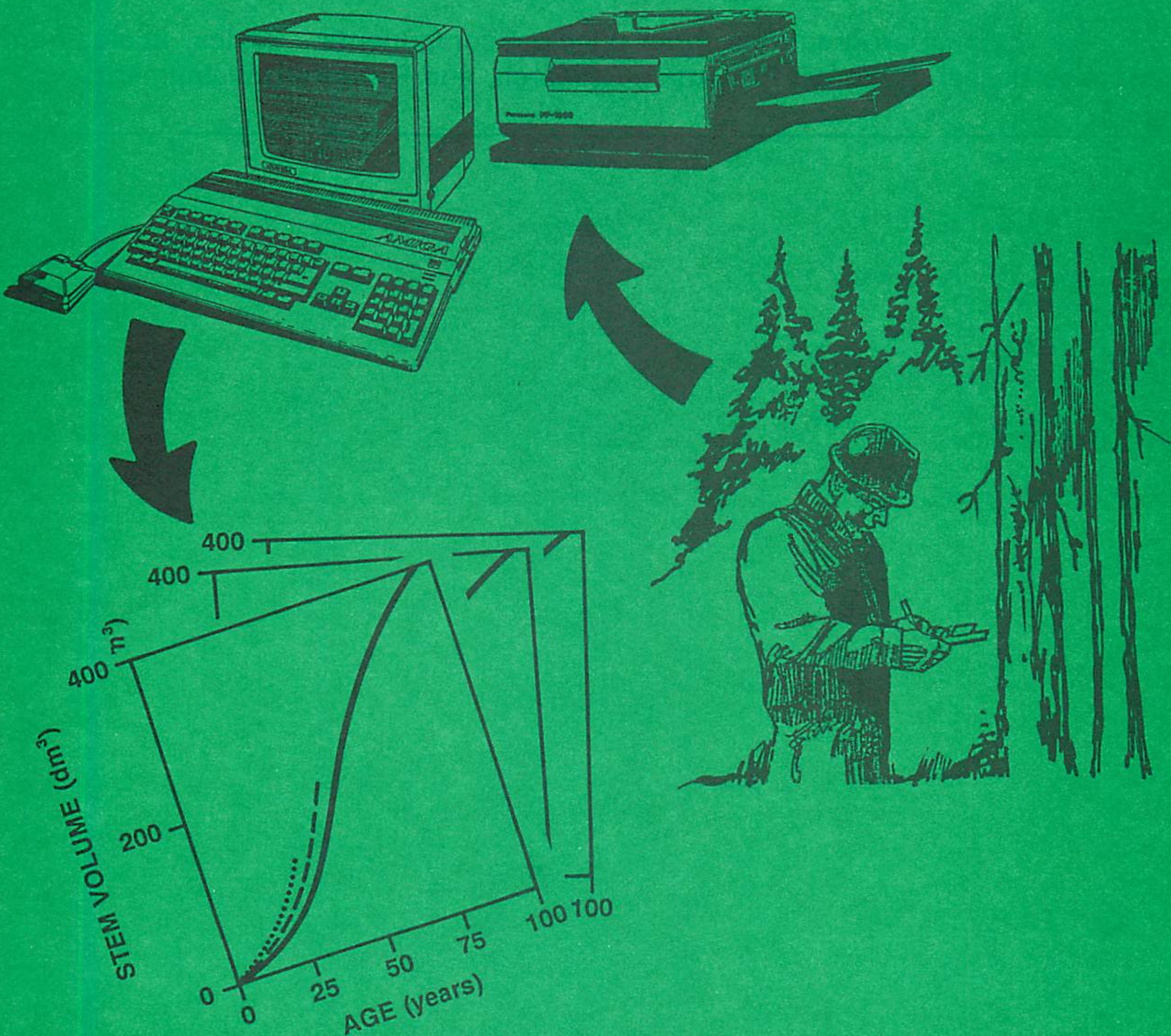
# Guidelines for editing permanent sample plot data

M.B. Karsh and M.B. Lavigne

Newfoundland and Labrador Region · Information Report N-X-287




Forestry Canada
Forêts Canada

**FORESTRY CANADA**
**NEWFOUNDLAND AND LABRADOR REGION**

The objective of Forestry Canada in this region is to provide the scientific, technological and economic information and services required for the improvement, protection and efficient utilization of the forest resources of the province of Newfoundland and Labrador. Its work is directed towards satisfying the requirements of the Provincial Government, the forest industry, federal agencies having forest related programs and other clients of Forestry Canada; many of its studies are undertaken jointly with client agencies. The region attempts to bring the results of research to potential users by publishing information reports, articles in scientific journals and by demonstrations of research results. It seeks to keep the public informed of its work by means of special displays, seminars and related activities.

The regional program has three principal components - Silvics, Ecology and Economics Research, Forest Protection Research, and Forestry Development.

# GUIDELINES FOR EDITING PERMANENT SAMPLE PLOT DATA

by

M.B. Karsh and M.B. Lavigne

INFORMATION REPORT N-X-287
1993

FORESTRY CANADA
NEWFOUNDLAND AND LABRADOR REGION

**CANADIAN CATALOGUING IN PUBLICATION DATA**

Karsh, M.B.

Guidelines for editing permanent sample plot data

(Information Report N-X-287)
Includes bibliographical references.
ISBN 0-662-20856-0
DSS Cat. No. Fo46-15/287E

1. Conifers - Newfoundland - Growth.
2. Tree planting - Newfoundland.
3. Forest site quality - Newfoundland.

I. Lavigne, M.B.
II. Canada, Forestry Canada, Newfoundland and Labrador Region.
III. Title
IV. Series: Information report (Canada, Forestry Canada, Newfoundland and Labrador Region); N-X-287.

SD396.K37 1993 634.9'75 C93-099627-5



ISSN NO. 0704-7657
CAT. NO. Fo46-15/287E
ISBN NO. 0-662-20856-0

Copies of this report are available at no charge from:

Forestry Canada
Newfoundland and Labrador Region
P.O. Box 6028
St. John's, Newfoundland
Canada A1C 5X8

Telephone No. (709) 772-4117

**OR**

A microfiche edition of this publication may be purchased from:

Micromedia Inc.
Place du Portage
165, Hôtel-de-Ville
Hull, Quebec
J8X 3X2

Printed in Canada

## ABSTRACT

Guidelines for editing data were developed for use with spacing trial data but they should have wider application. A logical sequence for finding and correcting errors in the data is described, with examples of how to correct data at each step of its sequence. Once the errors in the data are identified, those records are separated for correction at a later stage. The remaining data is assumed to be correct and is used to calculate the equations that were subsequently used to make corrections.

## RÉSUMÉ

Des principes de vérification de données, élaborés pour vérifier des données sur des essais d'espacement, pourraient aussi convenir à des applications plus larges. L'étude présente une suite logique permettant de déceler et de corriger les erreurs dans les données, ainsi que des exemples de la façon de corriger ces erreurs à chaque étape de la suite. Une fois que des erreurs de données ont été, reconnues, les enregistrements sont écartes pour etre corrigés plus tard. Les données restantes, présumées corrects, servent au calcul des équations utilisées ultérieurement pour faire les corrections.

## TABLE OF CONTENTS

## LIST OF TABLES



## LIST OF FIGURES

## ACKNOWLEDGEMENTS

The authors thank J.G. Donnelly for establishing the spacing trials and for providing tally sheets and suggestions on editing procedures. We are grateful to D. Harris for giving us data disk and input into guideline development. M. Osmond organized the collection of re-measurement data and provided additional comments on field conditions. Corner Brook Pulp and Paper and Abitibi-Price were responsible for spacing the trials.

# GUIDELINES FOR EDITING PERMANENT SAMPLE PLOT DATA

by

M.B. Karsh and M.B. Lavigne

## INTRODUCTION

In any large permanent sample plot program there are errors in the data returned from the field that must be detected. Moreover, it is not feasible to immediately return to the plots to rectify mistakes and therefore temporary corrections must be made in the office before analysing the data. Data from 111 plots of precommercially thinned spacing trials were prepared for analysis. We report on the procedures used for editing the data because there was a lack of helpful information in the literature and because the approach developed should be applicable to many permanent sample plot programs.

The editing sequences described refer to an initial set of observations with one set of re-measurements. If there are more than two sets of observations then the two most recent sets of measurements are compared until all adjacent sets of measurements are examined.

## SAMPLE DATABASE

This report deals with data on computer and assumes that double entry of data was done and that errors arising from key punching have been dealt with. The 111 plots were in seven spacing trials, located throughout the Island of Newfoundland (Donnelly *et al.* 1986). Three trials are in balsam fir [*Abies balsamea* (L.) Mill] stands, three trials are in black spruce [*Picea mariana* (Mill.) B.S.P.] stands and one trial is in a jack pine (*Pinus banksiana* Lamb.) stand. Results of the black spruce and balsam fir spacing trials have been reported on by Lavigne *et al.* (1987) and Lavigne and Donnelly (1989).

Heights, breast-height diameters, base of live crown, status and damage were recorded for every tree in a plot. All trees were tagged when plots were established. Crews used tally sheets with records of previous measurements when doing re-measurements.

## FINDING ERRORS

The sequence of steps followed to correct permanent sample plot data is shown in Figure 1. It was useful for sorting and clarity to convert the alphanumeric information used for species, tree condition and tree status to numeric codes (Table 1). The trees having incomplete records, negative increment of height or diameter or whose height-diameter relationship deviated beyond a threshold at either the first or second measurement were identified and separated for correction at a later stage (Figure 1). The remaining trees were assumed to be correct and were used to calculate

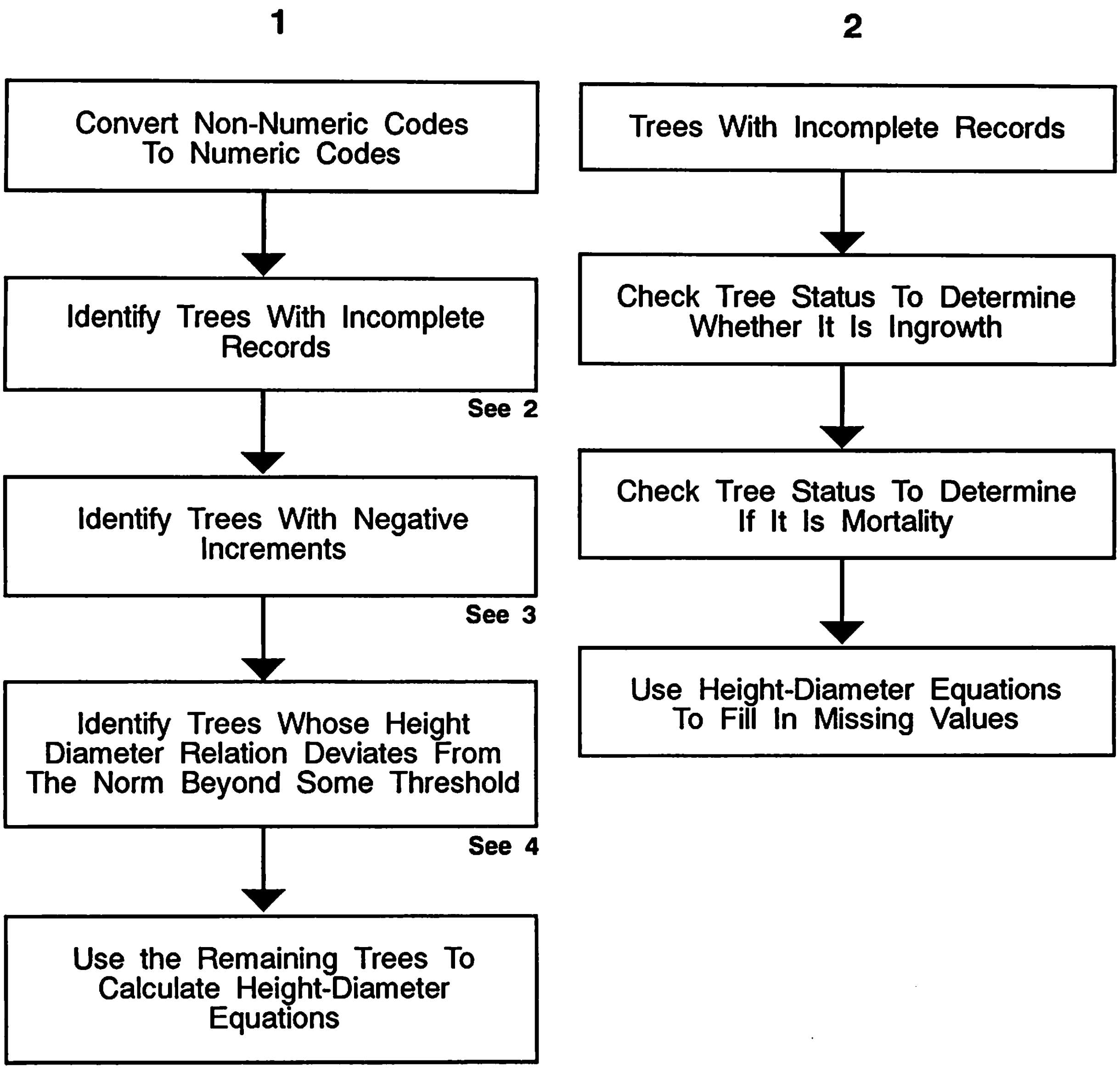



Figure 1. Sequence of steps to correct data.

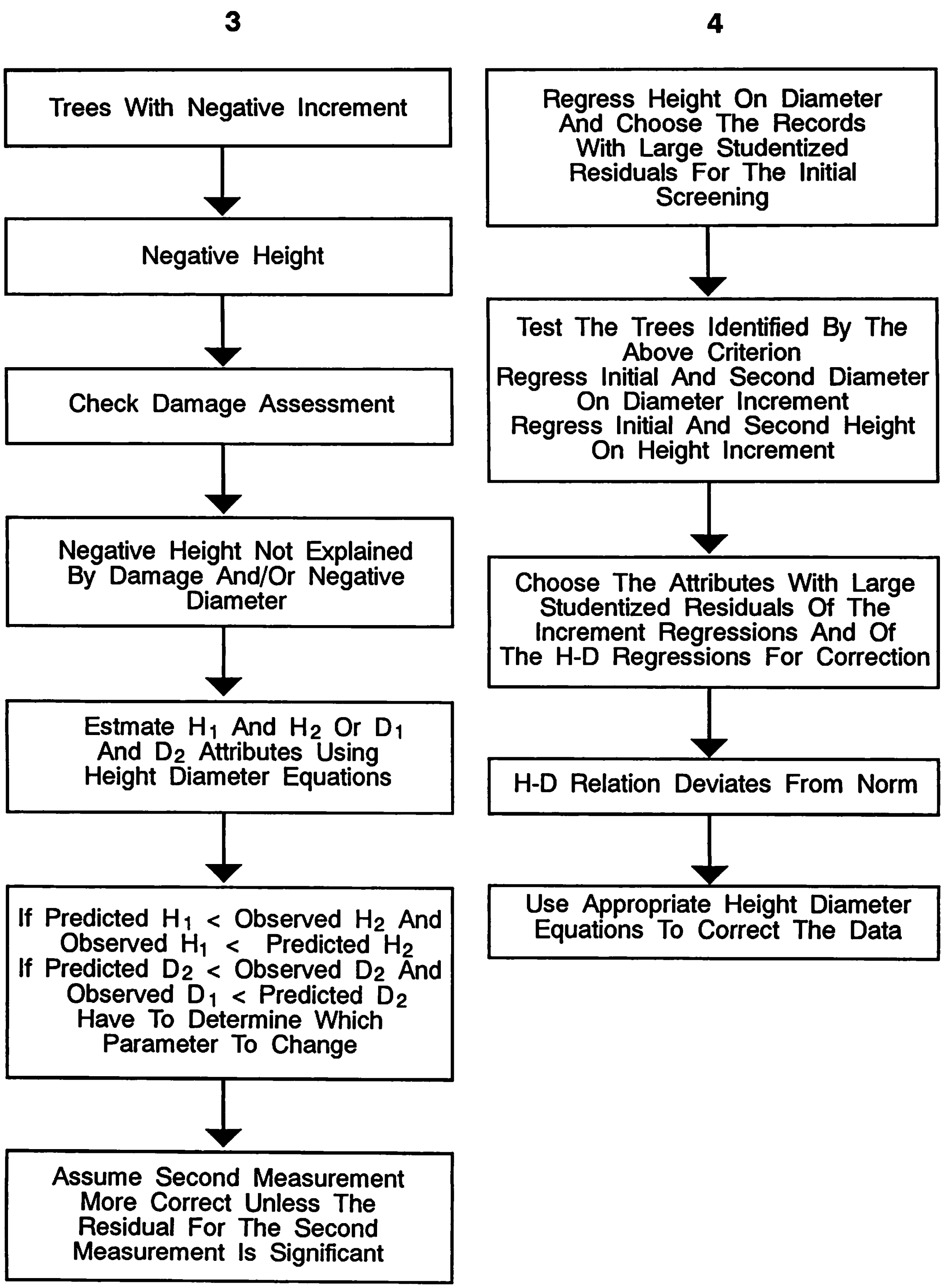



Figure 1. Concl'd.

Table 1. Numeric codes for status, class and damage categories.

| CODE | STATUS | CODE | CLASS | CODE | DAMAGE |
|---|---|---|---|---|---|
| 1 | Living | 1 | Ingrowth | 0 | None |
| 2 | Dead | 2 | Established | 1 | Multiple Leadering |
| 3 | Not Found | | | 2 | Stem Broken |
| 4 | Blowdown | | | 3 | Leaning |
| 5 | Missed | | | 4 | Crooked |
| 6 | Tag | | | 5 | Bent or Bent Stem |
| 7 | Retagged | | | 6 | Browsed |
| | | | | 7 | Browsed Lower Branches |
| | | | | 8 | Severely Browsed |
| | | | | 9 | Broken Top |
| | | | | 10 | Top Dead |
| | | | | 11 | Swelling Near Tag |
| | | | | 12 | Defoliated |
| | | | | 13 | Dead Leader |
| | | | | 14 | Forked |
| | | | | 15 | Broken |
| | | | | 16 | Deformed |
| | | | | 17 | New Leader |
| | | | | 18 | Lateral Takeover |
| | | | | 19 | Cut |
| | | | | 20 | Top Cut |
| | | | | 21 | Double Top |
| | | | | 22 | Suppressed |
| | | | | 23 | Dying |
| | | | | 24 | Yellow |
| | | | | 25 | Painted |
| | | | | 26 | Prostrate |
| | | | | 27 | Missing Leader |
| | | | | 28 | Cracked Stem |
| | | | | 29 | Insect Damage |

height-diameter equations that were subsequently used to make corrections.

To identify records whose height-diameter relationship of either the first or second measurement differed significantly from the norm the residuals of a regression of height on diameter were examined (Figure 1, part 4). A simple linear equation was adequate for our data (Figure 2), but other equations might be more appropriate with different data. Systat (V5.0) was used for the statistical analysis (SYSTAT INC. 1990).

Studentized residuals were selected because they have been used to detect outliers with good results (Draper and Smith 1981, Gauch 1984, Wilkinson 1990). A residual is the difference between the observed and the predicted value and a studentized residual is the residual divided by the standard error of estimate. A studentized residual exceeding 1.5 was sufficiently sensitive for identifying all the records requiring consideration as outliers. Trees having significant residuals in both measurement years were assumed to be part of the natural variability and therefore did not require correction (Figure 1, part 4).

Screening with studentized residuals was the first step in identifying trees with aberrant measurements. The trees identified by this criteria were tested for residuals of the relationship between increment and initial diameter or height, or final diameter or height before deciding to make a change. For the second step diameter and height increment were regressed against the initial and second measurement of diameter and height; $\Delta d = b_0 + b_1 d_1$, $\Delta d = b_0 + b_1 d_2$, $\Delta h = b_0 + b_1 h_1$, $\Delta h = b_0 + b_1 h_2$ and those records with large studentized residuals of one of these regressions, and of the height-diameter relationship were marked for correction (Figure 1, part 4).

## MAKING CORRECTIONS

### Correction Equations

Records that passed the screening process (Figure 1) were used to fit equations for correcting observations with errors. For each measurement the height-diameter relationship was fit to a straight line.

### Incomplete Records

When the entire first measurement was missing and the tree status was legitimately ingrowth, no correction was done (Figure 1, part 2). Trees that were too large at the second measurement to be legitimately described as ingrowth did require correction and the equations $\hat{h}_1 = b_0 + b_1 h_2 + b_2 d_2$ and $\hat{d}_1 = b_0 + b_1 h_2 + b_2 d_2$ were used to complete the record (Figure 3a). The coefficients were calculated using the "good" trees in a plot.

When the entire second measurement was missing and the tree status was dead, no correction was required (Figure 1, part 2). The equations $\hat{h}_2 = b_0 + b_1 h_1 + b_2 d_1$ and $\hat{d}_2 = b_0 + b_1 h_1 + b_2 d_1$ were used to estimate the missing data of the living trees.

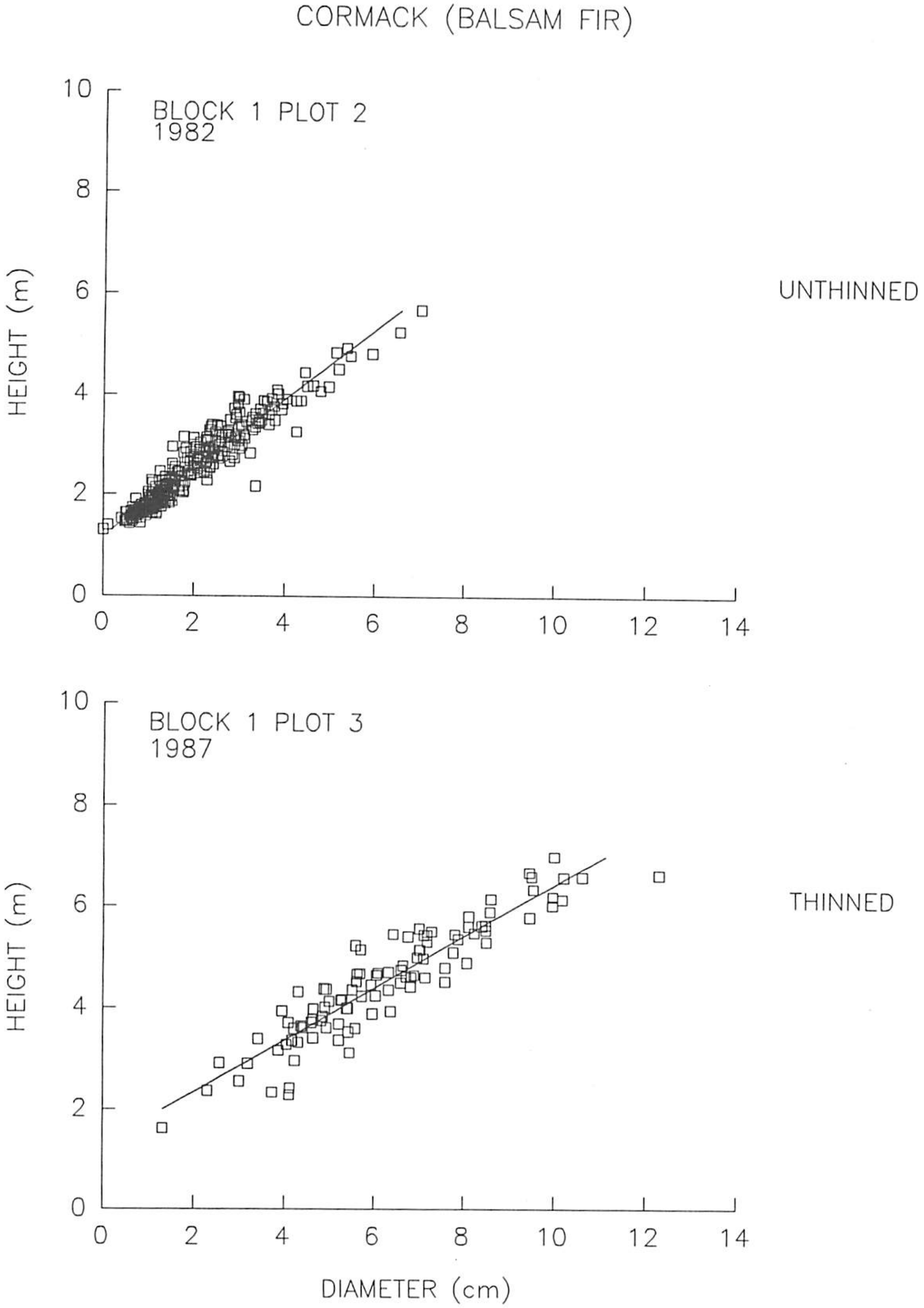



Figure 2. Height diameter relationship for (a) an unthinned plot and (b) a thinned plot of the spacing trial in a balsam fir stand in western Newfoundland.

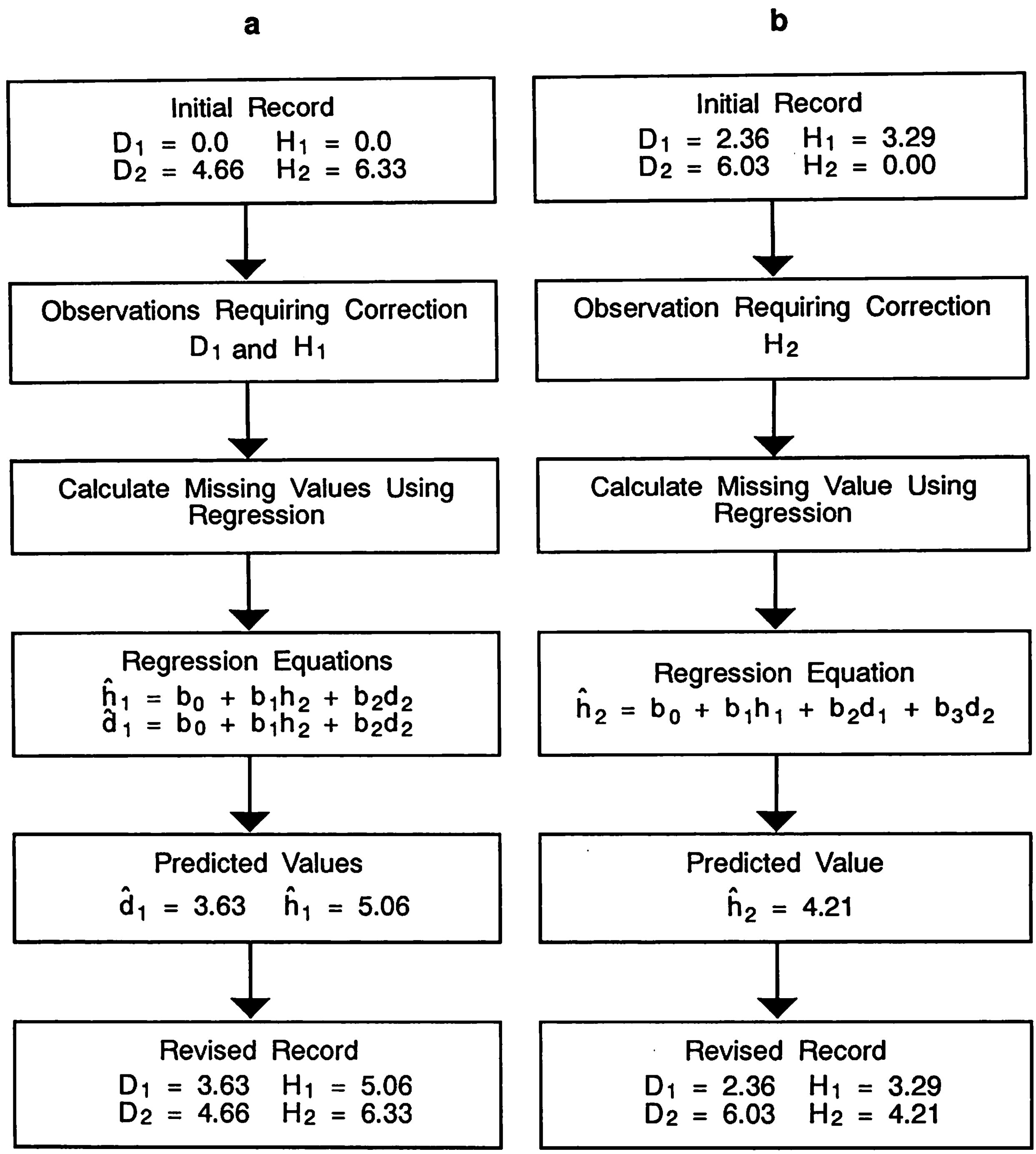



Figure 3. Making corrections for a tree with (a) missing initial measurements and (b) one missing second measurement.

If only one measurement of height or diameter was missing we assumed that the other three measurements were correct and used the equations $\hat{h}_1 = b_0 + b_1h_2 + b_2d_1 + b_3d_2$ and $\hat{d}_1 = b_0 + b_1d_2 + b_2h_1 + b_3h_2$, $\hat{h}_2 = b_0 + b_1h_1 + b_2d_1 + b_3d_2$ and $\hat{d}_2 = b_0 + b_1d_1 + b_2h_1 + b_3h_2$ to estimate the required parameters (Figure 3b).

__Negative Increments__

No correction was needed when a negative height increment could be explained by damage that occurred in the interval between measurements. If the tree was leaning a correction was required because the procedures used to measure height of leaning trees at the first measurement differed from the procedures used at the second measurement. If the second height for a leaning tree was less than the initial height then we assumed zero increment and let $h_2 = h_1$. Trees having negative height increments that could not be explained by damage required correction.

A negative diameter increment could occur if the second measurement was made at a different location from the first measurement as a result of some sort of damage to the main stem. Instances where this happened were noted in the damage assessment. All trees with negative diameter increments required correction.

For a negative height increment the two parameter equations $\hat{h}_1 = b_0 + b_1d_1 + b_2d_2$ and $\hat{h}_2 = b_0 + b_1d_1 + b_2d_2$ were used to estimate $h_1$ and $h_2$ (Figure 4). If $\hat{h}_1 < h_2$ __and__ $h_1 < \hat{h}_2$ then it was necessary to determine which parameter to change. We assumed that the second measurement was more correct, because the trees were larger and field staff had previous measurements to refer to, unless the residual for the second measurement was significant (Figure 1, part 3).

For a negative diameter increment the two parameter equations $\hat{d}_1 = b_0 + b_1h_1 + b_2h_2$ and $\hat{d}_2 = b_0 + b_1h_1 + b_2h_2$ were used to estimate $d_1$ and $d_2$. If $\hat{d}_1 < d_2$ __and__ $d_1 < \hat{d}_2$ then we had to determine which parameter to change. We assumed that the second measurement was more correct, because the trees were larger and field staff had previous measurements to refer to, unless the residual for the second measurement was significant.

For both a negative height and diameter increment the two parameter equations $\hat{h}_1 = b_0 + b_1h_2 + b_2d_2$ and $\hat{d}_1 = b_0 + b_1h_2 + b_2d_2$, $\hat{h}_2 = b_0 + b_1h_1 + b_2d_1$ and $\hat{d}_2 = b_0 + b_1h_1 + b_2d_1$ were used to estimate the required parameters. Deviation between predicted and observed were calculated and put on a relative scale by dividing by the respective standard errors of estimation to obtain the residuals.

__Trees with Aberrant H-D Relationship__

To correct a tree with aberrant measurements, height was first regressed on diameter. The studentized residuals were examined to identify the year with possible aberrant relations (Figure 5). To determine whether the height or diameter required correction

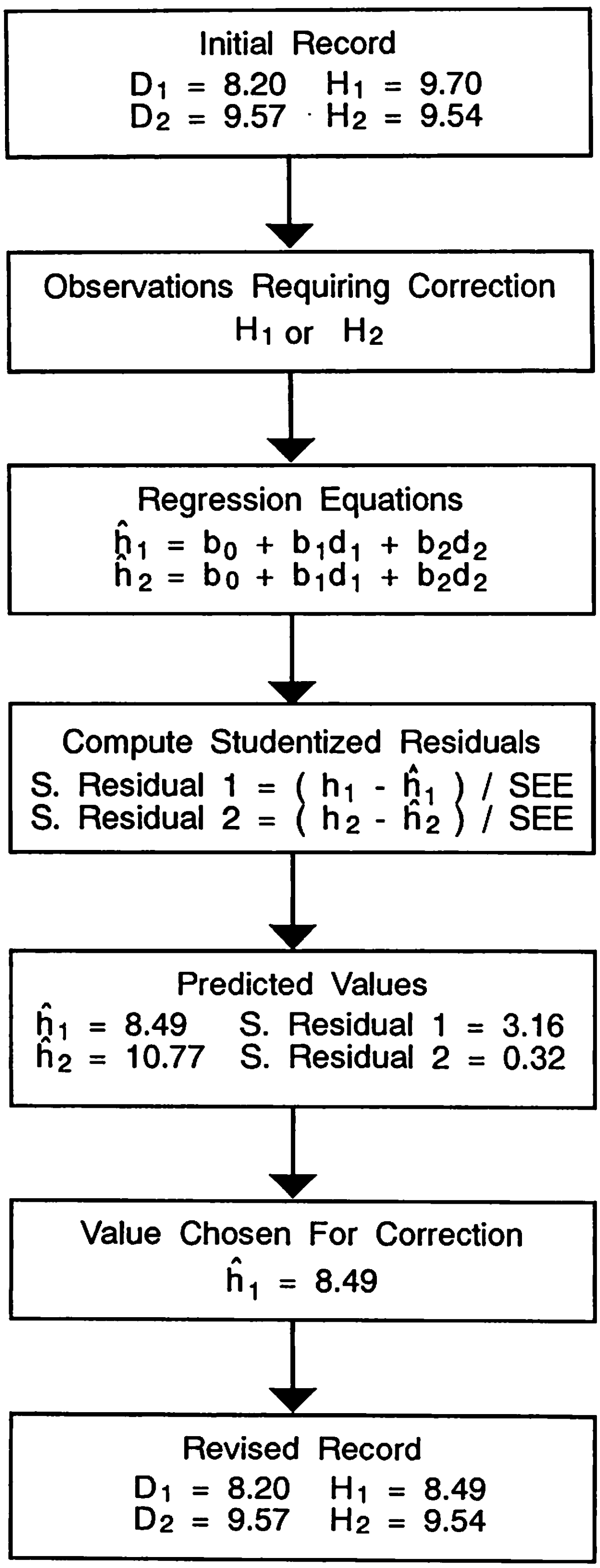



Figure 4. Making corrections for a tree with a negative height increment.

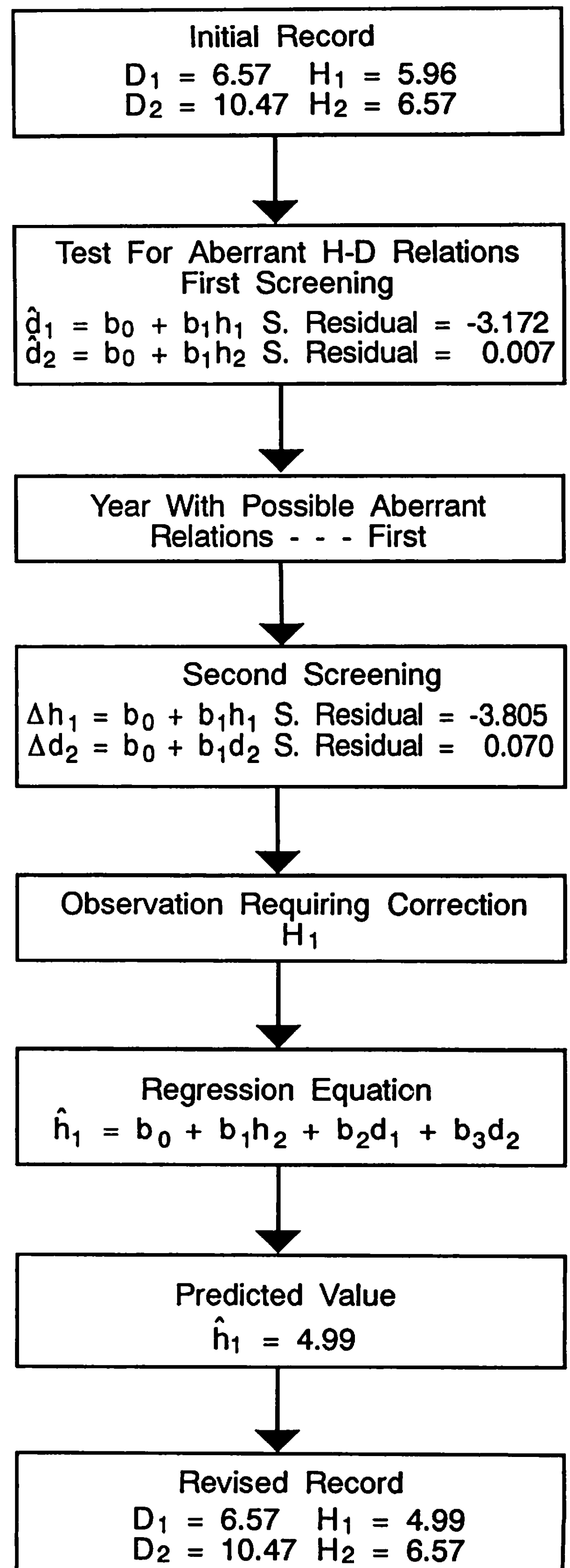



Figure 5. Making corrections for a tree with aberrant height-diameter relations.

we calculated increments for the trees in question, and compared the residuals of the attributes from the increment equations. The attribute having the largest studentized residual was chosen for correction. Sometimes both height and diameter required correction if their studentized residuals exceeded 1.5. The three and two parameter models were used for correcting the attributes as required (Figure 5).

### Special Cases

If the regression procedures were inadequate for very suppressed trees we assumed that there was zero increment and set both sets of measurements equal, giving preference to the second set of measurements.

When missing measurements were combined with ingrowth status one of the attributes was estimated using a one parameter model and then a predicted value was substituted to estimate the other attribute.

For low abundance species it may be necessary to combine plot data if the trial consists of numerous plots. When the number of species occurred in sufficient numbers they were always corrected on a plot by plot basis.

### File Maintenance

Corrections were made to the original computer file with the tree measurements for the plot. A column was added for each measurement year to this file to mark trees with changed values. In this column codes were used to indicate what attribute was changed and why it was changed (Table 2). These codes will be used to make lists of trees that require checking the next time field staff visit the plot.

## RESULTS AND DISCUSSION

Using these procedures approximately two plots of 100 trees were corrected per day. To speed up the process a comprehensive computer program could be written to do many of the editing procedures.

The majority of the errors occurred for records with unexplained negative increments (Table 3). Of the 111 plots the largest percent of trees with missing measurements and negative increments that had to be corrected was 17% and 22% respectively (Table 4). The change in estimated volume increment is shown in Table 5.

The editing sequences were tested on spacing trial data but should be applicable to other types of Permanent Sample Plot data. The editing sequences were capable of finding and correcting measurement and recording errors, missing measurements and negative growth increments. The approach taken to correct the data was unique because all missing or incorrect measurements were first identified and then the remaining trees were used to correct these records.

We highly recommend the use of the status, class, damage and leaning categories in any Permanent Sample Plot or re-measurement program.

Table 2. Description of changes made to attributes.

| Code | Description |
|---|---|
| 0. | No change |
| 1. | Diameter measurement missing |
| 2. | Height measurement missing |
| 3. | D and H measurements missing |
| 4. | Diameter changed to correct negative diameter increment |
| 5. | Height changed to correct negative height increment |
| 6. | D and H changed to correct negative increments |
| 7. | Diameter measurement exceed threshold |
| 8. | Height measurement exceed threshold |
| 9. | D and H measurement exceed threshold |
| 10. | Height measurement changed to correct leaning status |
| 11. | H changed because tree was very suppressed (regression inadequate) |
| 12. | D changed because tree was very suppressed (regression inadequate) |
| 13. | D and H changed because tree was very suppressed (regression inadequate) |
| 14. | Species code changed, previously unknown code |
| 15. | Class code changed to ingrowth (less than 1.3 m) |

Table 3. Summary of corrections.

| | |
|---|---|
| No. of trees | 17 037 |
| No. of trees with missing measurements | 1 775 |
| No. of trees with missing measurements explained by status or class | 1 450 |
| No. of trees with missing measurements that were corrected | 325 |
| No. of trees with negative increment | 793 |
| No. of trees with negative increment explained by damage | 74 |
| No. of trees with negative increment explained by leaning status | 89 |
| No. of trees with negative increment that were corrected | 720 |
| No. of trees with aberrant H-D relations that were corrected | 466 |
| No. of corrected records | 1 511 |

Table 4. Plot to plot variability of corrections for the 111 plots.

| | Min. | Max. |
|---|---|---|
| % of trees on plots with missing values that required correcting | 0 | 17.0 |
| % of trees on plots with negative increment that required correcting | 0 | 22.0 |
| % of trees with deviant H-D relation | 0 | 20.0 |

Table 5. Change in estimated stem volume increment resulting from editing.

| | Mean | Min. | Max. |
|---|---|---|---|
| volume increment of uncorrected data ($Dm^3$) | 5.15 | - 6.69 | 48.21 |
| volume increment of corrected data ($Dm^3$) | 5.20 | 0.00 | 45.47 |
| % difference | 0.92 | >100.00 | 6.03 |
| range | 0 - 27 | 0 - >100 | 0 - >100 |

Together they explained 63% of the missing measurements and negative increments. To reduce the error associated with unexplained negative increments field procedures should include more diligent use of previous measurements including investigations of apparent discrepancies.

## LITERATURE CITED

Donnelly, J.G.; M.B. Lavigne and R.S. van Nostrand. 1986. Precommercial thinning spacing trials established between 1979 and 1985. Can. For. Serv., Nfld. For. Centre. File Rep. Study 2401. 65p.

Draper, N. and H. Smith. 1981. Applied regression analysis. 2nd edition. John Wiley and Sons, Toronto. p.152-153.

Gouch, H.G. 1984. Multivariate analysis in community ecology. E. Beck, H.J.B. Birks and E.F. Connor (Eds.). Cambridge University Press, New York. 298p.

Lavigne, M.B., J.G. Donnelly and R.S. van Nostrand. 1987. A spacing trial of black spruce at North Pond: stemwood production during the first five years after thinning. Can. For. Serv., Nfld. For. Res. Centre., Info. Rep. N-X-262. 16p.

Lavigne, M.B. and J.G. Donnelly. 1989. Early results of a spacing trial in a precommercially thinned balsam fir stand in Western Newfoundland. Can. For. Serv., Nfld. For. Res. Centre., Info. Rep. N-X-269. 23p.

Wilkinson, L. 1990. Systat: The system for statistics. Evanston, IL: Systat, Inc. p.155.