Biometrika (1985), **72**, 3, *pp.* 527–37
Printed in Great Britain

Linear models for field trials, smoothing and cross-validation

By PETER J. GREEN
Department of Mathematical Sciences, University of Durham, Durham DH1 3LE, U.K.

Summary

Spatial methods for the analysis of agricultural field experiments are represented here as smoothing methods applied simultaneously with the estimation of treatment effects. Selection of both the form of the smoother and the degree of smoothing required may be based on cross-validation. Particular emphasis is placed in this paper on generalized least squares estimation in linear models, but the principle applies quite generally.

Some key words: Agricultural field experiment; Generalized least squares; Incomplete block design; Neighbour method; Papadakis method; Recovery of interblock information; Smoothing; Spatial model.

1. Introduction

The aim of many agricultural field experiments is to estimate treatment contrasts efficiently whilst avoiding bias due to trends in fertility and other environmental factors. Blocking methods are customarily used, even when blocks have no physical meaning in the experiment, but there has recently been increasing interest in adjusting for trends in a more continuous way leading to so-called 'spatial' or 'neighbour' methods that deliberately exploit the spatial context.

An early example is the method of adjustment using residuals from neighbouring plots due to Papadakis (1937); see also Bartlett (1978). Succeeding developments have been fostered by increased general interest in spatial methods and by enhanced computing power. Various recent proposals appear in the innovative paper by Wilkinson et al. (1983) and its accompanying stimulating discussion.

The intention of the present paper is to increase understanding of the proposed methods, and to aid their comparison, by representing them all as smoothing methods, whether they were originally conceived as such, for example Green, Jennison & Seheult (1983, 1985), or derived from explicit spatial stochastic models, for example Besag (1977), assumed covariance structure, for example Williams (1986) or other principles, for example Papadakis (1937) and Wilkinson et al. (1983).

We first demonstrate that the generalized least squares analysis of any linear model is a smoothing method. While this embraces only some of the spatial methods mentioned, it provides an important link with classical block-based analysis. Indeed, for incomplete block designs, Yates's analysis with recovery of interblock information (Yates, 1939, 1940) may be regarded as a prototype example for this discussion. The smoothing interpretation extends to analyses not based on least squares. Later we discuss the choice or estimation of the tuning constant or variance ratio controlling each method, advocating the criterion of cross-validation for this choice; this criterion also provides a means of selecting and assessing the method itself.

Design will not be considered here; no attempt is made at justification or robustification via randomization theory, so design plays no explicit role in analysis. It will, of course, affect efficiency (Martin, 1982; Williams, 1985).

2. GENERALIZED LEAST SQUARES

Consider the linear model

$$E(y) = D\tau + R\rho, \quad \text{var}(y) = \sigma^2 V(\phi) \tag{2·1}$$

for the vector of yields y from an experiment on n plots to compare a single set of t treatments. Here D is the design matrix for treatment effects τ and R that for any other fixed effects ρ to be fitted; we assume that the complete design matrix $X = [D \vdots R]$ has full rank. The spatial context is represented by appropriate choice of the variance matrix $\sigma^2 V(\phi)$ which is nonsingular and assumed known apart from the multiplier σ^2 and the parameter ϕ, usually scalar.

The presence of ϕ complicates an otherwise trivial estimation problem. If ϕ and hence V are known, generalized least squares leads to the estimating equations

$$X^{\mathrm{T}} V^{-1}\left\{y - X\begin{pmatrix}\tau\\ \rho\end{pmatrix}\right\} = 0,$$

from which ρ may be eliminated to yield the reduced equations for τ alone:

$$D^{\mathrm{T}}(I-S)(y-D\tau) = 0, \tag{2·2}$$

where it can be shown that (Pukelsheim, 1976)

$$S = I - V^{-1} + V^{-1} R(R^{\mathrm{T}} V^{-1} R)^{-} R^{\mathrm{T}} V^{-1} = I - \{(I-P_R) V (I-P_R)\}^{+}. \tag{2·3}$$

Here I is the $n \times n$ identity matrix, and, for any matrix A, A^{-} denotes any generalized inverse, A^{+} the Moore–Penrose inverse, and P_A the projector $A(A^{\mathrm{T}} A)^{-} A^{\mathrm{T}}$, which is invariant to the choice of generalized inverse. The second form for S emphasizes that not all of V need be specified, only the result after sweeping out the fixed effects in R; see Example 2 below.

Since V depends on ϕ, so do S and the generalized least squares estimate $\hat{\tau}$; these will be denoted $S(\phi)$, $\hat{\tau}(\phi)$ for emphasis where necessary. If the model is correct and ϕ known, this is an efficient analysis, and σ^2 may be estimated to quantify the precision of $\hat{\tau}$: however, the least squares principle does not of itself lead to an estimate of ϕ.

Example 1: *Incomplete block designs.* In the analysis of incomplete block designs, with recovery of interblock information (Cochran & Cox, 1957, Ch. 9–11), the assumed variance matrix is $\sigma^2 V(\phi) = \sigma^2(I + \phi P_Z)$; here Z is the design matrix for blocks, so for constant block size k, $P_Z = k^{-1} ZZ^{\mathrm{T}}$. This gives an error structure with two uncorrelated components: plot error with variance σ^2, and random block effects with variance $\sigma^2 \phi k^{-1}$. If the design is resolvable, R is taken to be the design matrix for fixed replicate effects; otherwise, R is a single column of ones to fit an overall mean only. Note that some authors, for example Nelder (1968), assume instead that replicate effects are random.

Example 2: *Least squares smoothing.* Green et al. (1983, 1985) describe a method of analysis for field experiments derived from smoothing based on a quadratic penalty function. The version providing one-dimensional adjustment is equivalent to a generalized least squares regression of Δy on ΔD with $\sigma^2(\lambda^{-1} I + \Delta\Delta^{\mathrm{T}})$ as assumed variance matrix for Δy, where Δ is a rectangular matrix taking second differences along lines of

adjacent plots, and λ is a tuning constant. The analysis is invariant to linear trends within lines of adjacent plots; thus we take R to be a design matrix for separate linear regressions in each such line, D fits all treatment contrasts, and V is any matrix such that $\Delta V \Delta^{\mathrm{T}} = \lambda^{-1} I + \Delta\Delta^{\mathrm{T}}$. The tuning constant λ, or λ^{-1}, plays the role of ϕ. This model was also proposed, independently, by Nelder (1983).

Other forms of least squares smoothing are discussed by Green et al. (1985). One is equivalent to the linear variance analysis of Patterson (1983) and Williams (1986), which uses V as above, but where Δ takes first differences between adjacent plots in the same replicate.

3. The connection with smoothing

Green et al. (1983) start from an explicit smoothing formulation, and later show its equivalence with generalized least squares. The connection holds quite generally, and is worth exploiting. It helps to stress that model (2·1) is only assumed in order to generate an analysis; it provides a different interpretation and possible improved algorithms, and it enables us to tie in other methods not equivalent to least squares for some linear model.

Since $V(\phi)$ is positive-definite, we may assume, after possibly rescaling σ^2 and V, that $V-I$ is nonnegative-definite. Rewrite model (2·1) as

$$y = D\tau + \xi + \eta, \tag{3·1}$$

where $\mathrm{cov}\,(\xi, \eta) = 0$, $E(\xi) = R\rho$, $\mathrm{var}\,(\xi) = \sigma^2(V-I)$, $E(\eta) = 0$, $\mathrm{var}\,(\eta) = \sigma^2 I$. Now consider the equations

$$\xi = S(y - D\tau), \tag{3·2}$$

$$\tau = (D^{\mathrm{T}} D)^{-1} D^{\mathrm{T}}(y - \xi). \tag{3·3}$$

Their simultaneous solution gives the generalized least squares estimate τ; see (2·2). But if we alternate between (3·2) and (3·3), from any initial estimates, we converge to a minimum of

$$\Omega(\tau) = (y - D\tau)^{\mathrm{T}}(I - S)\,(y - D\tau)$$

and hence to $\hat{\tau}$. For if (3·2) followed by (3·3) updates τ to τ^*, then

$$\Omega(\tau^*) = \Omega(\tau) - z^{\mathrm{T}}(I + S)\,z,$$

where $z = P_D(I - S)\,(y - D\tau)$; convergence is obtained if $(I-S)$ is nonnegative-definite and $(I+S)$ is positive-definite, and these are true for S of (2·3). Further, replacing S by $\alpha S + (1-\alpha) I$ does not affect the solution to (2·2), so that adjusting α may increase the speed of convergence: it is fastest when $\alpha \doteq 2$.

Generalized least squares estimates may thus be obtained by alternately performing an ordinary least squares regression (3·3) of $(y - \xi)$ and smoothing the residuals from fitted treatment effects (3·2). We term S a 'smoother' because its eigenvalues lie in $[0, 1]$, with not all of them equal to 0 or 1.

For the example of incomplete block designs,

$$V = I + \phi P_Z, \quad V^{-1} = I - (1+\phi)^{-1}\phi P_Z, \quad P_Z R = R,$$

so $S = (1+\phi)^{-1}(P_R + \phi P_Z)$. Thus the relevant 'smoother' involves a weighted average of the block means and the overall mean, or replicate means in the resolvable case. In neighbour methods, S corresponds more closely to the intuitive notion of smoothing.

We should clarify the status of ξ. In the model (3·1) this represents a 'trend' term incorporating both fixed and random effects. Under the additional assumption of joint normality for ξ and η, the conditional expectation of ξ given y is

$$(I - V^{-1})(y - D\tau) + V^{-1} R\rho,$$

whose generalized least squares estimate $S(y - D\hat{\tau})$ is produced by the alternating iteration described above. This generalizes ridge regression (Hoerl & Kennard, 1970): if τ and ρ are absent, and ξ has prior variance $\sigma^2 \lambda^{-1} W W^{\mathrm{T}}$, then $\xi = W\beta$, where β is estimated by $(W^{\mathrm{T}} W + \lambda I)^{-1} W^{\mathrm{T}} y$.

When the variance matrix is $\sigma^2 V = \sigma^2(I + \Sigma \phi_i W_i W_i^{\mathrm{T}})$, where $\{\phi_i\}$ and $\{W_i\}$ are known, i.e. one or more variance components besides white noise, an alternative representation as a smoothing problem is possible. Taking $\{W_i\}$ to be of full rank, we have implicitly

$$y = D\tau + R\rho + \Sigma W_i \beta_i + \eta, \tag{3·4}$$

where τ and ρ are fixed effects, and $\{\beta_i\}$ and η are uncorrelated zero-expectation random effects with $\mathrm{var}(\beta_i) = \phi_i \sigma^2 I$ and $\mathrm{var}(\eta) = \sigma^2 I$. The identity matrices may be of different orders. Minimization of the penalty function,

$$C = \Sigma \phi_i^{-1} \beta_i^{\mathrm{T}} \beta_i + \eta^{\mathrm{T}} \eta, \tag{3·5}$$

subject to the additive model (3·4) leads again to the estimate $\hat{\tau}$. Yet another equivalent formulation is to minimize the ordinary error sum-of-squares $\eta^{\mathrm{T}} \eta$ subject to (3·4) and upper bounds on $\{\beta_i^{\mathrm{T}} \beta_i\}$.

A very similar approach to smoothing is often followed in nonparametric regression problems (Wahba & Wold, 1975; Wahba, 1977). Here the model would be $y_i = \xi(t_i) + \eta_i$ and one possible penalty function is $\int \{\xi''(t)\}^2 dt + \Sigma \eta_i^2$.

Natural points of departure for generalizing the least squares smoother are the simultaneous equations (3·2) and (3·3) or the penalty function (3·5). There is no need for S to be symmetric for (3·2) and (3·3) to solve (2·2), so that asymmetric linear estimating equations such as those of Wilkinson et al. (1983) may be included. Papadakis's method, whether iterated or not, also fits this formulation naturally. As suggested by Green et al. (1985), alternative robust/resistant analyses may be obtained by use of a nonlinear smoother or treatment estimator in (3·2) and (3·3), or by amending the quadratic loss function (3·5).

4. CHOOSING ϕ

Since ϕ cannot be estimated by least squares principles, a wide variety of methods for choosing its value have been proposed. Yates's original proposal for incomplete block designs (1939, 1940) entailed equating two suitably chosen sums of squares to expectation, an approach also adopted by Williams (1986). Alternative estimators for block designs, based on normal-theory likelihood methods were given by Nelder (1968) and Patterson & Thompson (1971). We return to these criteria in §5. For Bayesian viewpoints, see Lindley & Smith (1972) and Box & Tiao (1973, Ch. 7).

An attractive, less model-dependent, alternative is to use the criterion of cross-validation, as described by Stone (1974). The idea is to treat each observation in turn as 'missing' and to 'predict' it from the model as fitted to the remaining observations for each given value of ϕ. The parameter ϕ is then chosen to minimize the mean squared error of prediction. Considerable use has been made of cross-validation and related

techniques in recent work on smoothing; see, for example, Wahba & Wold (1975), Craven & Wahba (1979) and Silverman (1985). In Stone's terminology, the term 'model' is abandoned as conveying a richer meaning than intended, and replaced by that of a 'prescription' or class of predictors.

Our prescription involves minimizing the weighted sum-of-squares

$$(y-D\tau-R\rho)^{\mathrm{T}}\, V(\phi)^{-1}(y-D\tau-R\rho),$$

or, equivalently,

$$(y-D\tau)^{\mathrm{T}}\{I-S(\phi)\}\,(y-D\tau),$$

over choice of τ, ρ, thereby estimating these, and choice of the 'missing' component(s) of y. This is equivalent to fitting a dummy covariate for each missing observation, gives generalized least squares estimates based on the available data, and provides the conventional 'missing value formula' in the case of uncorrelated data.

Introducing the dummy design matrix $\mathscr{E}$ for missing observations gives the augmented model

$$E(y) = D\tau + R\rho + \mathscr{E}\gamma, \quad \mathrm{var}\,(y) = \sigma^2\, V(\phi).$$

The reduced estimating equations for the cross-validation errors γ are $\mathscr{E}^{\mathrm{T}}\, M(y-\mathscr{E}\gamma) = 0$, where

$$\begin{aligned} M = M(\phi) &= V^{-1} - V^{-1}\, X(X^{\mathrm{T}}\, V^{-1}\, X)^{-}\, X^{\mathrm{T}}\, V^{-1} \\ &= (I-S) - (I-S)\, D\{D^{\mathrm{T}}(I-S)\, D\}^{-}\, D^{\mathrm{T}}(I-S) \\ &= \{(I-P_{D,R})\, V(I-P_{D,R})\}^{+}. \end{aligned} \tag{4·1}$$

Here, $P_{D,R}$ is the projector for the partitioned matrix $[D \vdots R]$. In particular, if a single observation, y_i say, is deemed to be missing, $\mathscr{E}$ consists of a single column of zeros with a one in row i, and the prediction error $\hat{\gamma} = \hat{\gamma}_{(i)}$, say, is

$$\hat{\gamma}_{(i)} = (\mathscr{E}^{\mathrm{T}}\, M\mathscr{E})^{-1}\, \mathscr{E}^{\mathrm{T}}\, My = \left(\sum_{j=1}^{n} M_{ij}\, y_j\right) \Big/ M_{ii}.$$

The cross-validation mean squared error is therefore

$$C(\phi) = n^{-1} \sum_{i=1}^{n} \hat{\gamma}_{(i)}^2 = \frac{1}{n} \sum_{i=1}^{n} \left\{\left(\sum_{j=1}^{n} M_{ij}\, y_j\right) \Big/ M_{ii}\right\}^2. \tag{4·2}$$

Thus, by analogy with Wahba (1977) and Craven & Wahba (1979), there is an algebraic form for $C(\phi)$ that can be computed without performing n separate regression calculations. However, there is still usually a much greater burden in evaluating $M(\phi)$ than in finding $\hat{\tau}(\phi)$. Some short-cuts are possible, see §§ 5 and 6, and also Craven & Wahba (1979), but to alleviate the problem, and to acquire a form of rotation-invariance, Wahba (1977) proposed an alternative criterion of generalized cross-validation, derived by replacing M_{ii} in (4·2) by its average, $n^{-1}\,\mathrm{tr}\,(M)$, to give

$$G(\phi) = n\{\mathrm{tr}\,(M)\}^{-2}\, y^{\mathrm{T}}\, M^2\, y. \tag{4·3}$$

Under certain conditions of balance or symmetry in designed experiments, the M_{ii} are equal. Choice of ϕ, whether scalar or vector-valued, can be made in practice by numerical minimization of (4·2) or (4·3): for scalar ϕ, we have found that both golden-section and quadratic interpolation on $\log \phi$ work well.

In the context of nonparametric regression, Silverman (1984) uses results of Utreras (1980, 1981) to amend the criterion further, by calculating the trace from eigenvalue approximations.

When using a smoother S not of the form (2·3), these algebraic simplifications are not available, but the principle of cross-validation may still be used. The prescription must define how to smooth across the gap caused by a missing observation.

A logical extension to cross-validation for selecting the parameter ϕ is to allow it to choose the form of the variance structure $V(\phi)$, presumably in practice from a small number of alternatives. Coupled with cross-validatory assessment of that choice (Stone, 1974), this may be the only reasonable way to choose between methods on the basis of an individual data set, rather than, for example, from uniformity data believed to have similar covariance structure.

Mr Robin Thompson has pointed out to me an interesting parallel between the present approach and that of restricted maximum likelihood, which may be stated in some generality as follows. Differentiating (4·3) with respect to ϕ and noting that $MX = 0$ and $MVM = M$ reveals that generalized cross-validation is equivalent to equating to their expectation under (2·1) certain sums of squares, namely $y^{\mathrm{T}} M^2 y$ and its derivative, or for vector ϕ all partial derivatives. The restricted maximum likelihood approach of Patterson & Thompson (1971), generalized to arbitrary V, does the same but with M^2 replaced by M. Some numerical comparisons will be made in §§ 5 and 6.

5. A SPECTRAL DECOMPOSITION

We now restrict attention to variance structures of the form $V = I + \phi W W^{\mathrm{T}}$ for some known matrix W, representing plot error with one other variance component: ϕ is the ratio of variances. This includes the incomplete blocks model, and least squares smoothing based on first or second differences.

From (4·1) we have

$$M^+ = (I-P_X)\,V(I-P_X) = (I-P_X)+\phi(I-P_X)\,WW^{\mathrm{T}}(I-P_X).$$

Since the two terms are symmetric and commute, they may be simultaneously diagonalized (Wilkinson, 1965, p. 52) and then, if we use the idempotence of $I-P_X$, it is straightforward to prove the existence of an orthogonal matrix U partitioned as $[U_1 \vdots U_2 \vdots U_3]$, where U_i has r_i columns, and a diagonal matrix Λ of positive eigenvalues $\{\lambda_j\}$, such that

$$P_X = U_1 U_1^{\mathrm{T}}, \quad (I-P_X)\,WW^{\mathrm{T}}(I-P_X) = U_2 \Lambda U_2^{\mathrm{T}}, \quad I-P_{X,W} = U_3 U_3^{\mathrm{T}}.$$

Then

$$M(\phi) = U_2(I+\phi\Lambda)^{-1} U_2^{\mathrm{T}} + U_3 U_3^{\mathrm{T}}, \tag{5·1}$$

so that M has eigenvalues 0 and 1 with multiplicities r_1 and r_3, and also $\{(1+\phi\lambda_j)^{-1}; j = 1, 2, \ldots, r_2\}$. Note that this explicitly demonstrates how M varies from $(I-P_X)$ to $(I-P_{X,W})$ as ϕ increases from 0 to ∞; for example, in the incomplete blocks model, these limits corresponding to ignoring block effects, and to fitting them as fixed, respectively. The decomposition used here is essentially that used by Patterson & Thompson (1971) for block models.

It follows from (5·1) that $dM/d\phi = \phi^{-1}(M^2 - M)$, so that by the remarks at the end of

§4, generalized cross-validation is equivalent to equating to expectation $y^{\mathrm{T}}M^r y$ for $r = 2$ and 3, and restricted maximum likelihood similarly for $r = 1$ and 2.

When the $\{\lambda_j\}$ are equal, all positive powers of M are convex combinations of $I-P_X$ and $I-P_{X,W}$ so that both criteria are algebraically equivalent to Yates's estimate of ϕ for incomplete block designs (Yates, 1939, 1940); in our notation, with $W = P_Z$, this entails equating to expectation the intrablock residual sum-of-squares, $y^{\mathrm{T}}(I-P_{X,W})\,y = S_{X,W}$, and the blocks sum-of-squares, within replicates if relevant, adjusted for treatments, $y^{\mathrm{T}}(P_{X,\,W}-P_X)\,y = S_X - S_{X,W}$.

The requirement of equality of the nonzero eigenvalues $\{\lambda_j\}$ is strong. It demands considerable implicit symmetry in the design; for example, it does not apply to all balanced incomplete block designs. If WW^{T} is proportional to a projector for some factor, the requirement is of first-order balance for this factor with respect to X, in the sense of James & Wilkinson (1971).

Some numerical examples are given in Table 1 for the same collection of balanced incomplete block designs considered by Jensen & Stone (1976) in their application of cross-validation to these designs. Note the exact or close agreement between cross-validation and the criteria of Yates (1940), Nelder (1968) and Patterson & Thompson (1971). When the eigenvalues $\{\lambda_j\}$ differ, the exact connection is broken; however if they do not differ too markedly, the argument above suggests approximate equivalence of the two criteria.

It will be noted from Table 1 that Jensen & Stone (1976) obtained different weights from a cross-validation argument based on a different prescription. They obtain separate predictors for the intrablock and interblock extremes of ϕ, that is ∞ and -1, regarding an entire block as missing in deriving the interblock predictor. The Stein estimates differ considerably from the rest, apparently because they utilize interblock information only through the treatment component: Stein's estimator (Stein, 1966) uses only $\hat{\tau}(-1)$, $\hat{\tau}(\infty)$ and $S_{X,W}$.

Table 1. *Balanced incomplete block designs: estimates of ϕ for four data sets, according to five different criteria*

	Data set				
	Davies (1954, p. 207)	Davies (1954, p. 216)	Quenouille (1953, p. 177)	John (1971, p. 226)	John (1971) fitting replicates
			n, t, k		
Estimate	12, 4, 3	20, 5, 4	30, 6, 3	36, 9, 3	36, 9, 3
Jensen & Stone (1976)*	13·08	−0·6357	2·938	−0·4444	—
Yates (1940)	21·40	−0·5033	3·320	0·1348	−0·1685
Cross-validation of §4	21·40	−0·5033	3·446	0·1975	−0·1685
Nelder (1968)	21·40	−0·5033	3·355	0·2102	−0·1685
Stein (1966)*	68·32	0·2640	9·123	0·4303	—

Where ϕ is estimated as negative, it would be customary to use $\phi = 0$, that is to use no block adjustment.

Rows marked * were given for these data sets by Jensen & Stone (1976).

Nelder's estimates are identical with those of Patterson & Thompson (1971) since the blocks are of equal size.

To derive an analogue of Yates's estimate for ϕ, valid for general known W in $V = I+\phi WW^{\mathrm{T}}$, note that in consequence of (5·1) S_X and $S_{X,W}$ are the limits, as $\phi \to 0$

and ∞ respectively, of the weighted error sum-of-squares

$$\Gamma(\phi) = y^{\mathrm{T}} M y = \min_{\tau,\rho}\{(y-D\tau-R\rho)^{\mathrm{T}} V^{-1}(y-D\tau-R\rho)\}$$

that would be the focus of attention if ϕ were known. Equating $\Gamma(0)$ and $\lim \Gamma(\phi)$ as $\phi \to \infty$ to expectation under model (2·1) therefore seems the natural analogue, although this procedure is not as cheap and convenient as in the incomplete blocks case, because the expectations in general involve nontrivial trace terms. Further, if D, R and W together span R^n, for example, least squares smoothing with first or second differences, $\Gamma(\phi) \to 0$ as $\phi \to \infty$. In this situation, Williams (1986) tacitly uses $\lim \phi\Gamma(\phi)$ as $\phi \to \infty$ as the second sum of squares: from (5·1) we see that this equals $y^{\mathrm{T}} N y$, where $N = \{(I-P_X) W W^{\mathrm{T}} (I-P_X)\}^+$. Provided that $(I-P_V)(I-P_R) = 0$, the result of Pukelsheim (1976), stated in (2·4), remains true when V is singular, if V^{-1} is replaced by V^+. Thus, replacing V by WW^{T} and R by X, we see that, if $(I-P_W)(I-P_X) = 0$, we have

$$N = \Delta^{\mathrm{T}}\Delta - \Delta^{\mathrm{T}}\Delta X(X^{\mathrm{T}}\Delta^{\mathrm{T}}\Delta X)^- X^{\mathrm{T}}\Delta^{\mathrm{T}}\Delta = \Delta^{\mathrm{T}}(I-P_{\Delta X})\Delta,$$

if Δ is chosen so that $(WW^{\mathrm{T}})^+ = \Delta^{\mathrm{T}}\Delta$; for example $\Delta = (W^{\mathrm{T}} W)^{-1} W^{\mathrm{T}}$, if W has full rank. Thus $y^{\mathrm{T}} N y$ is the residual sum-of-squares from ordinary regression of Δy on ΔX.

6. Application to least squares smoothing

The spectral decomposition of M given in (5·1) reduces the computation in least squares smoothing methods, whether in the form based on second differences as discussed in detail by Green et al. (1985) or in the generalizations described here.

All of the methods for choosing the tuning constant ϕ^{-1} proposed by Green et al. (1985) involve computing the decomposition $y = D\hat{\tau} + \hat{\xi} + \hat{\eta}$ for several values of ϕ^{-1}. The spectral decomposition (5·1), while incurring a set-up cost, permits the decomposition and various derived statistics to be computed very cheaply for subsequent values of ϕ^{-1}.

When $V(\phi) = I + \phi W W^{\mathrm{T}}$, note that

$$V^{-1} = I - W(\phi^{-1} I + W^{\mathrm{T}} W)^{-1} W^{\mathrm{T}} \to I - P_W$$

as $\phi \to \infty$. The fixed effects τ, ρ in model (2·1) are estimated by Ty where

$$T = T(\phi) = (X^{\mathrm{T}} V^{-1} X)^{-1} X^{\mathrm{T}} V^{-1}.$$

Note that $M = V^{-1}(I-XT)$. Now,

$$T(\infty)\{I-M(\phi)\} = \{X^{\mathrm{T}}(I-P_W)X\}^{-1} X^{\mathrm{T}}(I-P_W)[I-V(\phi)^{-1}\{I-XT(\phi)\}] = T(\phi),$$

since $(I-P_W)(I-V^{-1}) = 0$. Thus if we write $\hat{\eta} = \hat{\eta}(\phi) = M(\phi) y$, the decomposition of $y - \hat{\eta}(\phi)$ as $D\hat{\tau}(\phi) + \hat{\xi}(\phi)$ is obtained by linear transformations not depending on ϕ: only $M(\phi)$ need be recomputed for each ϕ, and that from the spectral decomposition (5·1).

Numerical examples of choices of ϕ are given in Table 2: the data sets are those used by Green et al. (1985).

7. A simulation study

To compare the use of cross-validation in choosing ϕ with other criteria, we performed a simulation experiment, making use of the algebraic results in §§ 4, 5 and 6. The form of

Table 2. *Choice of ϕ in second-difference least squares smoothing: data sets as described and analysed by Green et al.* (1985)

Estimate	Mildew $n, t = 38, 4$	Data set SB 77 ES 5 $n, t = 51, 17$	SB 77 ES 6 $n, t = 48, 16$
Yates	0·11	0·0162	0·0603
Tukey	1·86	0	0·0613
Restricted maximum likelihood	2·79	0·0041	0·0591
Generalized cross-validation	3·06	0·0051	0·0809
Cross-validation	3·40	0·0022	0·1113
Maximum likelihood	5·27	0·0238	0·1915

$V(\phi)$ assumed was that corresponding to ordinary least squares smoothing; see Example 2 in § 2. Three factors were varied in the experiment:

(i) the true variance structure for y: three alternatives, in each case white noise η plus correlated ξ, all jointly normally distributed with: (a) first differences, or (b) second differences of ξ independent and identically distributed, or (c) ξ as a 7-point equally weighted moving average;

(ii) the true roughness, measured by $\phi_{\text{true}} = (n\xi^{\mathrm{T}}\Delta^{\mathrm{T}}\Delta\xi)/\{(n-2)\eta^{\mathrm{T}}\eta\}$: two alternatives, 0·02 and 0·2;

(iii) the design and number of treatments: two alternatives, both with $n = 48$ plots: one adapted from a serially balanced design on 4 treatments, the other a triple lattice with 3 replicates of 16 treatments, block size 4.

For each of the resulting $12 = 3 \times 2 \times 2$ cases, 100 replicates were performed: ξ and η were drawn independently from (i) and scaled according to (ii), then the sum $y = \xi + \eta$ analysed by least squares smoothing assuming design (iii). Thus true treatment effects were set to zero.

Table 3. *Simulation for least squares smoothing: 10th and 50th percentiles of percentage efficiency for 5 criteria for choice of variance parameter, ϕ; variance structures,* (a), (b) *and* (c)

			10th percentile					50th percentile				
			Criterion					Criterion				
	ϕ_{true}	t	CV	GCV	Tukey	REML	ML	CV	GCV	Tukey	REML	ML
(a)	0·02	4	45	47	45	45	45	93	94	94	96	96
		16	79	77	78	80	78	97	97	96	97	97
	0·2	4	38	37	36	35	36	88	84	87	89	89
		16	73	73	76	78	76	95	94	93	93	93
(b)	0·02	4	40	39	36	37	39	87	85	84	88	86
		16	71	65	75	78	74	95	95	95	96	96
	0·2	4	18	17	19	17	19	79	81	79	78	77
		16	70	71	74	71	67	92	92	94	93	93
(c)	0·02	4	42	42	42	42	42	94	93	96	96	96
		16	77	80	82	85	82	97	97	98	98	97
	0·2	4	25	27	28	26	26	83	84	89	84	87
		16	71	71	71	75	70	88	90	90	93	90

Designs, each 48 plots in one line.
For $t = 4$: 4 2 3 2 1 4 3 4 1 2 1 3 4 2 1 2 3 4 3 2 4 1 4 1 3 2 3 1 2 4 2 4 3 1 3 1 4 2 3 2 1 4 3 4 1 2 1 3.
For $t = 16$: 8 4 16 12 7 11 3 15 1 13 5 9 6 2 10 14 5 12 15 2 3 14 8 9 11 16 6 1 7 10 4 13 11 2 13 8 6 4 15 9 16 5 3 10 7 12 14 1.
CV, cross-validation; GCV, generalized cross-validation; REML, restricted maximum likelihood.

Five criteria were compared: ordinary and generalized cross-validation, and three methods suggested by Green et al. (1985): 'Tukey's rule', in which $y^{\mathrm{T}} My/\{\mathrm{tr}\,(M)\}^2$ is minimized, and full and restricted maximum likelihood. The resulting choices of ϕ were extremely highly correlated. It was therefore sufficient to examine the criteria marginally: an appropriate measure of the efficiency of a particular choice ϕ^* is

$$\min_{\phi} \sum_{j=1}^{t} \{\hat{\tau}_j(\phi)-\tau_j\}^2 \Big/ \sum_{j=1}^{t} \{\hat{\tau}_j(\phi^*)-\tau_j\}^2,$$

where $\{\tau_j\}$ are the true values, here zero. Note that 100% efficiency cannot be attained. For each criterion, the empirical distribution of efficiency, from the 100 replicates, was constructed. Some of the results are presented in Table 3. They demonstrate rather close agreement between the criteria, and suggest no clear preferences. The superiority of generalized cross-validation over the ordinary version found by Craven & Wahba (1979) is not apparent here, presumably because of the well-conditioned nature of these designed experiments.

Since the other criteria are less readily adapted to a variety of smoothing methods, especially those not derived from least squares, these results support the use of cross-validation for choice of ϕ.

Acknowledgements

I am particularly grateful to Robin Thompson for making the connection with restricted maximum likelihood, to Chris Jennison for other constructive comments on an earlier version of this paper, and to the referees for suggestions that greatly simplified the algebraic details. I appreciate the support of the United States Army and the Mathematics Research Center, Madison, Wisconsin, where this work was completed.

References

Bartlett, M. S. (1978). Nearest neighbour models in the analysis of field experiments (with discussion). *J. R. Statist. Soc.* B **40**, 147–74.

Besag, J. E. (1977). Errors-in-variables estimation for Gaussian lattice schemes. *J. R. Statist. Soc.* B **39**, 73–8.

Box, G. E. P & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Reading, Mass: Addison-Wesley.

Cochran, W. G. & Cox, G. M. (1957). *Experimental Designs,* 2nd ed. New York: Wiley.

Craven, P. & Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31**, 377–403.

Davies, O. L. (1954). *The Design and Analysis of Industrial Experiments.* London: Oliver and Boyd.

Green, P. J., Jennison, C. & Seheult, A. H. (1983). Discussion of paper by G. N. Wilkinson et al. *J. R. Statist. Soc.* B **45**, 193–5.

Green, P. J., Jennison, C. & Seheult, A. H. (1985). Analysis of field experiments by least squares smoothing. *J. R. Statist. Soc.* B **47**. To appear.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal problems. *Technometrics* **12**, 55–67.

James, A. T. & Wilkinson, G. N. (1971). Factorization of the residual operator and canonical decomposition of nonorthogonal factors in analysis of variance. *Biometrika* **58**, 279–94.

Jensen, L. & Stone, M. (1976). Cross-validatory choice of weights for inter- and intra-block estimation in balanced incomplete block designs. *Biometrics* **32**, 677–81.

John, P. W. M. (1971). *Statistical Design and Analysis of Experiments.* New York: Macmillan.

Lindley, D. V. & Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc.* B **34**, 1–41.

Martin, R. J. (1982). Some aspects of experimental design when errors are correlated. *Biometrika* **69**, 597–612.

NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. R. Statist. Soc.* B **30**, 303–11.

NELDER, J. A. (1983). Discussion of paper by G. N. Wilkinson et al. *J. R. Statist. Soc.* B **45**, 200–2.

PAPADAKIS, J. S. (1937). Méthode statistique pour des expériences sur champ. *Bull. Inst. Amél. Plantes à Salonique* **23**.

PATTERSON, H. D. (1983). Discussion of paper by G. N. Wilkinson et al. *J. R. Statist. Soc.* B **45**, 178–80.

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–54.

PUKELSHEIM, F. (1976). Estimating variance components in linear models. *J. Mult. Anal.* **6**, 626–9.

QUENOUILLE, M. H. (1953). *The Design and Analysis of Experiments.* London: Griffin.

SILVERMAN, B. W. (1984). A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Am. Statist. Assoc.* **79**, 584–9.

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

STEIN, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In *Research Papers in Statistics.* Ed. F. N. David, pp. 351–66. Chichester: Wiley.

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions (with discussion). *J. R. Statist. Soc.* B **36**, 111–47.

UTRERAS, D. F. (1980). Sur le choix du parametre d'ajustement dans le lissage par fonctions spline. *Numer. Math.* **34**, 15–28.

UTRERAS, D. F. (1981). Optimal smoothing of noisy data using spline functions. *S.I.A.M. J. Sci. Statist. Comput.* **2**, 349–62.

WAHBA, G. (1977). A survey of some smoothing problems, and the method of generalized cross-validation for solving them. In *Applications of Statistics*, Ed. P. R. Krishnaiah, pp. 507–23. Amsterdam: North Holland.

WAHBA, G. & WOLD, S. (1975). A completely automatic French curve: fitting spline functions by cross-validation. *Comm. Statist.* **4**, 1–17.

WILKINSON, G. N., ECKERT, S. R., HANCOCK, T. W. & MAYO, O. (1983). Nearest neighbour (NN) analysis of field experiments (with discussion). *J. R. Statist. Soc.* B **45**, 151–210.

WILKINSON, J. H. (1965). *The Algebraic Eigenvalue Problem.* Oxford University Press.

WILLIAMS, E. R. (1985). A criterion for the construction of optimal neighbour design. *J. R. Statist. Soc.* B. To appear.

WILLIAMS, E. R. (1986). A neighbour model for field experiments. *Biometrika* **73**. To appear.

YATES, F. (1939). The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Ann. Eugen.* **9**, 136–56.

YATES, F. (1940). The recovery of inter-block information in balanced incomplete block designs. *Ann. Eugen.* **10**, 317–25.

[*Received July* 1984. *Revised December* 1984]