Journal of Agricultural Science (2006), **144**, 303–316. © 2006 *Cambridge University Press*
doi:10.1017/S0021859606006289 *Printed in the United Kingdom*

Estimation of economically optimum seed rates for winter wheat from series of trials

C. M. THEOBALD[1]*, A. M. I. ROBERTS[2], M. TALBOT[2] AND J. H. SPINK[3]

[1] *Biomathematics & Statistics Scotland and School of Mathematics, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JZ, UK*
[2] *Biomathematics & Statistics Scotland, The King's Buildings, Edinburgh EH9 3JZ, UK*
[3] *ADAS Rosemaund Research Centre, Preston Wynne, Hereford, HR1 3PG, UK*

(*Revised MS received 4 May 2006, First published online 31 July 2006*)

SUMMARY

The results of recent trials for winter wheat (*Triticum aestivum* L.) have influenced farming practice in the UK by encouraging the use of lower seed rates. Spink *et al.* (2000) have demonstrated that, particularly if sown early, wheat can compensate for reduced plant populations by increased tiller production.

Results from seed-rate trials are usually analysed separately for each environment or each combination of environment and variety, and not combined into a single model. They therefore address the question of what the best seed rate would have been for each combination, rather than answer the more relevant question of what rate to choose for a future site. The current paper presents a Bayesian method for combining data from seed-rate trials and choosing optimum seed rates: this method can incorporate information on seed and treatment costs, crop value and covariates. More importantly, for use as an advisory tool, it allows incorporation of expert knowledge of the crop and of the target site.

The method is illustrated using two series of trials: the first, carried out at two sites in 1997–99, investigated the effects of sowing date and variety in addition to seed rate. The second was conducted at seven sites in 2001–03 and included latitude and certain management factors. Recommended seed rates based on these series vary substantially with sowing date and latitude.

Two non-linear dose-response functions are fitted to the data, the widely used exponential-plus-linear function and the inverse-quadratic function (Nelder 1966). The inverse-quadratic function is found to provide a better fit to the data than the exponential-plus-linear and the latter function gives estimated optimum rates which are as much as 40% lower. The economic consequences of using one function rather than the other are not great in these circumstances.

The method is found to be robust to changes in the prior distribution and to other changes in the model used for dependence of yield on sowing date, latitude, variety and management factors.

INTRODUCTION

The use of lower seed rates for winter wheat in the UK has been encouraged by the results of trials reported by Spink *et al.* (2000), which were carried out at two sites in the English Midlands in harvest years 1997–99. These trials used four varieties, a wide range of seed rates and sowing dates from September to December: the results suggested that plant populations could be reduced, thus lowering costs and reducing the risk of lodging. The present paper examines this series and a subsequent series of trials, conducted using a single variety at the same sites and five others between the south coast of England and northern Scotland in harvest years 2001–03. The intention was to reach more general conclusions about how the dependence of yield on seed rate is related to latitude, and to agronomic factors such as slug control and position in the crop rotation.

In order to consider the evidence on optimum seed rates, and their dependence on treatments, varieties and covariates such as sowing date and latitude, a Bayesian method has been developed for combining

---

* To whom all correspondence should be addressed. Email: cmt@bioss.ac.uk

Table 1. *Trial sites*

Site	Latitude (degrees)	Harvest years	Sowing dates (from 1 Jan)	Treatments varied in Phase II		
				2001	2002	2003
Aberdeen	57·34	2001–03	266–280	PGR	N timing	N timing
Edinburgh	55·87	2001–03	280–290	Slug treatments	N timing	Slug treatments
High Mowthorpe	54·11	2001–03	270–319	PGR	Rotation	–
Sutton Bonington	52·83	1997–99	276–285	–	–	–
		2001–03	263–293	N timing	N timing	N timing
Rosemaund	52·13	1997–99	266–350	–	–	–
		2001–03	266–278	Slug treatments	PGR	Rotation
Bridgets	51·10	2001	279	N timing	–	–
Mamhead	50·62	2002–03	276–282	–	PGR	Rotation

data from seed-rate trials and choosing optimum rates. The data available are first described and several issues relating to the analysis of seed-rate data and the resulting recommendations on rates are identified and investigated. Analyses of the two sets of trials are presented, assuming exponential-plus-linear and inverse-quadratic dose-response functions. In addition to relating crop yields to seed rates, these analyses examine the effects of sowing date and variety (in the 1997–99 data) and latitude and treatments (in the 2001–03 series). They allow optimum seed rates to be determined for combinations of date and variety or of latitude and treatment under assumptions about the costs of the seed and of treatments and the value of the grain. The results can also be compared using the two dose-response functions, and it can be shown that the fit of the exponential-plus-linear function is poor for both trial series. In addition, the effects of changing the prior distribution for the unknown model parameters and of various extensions of the model are examined.

Table 2. *Treatment factors included in Phase II: the standard level is underlined, and any differences in cost for others are given in parentheses (per ha or per tonne of seed)*

Treatment	Level 1	Level 2	Level 3
Rotation	First	Second/third	Second/third +Latitude[1] (£150/t)
Slug treatment	None (−£13/ha)	Post drilling	Post drilling +Sibutol Secur (£80/t)
Nitrogen timing	Early	Normal	Late
PGR use	None (−£9·5/ha)	At tillering	At stem extension

[1] In this table 'Latitude' refers to the proprietary name of a fungicide.

THE DATA AVAILABLE

The data relate to two phases of work, as follows.

Phase I: Based at Rosemaund (Herefordshire, UK) and Sutton Bonington (Leicestershire, UK) in harvest years 1997–99. Varieties Cadenza, Haven, Soissons and Spark were used, with seed rates of 20, 40, 80, 160, 320 and 640 seeds/m^2. Sowing dates ranged from September to December: see Spink *et al.* (2000) for further details.

Phase II: Based at seven sites between the south coast of England and North East Scotland in 2001–03. A single variety, Claire, was used with seed rates of 40, 80, 160, 320 and 640 seeds/m^2. In addition to a wide range of latitudes, it included treatments related to rotational position, slug control, nitrogen timing and PGR use. Each treatment had a standard level (the level used in Phase I) and two others. Except at High Mowthorpe in 2003, one of these treatments was investigated each year at each site. The sowing dates chosen were intended to be typical of winter wheat sowings for each site.

Table 1 gives information on the sites (ordered from north to south). They include one that was replaced by another with similar latitude after the first year of Phase II. Table 2 lists the treatments used in this phase, and shows the costs assumed for them relative to the standard level.

The investigation of rotational position in Phase II contrasted first and third wheat at High Mowthorpe and Rosemaund, but first and second at Mamhead: the present analysis ignores the distinction between second and third wheats. A split-plot design was used with agronomic treatment applied to main plots and seed rate to subplots, replicated four times.

Rather than including year and site effects and their interactions in the models, the environments are related by including dependence on sowing date (in Phase I) and on the latitude of the trial site (in Phase II).

Mean yields have been analysed over blocks for each variety rather than individual plot values. Plot values were missing for 0·7% and 2·8% of plots in Phases I and II respectively: for these, variety means were calculated using the ANOVA Procedure in GenStat (Genstat 2002). There were 288 means in Phase I and 260 in Phase II, 90 of them relating to the standard combination of treatments for this trial series. These means are referred to in the present paper as 'yields'.

The assumed value of grain was £80/tonne, the cost of seed £300/tonne and the average mass of a seed 45 mg.

ISSUES ARISING IN THE ANALYSIS OF SEED-RATE DATA

Separate or combined analyses?

The usual method for combining information from seed-rate (or fertilizer) experiments over several environments is to fit a parametric dose-yield function, such as the exponential-plus-linear function, for each environment, compare the estimates for the various environments and combine them in an informal way. Fisher (1935) and Yates & Cochran (1938) recognized the need to combine information on variety yields over several environments: methods and models for this purpose are reviewed by Patterson (1997) and Smith *et al.* (2005). It is argued below that there are also advantages in a combined analysis of seed-rate data over environments. Combining a large number of non-linear regression analyses with common parameters has been too complex to attempt until recent years, but software now exists which makes it feasible.

Modelling seed-rate data from several environments within a single model that includes effects for the environments, rather than carrying out individual analyses, has the following advantages.

(1) Individual analyses are usually based on rather small data sets, and are thus likely to lead to unsatisfactory estimates of optimum rates. With an exponential-plus-linear model, for example, it is easy to obtain parameter estimates that imply the expected yield increases indefinitely with seed rate, possibly leading to no 'optimum' rate.

(2) One might wish to relate estimated optima from individual analyses to sowing date and to characteristics of the site, such as latitude and soil type, which can be expected to have substantial effects on the optimum seed rate. This is hampered by the sensitivity of the estimates to changes in the model fitted and the seed price assumed.

(3) If environmental characteristics such as those mentioned above can be included in a combined analysis, this offers the possibility of seed-rate recommendations that are specific to each target site.

(4) Individual analyses address the question 'What would the best seed rate have been in this environment?', whereas the grower has to choose the rate for a site elsewhere on a future occasion. For this task, some assumption is needed about how the target site is related to the trial sites. At the simplest, this relationship would be that the trial and target environments form a random sample from some population, so that we are led to model environment differences using a random-effects model. This model may be generalized, for example by incorporating covariate information on the environments.

(5) If the analysis for each environment includes factors such as the treatments listed in Table 2, and non-significant factors are omitted from the fitted models, there is the extra difficulty of combining information over different models.

One-stage or two-stage analysis?

The present paper follows the traditional practice in the analysis of European multi-environment field trials of first calculating mean yields over blocks for each variety and then analysing these means. Recent developments in mixed models, and increased computing power, permit the analysis of individual plot values and the inclusion of spatial effects: see, for example, Smith *et al.* (2001). A two-stage analysis is considered in the present paper for simplicity of exposition, but it should be noted that a one-stage Bayesian analysis of multi-environment trials including spatial effects is feasible (Besag & Higdon 1999).

Why use a Bayesian analysis?

The choice of seed rate is an example of decision-making in a situation of uncertainty. Theoretical investigation of how decisions should be made under uncertainty in order to be coherent has shown that the uncertainty should be expressed using a probability distribution, and that the possible consequences of the available decisions should be described by using a utility function; the optimum decision is that which maximizes the expected utility over the unknown parameters: see, for example, Raiffa & Schlaifer (1961) and Bernardo & Smith (1994). From this point of view, it is necessary to treat the unknown parameters in a statistical model as random variables, i.e. to adopt a Bayesian approach. In contrast, the usual approach to determining the optimum seed rate is to consider the optimum given the values of the unknown parameters and to replace these parameters by estimates based on the trial data: because of sampling variation in the estimates, the resulting 'optimum' is not the best rate.

The requirement that the unknown model parameters should be taken as random in a Bayesian analysis means that they have to be assigned a probability distribution, the *prior* distribution, reflecting the analyst's knowledge before the data are considered. The choice of prior distribution may be difficult, particularly for models with many parameters, but in the present context it provides the opportunity to incorporate expert knowledge of the crop. Such knowledge includes assessments of likely yields in relation to seed rate and of the extent of variation between environments. It may be based on experience of many more seasons, locations and varieties than are represented in the data. In particular, it can automatically exclude parameter values corresponding to unreasonable models, such as exponential-plus-linear models without finite maxima. The prior distribution is combined with the information in the data to form a *posterior* distribution for the parameters.

The Bayesian approach also requires that the utility of growing the crop be specified to take account of the crop value and the costs of seed, treatments and management. Then – given the data and the prior distribution – the optimum choice of rate (and possibly of variety or treatment) is that which maximizes the expected utility over the posterior distribution, known as the *posterior expected utility*. The present authors have implemented the necessary calculations using the WinBUGS program (Spiegelhalter *et al.* 2003), which is freely available from http://www.mrc-bsu.cam.ac.uk/bugs.

An accessible introduction to Bayesian statistics for biologists (in the context of conservation biology) is given by Marin *et al.* (2003), and a standard work on the Bayesian approach is Gelman *et al.* (2003). Lindley (1985) emphasizes the decision-making aspects of this approach.

Inclusion of non-standard treatments

Some managerial factors, such as rotational position and treatments for seed against Take-all and slugs, are decided before or at the same time as the choice of seed rate. Other treatments, including nitrogen timing and PGR use, are chosen later in the growing season after examining the crop. To assess the value of the latter treatments fully one should consider not only their cost and the resulting change in yield but also whether criteria for applying them were satisfied in each environment. Such knowledge is not available here, so the benefit of varying these treatments from the standard level cannot be assessed. Instead, the way in which they affect the expected profit margin if they are used regardless of the criteria is examined, assuming that seed rate is near the optimum for a standard set of treatments.

Dose-response functions

Several dose-response functions might be considered for relating crop yield to seed rate. It is assumed that in any environment the expected yield would increase to a maximum with seed rate and then decline, and the method proposed for incorporating prior information is limited to functions with this behaviour. An alternative would be to use cubic smoothing splines within a mixed-model framework: this approach is applied to seed rates for wheat by Walker *et al.* (2002) and Lemerle *et al.* (2004). It permits the inclusion of experimental and spatial effects, but does not ensure that the fitted functions have single maxima. The following two dose-response functions are considered, assuming their intercepts at zero seed rate to be zero.

(1) The exponential-plus-linear function, which may be expressed as $\beta(\rho^x-1)-\kappa x$ $(x>0)$, where x denotes the seed rate and β, ρ and κ are unknown parameters. This appears to be the function employed most often in studies of seed rate and of nitrogen fertilizer, and is used in Spink *et al.* (2000). However, at extremely high rates it either goes negative or carries on increasing, neither of which is realistic.

(2) The inverse-quadratic function (Nelder 1966). This is the ratio of a linear function and a quadratic, and may be written as $x/(\beta_0+\beta_1 x+\beta_2 x^2)$ $(x>0)$. It can be easily constrained to remain positive at all rates, have a maximum at a finite rate and tend towards zero at very high rates. Unlike a quadratic function of x, it changes more gradually above than below the inflection point.

In order to facilitate the choice of prior distributions for model parameters, the approach of Theobald & Talbot (2002) on yield response to applied nitrogen is followed by expressing these two dose-response functions in terms of parameters intended to have clear interpretations, and to measure distinct properties of the functions. It should be noted that, in contrast, the individual parameters in the above expressions for the exponential-plus-linear and inverse-quadratic functions do not have direct interpretations. There are many ways in which 'interpretable' parameters might be chosen, but two obvious choices are the maximum expected yield and the seed rate at which the maximum occurs. Note that the latter is distinct from the optimum rate, which depends on the relative prices of the seed and the crop. Because of the importance of susceptibility to lodging and the consequent yield reduction, the third parameter is chosen as a measure of how rapidly expected yield declines beyond the maximum. Thus, the following parameters are considered:

γ is the maximum expected yield.
δ is the seed rate giving maximum expected yield.
λ is the ratio of the expected yield at 2δ to that at δ.

By definition, λ is restricted to the interval (0, 1), so it is convenient to define an equivalent parameter η equal to the logit of λ, that is $\ln\{\lambda/(1-\lambda)\}$: then the range of η is unrestricted and λ is expressed as $\exp(\eta)/\{1+\exp(\eta)\}$. Also, a prior distribution for δ might be expected to be positively skewed, and this is modelled by taking $\ln\delta$ (which is unrestricted) to be Normally distributed. The two dose-response functions chosen may then be redefined in terms of the interpretable parameters γ, δ and η. For example, the inverse-quadratic function may be re-expressed as

$$\mathrm{E}(y\,|\,x,\gamma,\delta,\eta)=\frac{\gamma\delta x}{\delta x+2e^{-\eta}(x-\delta)^2}\quad(x>0)\qquad(1)$$

where E denotes expected value and y denotes the yield. The exponential-plus-linear function cannot be expressed directly in terms of γ, δ and η, but it can be given as

$$\mathrm{E}(y\,|\,x,\gamma,\delta,\eta)=\frac{\gamma\{e^{\omega}(1-e^{-\omega x/\delta})-\omega x/\delta\}}{e^{\omega}-1-\omega}\quad(x>0)\qquad(2)$$

where ω satisfies

$$e^{\eta}=\frac{e^{\omega}-e^{-\omega}-2\omega}{e^{-\omega}-1+\omega}.\qquad(3)$$

Although Eqn (3) has no obvious solution in ω, a good approximation is given by $1{\cdot}5\ln(1+e^{\eta})$. Negative values of Eqn (2) are also replaced by zero in modelling and the calculation of posterior expected utilities, so these expectations are always positive.

THE MODELLING PROCEDURE

The Bayesian method used in the current paper is adapted and extended from Theobald & Talbot (2002); it encompasses the trial data (including any covariates, varieties and treatments), prior information, future yields and costs. It may be summarized as follows.

(1) Choose a dose-response function.
(2) Express this function in terms of parameters that can be easily interpreted and can be treated as statistically independent *a priori*.
(3) Using these parameters, model the variation in yield between environments, possibly also incorporating covariates, varieties and treatment effects.
(4) Choose a prior distribution for all the model parameters to reflect knowledge of the crop varieties, the extent of variation between environments and the likely effects of treatments and covariates.
(5) Define the utility of sowing seed at any given rate in the target environment. The choice of utility made here is the value of the crop (per ha) minus the costs of the seed and of any non-standard treatments, although a non-linear function of yield might be used to reflect the grower's aversion to risk. Other costs, such as those of sowing seed and recording covariates, could also be included.
(6) Combine the prior distribution with the information in the data to find the posterior distribution of the model parameters and the posterior expected utility for any of the varieties at a sequence of possible seed rates. The corresponding sequence of posterior expected utilities must be found for different values of the covariates (where relevant) and for any non-standard treatments. If the values of the covariates are uncertain (by depending on the weather in the coming season, for example) make allowance for this by sampling from their estimated distribution.
(7) Identify an optimum rate for those combinations of covariate values and treatment or variety that are of interest: this may exclude treatments that might be applied after examining the crop. Possibly assess the economic benefits to be expected from using non-standard treatments.
(8) Repeat the calculations to assess the robustness of the optimum rates to changes in the prior distribution, in costs and in the dose-response function.

It should be noted that the averaging of the utility over the posterior distribution of the unknown parameters means that the procedure does not require point estimation of these parameters, nor would confidence intervals for optimum rates serve any purpose, since a single rate must be chosen for sowing.

Including environment, covariate and variety or treatment effects

In the context of fertilizer trials, Theobald & Talbot (2002) model variation between environments and varieties by allowing their interpretable parameters to vary according to the combinations of these factors: the environment effects for the trial and target environments are treated as arising from a common distribution, leading to a random-effects model. Their model may be modified to include dependence on varieties or treatments and on environment-specific covariates such as latitude. Thus, for the maximum-yield parameter γ_{jk} corresponding to the combination of environment j and variety k or level k of a treatment, it might be assumed that the effects of environments and of varieties or treatments are additive, or allow some pattern of interactions between them. Where a covariate is available, and one wishes to model dependence on it in order to predict yield at a target site, linear dependence in an additive model might be assumed:

$$\gamma_{jk}=\gamma_{ej}+\beta_{\gamma z}(z_j-\bar{z})+\tau_{\gamma k}\qquad(4)$$

Table 3. *Values defining the prior distributions of Normal parameters*

Parameters	Maximum expected yield (γ)		Log of rate for maximum ($\ln \delta$)		Logit of ratio of expected yields (η)	
	Expectation	S.D.	Expectation	S.D.	Expectation	S.D.
Variety effects	10·0	0·50	5·3	0·35	1·4	0·50
Non-standard treatments	0·0	0·50	0·0	0·05	0·0	0·05
Coefficients for sowing date	0·0	0·02	0·0	0·01	0·0	0·02
Coefficients for latitude	0·0	0·20	0·0	0·14	0·0	0·20

where z_j denotes the covariate value for environment j, γ_{ej} is that part of the environment effect not accounted for by the covariate, $\bar{z}$ is the mean of the z_j, $\beta_{\gamma z}$ is the corresponding regression coefficient, and $\tau_{\gamma k}$ may be interpreted as the effect on γ_{jk} of the variety or of the treatment level relative to the standard level. It may be necessary to extend this model, for example allowing quadratic dependence or separate coefficients for different varieties.

It should be noted that the model in Eqn (4) is hierarchical in the sense that the distribution for the γ_{jk} is defined in terms of higher-level parameters such as the γ_{ej}, whose distributions are themselves defined by other parameters. The prior distribution of the γ_{jk} is defined here in terms of the variance component $\sigma^2_{\gamma e}$ for the γ_{ej} and distributions for the regression coefficients and treatment effects. The γ_{ej} are also taken to have a common expectation μ_γ.

Changes in seed rate and differences between varieties or treatment levels can be expected to affect other aspects of the dose-response relation in addition to the maximum-yield parameter γ: these changes are allowed for by assuming models similar to Eqn (4) for the other two interpretable parameters. Applying such a decomposition to the parameters δ_{jk} and λ_{jk} would be problematic, because they are restricted to positive values and to values between 0 and 1 respectively: instead additive models like Eqn (4) are assumed for $\ln \delta_{jk}$ and η_{jk}. Finally, the parameter σ^2_y represents the residual variance of the yield, assumed to be the same for all environments, varieties and treatments.

Because of the non-linearity of the dose-response functions, the additive assumptions in Eqn (4) and the corresponding equations for $\ln \delta_{jk}$ and η_{jk} do not imply that environment, covariate and treatment effects are additive on the scale of yield, but they restrict the type of interaction which can be represented: for example, a treatment having substantial and opposite effects at different latitudes would not be modelled well. Possible generalizations of Eqn (4) are considered in the subsection '*Are the models for dependence on sowing date and latitude adequate?*'

The prior distribution for each variance component may be specified using a prior estimate and corresponding degrees of freedom: higher degrees of freedom imply greater confidence in the estimate. The remaining parameters are given Normal distributions specified using their prior means and standard deviations. The model parameters γ, δ and λ (or η) are intended to measure distinct aspects of the dose-response functions, so that it should be reasonable to assume that they are statistically independent *a priori*: this assumption could be relaxed if sufficient prior knowledge is available. Even if assumed independent *a priori*, they are not independent in the posterior distribution.

To examine how the covariates and the different varieties or non-standard treatments influence the dependence of yield on seed rate, Bayesian confidence intervals for the corresponding coefficients or treatment effects in the fitted models may be considered.

The prior distributions

The prior probability distributions assumed for the variety effects, treatment effects and regression coefficients are set out in Table 3. For example, in the 'Variety effects' row of the table, the value of 10·0 is the prior expected value of the maximum yield for all varieties; the value of 5·3 is approximately ln (200), implying that maximum yield is expected to occur at a rate of about 200 seeds/m^2; also 1·4 is approximately ln (0·8/0·2), suggesting that if the seed rate which gives maximum expected yield is doubled a yield of around 80 % of the maximum can be expected. The corresponding standard deviations quantify uncertainty about these prior expected values.

Effects for all the non-standard treatment levels are expressed as differences from the standard levels, and the sceptical view has been taken that the expected results of varying the treatments are around zero. For each non-standard level it has been assumed that the effect on the maximum expected yield γ has prior standard deviation 0·5 t/ha, and the effect on $\ln \delta$ has prior standard deviation 0·05 (so that the change in maximizing seed rate is of the order of 5 %).

Table 4. *Values defining the prior distributions of variance parameters*

Parameter	$\sigma^2_{\gamma e}$	$\sigma^2_{\gamma v}$	$\sigma^2_{\ln \delta,\, e}$	$\sigma^2_{\ln \delta,\, v}$	$\sigma^2_{\eta e}$	$\sigma^2_{\eta v}$	σ^2_{y}
Estimate	0·30	0·10	0·10	0·08	0·25	0·15	0·10
D.F.	10	10	5	5	5	5	50

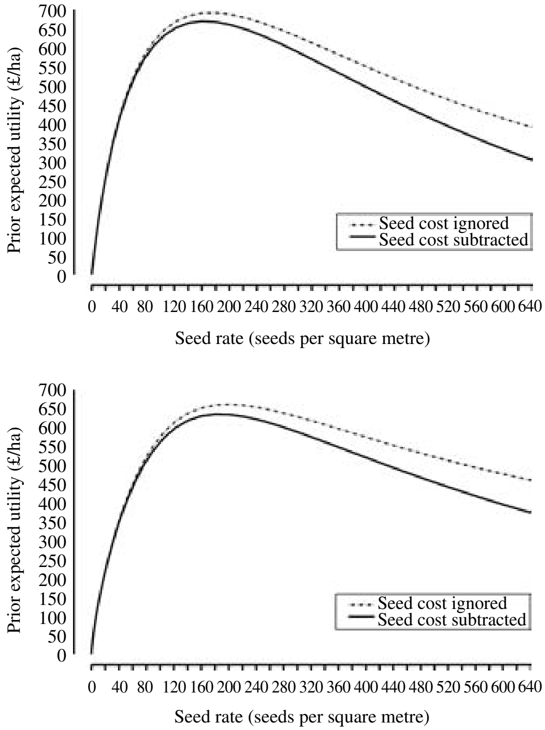

Fig. 1. Prior expected crop values with and without seed cost assuming exponential-plus-linear (above) and inverse-quadratic (below) dose-response functions.

The values given in Table 3 for the regression coefficients of γ, $\ln\delta$ and η on sowing date and latitude assume that the dependence is linear but that the effects are again around zero. It is supposed that differences of 50 days in sowing date and 5 degrees in latitude might produce effects of about 1 t/ha on maximum expected yield, a doubling or halving of the seed rate giving maximum expected yield and a 1 unit change in the logit of the ratio of the expected yields.

The prior distributions of the variance parameters in the model are given in Table 4. The reciprocal of each variance is assumed to follow a Gamma distribution that is specified by an estimate of that variance and a degrees-of-freedom parameter: a larger value of this parameter indicates greater prior precision.

One way to examine whether the prior distribution is reasonable is to look at the expected utility under this distribution: this is shown in Fig. 1 for the two dose-response functions, omitting any dependence on covariates or non-standard treatments. The corresponding 'optimum' seed rates are those at which the expected utilities (shown by the full lines) achieve their maxima, here 165 and 188 seeds/m^2 for the exponential-plus-linear and inverse-quadratic functions.

RESULTS

The modelling procedure defined above is used to examine the dependence of yield on varieties and sowing date for the Phase I trials and on treatments and latitude for those in Phase II. Equation (4) and analogous equations for the $\ln\delta_{jk}$ and η_{jk} are used initially, and then the question of whether these linear, additive models need to be extended is considered.

Phase I: effects of varieties and sowing date

Under both the exponential-plus-linear and inverse-quadratic dose-response models, only one of the three coefficients for sowing date has a Bayesian 95% confidence interval excluding zero: this is for the dependence of the seed rate giving maximum expected yield on sowing date, which is significantly positive. The estimated sizes of this effect under the two models correspond respectively to increases in the maximizing seed rate of about 2·2 and 2·1% from delaying sowing by one day. There are also estimated reductions in maximum expected yield/day of 0·015 and 0·012 t/ha respectively.

Table 5 shows the optimum seed rates and posterior expected utilities for the four varieties sown in Phase I at sowing dates of 30 September and 30 October. The posterior expected utilities are similar for the two dose-response functions, but about £50/ha lower for the later date. The differences in the optimum seed rates shown for the varieties are a consequence of including separate variety effects in the models. They prompt the question of whether such differences matter in practice. The answer is sought in the expected utilities for different varieties rather than in statistical significance. If the application to each variety of seed rates close to the median value in each column of optima are considered, e.g. 120, 210, 200 and 340 seeds/m^2, then the reduction in expected utility is no more than £4/ha for Haven, Soisson and Spark on either date. For Cadenza, which is more subject to lodging, the reduction is between £3/ha and £14/ha.

Figure 2 shows the posterior expected utilities against seed rate for the same dose-response functions, varieties and sowing dates. Comparison with

Table 5. *Optimum seed rates (seeds/m²) and corresponding posterior expected utilities (£/ha) for two sowing dates and four varieties with two dose-response functions and data from Phase I*

	30 Sep (day 273)		30 Oct (day 303)	
Variety	Exponential-plus-linear	Inverse-quadratic	Exponential-plus-linear	Inverse-quadratic
Cadenza	96, *836*	140, *836*	168, *789*	237, *788*
Haven	115, *861*	236, *856*	187, *814*	371, *799*
Soissons	118, *757*	180, *760*	205, *708*	305, *708*
Spark	141, *784*	248, *779*	241, *732*	397, *720*

Fig. 1 (noting the false origin in Fig. 2) indicates that the initial increases in yield are more rapid and the reductions beyond the maximum are more gradual than those predicted under the prior distribution. The optimum seed rates are consistently lower (by about 40% on average) for the exponential-plus-linear function than for the inverse-quadratic. This can be attributed to the difference in shape of the corresponding posterior expected utility functions: it is clear from Fig. 2 that the exponential-plus-linear function has a sharper elbow than the inverse-quadratic when they are fitted to the same data. The fit of the exponential-plus-linear function is later shown to be poor.

Phase II: effects of treatments and latitude

Under the exponential-plus-linear dose-response model, one coefficient has a Bayesian 95% confidence interval excluding zero, that for the dependence of the seed rate giving maximum expected yield on latitude: the coefficient is not significant in this sense under the inverse-quadratic model. The estimated sizes of this effect under the exponential-plus-linear and inverse-quadratic models correspond to increases in the maximizing seed rate of about 10·1 and 6·8%, respectively, per degree of latitude. The estimated reductions in maximum expected yield per degree are 0·036 and 0·029 t/ha, respectively.

Among the non-standard treatments defined in Table 2, the only set of treatments to give Bayesian 95% confidence intervals excluding zero relates to rotational position. The intervals for the reduction in maximum expected yield (t/ha) from second/third wheat (without the fungicide Latitude) rather than first wheat are (0·81, 1·49) and (0·91, 1·52) for the exponential-plus-linear and inverse-quadratic functions. With second/third wheat plus Latitude the corresponding intervals are similar, (0·90, 1·57) and (0·98, 1·60).

Table 6 shows the optimum seed rates and corresponding posterior expected utilities for Claire at two latitudes, those of Rosemaund (52·13) and Edinburgh (55·87), under the treatments that are fixed at the time of sowing. As in the trial data, sowing dates were assumed typical of winter-wheat sowings for those latitudes. The posterior expected utilities are £20/ha to £25/ha lower for the higher latitude, but similar for the two dose-response functions. The optimum seed rates are again consistently lower (now by about 19% on average) for the exponential-plus-linear function than for the inverse-quadratic.

Figure 3 shows, for the inverse-quadratic function, the posterior expected utilities for Claire at latitude 52·13 calculated under the standard combination of treatments and under each of the non-standard ones, assuming that only one treatment is varied at a time. The corresponding plot for the exponential-plus-linear function shows the same general difference in shape as in Fig. 2. The reductions in expected utility with seed rate beyond the optima are steeper for 'Sibutol slug treatment' and 'Second/third wheat + Latitude' than for other treatments because their cost increases in proportion to seed rate. The use of Latitude appears not to correct the large loss in expected utility from second/third rather than first wheat, and both variations from the standard slug treatment appear to reduce expected utility slightly. Of the remaining treatments, late nitrogen application appears to raise expected utility by about £9/ha at both latitudes, and PGR at tillering rather than at stem extension seems to increase it by about £5/ha.

Comparing the fit of the two dose-response functions

Figure 2 and Tables 5 and 6 suggest that there can be a substantial difference in the shape of exponential-plus-linear and inverse-quadratic functions fitted to the same data. The fit of the two functions are investigated using the method of *posterior predictive checking* described in Gelman *et al.* (2003). For each of the observed yields y in either data set, there is a theoretical expected value $E(y \mid x, \theta)$ given by the right-hand side of Eqn (1) or (2): here θ denotes the vector of unknown parameters in either model. For any value of θ, we can generate a value y^*, say, from the assumed distribution of y given θ under the model. As θ varies with respect to its posterior distribution, y and y^* will have similar distributions if the model is correct, but will show differences if it is

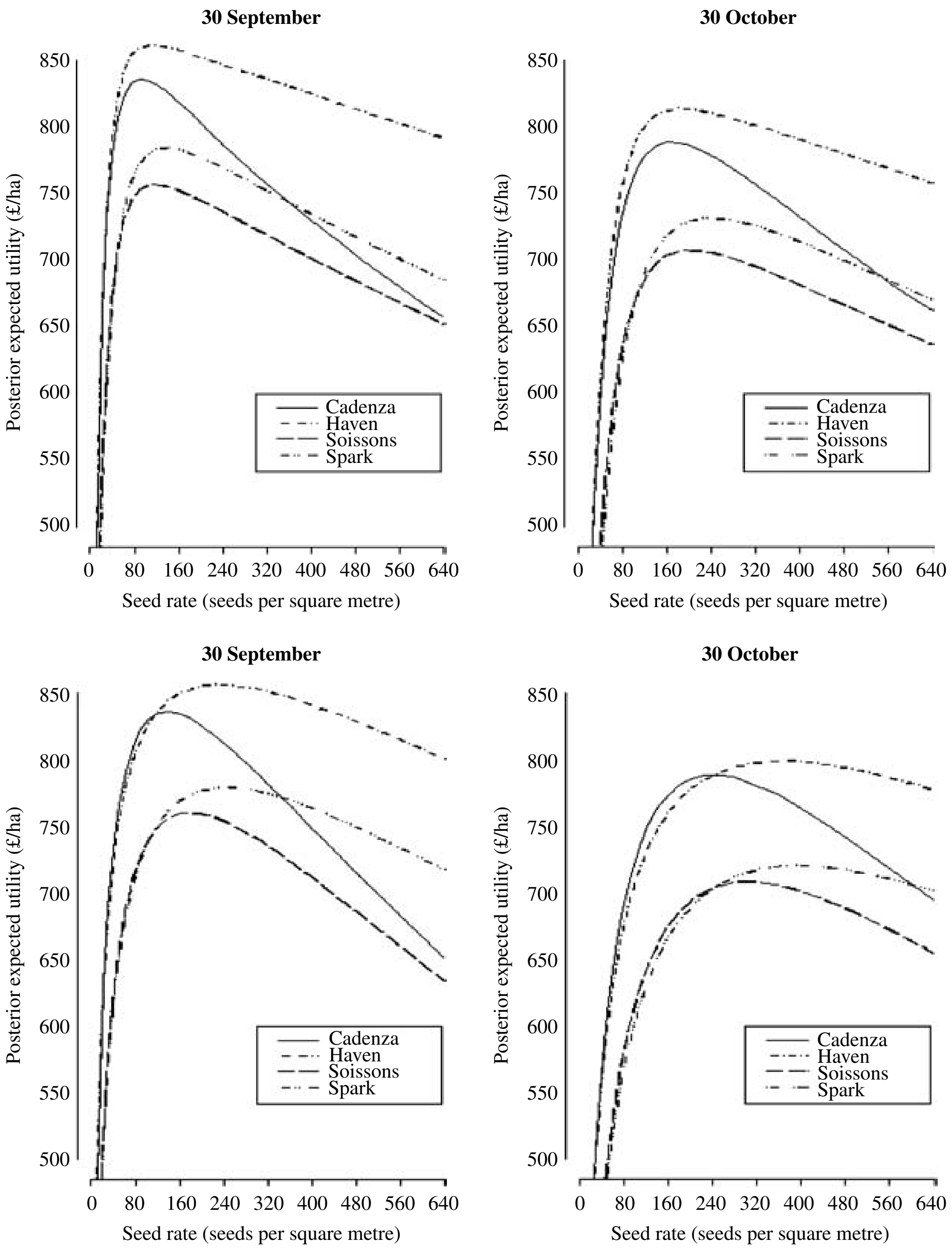

Fig. 2. Posterior expected utilities for Phase I data with four varieties and two sowing dates assuming exponential-plus-linear (above) and inverse-quadratic (below) dose-response functions.

not. Test quantities are therefore defined to measure those discrepancies between the distributions that are of particular interest. Here the main interest is in the dependence of yield on seed rate, so for each experimental phase the means $\bar{y}_x$, say, of all the yields observed at each seed rate x are used as test quantities. To examine how much each model conflicts with the data, the posterior probability that the corresponding mean $\bar{y}_x^*$ for the generated data exceeds $\bar{y}_x$ is found. Very small and very large

Table 6. *Optimum seed rates (seeds/m^2) and corresponding posterior expected utilities (£/ha) for Claire at two latitudes with standard and non-standard treatments, two dose-response functions and data from Phase II*

	Latitude 52·13 (Rosemaund)		Latitude 55·87 (Edinburgh)	
Treatment	Exponential-plus-linear	Inverse-quadratic	Exponential-plus-linear	Inverse-quadratic
Standard	166, *728*	213, *729*	225, *708*	276, *708*
Rotational position: second/third	160, *637*	203, *634*	215, *617*	261, *614*
Rotational position: second/third + Latitude[1]	152, *619*	183, *614*	202, *595*	233, *589*
Slug treatment: none	176, *721*	228, *723*	237, *700*	292, *701*
Slug treatment: + Sibutol Secur	152, *708*	188, *707*	205, *686*	243, *684*

[1] In this column 'Latitude' refers to the proprietary name of a fungicide.

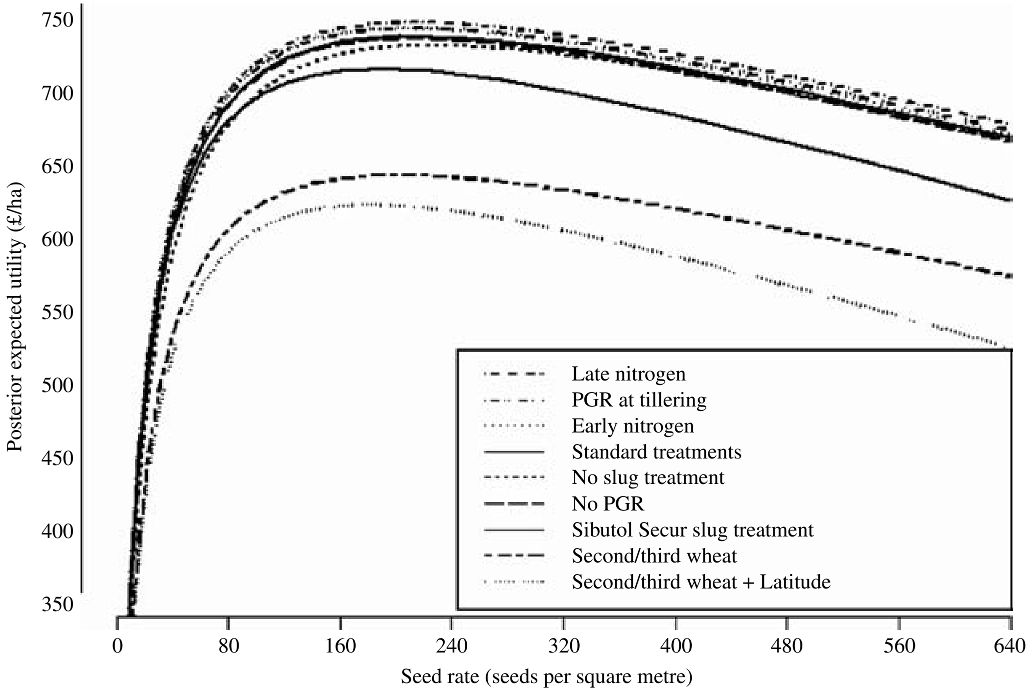

Fig. 3. Posterior expected utilities for Phase II data (variety Claire) under various treatments at latitude 52·13 assuming inverse-quadratic dose-response functions: the ordering of the treatments in the legend is according to their posterior expected utilities at 640 seeds/m^2.

probabilities indicate that the model over-estimates or under-estimates, respectively, the yields at rate x. It is straightforward to incorporate the calculation of these probabilities with those of the posterior distribution and the posterior expected utility in the WinBUGS program.

The resulting probabilities are shown in Table 7. For both phases, the exponential-plus-linear function appears to over-estimate the yields at 80 seeds/m^2 and to under-estimate them at the highest and lowest rates applied. This provides an explanation for the substantially lower optimum seed rates derived under this model in Tables 5 and 6. By contrast, there are no extreme probabilities for the inverse-quadratic function in either phase, and hence no evidence that this function fits the data badly.

Table 7. *Posterior probabilities for Phases I and II and two dose-response functions that predicted mean yields exceed observed mean yields at each seed rate*

	Phase I		Phase II	
Seed rate	Exponential-plus-linear	Inverse-quadratic	Exponential-plus-linear	Inverse-quadratic
20	0·982	0·279	–	–
40	0·140	0·742	0·986	0·627
80	0·002	0·363	0·0001	0·337
160	0·463	0·575	0·036	0·492
320	0·937	0·488	0·858	0·266
640	0·926	0·567	0·979	0·724

Sensitivity of the results to changes in the prior distribution

The posterior expected utilities, and hence the optimum rates, depend on the prior distributions assumed as well as on the dose-response function. To examine their sensitivity to changes in these distributions, the calculations may be repeated with different prior assumptions. Because of the evidence that the exponential-plus-linear function fits the data poorly, attention is concentrated on the inverse-quadratic. Note, though, that the evidence of poor fit for the exponential-plus-linear function remains under the various prior assumptions considered below.

There are many assumptions which might be changed: for illustration, the effects of altering (separately) the expectations in the prior distributions for γ, $\ln\delta$ and η are considered, replacing the values 10·0, 5·3 and 1·4 in Table 3 by 8·0, 6·0 and 2·94 respectively. The new expectation of 6·0 for $\ln\delta$ corresponds to the maximum yield occurring at a rate of about exp (6·0) or 400 seeds/m^2; the expectation of 2·94 for η equals ln (0·95/0·05), so that a doubling of the seed rate giving maximum expected yield has the effect of reducing the expected yield by about 5% rather than 20%.

For the Phase I data, the above change in the prior distribution of γ causes reductions in optimum rates of no more than 5%, but reduces posterior expected utilities by about £37/ha. Altering the prior distributions of δ increases optimum rates by no more than 6% and reduces posterior expected utilities by at most £4/ha. The change in the prior distribution of η reduces optimum rates by at most 8% and increases posterior expected utilities by at most £3/ha.

For Phase II, changing the prior distribution of γ reduces optimum rates by at most 3%, but lowers posterior expected utilities by about £50/ha. The effects of altering the prior distributions of δ and η are negligible, with changes in posterior expected utilities of at most £3/ha; optimum rates are increased by about 4% with the change in δ and reduced by about 2% with the change in η.

Thus the optimum rates based on both phases appear to be reasonably robust to these three changes in the prior distribution.

Are the models for dependence on sowing date and latitude adequate?

The above results have been obtained under the assumption that Eqn (4) and analogous equations for the $\ln\delta_{jk}$ and η_{jk} apply, so that the dependence of our interpretable parameters on sowing date and variety effects or on latitude and treatment effects is linear and additive.

Many non-additive generalizations of this model might be considered. For example, the maximum-yield parameter γ_{jk} may be allowed to include interaction between sowing date and variety in Phase I by replacing the common regression coefficient $\beta_{\gamma z}$ by separate coefficients for the varieties. These are given a common Normal prior distribution with expectation $\mu_{\beta\gamma z}$ and variance $\sigma^2_{\beta\gamma z}$, where $\mu_{\beta\gamma z}$ is Normal with expectation 0 and variance 0·0002, and $\sigma^2_{\beta\gamma z}$ has estimate 0·0002 with 5 D.F. Under the inverse-quadratic dose-response function, the largest effects of this alteration in the model on the optimum rates in Table 5 are to change the optima for variety Spark to 236 and 363 seeds/m^2 on 30 September and 30 October respectively; the other changes are by 7 seeds/m^2 or less. Very similar effects on the optimum rates are found in Phase I if the common regression coefficient is retained in Eqn (4), but Normal variety × environment interaction effects are included with expectation 0 and variance $\sigma^2_{\gamma ev}$, where $\sigma^2_{\gamma ev}$ has estimate 0·15 with 5 D.F.

In a more radical change from the additive model, non-additive models are also fitted in γ, $\ln\delta$ and η with no covariate dependence. For Phases I and II respectively, variety and treatment effects are assumed to be nested within environments: the variance components on the γ, $\ln\delta$ and η scales are given similar prior distributions to those listed in Table 4, with prior estimates 1, 0·1, 0·1 and 5 degrees of freedom each for both phases.

The posterior expected values of the combined effects of environment and variety for Phase I under this non-additive model are plotted against sowing date in the upper row of Fig. 4. The lower row shows the posterior expected values of the combined effects of environment and treatment for Phase II plotted against latitude: here the standard combination of treatments is distinguished from non-standard ones. The most obvious departure from linearity in these six plots is in the lower left one: the posterior expected values of the combined effects of environment and treatment show non-linear dependence on latitude. This plot reflects the dependence of mean

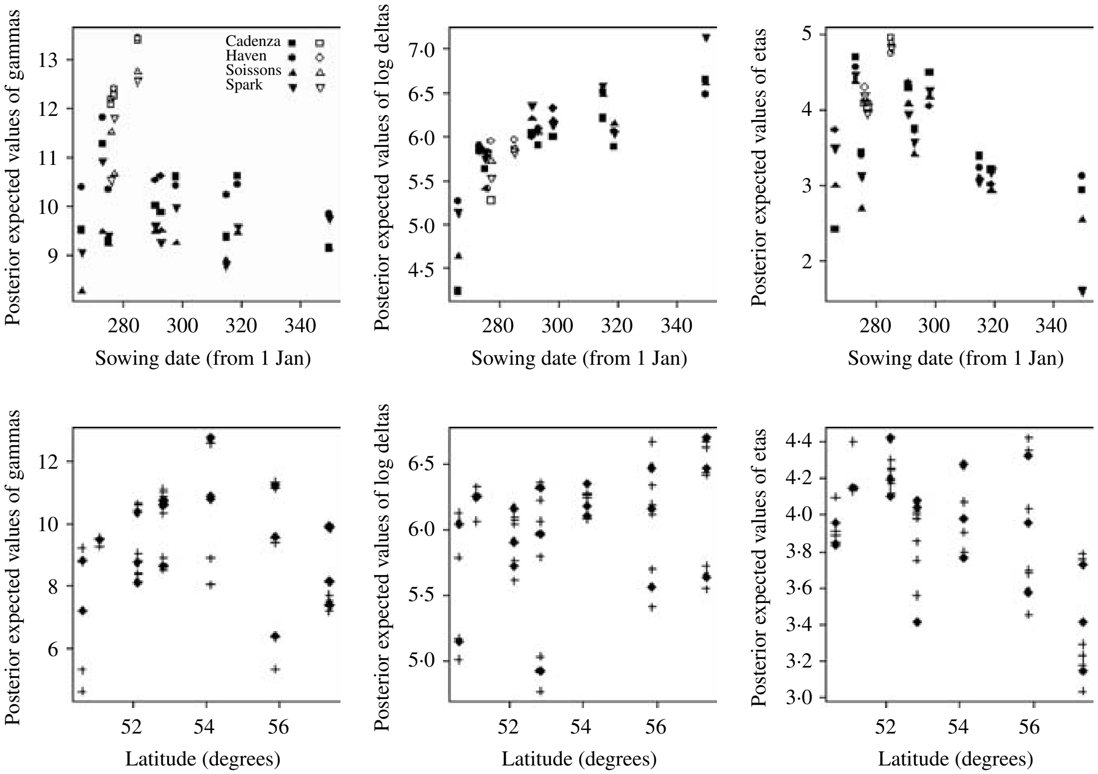

Fig. 4. Upper row: posterior expected values for Phase I data of the combined effects of environment and variety on the scales of the 'interpretable' parameters plotted against sowing date: filled and open symbols indicate Rosemaund and Sutton Bonington environments. Lower row: posterior expected values for Phase II data of the combined effects of environment and treatment plotted against latitude: diamonds indicate the standard combination of treatments and crosses the non-standard ones.

yield on latitude in Phase II, although this non-linear effect seems atypical of longer-term patterns in yield.

The effects of extending our linear additive model for the Phase II data are examined by including a squared term $(z_j - \bar{z})^2$ in latitude in Eqn (4). The corresponding regression coefficient is given a Normal prior distribution with expectation zero and S.D. 0·5, independently of the other parameters. The Bayesian 95% confidence interval for this coefficient includes only negative values, indicating significant curvature in the relationship. Nevertheless the optimum seed rates from this analysis differ from those in Table 6 by at most 1 seed/m^2, and the corresponding posterior expected utilities are changed by at most £2/ha.

Heterogeneity of residual variances

Another possible inadequacy of the model is that it assumes a common residual variance for the yield over the trials within each phase. The effect of allowing these variances to differ between trials is therefore examined, giving each variance the prior distribution used above for the common variance. For Phase I, the resulting changes to the optimum seed rates given in Table 5 are between −8 and 6%, while the largest change in posterior expected utility is £9/ha. For Phase II, the changes in optimum rates relative to Table 6 are between −1 and 3%; the posterior expected utilities for the later rotational positions are reduced by about £20/ha, but those for the remaining treatments are hardly affected.

DISCUSSION

A method has been developed for combining the information from seed-rate experiments which can include effects for environments, varieties and treatments, and which allows dependence on covariates such as sowing date and latitude to be modelled. For the dependence of yield on seed rate, it assumes a hierarchical nonlinear regression model. Approximate non-Bayesian methods for fitting such models are available (e.g. Davidian & Giltinan 1995) and that

Makowski *et al*. (2001) use this approach for modelling the response of winter wheat to applied nitrogen across trials. The adoption in the present paper of a Bayesian approach permits expert knowledge of the crop to be incorporated into the analysis of the data. A common objection to such an approach is that prior distributions may be difficult or impossible to specify. While not an essential part of the Bayesian approach, the specification of dose-response functions in terms of 'interpretable' parameters is intended to facilitate the choice of these distributions.

By treating trial environments as drawn from a population, the concept of an optimum seed rate can be reformulated so that it applies to a future target site rather than to the individual trial environments. If a covariate is included in the model then the optimum rate depends on its value, and the target environment is considered as sampled from a population with the same value.

The present formulation includes a utility function meant to represent the value of the crop minus the variable costs associated with different seed rates and non-standard treatments. Decisions about optimum rates and treatments are based on the posterior expected utility for any seed rate, which is the mean of the utility over the posterior distribution. This distribution is intended to represent the variation to be expected between environments in the response of yield to changes in seed rate. Thus, the optimum rate is the rate expected to give the highest return in the coming season, given the treatments already chosen, the variety to be sown, the covariate values for the site and knowledge of variation over environments.

Recommendations for seed rates are conventionally based on estimated optima for trial environments, with some upward adjustment as insurance against adverse growing conditions. Since the posterior expected utility represents an average over the possible future environments, it automatically includes some insurance against unfavourable conditions. The success of the method in achieving an appropriate balance over possible environments depends on how representative conditions in the trials are of those that might be experienced in future, and also on the appropriateness of the dose-response model and the prior distribution.

The results of the analyses appear reasonable in the sense that they depend on covariates and treatments in sensible ways. The optima are sensitive to the choice of dose-response function, but the low optima obtained with the exponential-plus-linear function can be at least partly attributed to the poor fit of this function.

It may seem disappointing that the choice of dose-response function has a large effect on the calculation of optimum seed rates. The posterior expected utility can be used to examine the consequences of basing optimum rates on one model when the other is in fact appropriate. For example, under the standard set of treatments in Phase II, the optimum rates at latitudes 52·13 and 55·87 from the exponential-plus-linear model are (from Table 6) 166 and 225 seeds/m^2 respectively: the expected utilities of these rates at the two latitudes under the inverse-quadratic model are £726/ha and £706/ha, only slightly lower than the maxima under this model of £729/ha and £708/ha. Hence, the difference between the two optima in terms of expected margin is less disappointing than it first appears. Indeed the posterior expected utility under the inverse-quadratic is within £5/ha of its maximum for rates in the intervals (156, 290) and (210, 360) for these two latitudes. The widths of these intervals arise because the expected utility curves are fairly flat near their maxima: the flatness is explained by the ability of wheat to compensate for low seed rates by increased production of tillers, and as a consequence of each curve being based on an average over many inverse-quadratic functions with different parameters rather than on a single estimated inverse-quadratic.

The optimum seed rates derived under the inverse-quadratic dose-response function appear to be fairly robust to changes in the prior distribution. Extending the model in order to incorporate differing residual variances between trials, interaction between sowing date and variety in the maximum-yield parameter, or quadratic dependence of maximum expected yield on latitude also has little effect on optimum rates.

Guidance on seed rate should make reference to sowing date and latitude. The results presented in Tables 5 and 6 show that the optima from models including these covariates depend substantially on their values. Communicating the results of data analyses would become difficult if several covariates were included, such as applied nitrogen or soil type in addition to date and latitude: the inclusion of additional responses, such as quality characteristics, would add to the burden. It would be better to make advice available via an interactive system similar to HGCA's RL *Plus* (http://www.hgca.com/varieties/rl-plus/index.html).

Whereas the values of latitude and sowing date are known at the time of sowing, one may wish to model the effects of covariates whose values are not available until later, such as accumulated temperature over the growing season or the plant population on some date. While such uncertain covariates are useful for modelling crop growth, their application to choosing seed rate requires appropriate allowance to be made for the uncertainty in their values for the target site in the coming season. This may be achieved by repeating yield predictions at covariate values sampled from an estimated joint distribution for the site: Theobald *et al.* (2002) use this method to allow for uncertainty in a measure of accumulated average daily temperatures when predicting yields of maize. Alternatively, the covariates can be replaced in the model by long-term mean or median values for the site. Either method can be expected to lead to a reduction in explanatory power relative to the values for the current season.

By contrast with the influence of latitude and sowing date, Table 6 suggests that little adjustment to seed rates needs to be made for the non-standard treatment levels used in Phase II, except perhaps where the cost of the treatment increases with seed rate. Table 5 provides evidence that rates may depend on whether the variety sown is prone to lodging.

An alternative to providing guidance on seed rates is to recommend optimum plant populations, and to allow growers to infer appropriate seed rates according to local knowledge of previous emergence and current growing conditions. While measurement of plant populations may be essential to the agronomist in understanding the process of crop growth, we see several difficulties with emphasizing them rather than seed rates in providing guidance. First, the grower's direct concern is how much seed to sow (Gooding *et al.* 2002), even if he is willing to interpret guidance in the light of his knowledge of previous emergence. Also, no particular seed cost can be associated with a given plant population, and the population varies over the season, particularly if winter kill is a possibility.

We are grateful to HGCA and the Environment and Rural Affairs Department of the Scottish Executive for funding this study, and to Roger Williams (HGCA), Roger Sylvester-Bradley and Antony Wade (both of ADAS) for valuable discussions.

REFERENCES

BERNARDO, J. M. & SMITH, A. F. M. (1994). *Bayesian Theory*. Chichester: John Wiley & Sons.

BESAG, J. & HIGDON, D. (1999). Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **61**, 691–746.

DAVIDIAN, M. & GILTINAN, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. London: Chapman & Hall.

FISHER, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver & Boyd.

GELMAN, A., CARLIN, J. B., STERN, H. S. & RUBIN, D. B. (2003). *Bayesian Data Analysis*, 2nd edition. London: Chapman & Hall.

GENSTAT (2002). *GenStat for Windows. Release 6.2*. Oxford: VSN international Ltd.

GOODING, M. J., PINYOSINWAT, A. & ELLIS, R. H. (2002). Responses of wheat grain yield and quality to seed rate. *Journal of Agricultural Science, Cambridge* **138**, 317–331.

LEMERLE, D., COUSENS, R. D., GILL, G. S., PELTZER, S. J., MOERKERK, M., MURPHY, C. E., COLLINS, D. & CULLIS, B. R. (2004). Reliability of higher seeding rates of wheat for increased competitiveness with weeds in low rainfall environments. *Journal of Agricultural Science, Cambridge* **142**, 395–409.

LINDLEY, D. V. (1985). *Making Decisions*, 2nd edition. Chichester: John Wiley & Sons.

MAKOWSKI, D., WALLACH, D. & MEYNARD, J.-M. (2001). Statistical methods for predicting responses to applied nitrogen and calculating optimal nitrogen rates. *Agronomy Journal* **93**, 531–539.

MARIN, J. M., MONTES DIEZ, R. & RIOS INSUA, D. (2003). Bayesian methods in plant conservation biology. *Biological Conservation* **113**, 379–387.

NELDER, J. A. (1966). Inverse polynomials, a useful group of multi-factor response functions. *Biometrics* **22**, 128–141.

PATTERSON, H. D. (1997). Analysis of series of variety trials. In *Statistical Methods for Plant Variety Evaluation* (Eds R. A. Kempton & P. N. Fox), pp. 139–161. London: Chapman & Hall.

RAIFFA, H. & SCHLAIFER, R. (1961). *Applied Statistical Decision Theory*. Boston: Harvard University.

SMITH, A. B., CULLIS, B. R. & THOMPSON, R. (2001). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* **57**, 1138–1147.

SMITH, A. B., CULLIS, B. R. & THOMPSON, R. (2005). The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. *Journal of Agricultural Science, Cambridge* **143**, 449–462.

SPIEGELHALTER, D. J., THOMAS, A., BEST, N. G. & LUNN, D. (2003). *WinBUGS User Manual, Version 1.4*. Cambridge: MRC Biostatistics Unit.

SPINK, J. H., SEMERE, T., SPARKES, D. L., WHALEY, J. M., FOULKES, M. J., CLARE, R. W. & SCOTT, R. K. (2000). Effect of sowing date on the optimum plant density of winter wheat. *Annals of Applied Biology* **137**, 179–188.

THEOBALD, C. M. & TALBOT, M. (2002). The Bayesian choice of crop variety and fertilizer dose. *Journal of the Royal Statistical Society, Series C: Applied Statistics* **51**, 23–36.

THEOBALD, C. M., TALBOT, M. & NABUGOOMU, F. (2002). A Bayesian approach to regional and local-area prediction from crop variety trials. *Journal of Agricultural, Biological, and Environmental Statistics* **7**, 403–419.

WALKER, S. R., MEDD, R. W., ROBINSON, G. R. & CULLIS, B. R. (2002). Improved management of *Avena ludoviciana* and *Phalaris paradoxa* with more densely sown wheat and less herbicide. *Weed Research* **42**, 257–270.

YATES, F. & COCHRAN, W. G. (1938). The analysis of groups of experiments. *Journal of Agricultural Science, Cambridge* **28**, 556–580.