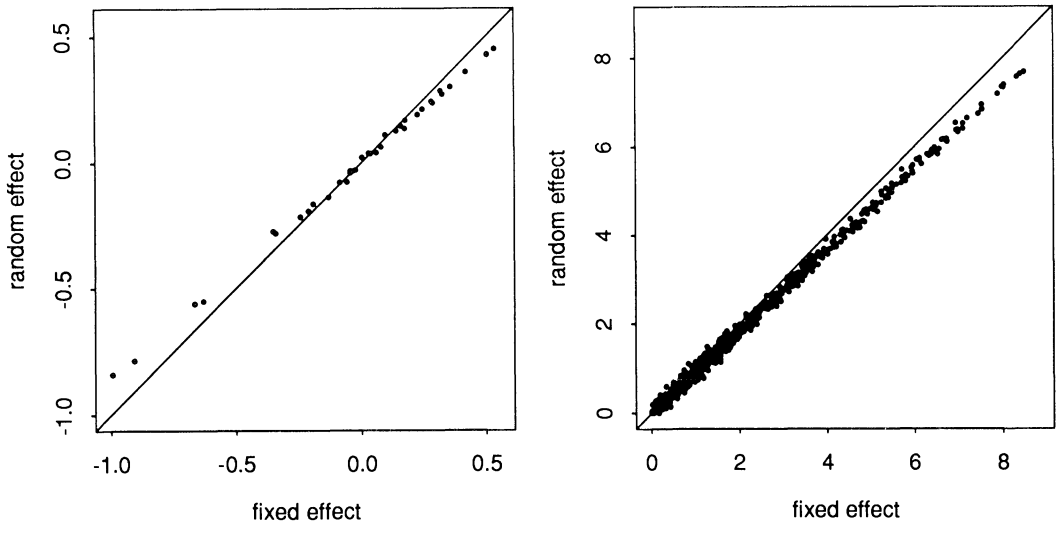

Figure 3. Comparison between fixed and random effects models, trial ARC 2. Posterior means for varieties (left) and means divided by standard deviations for treatment differences (right). The vertical and horizontal axes indicate random and fixed effect models, respectively.

3. Modelling Outliers and Fertility Jumps

3.1 *The Robust Model*

We now extend our model to allow for the possibility of outliers and fertility jumps. The term *outlier* is used to refer to a single observation whose value was generated by a different mechanism,

whereas a fertility jump refers to a sudden, relatively large shift in the fertility level compared to the majority of the observations. Both occurrences are considered rare, but when present can greatly influence the resulting inferences. We retain the decomposition in equation (1), and the treatments τ and design matrix D, but assume that a fertility jump modifies the distribution of F and an outlier modifies the distribution of ε.

The field fertility is modelled as in equations (2) and (3). However, given that there is a fertility jump at plot t, we assume that the fertility at plot t is independent of the fertility at plot $t-1$. We apply the same justification as was used in Section 2 when plot t is at the beginning of a block, and hence apply the same distributional assumptions. Thus equation (3) is used if plot t is either at the beginning of a block or at a fertility jump, with equation (2) being used otherwise.

We model outliers by modifying the distribution of the measurement errors ε. While retaining normality and independence of the ε_t, we inflate the variance of ε_t by a constant factor, k^2_{out}, when observation t is an outlier. We shall henceforth assume this variance inflation factor to be 100 (see Section 5 for a discussion of this value). Thus, while an outlier observation y_t still has a distribution centred on $(D\tau)_t + F_t$, its distribution is considerably more spread out.

To avoid a possible identifiability problem, we assume that a fertility jump cannot occur at the beginning of a block. We also make the assumption that a fertility jump and an outlier cannot both be present at the same plot. This simplifies our model by allowing at most three possible cases for each plot: a fertility jump, an outlier, or neither. The fourth possibility of both a fertility jump and an outlier should be uncommon, and we note in this case that the distribution of y_t is affected little by the exact value of the fertility at plot t. Hence in this case we redefine the fertility jump at plot t to be at plot $t+1$. While this results in an incorrect estimate of the plot fertility at plot t, we note this and consider it to be of little consequence, as the posterior distribution of τ, which is of primary interest, will be affected negligibly.

Formally, we shall let c_t denote the condition of plot t. Thus $c_t = 1$ if plot t has neither an outlier nor a fertility jump, $c_t = 2$ if plot t has an outlier, and $c_t = 3$ if plot t is at a fertility jump. Then $c = (c_1, \ldots, c_n)$ denotes the condition of all the plots. We define the submodel M_c to be the model with plot conditions c.

Our model for Y consists of a mixture of these submodels where each submodel M_c is weighted according to the prior probability of model M_c. The likelihood for Y is therefore

$$p(Y) = \sum_c p(Y|M_c)p(M_c)$$

and the posterior distribution of τ, the quantity of interest, is given by

$$p(\tau|Y) \propto \sum_c p(\tau|Y, M_c)p(M_c|Y).$$

We assume that the prior specifies the conditions of the different plots to be statistically independent and that the conditions 1, 2, and 3 have prior probabilities 96%, 2%, and 2%, respectively. These choices are discussed in Section 5.

3.2 *Comparison with Robust Time Series*

In the absence of the term $D\tau$, we obtain a robust model similar to those employed in robust nonseasonal time series (Harrison and Stevens, 1976; West and Harrison, 1990). Estimation under such models is aided by the fact that an observation is most correlated with observations on nearby plots, the *neighbours*. Thus the condition of a plot can be well estimated with knowledge only of the observations on the neighbouring plots. Estimation can then proceed recursively by updating the plots in field order, so that a plot is considered at about the same time as its neighbours.

In the presence of the term $D\tau$, the neighbourhood structure described above is more complex, with the neighbouring plots being not only those nearby in the field, but also those receiving a similar treatment combination. A recursive procedure is therefore not as efficient, as it becomes difficult to update the plots in such a way that a plot is updated at about the same time as its neighbours. The simplest solution to this problem would be to condition on an estimate of τ, and then estimate the condition of the plots as before. Figure 4 compares the results of using maximum likelihood on the full model to estimate τ with the exact Bayesian solution in some simple cases.

In Figure 4 we assumed that $F = 0$ and $\sigma^2_{\text{obs}} = 1$, so that the observations are independent with a variance of 1, or of 100 if conditioned to be an outlier. The (scalar) mean τ was assumed to have the improper uniform prior and only the first observation had positive prior probability (2%) of being an outlier. The approximation is good with nine replicates, but seriously underestimates the

uncertainty about whether the observation is an outlier or not with three replicates. The problem occurs because the likelihood is the mixture of the two normal likelihoods corresponding to the two submodels, and with few replicates the critical factor is which of these likelihoods dominates the other. Maximum likelihood cannot be expected to perform well when the log-likelihood is so nonquadratic, nor is it trivial to ensure that the global maximum and not just a local maximum is achieved.

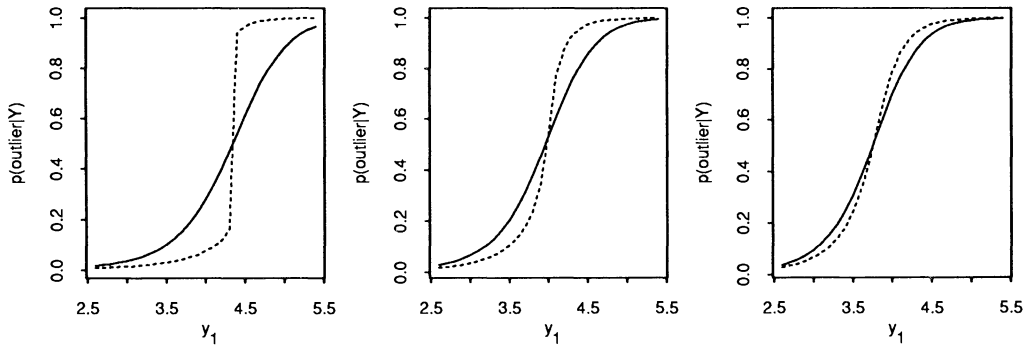

Figure 4. Effect on posterior probability of an outlier when conditioning on the MLE for τ under the simple model with $F = 0$ and $\sigma^2_{\text{obs}} = 1$. Only the first observation y_1 is nonzero. The three graphs assume (left to right) 3, 5, and 9 replicates, respectively, with the true posterior (solid) and approximation (dotted).

3.3 *Estimation for the Robust Model*

The oversimplified model used for Figure 4 suggests that the posterior probability of an observation being an outlier is well approximated by zero unless its cross-validated residual is at least 3. Since such a large residual is rare under the nonrobust model, this suggests that most of the submodels of our model will have very low posterior probability. We therefore propose to calculate only the submodels that have nonnegligible posterior probabilities M_c, $c \in \mathcal{C}$, and approximate the posterior probabilities for the other submodels by zero. We defer the estimation of $\mathcal{C}$ to Section 3.4, but now take it as given.

We approximate the posterior distribution of τ as follows:

$$p(\tau|Y) \approx \sum_{c\in\mathcal{C}} \int p(\tau, \theta, M_c|Y)\, d\theta$$

$$\approx \sum_{c\in\mathcal{C}} p(M_c|Y)p(\tau|\hat{\theta}_c, M_c, Y). \tag{6}$$

In equation (6), $\hat{\theta}_c$ is the value of θ that maximises $p(\theta|M_c, Y)$, and

$$p(M_c|Y) \propto p(Y|M_c)p(M_c)$$

$$= p(M_c) \int p(Y|\theta, M_c)p(\theta|M_c)\, d\theta$$

$$\approx p(M_c)p(Y|\hat{\theta}_c, M_c)p(\hat{\theta}_c|M_c), \tag{7}$$

where $p(\theta|M_c)$ is the prior for θ under submodel M_c, and is typically assumed to be the same for every submodel. Thus each submodel is estimated separately, as in Section 2.2, equation (5), and the results combined under the assumption that integrating θ out of $p(Y, \theta|M_c)$ can be well approximated by maximising it with respect to θ.

The simpler approximation that conditions on a single value $\hat{\theta}$ of θ that maximises $\Sigma_{c\in\mathcal{C}}\, p(\theta, M_c|Y)$,

$$p(\tau|Y) = \int \sum_c p(\tau, \theta, M_c|Y)\, d\theta$$

$$\approx \int \sum_{c \in \mathscr{C}} p(\tau|\theta, M_c, Y)p(M_c|\theta, Y)p(\theta|Y)\, d\theta$$

$$\approx \sum_{c \in \mathscr{C}} p(\tau|\hat{\theta}, M_c, Y)p(M_c|\hat{\theta}, Y), \tag{8}$$

applies the same approximation as in Section 2.2 to the complete model. A comparison of these two approximations is illustrated in Figure 5, where we again assume the simplified model with zero fertility, so that $\theta = \sigma^2_{\text{obs}}$, and the observations are independent. Observations receiving the first treatment were all assumed to be zero except for y_1, whereas observations on the other treatments were such that their residual sum of squares equalled their degrees of freedom (so $\hat{\sigma}^2_{\text{obs}} \approx 1$).

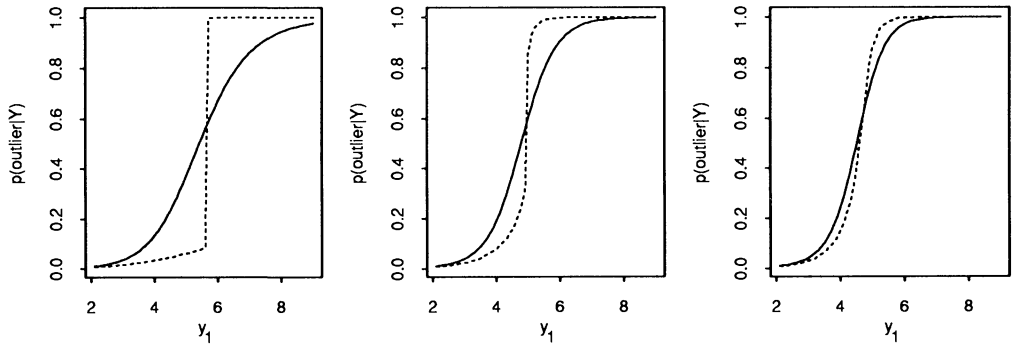

Figure 5. Effect on posterior probability of an outlier when conditioning on the MLE for σ^2_{obs} under the simple model with $F = 0$. The only nonzero observation for the first treatment is y_1. The graphs (left to right) correspond to trials with 5, 10, and 20 treatments, respectively, replicated 3 times. The true posterior (and approximation (7)) is solid and approximation (8) is dotted.

In this case approximation (7) is very good (and is exact if the prior $p(\sigma^2_{\text{obs}}) \propto 1$ is used in (7) in place of the assumed $p(\sigma^2_{\text{obs}}) \propto 1/\sigma^2_{\text{obs}}$). Approximation (8) underestimates the uncertainty about whether the first plot contains an outlier or not, especially for small trials.

3.4 *Determining the Conditioning Set of Outliers and Jumps*

Due to the large number of possible submodels, it is likely that any practical choice of $\mathscr{C}$, the conditioning set of anomalies, will in total comprise a small percentage of the posterior probability, the majority of the probability being spread very thinly over the very large number of remaining submodels. Nevertheless, the posterior for the treatment difference $\tau_i - \tau_j$ will primarily be influenced by the presence of outliers on any plots with treatments i or j applied, or by fertility jumps on plots close to these plots. The marginal posterior probabilities of outliers and fertility jumps on these plots are of primary importance when calculating the posterior for this treatment difference. These marginal posteriors can be reasonably approximated by a few submodels since to a first approximation the posterior occurrences of outliers and fertility jumps can be assumed independent. This approximation is most likely to be violated when the addition of an outlier or fertility jump greatly modifies the estimate of θ, influencing the probabilities of future outliers or fertility jumps through different distributional assumptions (likely only in small trials or when the original outlier or fertility jump has very high posterior probability), or when the outliers or fertility jumps in question are close under the neighbourhood structure.

To determine the set of submodels M_c, $c \in \mathscr{C}$, that have nonnegligible posterior probability, we adopt a recursive procedure beginning with the submodel $c_t = 1$ for all t. This submodel corresponds to no outliers or fertility jumps and often has high posterior probability; we adopt the convention of always including it in $\mathscr{C}$.

Given that we have just included the submodel M_c in $\mathscr{C}$, we then consider the submodels $M_{c'}$ in $\mathscr{C}_{c'}$ that have the same outliers and fertility jumps as M_c plus one extra outlier or fertility jump. The submodels in $\mathscr{C}_{c'}$ are then approximately ranked according to their posterior probability, and estimated in turn until a submodel from $\mathscr{C}_{c'}$ is rejected from $\mathscr{C}$ due to its low posterior probability. Here the submodel $M_{c'}$ is rejected if its posterior probability is less than a proportion α of the sum of the posterior probabilities for the submodels previously considered from $\mathscr{C}_{c'}$ and M_c. The

proportion α is taken to be small, and in our examples we took $\alpha = 2\%$. Choosing α to be close to the (usually small) prior probability of an outlier or a fertility jump seems reasonable because it implies roughly that when the data alone provide evidence against a plot being an outlier or a jump, that possibility is ignored. The recursion is continued by considering those submodels with an extra outlier or fertility jump from those submodels $M_{c'}$ accepted into $\mathcal{C}$.

There are various ways of determining the ranking of $\mathcal{C}_{c'}$. Likely outliers can be detected by the estimated measurement errors with large magnitude from the submodel M_c (note that $\hat{\theta}_{c'}$ is likely to be close to $\hat{\theta}_c$ due to the similarity of the submodels). Similarly, likely fertility jumps may be detected by values of $\hat{F}_t + \hat{\varepsilon}_t - (\hat{F}_{t-1} + \hat{\varepsilon}_{t-1})$ from the submodel M_c. A more refined estimate for a selection of submodels could be based on $p(M_{c'}|\hat{\theta}_c, \hat{\tau}_c, Y)$, where $\hat{\tau}_c$ is the estimate of τ from M_c, or even better based on $p(M_{c'}|\hat{\theta}_c, Y)$. These rankings for $\mathcal{C}_{c'}$ are likely to be progressively more accurate but require progressively more computation. They all require considerably less computation than our final estimate, $p(M_{c'}|\hat{\theta}_{c'}, Y)$, which requires a numerical optimisation and should be required for only a small subset of submodels in $\mathcal{C}_{c'}$.

4. Examples

We now consider three trials as examples. All are early grain variety trials conducted by the Agricultural Research Council of Great Britain and consist of 3 or 4 blocks with each variety applied to exactly one plot in each block. The first two trials (ARC 8 and ARC 6) illustrate, respectively, the case where there is uncertainty about whether a single outlier or fertility jump is present, while the third trial (ARC 1) illustrates the occurrence of many possible outliers and fertility jumps. Further examples can be found in Taplin (unpublished Ph.D. dissertation, Department of Statistics, University of Washington, 1990), where estimation of the general model given in equations (2) and (3) often resulted in either $\hat{\lambda}_2 = 0$ or $\hat{\lambda}_1 = \hat{\lambda}_2$. We therefore shall summarise the results for the two extreme cases $\lambda_2 = 0$ and $\lambda_1 = \lambda_2$. In any case we shall see that submodel uncertainty has a greater influence on the posterior for τ than model uncertainty. Cullis and Gleeson (1989) found that the simple I(1) fertility model was satisfactory in 90% of 1,019 variety trials conducted in Australia during 1987–1989.

4.1 *A Possible Outlier: Trial* ARC 8

Figure 6 displays the treatment-adjusted data (i.e., $Y - D\hat{\tau}$, where $\hat{\tau}$ is the estimate of τ under the submodel $c_t = 1$ for all t) in field order for the three blocks with the treatment number as the plotting symbol. The lines are the estimated fertilities from the $\lambda_1 = \lambda_2$ model (solid) and the $\lambda_2 = 0$ model (dotted). While the posteriors for τ were very similar under the two models, the estimated fertility is smoother under the $\lambda_1 = \lambda_2$ model, with correspondingly larger estimated measurement errors.

The large magnitude of the estimated measurement error on plot 23 suggests that this value is atypical, and Figure 7 is similar to Figure 6 but assumes the submodel $c_{23} = 2$, $c_t = 1$ for all $t \neq 23$. This is almost equivalent to removing this observation from the analysis. Table 1 summarises the estimation of these two submodels under the two models (estimation of the model with general λ_1 and λ_2 results in $\hat{\lambda}_2 = 0$).

Both models achieve similar log-likelihoods, and this together with the fact that for either submodel the models achieve similar posteriors for τ suggests that for this trial submodel uncertainty is more important than model uncertainty. With a prior probability of an outlier of 2%, this translates [by approximation (7)] into posterior probabilities of an outlier on plot 23 of 82% and 79% for the $\lambda_2 = 0$ and $\lambda_1 = \lambda_2$ models, respectively. Approximation (8) however results in posterior probabilities of 98% and 97%, respectively, and essentially conditions on plot 23 containing an outlier.

For comparison, a complete Bayesian analysis with prior $p(\theta) \propto 1/\sigma^2$ resulted in a posterior probability of $c_{23} = 2$ of about 68 ± 4% (calculated using importance sampling as in Section 2.2). Thus approximation (7) also underestimates the uncertainty about the condition of plot 23, but not as much as approximation (8).

Finally, we note that it is the possibility of $c_{23} = 1$ or 2 that is the primary issue for this trial. The next most likely outlier appears to be on plot 33 with a posterior probability of less than 3%, and the most likely fertility jump occurs at plot 43, with a posterior probability of about 6%. These possibilities have little influence on the posterior for τ. The effect on the posterior of τ of uncertainty about the condition of plot 23 will be illustrated in Section 5.

4.2 *A Possible Fertility Jump: Trial* ARC 6

Whereas the trial ARC 8 was dominated by a large measurement error relative to the fertility trend, the trial ARC 6 provides the contrasting situation where σ^2_{obs} is very small. Here we shall assume that $\sigma^2_{\text{obs}} = 0$ for trial ARC 6 and illustrate a possible fertility jump. Figure 8 graphs the treatment-

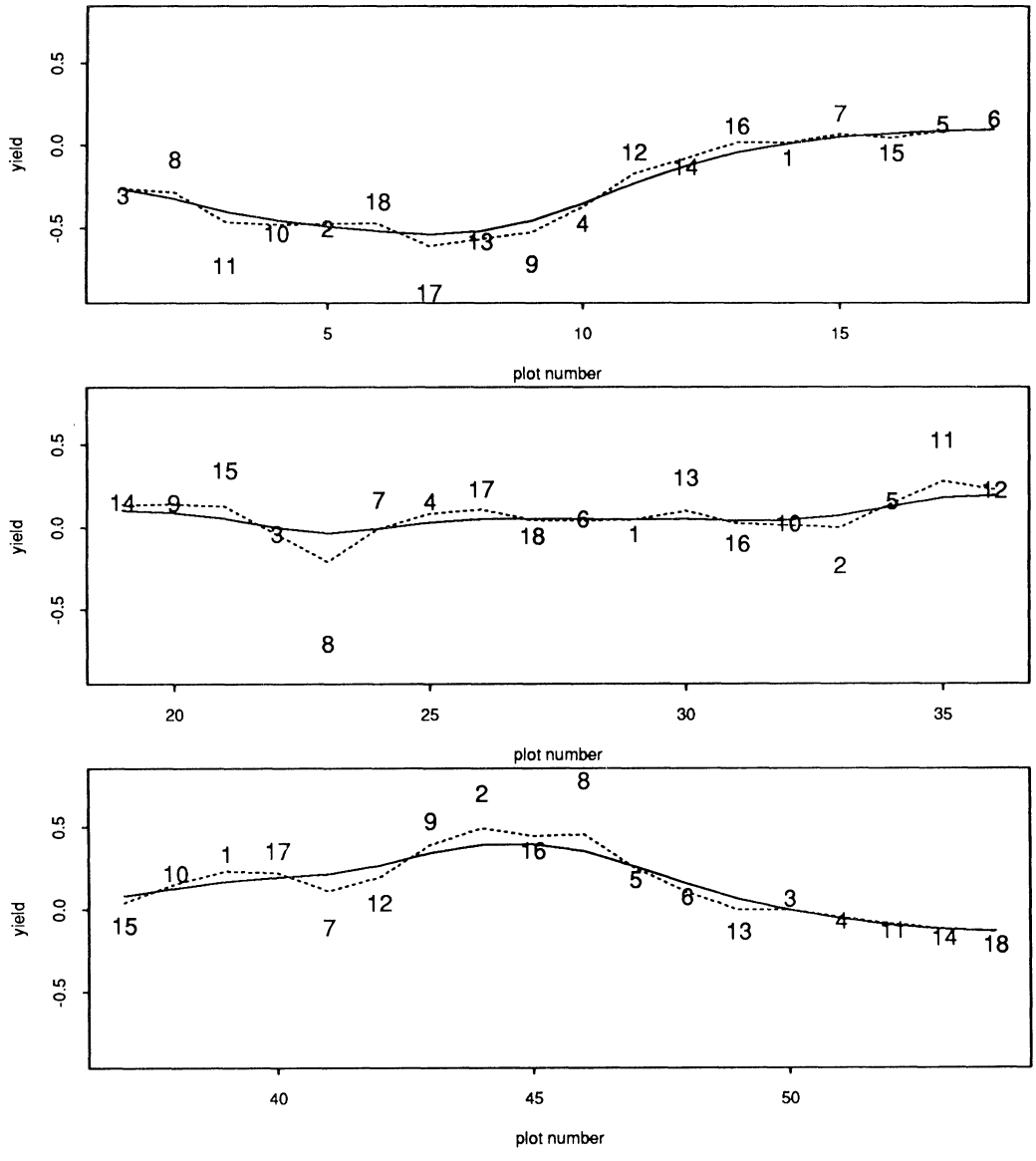

Figure 6. Trial ARC 8, treatment-adjusted yield graphed with symbol treatment number under submodel $c_t = 1$ for all t. Lines are estimated fertility from $\lambda_1 = \lambda_2$ model (solid) and $\lambda_2 = 0$ model (dotted).

adjusted yield (under the model with $\lambda_1 = \lambda_2$) against plot number in the same manner as for trial ARC 8. Note that here there are four blocks, and since $\sigma^2_{\text{obs}} = 0$ the estimated fertility equals the treatment-adjusted yield.

Table 2 summarises the estimation of the two submodels $c_t = 1$ for all t and $c_5 = 3$, $c_t = 1$ for all $t \neq 5$, under the two models with $\sigma^2_{\text{obs}} = 0$. We have chosen to force $\sigma^2_{\text{obs}} = 0$ because it was usually estimated to be zero anyway. Only when $\lambda_1 = \lambda_2$ and $c_5 = 3$ was σ^2_{obs} estimated to be nonzero, and even in this case it was negligible.

Unlike in the previous example, the choice of model now influences the evidence for the submodels; the fertility jump at plot 5 is more likely under the model with $\lambda_1 = \lambda_2$. Assuming a prior probability of a fertility jump of 2%, the posterior probability that $c_5 = 3$ is 63% when $\lambda_1 = \lambda_2$ compared to 16% when $\lambda_2 = 0$. This increased model influence is to be expected due to the lack of measurement error, making the distribution of Y more dependent on the distribution of F.

We note that while the model $\lambda_2 = 0$ was .6 unit of log-likelihood superior to the model $\lambda_1 = \lambda_2$ when conditioning on $c_5 = 1$, it is .2 unit inferior when taken unconditionally. Thus, when the possibility of fertility jumps is introduced, there is quite a shift in evidence about which fertility model is appropriate. Furthermore, under the submodel $c_t = 1$ for all t, it is the posterior for the

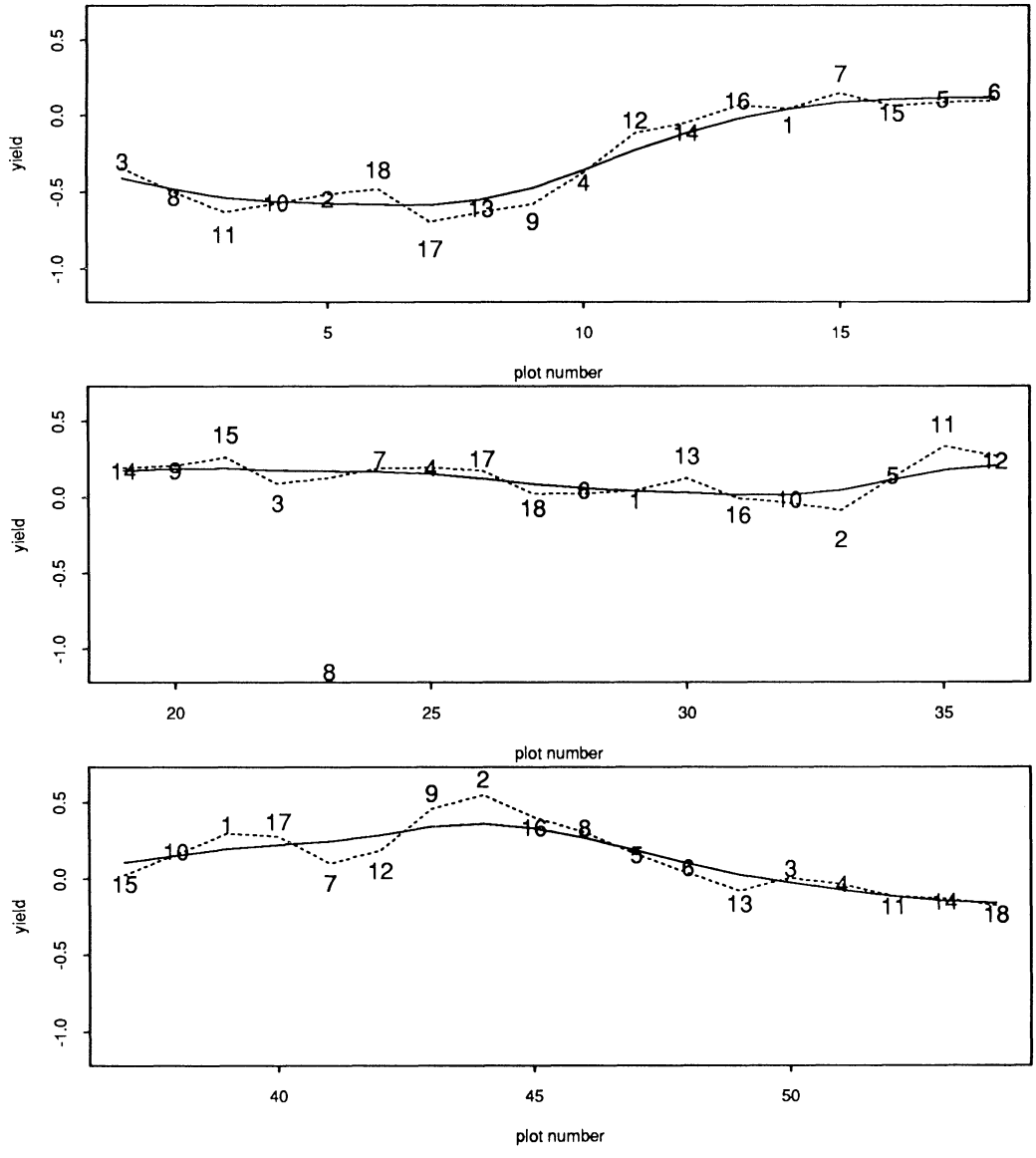

Figure 7. Trial ARC 8, treatment-adjusted yield graphed with symbol treatment number under submodel with $c_{23} = 2$, $c_t = 1$ for all $t \neq 23$. Lines are estimated fertility from $\lambda_1 = \lambda_2$ model (solid) and $\lambda_2 = 0$ model (dotted).

Table 1
Comparison of submodels for trial ARC 8

	Submodel with $c_{23} = 1$				Submodel with $c_{23} = 2$			
Model	Log-like.	σ^2	*%fert*	λ_1	Log-like.	σ^2	*%fert*	λ_1
$\lambda_2 = 0$	.1	.16	64	.81	5.5	.14	82	.83
$\lambda_1 = \lambda_2$	.0	.16	51	.71	5.2	.10	68	.77

treatment difference $\tau_{17} - \tau_{16}$ that differs most between the two models. It is also this treatment difference that is most influenced by the introduction of the fertility jump at plot 5 (since treatments 17 and 16 were applied to plots 5 and 4, respectively), and we shall show in Section 5 that allowing for the possibility of fertility jumps makes the treatment posteriors under the different models more similar.

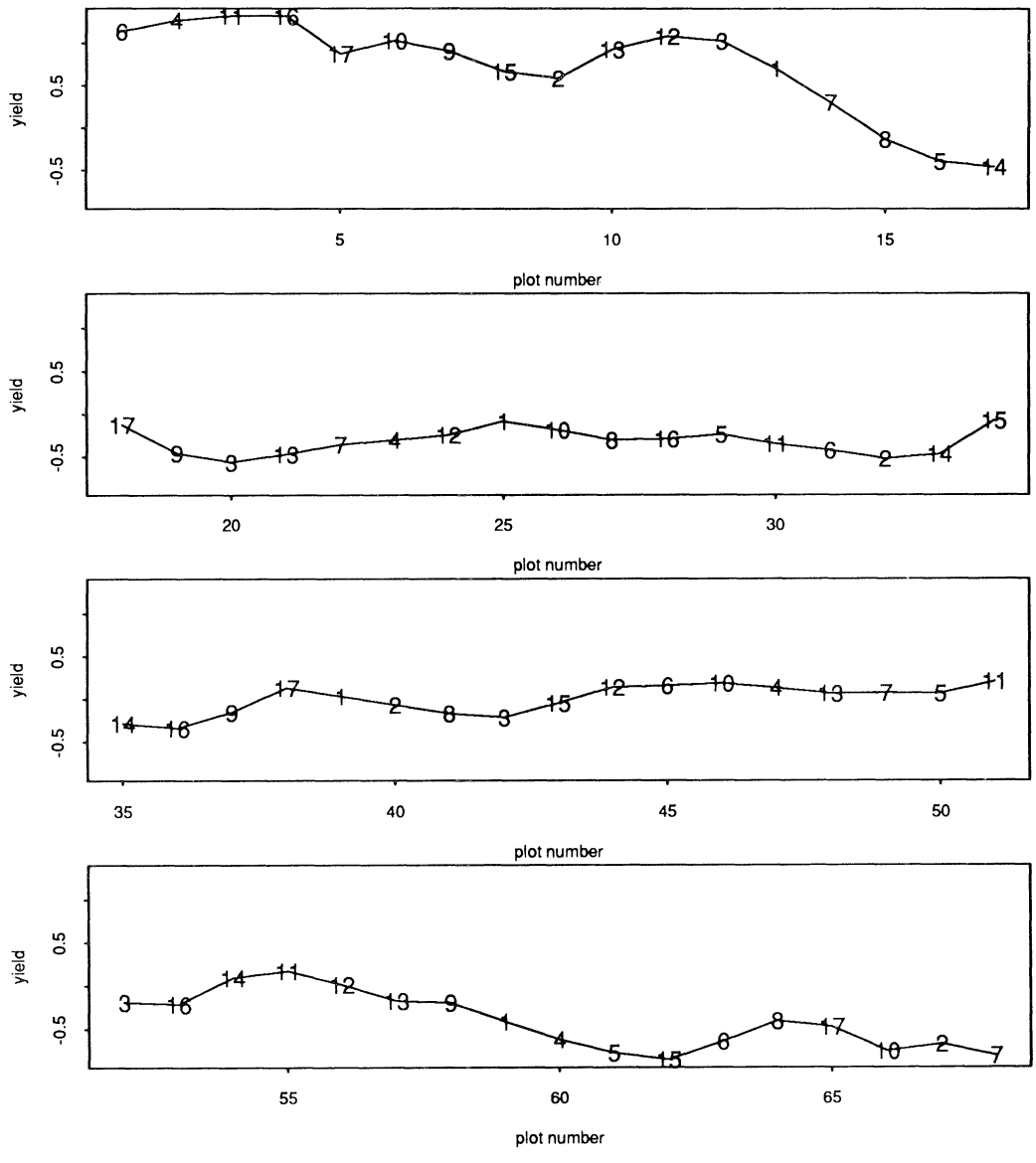

Figure 8. Trial ARC 6, treatment-adjusted yield graphed with symbol treatment number under submodel $c_t = 1$ for all t. Line is estimated fertility from $\lambda_1 = \lambda_2$ model.

Table 2
Comparison of submodels for trial ARC 6

	Submodel with $c_5 = 1$			Submodel with $c_5 = 3$		
Model	Log-like.	σ^2	λ_1	Log-like.	σ^2	λ_1
$\lambda_2 = 0$	.6	.36	.94	2.8	.50	.97
$\lambda_1 = \lambda_2$	.0	.30	.67	4.4	.36	.72

4.3 *Multiple Possible Outliers and Fertility Jumps: Trial* ARC 1

The treatment-adjusted data are plotted in Figure 9 together with the estimated fertility from the $\lambda_1 = \lambda_2$ model. A dominant feature here is the inability of the estimated fertility to increase as quickly as the treatment-adjusted data over plots 1–10. A fertility jump at plot 5 appears more appropriate, and Figure 10 displays the results conditional on the submodel $c_5 = 3$, $c_t = 1$ for all $t \neq 5$. This fertility jump is well supported by the data, with a posterior probability of over 95% from a prior of only 2%.

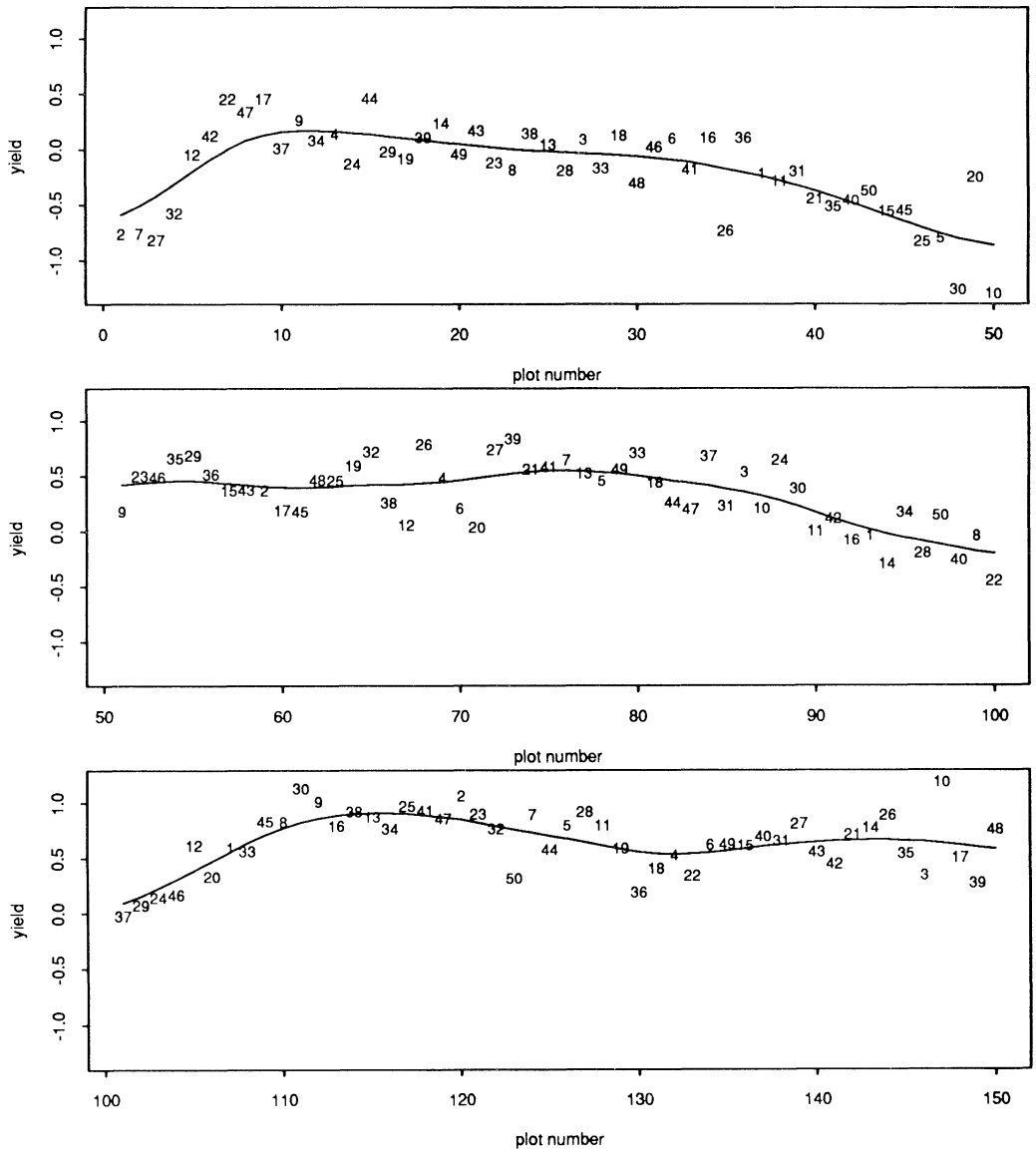

Figure 9. Trial ARC 1, treatment-adjusted yield graphed with symbol treatment number under submodel $c_t = 1$ for all t. Line is estimated fertility from $\lambda_1 = \lambda_2$ model.

We shall restrict attention here to the smoother model $\lambda_1 = \lambda_2$ because while it is superior to the $\lambda_2 = 0$ model by only .1 unit of log-likelihood when conditioned on $c_5 = 1$, it is superior by 1.4 units when $c_5 = 3$. Due to the dominance of $c_5 = 3$ over $c_5 = 1$ (under either model), the data strongly favour the $\lambda_1 = \lambda_2$ model under the robust model. Estimation of the model with general λ_1 and λ_2 results in $\hat{\lambda}_1 = \hat{\lambda}_2$.

Table 3 shows the 13 submodels with highest approximate posterior probability, and their posterior probabilities conditional on $\mathcal{C}$ only containing these submodels. Table 4 shows the marginal posterior probabilities of the 7 most likely outliers and fertility jumps from our analysis (with $\alpha = 2\%$; see Section 3.4). Note that conditioning on a single submodel is not appropriate, as the submodel with highest posterior probability conditions on a fertility jump at plot 5 only, whereas the marginal posterior probability of an outlier on plot 49 is 63%.

5. Sensitivity to Model Specification

We have conditioned on values for p_{out}, p_{jump}, and k_{out} chosen prior to the collection of the data. We have not determined these parameters from the data because the likelihood function can be quite flat, indicating that the data contain little information concerning these parameters. Nevertheless,

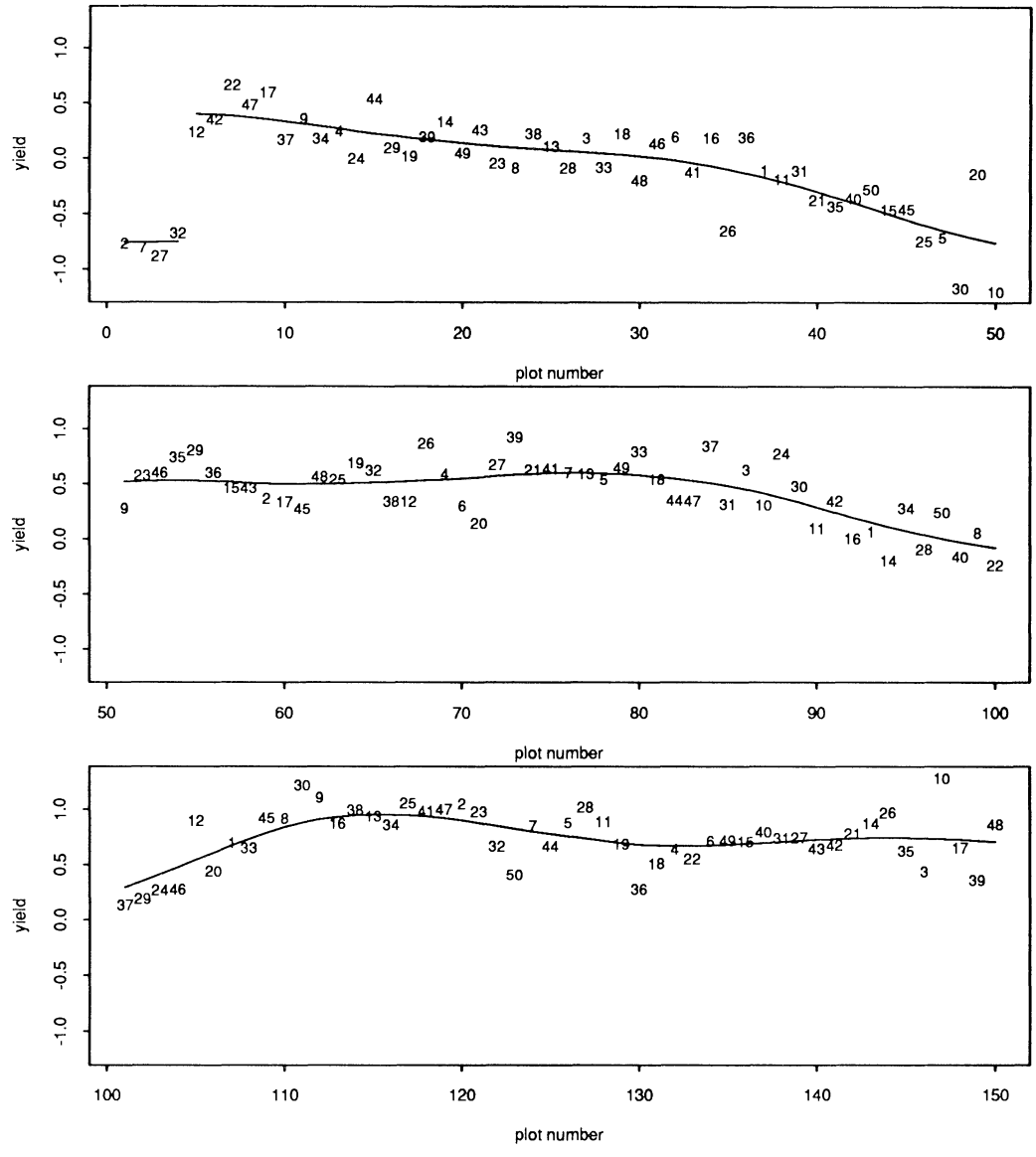

Figure 10. Trial ARC 1, treatment-adjusted yield graphed with symbol treatment number under submodel $c_5 = 3$, $c_t = 1$ for all $t \neq 5$. Line is estimated fertility from $\lambda_1 = \lambda_2$ model.

Table 3
The 12 best submodels for trial ARC 1 *and their approximate posterior probabilities*

Submodel		Posterior probability (%)
Jumps	Outliers	
5		15.6
5	49	11.0
5, 105	49	10.0
5, 48, 105	49	9.2
5, 46, 105	49	7.7
5	35, 49	7.2
5, 105		5.6
5, 105	35, 49	4.8
5	147	3.5
5, 47, 105	49	3.3
5, 48, 105	35, 49	3.1
5	35	3.1

Table 4
Marginal posterior probabilities of outliers and fertility jumps for trial ARC 1

Outlier or fertility jump	Posterior probability (%)
$j(5)$	97
$j(46)$	9
$j(48)$	14
$j(105)$	52
$o(35)$	27
$o(49)$	63
$o(147)$	17

the values of these parameters can have a substantial effect on the posterior for τ, especially if they model anomalies as very rare when they are present.

5.1 p_{out} *and* p_{jump}

We first consider how the posterior for τ is influenced by our prior choice for p_{out} and p_{jump}. This information is easily available under the approximation (6) and (7). We note that under this approximation the submodel priors influence the posterior for τ only through the mixing proportions of the posteriors for τ based on each submodel. Thus we can easily re-evaluate the posterior for τ under different prior probabilities for anomalies (assuming the same $\mathcal{C}$).

The left graph in Figure 11 displays the posterior for the treatment differences $\tau_8 - \tau_7$ in trial 8 as a function of p_{out}. It assumes the model with $\lambda_2 = 0$ and that $\mathcal{C}$ contains only the two submodels $c_t = 1$ for all t and $c_{23} = 2$, $c_t = 1$, for all $t \neq 23$. The .5%, 2.5%, 97.5%, and 99.5% quantiles are graphed along with the posterior mean. Note the insensitivity of the posterior to the exact value of p_{out} between .5% and 5% compared to the extremes of 0% and 100% (which correspond to the two submodels).

The other two graphs in Figure 11 refer to the effect of p_{jump} on the posterior for $\tau_{17} - \tau_{16}$ in trial ARC 6. Only the two submodels $c_t = 1$ for all t and $c_5 = 3$, $c_t = 1$ for all $t \neq 5$ are included in $\mathcal{C}$.

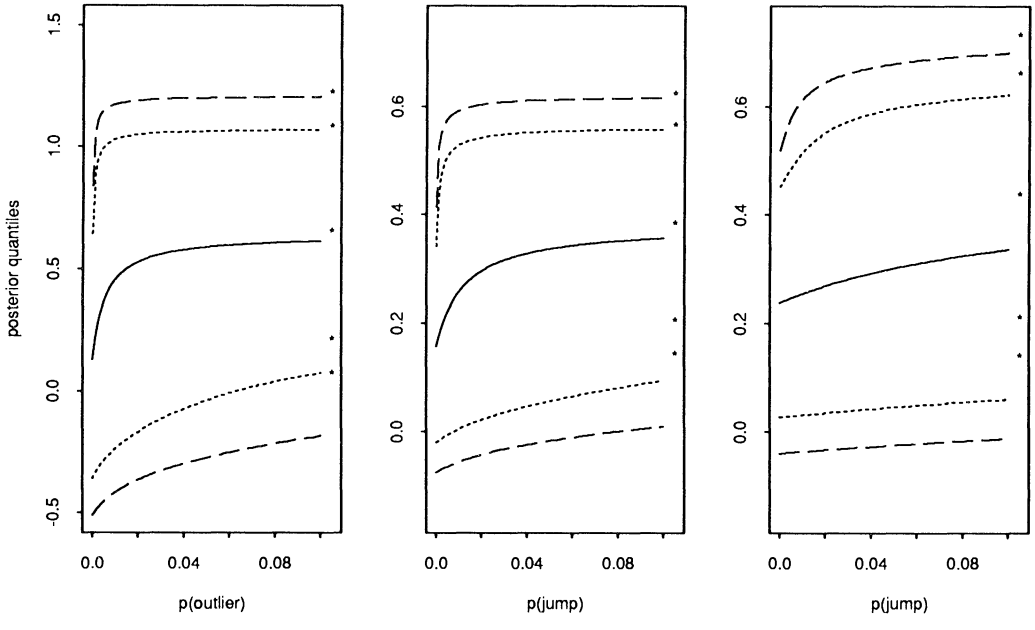

Figure 11. Effect on posterior for a treatment contrast of p_{out} and p_{jump}. Lines represent posterior mean (solid) and 2.5%, 97.5% (dotted), and .5%, 99.5% (dashed) quantiles. Asterisks represent values when p_{out} or $p_{\text{jump}} = 100\%$. Graphs from left refer to trial ARC 8 ($\lambda_2 = 0$), trial ARC 6 ($\lambda_1 = \lambda_2$), and trial ARC 6 ($\lambda_2 = 0$).

The centre graph assumes the model $\lambda_1 = \lambda_2$, whereas the model with $\lambda_2 = 0$ is assumed on the right. These graphs illustrate how similar the posteriors from different models can be after taking into account the submodel uncertainty. The indicated quantiles are similar when $p_{\text{jump}} = 2\%$ despite their being quite different for each submodel.

5.2 k_{out}

We now examine the influence of k_{out} (used to model outliers) on the posterior probabilities of the submodels. Unfortunately the situation is not as simple here, as each submodel must be refitted (i.e., optimised over θ) with every new value of k_{out}. In general this will require considerable computation, so instead we provide a simple approximation.

By definition of an outlier we require $k_{\text{out}} > 1$, and in practice we expect it to be at least as large as 5 (which will approximately remove an outlier from the analysis). If we approximate the likelihood by conditioning on estimates of τ and θ, then the ratio of the likelihoods for M_c under $k_{\text{out}} = k_{\text{out1}}$ to k_{out2} is approximately

$$\left(\frac{k_{\text{out2}}\hat{\sigma}_{\text{obs,2}}}{k_{\text{out1}}\hat{\sigma}_{\text{obs,1}}}\right)^n \exp\left(\frac{-1}{2k_{\text{out1}}^2\hat{\sigma}_{\text{obs,1}}^2}\sum_{i=1}^{n}\hat{\varepsilon}_{1,o(i)}^2\right)\Bigg/\exp\left(\frac{-1}{2k_{\text{out2}}^2\hat{\sigma}_{\text{obs,2}}^2}\sum_{i=1}^{n}\hat{\varepsilon}_{2,o(i)}^2\right), \tag{9}$$

where M_c conditions plots $o(1), o(2), \ldots, o(n)$ only as outliers, $\hat{\varepsilon}_{j,o(i)}$ is the estimated measurement error on plot $o(i)$ under $k_{\text{out}} = k_{\text{out}j}$, and $\hat{\sigma}_{\text{obs},j}$ is the estimate of σ_{obs} when $k_{\text{out}} = k_{\text{out}j}$. This approximation results by considering only the plots with outliers, since the distribution at these plots is influenced most by changing our model for outliers. The approximation is reasonable if $k_{\text{out}}^2\sigma_{\text{obs}}^2$ is large compared with σ^2 (so that outliers are approximately independent of the other plots), which will typically be the case.

Expression (9) still depends on the $\hat{\sigma}_{\text{obs},j}$, and to avoid the estimation of these for every value of k_{out}, we could condition on the same value of $\hat{\sigma}_{\text{obs}}$ obtained from, say, $k_{\text{out}} = 10$. The resulting approximation,

$$\left(\frac{k_{\text{out2}}}{k_{\text{out1}}}\right)^n \exp\left(\frac{-1}{2k_{\text{out1}}^2\hat{\sigma}_{\text{obs}}^2}\sum_{i=1}^{n}\hat{\varepsilon}_{1,o(i)}^2\right)\Bigg/\exp\left(\frac{-1}{2k_{\text{out2}}^2\hat{\sigma}_{\text{obs}}^2}\sum_{i=1}^{n}\hat{\varepsilon}_{2,o(i)}^2\right), \tag{10}$$

is very good and easily used to evaluate the effect of changing k_{out}. Note that this assumption of similar θ is more reasonable here than when comparing different submodels as practical values of k_{out} will all essentially remove outliers from the analysis. Furthermore, for values of $|\hat{\varepsilon}_{j,o(i)}|/\hat{\sigma}_{\text{obs}}$ of about 4 or 5 that create submodel uncertainty, the ratio of exponentials in equation (10) is approximately unity (especially for larger k_{out}), which results in the ratio of likelihoods being approximately $k_{\text{out2}}/k_{\text{out1}}$.

Thus, this approximation suggests using the $(p_{\text{out}}, k_{\text{out}})$ combination of $(p_{\text{out2}}k_{\text{out1}}/k_{\text{out2}}, k_{\text{out1}})$ instead of $(p_{\text{out2}}, k_{\text{out2}})$, and so Figure 11 can again be used. For example, if we wished to use $k_{\text{out}} = 5$ instead of $k_{\text{out}} = 10$, we could use Figure 11 after scaling the horizontal axis by a factor of one-half (so .10 becomes .05). This approximation also illustrates how, for larger k_{out}, it is the ratio of p_{out} to k_{out} that is of primary importance and not the individual values.

6. Discussion

We have proposed a procedure for modelling outliers and fertility jumps when estimating treatment effects in agricultural field trials. As a first stage, we have specified a state-space modelling framework for agricultural trials in the absence of such anomalies that includes many previously proposed models as special cases. One special case that has not received much previous attention, the so-called partial second-differencing model when $\lambda_1 = \lambda_2$, seems particularly interesting because of its properties and its success in examples. The overall procedure seems to be successful in accommodating outliers and fertility jumps in a routine way, and it also allows us to take account of uncertainty about model specification and the presence of anomalies when making inference about the treatment effects. The framework allows for both random and fixed effects modelling in a natural way, and permits the incorporation of prior information very easily.

The procedure we have used to model fertility jumps can be employed for any stationary fertility model by again assuming independence between fertilities across the jump. For the first-difference, or random walk, fertility model, the $F_t - F_{t-1}$ are iid $N(0, \sigma_{\text{grad}}^2)$, and we could model a fertility jump at plot t by multiplying the variance σ_{grad}^2 by a constant k_{grad}^2 (=100, say). While this has the desired effect, examples suggest that after accounting for possible fertility jumps, a smoother fertility model

is likely to be more appropriate. Modelling fertility jumps under models such as the ARIMA(1, 1, 0) or second-differencing models is not so straightforward.

There has been little attempt to model fertility jumps in agricultural field trials, although the occurrence of level shifts in general time series is often recognised. Outliers are usually dealt with by replacing these observations with missing values. The decision to designate an observation as an outlier may be made by a subjective examination of residuals, or by a formal statistical test [see, for example, Kitagawa (1979)]. Note that this is approximately equivalent to conditioning on a single submodel, which we have shown to be unsatisfactory in some circumstances.

The posterior distribution of τ can also be approximated using the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990). Here we define the state $X = (F^t, \theta^t, \tau^t, c^t)$, and from an initial value of X replace X_i with a realisation from the distribution of $(X_i|X_j\ (j \neq i), Y)$. All of these distributions are trivial except perhaps when X_i is an element of θ (Taplin, unpublished dissertation cited previously). After enough replacements, X will be approximately a realisation from the posterior $p(X|Y)$, and hence a stochastic approximation can be formed by such repeated sampling.

This procedure is efficient at determining $\mathscr{C}$, as it visits the submodels in time proportional to their posterior probability. For well-designed agricultural field trials, however, it appears relatively easy to determine $\mathscr{C}$. Furthermore, the Gibbs sampler is not as efficient at approximating $p(\tau|Y, M_c)$ as the approximation $p(\tau|Y, M_c, \hat{\theta})$ used in equation (6), which uses the normality assumptions to better advantage. Another disadvantage to the Gibbs sampler is that the estimation must be repeated for different submodel priors, so that a sensitivity analysis such as Figure 11 in Section 5 is not as immediately available. Also, by examining the submodels included in $\mathscr{C}$, the data, and using the intuition behind what the submodels represent, it may be possible to determine whether enough submodels have been included in $\mathscr{C}$. On the other hand, the partial results available to date (Raftery and Lewis, 1992) suggest that the Gibbs sampler would have to be run for a large number of iterations (perhaps on the order of 4,000–5,000) to obtain accuracy comparable to that obtained with the approximations we have used here. Thus, the Gibbs sampler may well be computationally inefficient compared to the approximations used here.

Based on our examples we suggest the following points: (i) It is better to account for uncertainty about outliers and fertility jumps than to condition on a single set of such anomalies. (ii) Possible outliers and fertility jumps can have a greater influence on our assessment of the treatment effects than the choice of any (reasonable) fertility model. (iii) Accounting for possible outliers and fertility jumps can reduce the disparity between treatment estimates from different fertility models. (iv) The possibility of outliers and fertility jumps should be considered *while* deciding on an appropriate fertility model (if one such fertility model is to be conditioned on).

Acknowledgements

This research was supported by ONR Contracts No. N00014-88-K-0265 and N00014-91-J-1074. The authors are grateful to Julian Besag and Doug Martin for very helpful discussions, and to the Scottish Agricultural Statistics Service for permission to use data from their ARC variety trials (1976–1979).

Résumé

Nous montrons comment une analyse bayésienne d'un modèle de fertilité, incorporant plusieurs modèles recommandés antérieurement, peut prendre en compte l'incertitude dans le choix du modèle de fertilité qui fournit la meilleure approximation dans chaque essai donné. Nous montrons aussi comment il est possible de tenir compte de l'incertitude due à des anomalies telles que des données suspectes ou des variations importantes de fertilité. Nous argumentons que cette attitude est préférable pour choisir un modèle "approprié." Nous montrons, par des exemples, comment la prise en compte d'anomalies possibles peut à la fois influencer le choix d'un modèle particulier de fertilité et réduire l'influence des estimations des traitements dans le choix du modèle.

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *Journal of the Royal Statistical Society, Series B* **36,** 192–236.

Besag, J. and Kempton, R. (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics* **42,** 231–251.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*, 2nd edition. Oakland, California: Holden-Day.

Cullis, B. R., Fisher, J. A., Read, B. J., and Gleeson, A. C. (1990). A new procedure for the analysis

of early generation variety trials. *Applied Statistics* **38,** 361–375.
Cullis, B. R. and Gleeson, A. C. (1989). Efficiency of neighbour analysis for replicated variety trials in Australia. *Journal of Agricultural Science, Cambridge* **113,** 233–239.
Cullis, B. R., McGilchrist, C. A., and Gleeson, A. C. (1991). Error model diagnostics in the general linear model with applications to the analysis of repeated measurements and field experiments. *Journal of the Royal Statistical Society, Series B* **53,** 409–416.
Draper, N. R. and Guttman, I. (1980). Incorporating overlap effects from neighbouring units into response surface models. *Applied Statistics* **29,** 128–134.
Gelfand, A. E. and Smith, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85,** 398–409.
Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6,** 721–741.
Gleeson, A. C. and Cullis, B. R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics* **43,** 277–287.
Green, P. J., Jennison, C., and Seheult, A. H. (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B* **47,** 299–315.
Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with Discussion). *Journal of the Royal Statistical Society, Series B* **38,** 205–247.
Jenkyn, J. F., Bainbridge, A., Dyke, G. V., and Todd, A. D. (1979). An investigation into interplot interactions, in experiments with mildew on barley, using balanced designs. *Annals of Applied Biology* **92,** 11–28.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of American Society of Mechanical Engineers, Journal of Basic Engineering* **82,** 35–45.
Kitagawa, G. (1979). On the use of AIC for the detection of outliers. *Technometrics* **21,** 193–199.
Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.
Martin, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Communications in Statistics—Theory and Methods* **19,** 55–81.
McGilchrist, C. A. (1989). Bias of ML and REML estimators in regression models with ARMA errors. *Journal of Statistical Computation and Simulation* **32,** 127–136.
Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications (with Discussion). *Journal of the American Statistical Association* **78,** 47–65.
Patterson, H. D. (1983). Contribution to Discussion of Wilkinson et al. (1983). *Journal of the Royal Statistical Society, Series B* **45,** 178–180.
Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58,** 545–554.
Raftery, A. E. and Lewis, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4*, J. M. Bernardo et al. (eds). New York: Oxford University Press.
West, M. and Harrison, P. J. (1990). *Bayesian Forecasting and Dynamic Models*. New York: Springer.
Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41,** 434–449.
Williams, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika* **39,** 151–167.

Received November 1991; *revised November* 1992 *and March* 1993; *accepted March* 1993.