Correlation of height growth in black spruce with site factors of the prove

By M. A. K. KHALIL and A. W. DOUGLAS*)

(Received August 1978 / June 1979)

Summary

A black spruce provenance study, based on 26 Newfoundland provenances, was started in 1968 in a nursery at Pasadena in western Newfoundland, using bulked open-pollinated seed. The five-replicated randomized complete block design with 100-seed plots was used. Four-year heights of eight randomly selected seedlings in each plot were determined in October 1971. The effects of nine environmental factors of the sites of the provenances on height growth were studied to identify the subject of those which contributed most to height growth and to assess the usefulness of the subset as a predictor.

Analysis of variance was performed with multiple and equal number of observations per plot, using the mixed model. A stepwise zig-zag regression procedure was adopted using the mean provenance height as the dependent variable and the nine environmental factors of the sites of the provenances with all their pair-wise cross-products and squares as independent variables.

No independent variable is correlated with the dependent variable in a simple linear manner. The significant subset of independent variables comprised the following:

X^2_1 — Square of the latitude.
X_2X_7 — Interaction of longitude and mean precipitation during the period May—September.
X_4X_5 — Interaction of mean number of frost-free days and mean number of days during the annual growing period.
X_4X_9 — Interaction of the number of frost-free days during the annual growing period and the day length on June 21.

Key words: Picea mariana (MILL.) B.S.P., early testing, juvenile growth.

Sommaire

Une étude de provenances d'Epinette noire, fondee sur 26 provenances de Terre-Neuve fut amorcee en 1968 dans une pépinière a Pasadena dans l'ouest de Terre-Neuve, en utilisant en grandes quantites des graines issues de la pollinisation libre. Le «design» du bloc complet randomisé à cinq repetitions avec placeaux de 100 graines chacun fut employe. La hauteur de huit semis de quatre ans choisis au hasard dans chaque échantillonnage fut déterminée en octobre 1971. On etudia l'influence de neuf facteurs environnementaux sur les stations des provenances, touchant la croissance en hauteur, afin d'identifier le sous- facteur qui, parmi les autres, contribua plus a la croissance en hauteur et afin d'evaluer l'utilité de ce sous-facteur comme element de prédiction.

On proceda a l'analyse de variance par multiples et un nombre egal d'observations par placeau, a l'aide du modele mixte. On adopta un processus de regression en zigzag par etapes en utilisant la hauteur moyenne des provenances comme variable dépendante et les neuf facteurs environnementaux des stations des provenances avec tous leurs produits croisés accouples et leurs carres comme variables independantes.

Aucune variable independante n'est corrélative a la variable dépendante d'une simple maniere linéaire. Le sous-facteur significatif des variables independantes comprenait les éléments suivants:

X^2_1 — Carre de la latitude.
X_2X_7 — Interaction de la longitude et de la precipitation moyenne de mai a septembre.
X_4X_5 — Interaction du nombre moyen de jours sans gel et du nombre moyen de jours durant la période de croissance annuelle.
X_4X_9 — Interaction du nombre moyen de jours sans gel durant la période de crossance annuelle et de la photopériode du 21 juin.

Zusammenfassung

Ein Provenienzversuch mit 26 Herkünften der nordamerikanischen Schwarzfichte [*Picea mariana* (MILL.) B. S. P.] aus Neufundland wurde mit der Vegetationsperiode 1968 in einer Baumschule in Pasadena (westliches Neufundland) begonnen, wobei frei abgeblühtes Samenmaterial mehrerer Bäume zur Verwendung kam. Die Untersuchung wurde in Form eines vollständigen randomisierten Blockversuchs in 5 Wiederholungen mit Parzellen zu je 100 Samen durchgeführt. Im Oktober 1971 wurde in jeder Parzelle an einer Zufallsauswahl von acht Sämlingen die Höhe gemessen. Die Auswirkung von neun Umweltfaktoren der Herkunftsorte auf das Höhenwachstum wurde untersucht, um diejenigen bestimmen zu können, die am meisten zum Höhenwachstum beitragen und die Brauchbarkeit dieser Faktoren zur Prognose zu beurteilen.

Unter Benutzung des gemischten Modells wurde eine Varianzanalyse mit Mehrfachbeobachtungen gleichen Umfanges in jeder Parzelle durchgeführt. Hierfür wurde ein schrittweises Zickzack-Regressionsverfahren gewählt, bei dem die mittlere Provenienzenhöhe die abhängige Variable darstellt und die neun Umweltfaktoren der Herkunftsorte mit sämtlichen ihrer paarweisen Vektorprodukte und (Quadrate die unabhängigen Variablen bilden.

Keine der unabhängigen Variablen ist mit der abhängigen auf einfache lineare Weise korreliert. Die signifikante Untergruppe der unabhängigen Variablen setzt sich wie folgt zusammen:

X^2_1 — Quadrat der geographischen Breite.
X_2X_7 — Wechselbeziehung zwischen geographischer Länge und mittlerer Niederschlagsmenge im Zeitraum Mai—September.
X_4X_5 — Wechselwirkung zwischen der mittleren Anzahl frostfreier Tage und der mittleren Zahl aller Tage innerhalb der jährlichen Wachstumsperiode.
X_4X_9 — Wechselwirkung zwischen der mittleren Anzahl frostfreier Tage der jährlichen Wachtumsperiode und der Tageslänge am 21. Juni.

Introduction

Black spruce (*Picea mariana* (MILL.) B.S.P.) occupies a very prominent position in Canadian forestry due to its excellence as pulpwood, its widespread distribution, occupy-

*) *The authors are Research Scientist, Newfoundland Forest Research Centre, Canadian Forestry Service, Environment Canada, St. John's Newfoundland, Canada and Director of Applied Statistics and Scientific Computing Branch, Computing and Applied Statistics Directorate, Environment Canada, Ottawa, Canada.*

ing very variable sites, relative tolerance to insect pests and diseases and relative ease in artificial regeneration. Consequently, several national and regional projects on the genetic improvement of the species are in progress. However, long periods of time to sexual and economic maturity are serious challenges to obtaining useful results from such studies. The conventional early tests, based on infantile-juvenile-mature correlations for specified characters of progenies as reported, among others, by Krug (1963), Langlet (1959), Nanson (1968) and Schmidt (1972) also require long periods of time for establishing criteria for early detection of superior provenances. The need for alternative methods clearly exists.

A regional black spruce provenance study is one of the three genetic studies on this species currently in progress at the Newfoundland Forest Research Centre of the Canadian Forestry Service. Preliminary results of this study reported earlier (Khalil, 1975) have indicated strong genetic control on juvenile height growth. The ecotypic nature of variation in this character, corresponding with Rowe's forest sections (Rowe, 1972) indicates selection pressures exerted by widely different climatic conditions prevailing in the various forest sections during the micro-evolution of black spruce on the Island. The data were analysed to study the effect of selected environmental factors of the sites of the provenances on height growth in an attempt to isolate which subset of these contributed most to height growth and to assess the usefulness of the subset as a predictor.

Material and Methods

The provenance

This study is based on 26 provenances from insular Newfoundland, all in the Boreal forest region (Rowe, 1972). The sampling covered wide north-south, east-west and coastal-inland distribution so as to include the maximum variety of geographical and climatic conditions. Open-pollinated seed was obtained from each of the seven dominant or co-dominant trees in each provenance. The parent trees were spaced at least 30 m apart to minimize inbreeding. The cleaned seed from the seven trees of each provenance was mixed in equal quantities by weight. This bulked seed of each provenance constituted its sample of seed. The bulked seed was sown on May 27—29, 1968 in the Canadian Forestry Service Research Nursery at Pasadena near Deer Lake in western Newfoundland (lat. 49⁰ 00' N, long. 57⁰ 35' N) in a five-replicated randomized complete block design experiment with 100-seed plots.

The data

Total heights of eight randomly selected seedlings in each plot were measured at the age of four years in October 1971. The provenance mean values were calculated based on the 40 seedlings of each provenance. Values of the following environmental variables at the origin of the parent trees were obtained.

$X^{*})_1$ — Latitude (⁰ N).
$X^{*})_2$ — Longitude (⁰W).
$X^{*})_3$ — Altitude (m).
$X^{**})_4$ — Mean number of frost-free days.
$X^{**})_5$ — Mean number of degree-days during the annual growing period.
$X^{**})_6$ — Mean temperature during the period May—September (⁰C).
$X^{**})_7$ — Mean precipitation during the period May-September (mm).
$X^{**})_8$ — Mean maximum July temperature (⁰C).
$X^{**})_9$ — Day length on June 21.

*) Anon. (1971).
**) Chapman and Brown (1966).

The values of variables $X_4 — X_9$ are based on observations over a very large number of years.

The statistical analysis

The data on the height of seedlings were analysed by analysis of variance with multiple and equal number of observations per plot using the mixed model (Steel and Torrie 1960). Multiple regression analysis was performed on the data, using the mean heights of provenances over all replications as the dependent or response variable and the nine afore-mentioned geographic and meteorological factors at the origin of the parent trees as independent variables. A preliminary analysis was done using the first order prediction equation shown in equation (1), without interactions or higher order effects of the independent variables:

$$Y = \hat{a} + \Sigma\hat{\beta}_i X_i \quad (1)$$

where $\hat{Y}$ is the predicted value of the dependent variable (Y), $\hat{a}$ is the estimated intercept and $\hat{\beta}_i$ is the estimated partial regression coefficient of the independent variable X_i (where i = 1, . . ., 9). The multiple correlation coefficient, residual mean square and the lack of fit test indicated that the first order prediction regression equation did not adequately represent the correlation between the dependent and the nine independent variables. The multiple regression analysis was repeated using the predictive equation shown in equation (2) which includes the nine independent variables and all their pairwise cross-products and squares, giving a total of 54 independent variables:

$$\hat{Y} = \hat{a} + \Sigma_i \hat{\beta}_i X_i + \Sigma_{i<j} \hat{\beta}_{i+9} X_i X_j + \Sigma_i \hat{\beta}_{i+45} X^2_i \quad (2)$$

(The various symbols used in this equation have the same meaning as explained above and X_j indicates an X-variable with a larger subscript than X_i.) A stepwise zig-zag regression procedure (two steps ahead, one back) was adopted as a mean of identifying the most important subset of the independent variables (Wilson, 1975; Blair, Brown and Wilson, 1971).

To determine at what stage of the stepwise regression procedure to stop a lack of fit term, which measures directly how well the equation fits the data, was calculated at each stage and tested for significance. This was continued until the test, which is a conservative one, yielded a non-significant result. Because each of the 26 provenances had only one value for each member of the set of nine X-variables associated with it, the regression results corresponded to a partitioning of the provenance sums of squares and degrees of freedom obtained from the analysis of variance into a component due to regression with degrees of freedom equal to the number of X-variables included at the particular stage of interest and a component due to lack of fit. These components could then be tested by an F-test using the Experimental Error mean square, which because it was based on individual observations, had to be divided by 40 to bring it to the same terms as the regression results which were based on provenance means. Following this criterion the zig-zag procedure was stopped at stage 5 which included four independent variables and established the general equation shown in equation (3):

$$\hat{Y} = \hat{a} + \hat{\beta}_1 X^2_1 + \hat{\beta}_2 X_2 X_7 + \hat{\beta}_3 X_4 X_5 + \hat{\beta}_4 X_4 X_9 \quad (3)$$

Results and Discussion

Equation (4) was finally obtained as the regression equation:

$$\hat{Y} = 184.3582 - 0.0339X^2_1 - 0.0108X_2X_7 + 0.0011X_4X_5 - 0.0007X_4X_9 \quad (4)$$

The multiple regression coefficient was very high ($R = 0.8847$; $R^2 = 0.7827$) and the residual mean square was 5.9839. *Table 1* shows the test of significance of the regression equation.

Table 1. — Test of significance of the regression equation.

Source of variation	D.F.	Sum of squares	Mean squares	F
Provenances	25			
Regression	4	712.2400	178.0600	29.76***
Lack of fit	21	197.7371	9.4161	1.57^{NS}
Experimental error	100	23,935.67	5.9839†	

† - Sum of squares divided by (40)(100), see text.

*** - Statistically significant, $P(F \geq 29.76) < 0.001$.

NS - Statistically non-significant, $P(F \geq 1.57) > 0.05$.

If the zig-zag regression procedure had been continued to the last step (which was step 31) it would have included 15 independent variables. The multiple correlation coefficient would have increased to 0.9720 ($R^2 = 0.9447$) and the residual mean square would have been reduced to 5.0350. As these changes would have been insignificant equation (4) was taken as the best fit for the data.

It is interesting that no independent variable is correlated with the response variable in a simple linear manner. Only one such variable, latitude, occurs alone and is correlated with height growth in the quadratic power. All other independent variables are correlated with the response variable through their first order interactions as is shown by the regression equation.

Although with additional data the form of the regression equation may change it is evident that the geographic and meteorological variables tested are highly correlated with total height at this age. The equation presented can therefore be used quite satisfactorily to predict average response under similar experimental conditions.

The biological interpretation of the effect of geographic and meteorological factors on response is complex. First, although these four variables are the ones most closely associated with the response variable, they probably affect the latter through other variables by a very complicated process. The complexity of the correlation is created, at least partially, by the complex pattern of the climate of insular Newfoundland, with its small land mass, surrounded by a variety of maritime influences. The climate is strongly influenced by the Labrador and Gaspe currents (Hare, 1952). The Labrador current virtually encircles the Island and brings the Arctic Sea ice which often persists into May and June and delays the rise of air temperature in spring over coastal areas. This effect is more pronounced on the Northern Peninsula and the northeast coast than elsewhere and low temperatures persist in these areas till late June or early July. The important effects of the current are delayed spring and lower temperatures in the north than in the south in winter and spring. The temperature pattern is altered in summer by frequent southwest winds. These winds pass over the cold Labrador current and cool the southern coast. They warm up as they move over the Island and produce the paradoxical trend of increasing temperatures from south to north in summer. However, the Northern Peninsula continues to be under the influence of the Labrador current and remains colder than the rest of the Island throughout the year. Consequently, the meteorological factors represented by the independent variables X_4—X_6, X_8 and X_9 have the highest values in central Newfoundland and decrease towards the coast. Only the mean precipitation during the period May—September increases from north to south. Thus, the genetic changes produced in black spruce of insular Newfoundland by natural selection during micro-evolution do not follow a north-south or east-west trend but a complex inland-coastal trend. This is reflected in the absence of correlation of any of the independent variables with the response variable except through their interaction with other independent variables or in the quadratic form.

Conclusion

The following four independent variables at the origin of the parent trees out of the 54 tested are significantly correlated with the four-year height growth of black spruce:

- X^2_1 — Square of the latitude.
- X_2X_7 — Interaction of longitude and mean precipitation during the period May—September.
- X_4X_5 — Interaction of mean number of frost-free days and mean number of days during the annual growing period.
- X_4X_9 — Interaction of the mean number of frost-free days during the annual growing period and the day length on June 21.

Literature Cited

Anon.: 1:50,000 Grid maps of Newfoundland. Surveys and Mapping Branch, Dep. of En., Mines & Res., Can. (1971). — Blair, E. A., Brown, D. M. and Wilson, A. L.: Regarding zig-zag regression, Am. Statist. 25 (3): 56 (1971). — Chapman, L. D. and Brown, D. M.: The climate of Canada for agriculture. The Can. Land Invent. Rep. No. 3, Dep. For. & Rural Dev., 24 pp. (1966). — Hare, F. K.: The climate of the island of Newfoundland. A geographical analysis. Geographical Bulletin No. 2, Can. Dep. Mines & Tec. Sur. 36—88 (1952). — Khalil, M. A. K.: Genetic variation in black spruce (*Picea mariana* (Mill.) B.S.P.) in Newfoundland. Silvae Genetica 24 (4): 88—96 (1975). — Krug, H.: Beitrag zur Frühdiagnose der Ertragsbildung von Kartoffelpflanzen unter besonderer Berücksichtigung der photoperiodischen Reaktion. Der Züchter (Special issue No. 6) (1963). — Edited translation by Multilingual Services Division, Translation Bureau, Department of the Secretary of State, Government of Canada (1963). — Langlet, O.: A cline or not a cline — a question of Scots pine. Silvae Genetica 8 (1): 13—22 (1959). — Nanson, A.: The study of early tests, particularly by means of multiple correlations and regressions. Forstpflanzen-Forstsamen 3, 1967. Transl. For. Comm. Lond. No. 406, 15 pp. (1968). — Rowe, J. S.: Forest regions of Canada. Dep. of the Environ., Can. For. Serv. Publ. No. 1300, Inf. Can., 172 pp. (1972). — Schmidt, W.: Juvenile-mature correlations. Early tests. In: Proc. of a Meeting of Working Party on Progeny Testing. IUFRO, Macon, Georgia, U.S.A., Oct. 25—27, 1972: 154—158 (1972). — Wilson, A. L.: A regression manual. Applied Stat. & Sc. Comp. Br., Environ. Can. (1975).