Cubic smoothing splines: a new method for analysing multi-site experiments

Richard W. Medd[1], Steve R. Walker[2] and Brian R. Cullis[3]
[1] NSW Agriculture, Orange Agricultural Institute, Forest Road, Orange, New South Wales 2800, Australia
[2] Department of Primary Industries, PO Box 2282, Toowoomba, Queensland 4350, Australia
[3] NSW Agriculture, Wagga Wagga Agricultural Institute, Pine Gully Road, Wagga Wagga, New South Wales 2650, Australia

Summary The effectiveness of crop competition for better weed control and reducing herbicide rates was determined for wild oats (*Avena ludoviciana* Durieu). Three experiments, previously broadcasted with weed seeds were sown with three wheat densities, and emerged weeds were treated with four herbicide doses (0–100% of recommended rate). The measured crop and weed traits were first analysed across experiments for treatment effects. Grain yield and weed seed production data were then analysed collectively using cubic smoothing splines to model the response surfaces, which have the advantage of allowing joint analysis of multi-site experiments and their associated complex designs and variance structures. Other reasons for the choice of semi-parametric methods over conventional non-linear regression methods are outlined.

Maximum crop yield and reduction in seed production of *A. ludoviciana* was achieved with 130 wheat plants m^{-2} applied with 75% herbicide rate. Alternatively, these benefits were achieved by increasing crop density to 150–160 plants m^{-2} applied with 50% herbicide rate. At high crop density, application of the 100% herbicide rate tended to reduce yield, and this impacted adversely on suppressing weed seed production. Thus, more competitive wheat crops have the potential for improving weed control and reducing herbicide rates.

Keywords Weed seed production, crop density, crop competitiveness, reduced herbicide, wild oats.

INTRODUCTION

Development of weed management tactics invariably involves experimentation over several seasons and geographic sites. The synthesis of such data sets has traditionally involved separate analysis of individual experiments to identify common trends, or non-linear regression. Herein we consider cubic smoothing splines, a new method of statistical analysis, which permits analysis of complex data sets from multi-site experiments in order to identify common trends and develop robust principles.

The data set was generated from field experiments conducted in the sub-tropical grain region of Australia (Walker *et al.* 2002), to investigate whether increased seeding density of wheat improves control of wild oats (*Avena ludoviciana* Durieu) and paradoxa grass (*Phalaris paradoxa* L.), and whether there is a trade-off between crop density and herbicide rate, with a view to reducing dependence on herbicides. The impact of these strategies was assessed both in terms of crop productivity and weed seed production. Only findings for *A. ludoviciana* are considered in this paper.

MATERIALS AND METHODS

Experimental design and measurements Three experiments were conducted during 1996 to 1998 at two locations on the Darling Downs in southern Queensland (Walker *et al.* 2002). These were artificially infested by hand-broadcasted wild oats (≈400 seeds m^{-2}) one month before sowing of crops. Experiments involved factorial combinations of three wheat densities by four herbicide rates, with three replications in randomised complete blocks. Treatments were three wheat seeding rates, which aimed for 50 (low), 100 (medium) and 150 (high) plants m^{-2}, and four herbicide doses, which were 0, 25, 50 and 100% of the recommended rate.

Tralkoxydim (as Achieve® 400 g a.i. kg^{-1}, Crop-Care Australasia) was applied at 200 g a.i. ha^{-1} for the 100% treatment.

Crop and weed plant densities were assessed (two quadrats, 1 × 0.5 m) before herbicide application in the untreated plots. Fertile tiller densities of *A. ludoviciana* and crop were measured (three quadrats, 1 × 0.5 m) several weeks prior to harvest. Seeds were counted on 10 weed tillers, chosen randomly from those within the three quadrats sampled from each plot. Weed seed production per unit area (m^2) was calculated from these variables. Grain yield (7 rows × 10 m) was measured at crop maturity.

Analysis of measured crop and weed effects All experiments were considered in a joint analysis of variance, that preserved the spatial integrity of the individual treatments in each of the experiments. This facilitated use of more sophisticated modelling of within experiment heterogeneity using the spatial analysis methods, as outlined by Gilmour *et al.* (1997) for the analysis of individual experiments. This has been extended to the analysis of multi-environment trials, such as these data, as outlined by Smith *et al.* (2001). The treatment effects were partitioned for herbicide rate, crop seeding rate and experiment. The statistics for assessing treatment effects from this analysis are approximate anti-conservative F-statistics, with values that differ significantly at the P = 0.01 level only being determined.

Prior to analyses, data for weed seed production were transformed to log(y + 0.25). The normality plots and mean/variance dependence of the residuals were substantially improved using these transformations.

Response surface modelling of crop yield and weed seed production The effects of the actual crop density achieved (instead of seeding rate) and herbicide rate on grain yield and weed seed production were considered further by a semi-parametric regression approach. This technique has been widely advocated in statistical literature and in particular Verbyla *et al.* (1999) illustrated how cubic smoothing splines may be incorporated into a mixed model framework. This greatly enhances the accessibility of the smoothing splines methodology and allows for joint modelling of other sources of variation in the analysis of complex designed experiments, such as these data.

Cubic smoothing splines and extensions A cubic spline is a piecewise cubic function that is constrained so that the function and its first two derivatives are continuous at the break points (knots) between one cubic segment and the next. Formulae for evaluating a cubic spline at any value of the explanatory variable are available (for example, Green and Silverman 1991, White *et al.* 1999). Natural cubic splines are cubic splines with the additional constraint that they must be linear for values of the explanatory variable which are less than the smallest knot and greater than the largest knot. Given data, the cubic smoothing spline (or simply the smoothing spline) is the natural cubic spline that minimises a penalised least squares criterion. This criterion depends on a parameter known as the smoothing parameter. The smoothing parameter determines the trade-off between fidelity to the data and smoothness. Using this penalty the smoothest cubic spline is a straight line, while the roughest cubic spline interpolates the data, if knot points match the unique values of the explanatory variable. Since Verbyla *et al.* (1999) formulated cubic splines within the mixed model framework, they suggested that a natural approach for estimating the smoothing parameter is Residual Maximum Likelihood (REML). The mixed model formulation of a cubic spline used by Verbyla *et al.* (1999) has two components. The first component is simply the linear regression of the response on the explanatory variable, denoted in the following by $lin(x)$. The second component is a random component, which can be considered as the deviation from this linear regression, and is denoted by $spn(x)$. The fit of the cubic spline can be represented by $y = 1 + lin(x) + spl(x)$. A natural extension of this model, which jointly models the effects of both crop density (D) and herbicide rate (H) on a response variable (y), is given by $y = (1+ lin(D)+ spl(D))*(1 + lin(H) + spl(H))$.

The experiments were analysed simultaneously, and we examined the variation of the response surface between experiments.

Following Verbyla *et al.* (1999) variation between experiments was examined in an analysis of variance decomposition of the cubic splines (Walker *et al.* 2002). The approximate F-statistics were calculated observing marginality. Residual maximum likelihood ratio tests were conducted to assess the significance of the deviations from linearity between experiments.

RESULTS

Measured crop response Grain yield differed significantly due to the herbicide rate applied, which interacted with seeding rate and experiment. Seeding rate by itself did not affect yield.

Crop yield response surface modelling With the cubic smoothing spline modelling of grain yield, the linear effect on yield of herbicide rate was significant, whereas crop density had no significant effect. These respective effects differed significantly between experiments. The overall non-linearity of the cubic smoothing spline was very significant.

The perspective plots of the smoothing splines for yield, averaged across experiments, illustrate the non-linear response to the interaction between herbicide rates and crop density (Figure 1a). At low crop density, yield increased markedly, as it did with higher crop densities and increasing herbicide rate to an optimum, then declined at the highest crop densities and highest herbicide rates.

Measured weed response Weed densities ranged between 10 and 89 plants m^{-2} and seed production was strongly affected by both crop seeding rate and herbicide rate, and these effects did not interact significantly with each other or with experiments. This was due to significant effects on weed fecundity and tiller density, both of which decreased markedly with increasing crop seeding rate and herbicide rate.

Weed response surface modelling Overall, there were very significant linear effects of herbicide rate and crop density on weed seed production, with no evidence of any interaction between herbicide rate and crop density. There was a significant interaction between experiment and herbicide rate.

The non-linear components of the cubic smoothing spline for crop density and herbicide rate effects were very significant, with no variation in the non-linearity between experiments (Walker *et al.* 2002). Perspective plots of the smoothing splines for weed seed production, averaged across experiments, illustrate that increasing crop density reduces weed seed production but this varied with herbicide rate with little reduction without herbicide (Figure 1b). The high inputs of crop density and herbicide minimised seed production non-linearly, with a trend towards a slight increase in seed production for the full herbicide rate applied at high crop density.

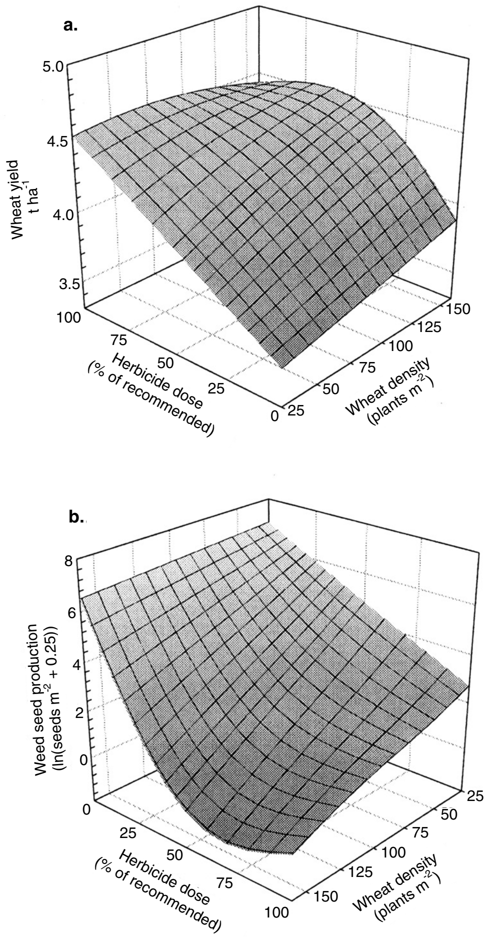

Figure 1. Perspective response surface plots of cubic smoothing spline for a. grain yield and b. weed seed production, in relation to crop density and herbicide rate averaged across experiments for *A. ludoviciana.*

Trade-off between crop density and herbicide rate for yield and weed seed production Relationships between crop density and herbicide rate from the cubic smoothing spline analyses are plotted as contours for grain yield (as 90, 95 and 99% of maximum yield across experiments), and reduction in seed production (as 95, 99 and 99.9% of maximum across experiments) (Figure 2), across all wild oat densities. Yields of ≥95% maximum yield were achieved with a minimum crop density of approximately 80–100 plants m^{-2} with the application of herbicide at the 50% rate. This was also achieved with 30% reduced herbicide rate, by increasing wheat density to about 160 plants m^{-2}. However to satisfy 95% yield expectation as well as 99% reduction in weed seed production, the trade-off for these requirements ranged from the combination of 20–30% herbicide rate and 160 crop plants m^{-2} to 100% herbicide rate and 40–60 crop plants m^{-2}. Alternatively, the trade-off to improve weed control to ≥99.9% reduction in weed seed production ranged from approximately 130 plants m^{-2} and 75% rate. There was little trade-off with crop density by increasing herbicide rate above 75%. A ≥99.9% reduction in weed seed production represents approximately five seeds m^{-2} returning to the weed seed bank.

DISCUSSION

The questions addressed in this work were can weeds be better controlled by increasing crop density, and if so, could this lead to a reduction in herbicide required? The answer to both of these questions is yes, but with some qualification. Conservation of crop yield is a

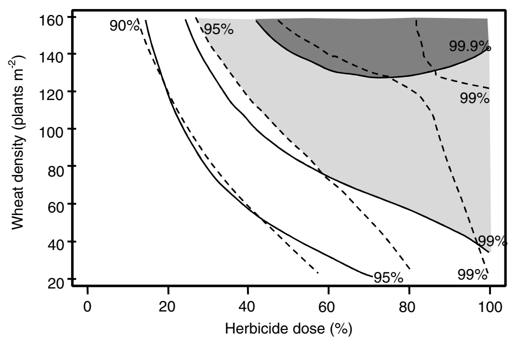

Figure 2. Contour plots of grain yield as % of maximum (dashed line), and % reduction in seed production relative to untreated (solid line). Light shade indicates zone of 'trade-off' between crop density and herbicide dose to achieve ≥99% reduction in seed production and dark shade ≥99.9% reduction in seed production and ≥95% of crop yield.

foremost objective of weed management. Clearly, increasing wheat density alone failed to achieve maximum yield potential in these experiments, however there appears to be a strong case for lifting wheat density for better weed management. The combination of high wheat density and application of herbicide at reduced rates reduced seed production to very low numbers and maximised yield.

To reach these conclusions we chose to use semi-parametric methods in preference to conventional non-linear regression methods in order to show the robustness of the trade-off principles drawn from multi-site experiments. The choice of the cubic spline response model was preferred to non-linear regression modelling, which is more widely accepted in the biological literature, for several reasons. Firstly, the models proposed by Brain *et al.* (1999), for example, are quite complex non-linear models, involving numerous parameters. We felt it may be both difficult, given the complexity of the experiments and the extra sources of variation, and perhaps inappropriate, given the restricted number of levels of the quantitative treatment factors, to fit these models. Secondly given the mixed model formulation of cubic smoothing splines, we could account for the error variance heterogeneity between experiments and the spatial variation within experiments in a relatively straightforward manner. Lastly, use of semi-parametric models, such as cubic smoothing splines, offers the data analyst the opportunity to explore and present the relationship between design variables and the response without presumptions, other than 'smoothness', about the exact form of such a relationship (Green and Silverman 1991, Hastie and Tibshirani 1990). In other words, the semi-parametric models are 'allowed' to fit the data without imposing constraints on the form of the relationship, as is often the case for non-linear models. A downside of the semi-parametric methods is that it is not possible to assign biological meaning to parameters, as is sometimes possible with non-linear methods.

Smoothing higher dimensional data has been widely considered in the area of spatial and image analysis. In these applications so-called thin plate splines have been advocated and widely used. Their basis relies on the assumption of isotropy of the penalty function which controls the departure of the response surface from an additive linear model on the explanatory variables. This assumption does not seem plausible for the analysis of multi-factorial experiments such as these data. Our extension for two dimensional smoothing is a simple and pragmatic one, though there may be more appealing alternatives.

ACKNOWLEDGMENTS

This project was supported by funds from the Grains Research and Development Corporation. We greatly appreciate the statistical advice provided by Dr. Arthur Gilmour and thank Jenni Tarleton for preparation of Figure 2.

REFERENCES

Brain, P., Wilson, B.J., Wright, K.J., Seavers, G.P. and Caseley, J.C. (1999). Modelling the effect of crop and weed on herbicide efficacy in wheat. *Weed Research* 39, 21-35.

Gilmour, A.R., Cullis, B.R. and Verbyla, A.P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural Biological, and Environmental Statistics* 2, 269-93.

Green, P.J. and Silverman, B.W. (1991). 'Nonparametric regression and generalized linear models – a roughness penalty approach'. Monographs of Statistics and Applied Probability 58, 182 pp. (Chapman and Hall, London).

Hastie, T.J. and Tibshirani, R.J.(1990). 'Generalised additive models', 335 pp. (Chapman and Hall, London).

Smith, A.B., Cullis, B.R. and Thompson, R. (2001). Analysing multi-environment trial data using multiplicative mixed models. *Biometrics* 58**,** 1138-47.

Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using splines (with discussion). *Applied Statistics* 48, 269-311.

Walker, S.R., Medd, R.W., Robinson, G.R. and Cullis, B.R. (2002). Improved management of *Avena ludoviciana* and *Phalaris paradoxa* with more densely sown wheat and less herbicide. *Weed Research* (in press).

White, I.M.S., Thompson, R. and Brotherston, S. (1999). Genetic and environment smoothing of lactation curves with cubic splines. *Journal of Dairy Science* 82, 632-8.