age. — Oxford University Press, London (1961). — BECKMAN, L., SCANDALIOS, J. G. and BREWRAKER, J. L.: Genetics of leucine aminopeptidase isozymes in maize. Genetics 50: 899–904 (1964). — BECKMAN, L., BECKMAN, G., CHRISTODOULOU, C. and IFEKWUNIGWE, A.: Variations in human placental alkaline phosphatase. — Acta Genet. Statist. Med. 17: 406—412 (1967). — BERGMANN, F.: The genetics of some isozyme systems in spruce endosperm (Picea abies). — Genetica 6: 353–360 (1974). — BHAT, N. R.: An improved genetical map of Punnett's B chromosome in Sweet Pea, Lathyrus odoratus L. — J. Genet. 48: 343–358 (1948). — BHAT, N. R.: A case in rice, *Oryza* sativa L. supporting Kosambis formula for estimating map lengths. — Amer. Nat. 84: 71-80 (1950). — CLEGG, M. T., ALLARD, R. W. and KAHLER, A. L.: Is the Gene the Unit of Selection? Evidence from Two Experimental Plant Populations. — Proc. Nat. Acad. Sci. USA 69: 2474–2478 (1972). — FISHER, R. A.: The Design of Experiments. — Oliver and Boyd, Edinburgh (1937). — GRELL, R.: The meiotic origin of temperature induced crossovers in Drosophila melanogaster females. — Genetics 54: 411–421 (1966). — GRIFFIN, B.: Accomodation of linkage in mass selection theory. — Austr. J. Biol. Sci. 13: 501—526 (1960). — HULTÉN, E.: Atlas of the distribution of vascular plants in NW Europe. — Generalstabens litografiska anstalts förlag, Stockholm (1950). — JENSEN, J. and HELMS JÖRGENSEN, J.: The barley chromosome 5 linkage map. I. Literature survey and map estimation procedure. — Hereditas 80: 5—16 (1975). — JONSSON, J. and BLOMQUIST, G.: Separation of Isoenzymes of Acid phosphatases from Needles of Pinus sylvestris by Partition in Aqueous Polymer Two Phase Systems. — Acta. Chem. Scand. B 30: 83–85 (1976). — KOJIMA, K. and LEWONTIN, R. C.: Evolutionary significance of linkage and epistasis. In Mathematical topics in population genetics (Ed. by K. KOJIMA). Springer Verlag Berlin, Heidelberg, New York (1970). — KORMUTAK, A.: Karyological structure of some Pinus species — Biologia (Bratislava) 30: 545—550 (1975). — KOSAMBI, D. D.: The estimation of map distance from recombination values. Ann. Eugen, London 12: 172–175 (1944). — LAWRENCE, S. H., MELNICK, P. J. and WEIMER, H. E.: A species comparison of serum proteins and enzymes by starch gel electrophoresis. — Proc. Soc. Exp. Biol. Med. 46: 572–575 (1960). — LEWONTIN, R. C.: The interaction of selection and linkage. I General consideration of heterotic models. Genet. 49: 49-67 (1964). — LUNDKVIST, K.: Analysis of linkage in Picea abies by means of isozyme studies. — Proceedings, Joint IUFRO Meeting S.02.04.1-3, Stockholm, Special Reports: 468 (1974). — MANGELSDORF, P. C.: The mutagenic effect of hybridizing maize and teosinte. — Cold Spring Harbor Sump. Quant. Biol. 23: 409—421 (1958). — MATHER, K.: The measurement of linkage in heredity. — Methuen & Co Ltd, London (1951). — NILAN, R. A.: The cytology and genetics of barley 1951–1962. — Monograph. Suppl. No. 3. Res. stud. Wash. state Univ. 32 (1), 278 p. (1964). — OWEN, A. R. G.: The theory of genetical recombination. — Advances in genetics 3: 117–159 (1950). — OWEN, A. R. G.: The analysis of multiple linkage data. — Heredity 7: 247—264 (1953). — POST, L. VON: Ur de sydsvenska skogarnas regionala historia under postarktisk tid. — Geol För Förh 46: 1-2: 83–128 (1924). — RUDIN, D. and RASMUSON, B.: Genetic variation in esterases from needles of Pinus sylvestris L. — Hereditas 73: 89–98 (1973). — RUDIN, D.: Inheritance of glutamate oxalate-transaminases (GOT) from needles and endosperms of Pinus sylvestris. — Hereditas 80: 296–300 (1975). — RUDIN, D.: Leucine-amino-peptidases (LAP) from needles and endosperms of Pinus sylvestris L. — a study of inheritance of allozymes. — Hereditas (in press) (1977). — RUNQUIST, E. W.: Meiotic investigation in Pinus sylvestris L. — Hereditas 60: 77–128 (1968). — SAYLOR, L. C.: Karyotype analysis of the genus Pinus subgenus Pinus. — Silvae Genet. 21: 155—163 (1972). — SCHWARTZ, M. R., NISSELBAUM, J. S. and BODANSKY, O.: Procedure for staining zones of activity of glutamatic oxaloacetic transaminase following electrophoresis with starch gel. — Am. J. Clin. Pathol. 40: 103–106 (1963). — SHAW, C. R. and PRASAD, R.: Starch Gel Electrophoresis of Enzymes — A compilation of Recipes. — Biochem. Genet. 4: 207–320 (1970). — SIMONSEN, V. and WELLENDORF, H.: Some polymorphic isoenzymes in the seed endosperm of sitka spruce (Picea sttchensis (BONG.) CARR.). — Forest Tree Improvement 9: 1–21 (1975). — STEPHENS, S. G.: The cytogenetics of speciation in Gossypium. I. Selective elimination of the donor parent genotype in interspecific backcrosses. — Genetics 34: 627—637 (1949). — STEPHENS, S. G.: The internal mechanism of speciation in Gossypium. — Bot. Rev. 16: 115—149 (1950). — WALLACE, M. E.: The use of affinity in chromosome mapping. — Biometrics 13: 98–110 (1957). — WRIGHT, J. W.: Introduction to Forest Genetics. — Academic Press. New York, San Francisco, London (1976).

An Analysis method to improve statistical efficiency of a randomized complete block design1)

By J. W. WRIGHT2)

(Received March 1978)

Summary

A procedure, called the "moving average" method is described whereby statistical efficiency can be made as high for experiments following randomized complete block designs as for those following more complex lattice or incomplete block designs. Briefly, the procedure is to consider each plot in relation to few of its neighbors and adjust its mean according to the performance of those neighbors. The method is most useful in experiments having large blocks with large amounts of within-block variation and effectively converts them into experiments consisting of numerous small, uniform blocks. As used in five plantations, the moving average method reduced error sums of squares by 30—50% and changed seedlots means by 0—16%.

Key words: Moving average method.

Zusammenfassung

Ein Verfahren, bezeichnet als „moving average"-Methode, wird beschrieben, nach dem die Auswertung von randomisierten vollständigen Blockanlagen mit gleich hoher statistischer Effizienz möglich ist wie diejenige von den komplexen lateinischen oder unvollständigen Blockanlagen. Bei diesem Verfahren wird jede Versuchsparzelle mit wenigen Nachbarzellen in Beziehung gebracht und das Parzellenmittel entsprechend korrigiert. Diese Methode eignet sich besonders für Versuchsanlagen mit großen Blocks und großem Variationsanteil für die Ursache ,,innerhalb Blocks" und wandelt diese in zahlreiche kleine gleichmäßige Blocks um. Durch die Anwendung des beschriebenen Verfahrens wurde bei der Auswertung von 5 Versuchsanlagen mit Pinus resinosa-Herkünften die Summe der Abweichungsquadrate für den Versuchsfehler um 30 bis 50% verringert. Die Herkunft-Mittelwerte änderten sich dabei um 0—16%.

Introduction

A randomized complete block design, in which each seedlot is reprsented by one plot in each block, is the simplest replicated design to install and analyze. In large experi-

---

1) Approved for publication as Journal Article No. 8465 of the Michigan Agricultural Experiment Station. The work reported here was supported in part by regional research funds from the U. S. Dept. of Agriculture as part of regional project NC-99 entitled "Improvement of Forest Trees through Selection and Breeding".

2) Professor of Forestry, Michigan State University, East Lansing, Mich. 48823, U.S.A.

ments, lattice or incomplete block designs in which block size is limited, may be much more efficient statistically. They are, however, much more difficult to install and analyze. The present paper describes a technique by which statistical efficiency can become as great for a randomized complete block experiment as for experiments involving more sophisticated designs.

The proposed method involves extra work, so its use is justified only in certain circumstances. I will describe a plantation in which it proved very useful. This is an 83-seedlot provenance test of red pine *(Pinus resinosa)* established in 1963 in northern Michigan by Maurice W. Day. The plantation follows a randomized complete block design with 10 replications, 4-tree row plots, and a single border row surrounding the experimental trees, and an 8 × 8 ft (2.4 × 2.4 m) spacing. The blocks were contiguous and as nearly square as possible. The plantation (excluding border rows) was 113 rows (896 ft = 274 m) long north-south and 24 trees (184 ft = 56 m) wide east-west.

The planting site was a level old field believed to be uniformly favorable for the growth of red pine. There were some bare-soil areas near the north end as the result of wind erosion but elsewhere the ground cover was light and uniform. Growth was uniform and good in an older red pine plantation 30 ft to the west.

The planting site proved to be generally favorable but not uniform. Within 3 years of planting there were pronounced north-south differences in foliage color and growth rate. As of 1977, 15 years after planting, average height varied from 16 ft (4.9 m) at the north end to 26 ft (7.9 m) at the south end.

The problem would have been minor if the site-induced differences in growth rate had been smaller or had followed block boundaries. In that case, ordinary calculation methods would have sufficed, as between-block differences would have been substracted from the error sum of squares in calculating the analysis of variance. That was not the case, however. There were slight gradients in site quality at the ends of the plantation and very abrupt differences in the middle. Some of these abrupt differences occurred in the middle of a block. In block 5 for example, one tier of 12 plots averaged 49% taller than another tier 32 ft (10 m) to the east.

The problem was serious, and led to the development of a technique for improving the statistical efficiency of a randomized complete block design. The procedure, which may be called the "moving average method" is described in this paper.

Description of the "Moving average Method"

Briefly, the procedure is to consider each plot in relation to a few of its neighbors, thus converting an experiment consisting of a few large blocks into one containing numerous small blocks. Detailed steps follow.

(a) Measure heights and record plot means on field data sheets.

(b) Prepare table summarizing height by seedlots and block, compute seedlot totals and means. Up to this point the procedure is the same as for an ordinary analysis of variance.

(c) Prepare another table following the same format as in (b) above and insert the deviation of each plot from the seedlot mean.

(d) Return to the field data sheets and insert these deviations for all plots, paying attention to the sign of each deviation.

(e) For a particular plot, calculate the average deviation for the two[3] plots on either side. For example, to obtain the average deviation applicable to the 11th plot, calculate the total deviation for plots 9, 10, 12 and 13, then divide the total by 4. Near the plantation's edge the average deviation must be based on two or three rather than the desired four neighboring plots. That is, base the average deviation for the first plot on plots 2 and 3; for the second plot on plots 1 3 and 4. For convenience, round off each average deviation to the nearest whole number; this usually has negligible consequences.

Each average deviation is a measure of the site quality of a small "block" applicable to but not including the plot in question.

(f) Reverse the sign of each average deviation, converting it into a correction factor, and add that factor to the actual mean for each plot. This effectively causes the entire plantation to be considered as of uniform site quality.

(g) Insert the corrected plot means into a third table following the same format as in (b) above, again summarizing height by seedlot and block.

(h) Calculate sums of squares and mean squares in the normal manner, with degrees of freedom as shown below:

Source of variation	Degrees of freedom
Seedlot	No of seedlots — 1
Regression on site quality	1
Total	Total no. of plots — 1
Error	Remainder

The block sum of squares is ignored because the correction factor eliminated differences among blocks. Mathematically, the correction procedure amounts to calculating each plot mean as a deviation from the regression line expressing site quality, with 1 degree of freedom subtracted for the regression.

Time required and statistical consequences

Doing all calculations by hand or with a small desk calculator, use of the moving average method requires about three times as much time as an ordinary analysis of variance. For the 83-seedlot, 10-replicated red pine provenance test, 6 hours were required. Parts of the procedure (steps b, g and h above) can be done by computer, using presently available programs. Programs could be written to handle the entire procedure by computer if there was enough demand.

Presumably the seedlot means calculated after the application of the correction factors are better estimates of the seedlots' true growth potentials than are the actual means. The most important genetic consequence is the change in seedlot means caused by the correction process. In the red pine experiment, the changes were:

40 means were changed by 0— 4%
35 means were changed by 5— 8%
3 means were changed by 9—12%
5 means were changed by 13—16%

Due to shortages of planting stock, some seedlots were omitted from some blocks, and the experiment included only 768 instead of the planned 830 plots. Thus, some seedlots were underrepresented and their means were changed the

Table 1. — Comparison of analyse of variance performed on actual (Act.) and adjusted (Adj.) data derived from 1973 measurements of height of a red pine provenance test.

Source of variation	Degrees of freedom		Sums of squares		Mean squares		F values	
	Act.	Adj.	Act.	Adj.	Act.	Adj.	Act.	Adj.
Seedlot	82	82	8,690	8,479	106	103	1.7	3.2
Block	9	--	36,698	--	4076	--		
Regression on site quality	--	1	--	56,683		56,683		
Error	676	684	41,518	21,744	61	32		
Total	767	767	86,906	86,906				

---

[3] The average deviation can also be based on the one or three of four plots on either side of the one in question.

most. However, there were large changes for seedlots represented in all 10 blocks. Overall, the changes in seedlot means seemed well worth the extra effort involved in the adjustment calculations.

The other important consequences concerned the analysis of variance *(Table 1)*. In the red pine experiment, the adjustments had little effect on the seedlot sums of squares but reduced the error sum of squares by almost 50%. As a result, the F value for seedlot rose from a barely significant 1.7 to a respectable 3.2, and the coefficient of variation was reduced to 10% of the mean.

General Applicability

The moving average method has been used with data from four other plantations having pronounced site differences within and between blocks. In each case, the improvement in statistical precision was considerable although not as marked as in the red pine provenance test. The method has also been tried with data from plantations without obvious large differences in site quality within blocks. In such cases, the adjustments amounted to random changes in plot means and had little effect on seedlot means, sums of squares or mean squares. In other words, unwarranted use of the proposed new method has done neither harm nor good.

Effectively, the moving average method causes an experiment such as the red pine provenance test to be considered as consisting of a very large number of small, incomplete blocks. Statistical efficiency becomes just as great as if a very efficient lattice or incomplete block design had been used when the experiment was installed.

In Michigan there are approximately 250 replicated genetic test plantations, of which 5 need to be analyzed by the rather laborious moving average method and the remainder can be analyzed by the simpler methods applicable to the randomized complete block design.

It would have been possible to achieve high statistical precision in all plantations by routine use of balanced incomplete block designs. However, the extra effort would have reaped commensurate benefits only 2% of the time. Availability of the running average method makes it possible to use simple designs but attain the high precision inherent in the more complex designs if there is a need.

I requested views as to the legitimacy of the moving average method from three trained statisticians. One had reservations but the other two considered it valid. The one with reservations considered it preferable to base the height adjustments on some measure of site quality rather than on height itself. In other words, he favored analysis of covariance.

Multivariate Classification in Provenance Research

A comparison of two statistical techniques*

By E. R. Falkenhagen and St. W. Nash

Forest Research Institute, (P.O. Box 727)
Pretoria, Transvaal, South Africa
and
Department of Mathematics
University of British Columbia
Vancouver, British Columbia, Canada
respectively.

(Received June 1976 / January 1978)

Summary

The relationships of two kinds of multivariate statistical analysis are discussed and illustrated with data from a forest tree provenance study. Canonical correlation analysis of the biological data, taking into account geographical data, separated provenances and ecological regions more effectively than did canonical variate analysis (discriminant analysis). In each case Mahalanobis distances between provenances could be calculated. Distances under the first analysis were greater than under the second, but the pattern was fairly consistent from one analysis to the other.

Key words: *Picea sitchensis* (Bong.) Carr., provenances, canonical correlation analysis, discriminant analysis, Mahalanobis distance.

Résumé

Titre: Classification multivariable en recherches sur les provenances d'arbres forestiers: comparaison de deux techniques statistiques.

Les relations entre deux analyses statistiques multivariables sont discutées et illustrées à l'aide de données accumulées lors d'un test de provenances d'arbre forestier. L'analyse des corrélations canoniques des données biologiques, tenant compte des coordonnées géographiques, a séparé les provenances et les régions écologiques de façon plus effectice que l'analyse de variables canoniques (appellée aussi, analyse discriminante). Dans tous les cas, les distances de Mahalanobis entre les provenances ont pu être calculées. Lors de la première analyse, les distances s'avérèrent plus grandes que lors de la seconde, mais leur relations restèrent passablement constantes d'une analyse à l'autre.

Zusammenfassung

Titel: Multivariable Klassifikation in der Herkunftsun-

---

* This paper grew out of research for a dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Graduate Studies, University of British Columbia, by the first author.