REGRESSION AND SPLINE METHODS FOR REMOVING ENVIRONMENTAL VARIANCE IN PROGENY TESTS

B. C. Bongarten and J. F. Dowd [1]

Abstract.--Regression and spline techniques were compared to blocking for their ability to model and eliminate environmental variation in progeny tests. With the regression and spline techniques, environmental gradients were determined by modeling trait response surfaces with respect to row and column position in the test. Deviations from the modeled surfaces were assumed to represent random genetic and environmental effects, and served as data for determination of variance components.

Fourth year height and diameter in a sycamore progeny test were used to demonstrate the method. Results show that the plot within family variance may be substantially reduced without affecting genetic variation using spline techniques. As a result, heritabilities were substantially higher than when blocking was used. Spline methods for eliminating environmental variation obviate the need for blocking altogether, and, thus, facilitate the establishment of progeny tests with large numbers of genetic entries. These methods will be most effective when single tree plots are employed.

Additional keywords: heritability, genetic selection.

The efficiency of progeny tests is largely dependent on the degree to which environmental variation can be separated from genetic variation. In progeny tests, environmental variation may be divided into two components: random variation (with respect to position in the progeny test), resulting from differences in microsite, stock condition, handling, etc., and non-random variation, resulting from environmental gradients in the site. In this work we address methods for eliminating the non-random sources of environmental variation.

This task is normally accomplished by blocking. However, blocking is inefficient because environmental gradients are not normally stepwise, but continuous. The erroneous assumption that environmental gradients are stepwise induces large family x block interactions because a family plot may occur on an "above-average" site in one block, and a "below-average" site in another block. More realistic modeling of environmental gradients should provide a means to reduce this source of variation and, thereby, improve heritabilities and the accuracy of among and within family selection.

---

[1] Associate and Assistant Professors, respectively, School of Forest Resources, The University of Georgia, Athens, GA 30602

Elimination of environmental trends in progeny test data may be accomplished by subtracting non-random environmental effects from the actual data. Non-random environmental effects may be estimated by fitting a smooth surface through the actual data, assuming that all genetic variation is random so that any trends are due solely to environmental causes.

Options for modeling trends include, among other things, polynomial regression and various spline techniques. Polynomials are simple, but they force the number of peaks and inflections and have limited flexibility. Therefore, they are unsuited to most situations. Splines, taken to their extreme, explain all variation. However, it is possible to fit splines to widely spaced "knots" using least-square methods. This technique is more flexible, but its effectiveness is dependent on determining the proper knot spacing. If knots are too close, some genetic variation will be eliminated; if they are too far removed, environmental trends will not be fully described.

Here, our objectives are to (1) demonstrate how these methods can reduce environmental variation, and therefore, increase heritability, (2) show how piece-wise spline knot spacing can affect genetic and environmental variances and (3) discuss how to apply these methods to maximum advantage.

MATERIALS AND METHODS

The example presented in this paper is from a sycamore (*Platanus occidentalis* L.) progeny test located in Putnam County, Georgia. The site is a level Piedmont bottomland. Before planting the site had been cleared of timber, stumped and cultivated. Subsoil variations have resulted in conspicuous gradients in site quality.

The progeny test was planted as a randomized complete block with 221 open-pollinated families, four blocks and two-tree plots. Tree spacing is 3 x 3 m. A double border row surrounds the perimeter. Heights and breast height diameters four years from planting were the data analyzed. Trees had just reached crown closure at the time of measurement.

The data were first analyzed using the planned ANOVAs for a randomized complete block design, and then using ANOVAs for a completely randomized design, after the environmental gradient effects estimated from the regression or spline models had been subtracted.

Models for estimating environmental gradient effects were constructed as follows. Curves were fit through the each row, independently, by least squares, using plot means. This procedure was then repeated for each column. For each plot, row and column estimates were averaged to produce a first approximation of the surface. The surface was smoothed by iteratively applying the row and column fitting procedure to the estimated values of the previous iteration. Less than ten iterations were necessary to provide a smooth approximation.

First, cubic polynomial regressions were used to model row and column curves. Then, piece-wise splines were used. In the piece-wise splines, piece lengths, hereafter called segments, were 6, 8, 10, 12 and 18 plots. This resulted in 5, 4, 3, 2 and 1 segment(s), respectively, for most rows and columns. Each segment was fit to a cubic polynomial by least-squares methods, subject to the constraints that the first and second derivatives of the curve segments must be the same at their unions. The resulting splines are, thus, referred to as least-squares piece-wise cubic splines.

Spline fitting was performed on a personal computer using programs adapted from deBoor (1978). These procedures will be detailed in a future paper. Cubic polynomial surfaces were fit with SAS (1982) procedure REG and with personal computer programs. Statistical analysis was conducted with SAS (1982) procedure VARCOMP.

For heritability calculations, block variation was excluded from the phenotypic variance term in the randomized complete block analysis, and the covariance among family members was assumed to be 0.25. Standard errors for heritabilities were calculated using the methods outlined by Namkoong (1979).

RESULTS

The results of environmental gradient modeling are portrayed by examining a slice through the progeny test along column Q (Figures 1-3). In Figure 1, curves resulting from the first approximation of the environmental gradient for height, determined by cubic polynomial regression and piece-wise splining with segment length of ten plots (spline 10), are shown. Highly irregular surfaces are produced from the first approximation. However, smooth surfaces are produced after ten iterations of surface fitting (Figure 2). From the smooth curves, it is evident that the piece-wise spline model was more "flexible" than the cubic polynomial model, producing peaks and valleys with greater amplitude and frequency. The effect of segment length on surface characteristics is shown in Figure 3. Spline surfaces constructed using short segment lengths have more flexibility than those using long segment lengths (e.g. compare the curve produced when segment length was six plots (spline 6) to that produced when segment length was 18 plots (spline 18)).

Greater flexibility means that more varition is explained by the the surface, leaving less for genetic and residual environmental variation. The effect of several environmental gradient models on family and residual environmental variance is shown in Figure 4. As expected, residual environmental variance was minimized when gradients were modeled using short segments in a piece-wise spline, and increased as the segments were lengthened (note that the polynomial regression is similar to a least-squares piece-wise spline where segment lengths equal the numbers of plots in each row or column). All of the smooth surface models resulted in less residual environmental variation than blocking.

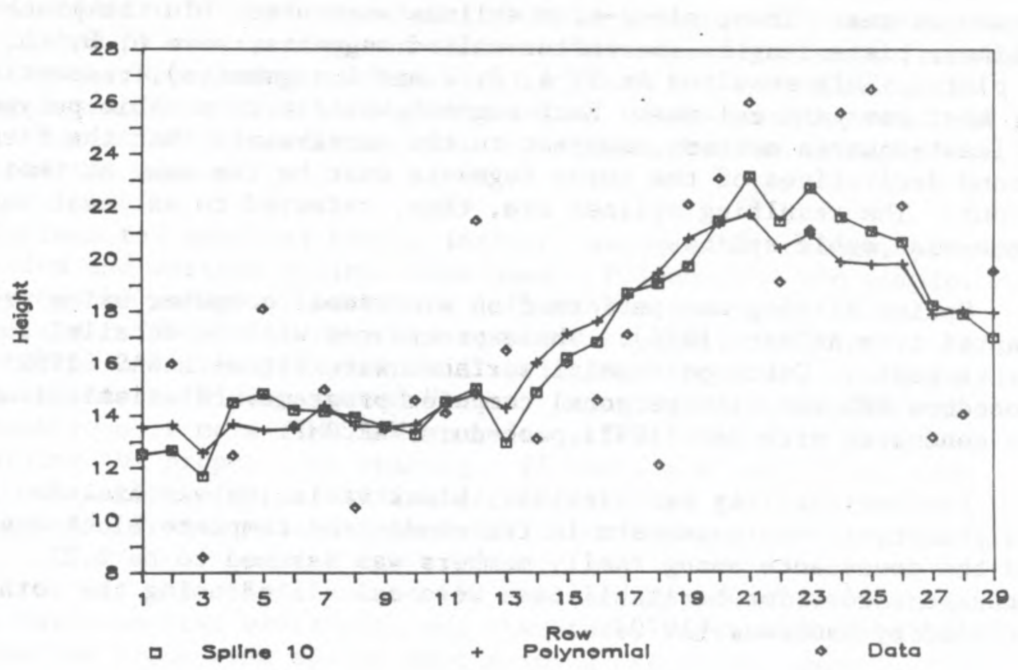

Figure 1. Least-squares spline and polynomial interpolation of environmental trends after one iteration. "10" refers to the number of data per segment.

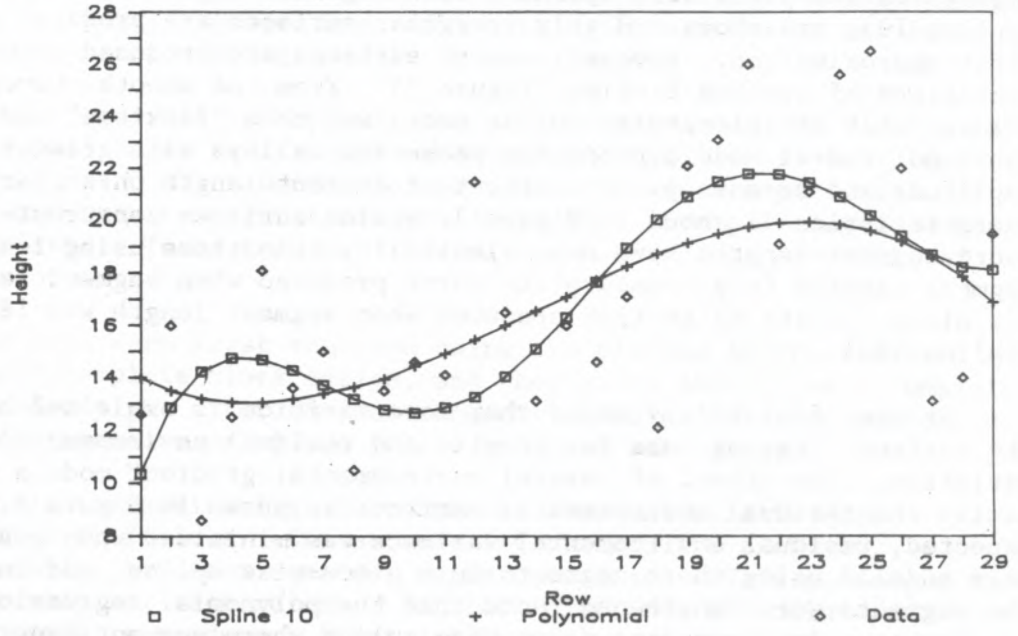

Figure 2. Least-squares spline and polynomial interpolations of environmental trends after ten iterations. "10" refers to the number of data per segment.

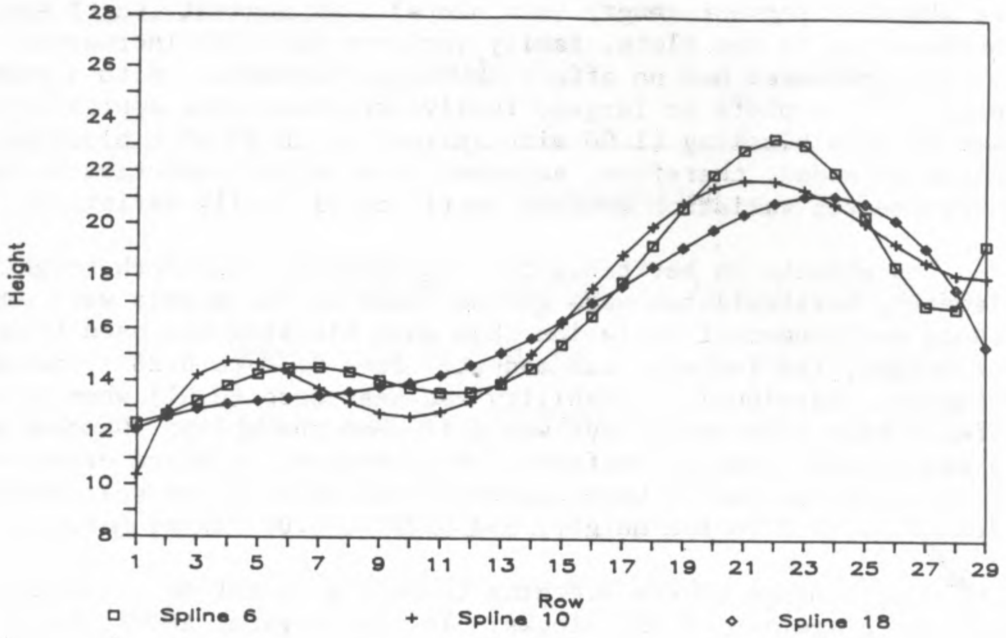

Figure 3. Least-squares spline interpolations of environmental trends after ten iterations using three segment lengths. Numbers refer to the number of plots per segment.

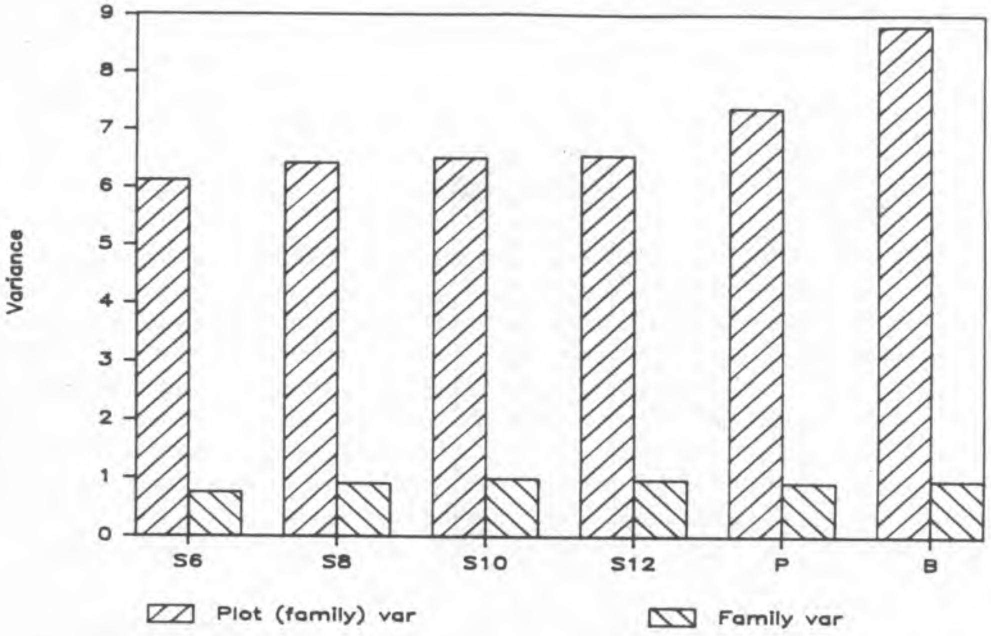

Figure 4. Residual environmental (plot within family) and family variances after environmental gradient effects were removed with splines (S), cubic polynomials (P) and blocking (B). Numbers refer to number of data per segment.

Family variance was also minimized (0.70) with the spline model having the shortest segment length (six plots). As segment length was increased up to ten plots, family variance was also increased. However, further increases had no effect on family variance. With a segment length of ten plots or larger, family variances were approximately the same as with blocking (1.00 with splines vs. 0.98 with blocking). The spline 10 model, therefore, achieved substantial reduction in residual environmental variation without sacrifice of family variation.

The effects on heritabilites are obvious. For both height and diameter, heritabilites were greater when spline models were employed to reduce environmental variation than when blocking was used (Figure 5). For height, the increase was modest: from 0.23 to 0.28. However, for diameter, individual heritability was near zero (0.01) when block effects were eliminated, but was 0.17 when the spline 10 model was used to remove site gradient effects. Furthermore, standard errors of the heritability estimates were somewhat reduced with the spline 10 model (from 0.08 to 0.06 for height, and 0.09 to 0.06 for diameter).

Shortcomings of the blocking technique cannot be attributed to improper placement of the blocks. In the original ANOVA, block effects were highly significant and removed substantial amounts of variation (Table 1). However, because environmental gradients are not stepwise, family x block interaction will be inflated compared to the true family x environment interaction. The success of the environmental gradient modeling techniques results from their ability to provide closer estimates of the true family x site interation variance (the plot (family) effect in the Table 1 ANOVA).

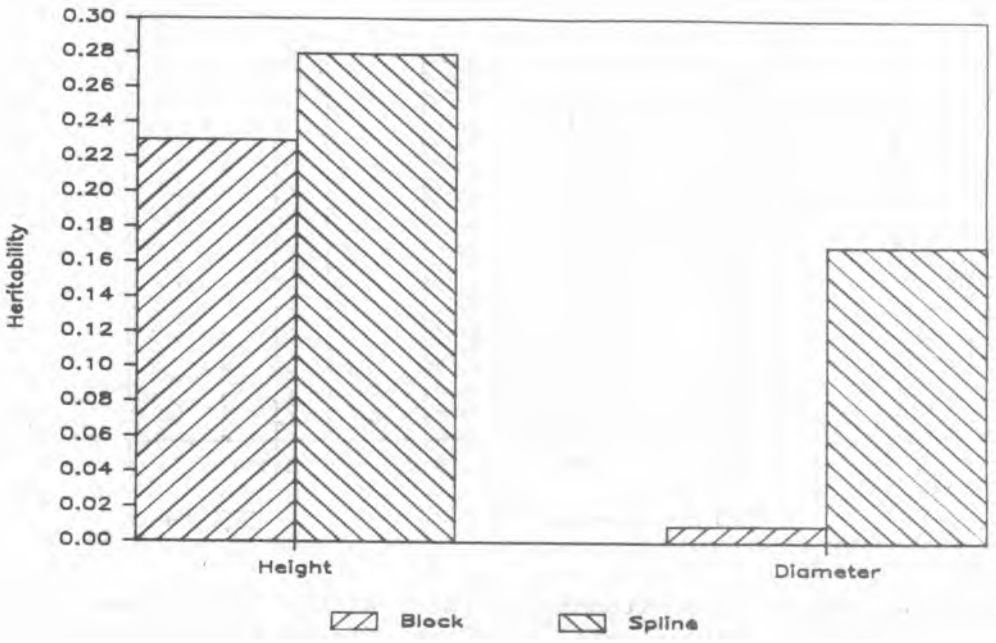

Figure 4. Heritabilities determined after site gradient effects were removed by blocking and by piece-wise least-squares cubic splines with a ten plot segment length.

Table 1. Analysis of variance with blocking and with environmental gradient effects removed by the spline 10 model.

Source	d.f.	Height		Diameter	
		MS	F	MS	F
			with blocking		
Block	3	928.78	138.21**	12.041	74.32**
Family	220	30.70	1.32*	0.573	1.01
Family x block	610	23.25	3.46**	0.565	3.49**
Tree(plot)	790	6.72		0.162	
			with splines		
Family	220	26.71	1.40**	0.525	1.23*
Plot(family)	613	19.11	2.84**	0.426	2.63**
Tree(plot)	790	6.72		0.162	

*Significant at α = 0.05.
**Significant at α = 0.01.

DISCUSSION

In this paper we have shown one example of how environmental gradient effects may be removed with regression methods (especially least-squares piece-wise splines) better than with blocking. However, the advantages of this approach will vary with site conditions and experimental design. For example, it will be most effective on sites that are environmentally heterogeneous and for traits that are environmentally sensitive. This will almost always be the case in hilly regions and bottomlands. Even in cases where environmental gradients are obvious, blocks are never uniform. Gradients typically occur in more than one dimension, and over distances too small for acceptable blocking.

Once blocking is eliminated from the experimental design, the size of a test need no longer be limited by block size, and the number of allowable genetic entries becomes unlimited. This makes analysis less "messy" and more accurate because test sites may contain the whole complement of families or clones.

Advantages of the spline approach will also be most apparent when single tree plots are employed. Small plots maximize the number of data points from which the model of environmental gradients will be based (only one value per plot is allowable since observations must be genetically random), and should result in the most accurate modeling of these gradients.

Spline models have the advantage of handling border effects well. One or two border rows are probably insufficient to eliminate border effects, but additional rows are too costly. Since least-squares spline models account for border effects, small borders contribute little to environmental variance. However, the trees that comprise border rows must be derived from the same population as the test trees, must be handled in the same manner as the test material and must be measured.

As a final point, the methods describe in this paper are compatible with the statisical and computational equipment presently available. The programs used for this analysis were run in less than two minutes on a personal computer, and the programs are general enough that they can be adapted to most situations with little difficulty.

LITERATURE CITED

de Boor, C. 1978. A practical guide to splines. Springer-Verlag, New York. 392 p.

Namkoong, G. 1979. Introduction to quantitative genetics in forestry USDA, Forest Service Tech. Bull. No. 1588. 342 p.

SAS Institute. 1982. SAS users guide: statistics, 1982 edition. SAS Institute, Inc., Cary, N.C. 584 p.