UNIVERSITY OF COPENHAGEN

Københavns Universitet

Statistical Analysis of Data from Provenance Trials

Ræbild, Anders; Hansen, Christian Pilegaard; Kjær, Erik Dahl

Publication date:
2002

Document version
Publisher's PDF, also known as Version of record

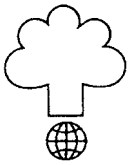
Citation for published version (APA):
Ræbild, A., Hansen, C. P., & Kjær, E. D. (2002). *Statistical Analysis of Data from Provenance Trials*. Danida Forest Seed Centre. Technical Note no. 63

Download date: 13. Nov. 2026

Statistical analysis of data from provenance trials

by

Anders Ræbild, Christian Pilegaard Hansen & Erik Dahl Kjaer



Guidelines & Technical Notes No 63

Danida Forest Seed Centre

September 2002

Citation:
A. Ræbild, C.P. Hansen & E.D. Kjaer. 2002. Statistical analysis of data from provenance trials. Guidelines & Technical Notes No.63. Danida Forest Seed Centre, Humlebaek, Denmark.

Reproduction is allowed with citation

ISSN 0902-3224

Cover photo:
Assessment of a species and provenance trial at Bebedouro, N.E. Brazil. Plot of *Prosopis pallida*. Assessment of biomass production by destructive sampling. Phot: Lars Graudal, DFSC. 1992.

This publication can be requested from:

Danida Forest Seed Centre
Krogerupvej 21. DK-3050 Humlebaek, Denmark
Phone: +45-49190500
Fax: +45-49190258
Email: dfsc@sns.dk
Web Site: www.dfsc.dk

and/or be downloaded from the DFSC homepage:
www.dfsc.dk/publications/

Technical Editor: Melita Jørgensen

Print:
Toptryk A/S, Graasten

Guidelines & Technical Notes' is a continuaton of the series of Technical Notes from Danida Forest Seed Centre.

DFSC publications are distributed free of charge

Danida Forest Seed Centre (DFSC) is a Danish non-profit institute which has been working with development and transfer of know-how in management of tree genetic resources since 1969. The development objective of DFSC is to contribute to improve the benefits of growing trees for the well-being of people in developing countries. DFSC's programme is financed by the Danish International Development Assistance (Danida).

Preface

This note describes the analysis of randomised block designs of provenance trials. It describes the methodology applied by DFSC in the analyses of several series of provenance trials, notably the *Pinus kesiya* and the Arid Zone series, and which has been applied in connection with the Neem network meeting in India 2001 and a workshop on analysis of Teak trials in Côte d'Ivoire 2002. It is intended to serve as an inspiration for the analysis of other trials involving multiple seedlots. Descriptions of the software programmes used for the analyses have not been included, since it would be too time- and space-consuming to include even the most commonly applied software packages. However, examples of the analyses made with the SAS® software are available from DFSC.

Acknowledgement

Several people have read and commented on various drafts of the note. In particular, we wish to thank participants at the workshops mentioned above, and Hubert Wellendorf, Jan Svejgård, Lars Graudal and Marcus Robbins.

Contents

Preface i
Acknowledgements i
Contents iii

1. Objectives of the note 1

2. Data entry and first check of data 2

3. Calculation of plot average values 3

4. Plotting the data 5

5. The statistical model 7

6. Verification of the model assumptions 8
   Residuals, Student's residuals and Cook's distances 8
   Independence 8
   Normality 9

7. What to do when the assumptions of the model are not fulfilled 13

   Independence 13
   Deviations from normality 13
   Variance heterogeneity 13
   Outliers 13

8. Test of differences between seedlots 15

9. Correction for multiple comparisons 16

10. Average performance of the seedlots: Least square means 17

11. Best linear unbiased predictors 19

12. When the trial is a mixture of different species 20

13. A note on the interpretation of provenance trials 21

14. References 22

1. Objectives of the note

The term "provenance" can be defined as the original geographic source of seed, pollen or propagules and usually refers to the original native source of a population. A population removed from its source and grown elsewhere may develop into a "land race" after growing for some generations at the new site (Schmidt 1997). The term "seedlot" is more general, since a seedlot is characterised by the site of the collection of the seed, and by the year of collection. For practical purposes we will in this note use "provenance" and "seedlot" as synonyms, using them to describe any genetic material of trees with a specific origin.

Often one of the first steps in the domestication of tree species is the establishment and analysis of provenance trials. This is because such trials may reveal whether trees with certain geographical origins are superior to others. Normally, the objectives of provenance trials and their analysis can be summarised in four points:

- to examine if there are any statistically significant differences between the seedlots;
- to estimate the differences between the tested seedlots;
- to identify superior seed sources in selected traits;
- to investigate more general patterns of genetic variation;

The note will only deal with the first three of the four objectives. Finding general patterns of variation between provenances or groups of provenances often involve sophisticated, multivariable analyses and is not described here.

The content of the note covers the sequence described in Fig. 1. The first step is to examine and treat the raw data, after which the model is formulated and applied. Emphasis is put on the verification of model assumptions, as this is an important subject often neglected. We present examples from trials where the model assumptions have not been fulfilled, and possible solutions to the problems are outlined. Finally, we describe the tests needed to examine the differences between the seedlots as well as the calculation of least square means and best linear unbiased predictors.

Assessment of a species and provenance trial at Bandia, Senegal, established in 1984. Plot of *Prosopis juliflora.* Phot: Lars Graudal, DFSC. 1993.

2. Data entry and first check of data

Once the field trial has been assessed, the data must be entered into a computer format. Usually spreadsheets are used for this purpose, depending on which software is used for the subsequent data analysis.

At this point it is important to look carefully at the data in order to find mistakes. For example, in trials where the height varies between 5 and 10 m, a height of 90 m for a tree is easily spotted as being wrong. Later, after average plot values have been calculated, it is much more difficult to find obvious mistakes such as diameters recorded in a wrong scale (e.g. cm instead of mm), or recorded values for stem form that do not correspond to the applied scoring scale. Typing errors that have occurred during entering of the data to the spreadsheet may also be identified.

Simple plots of the raw data are very useful in this respect, but data can also be checked by using automatic data verification in the spreadsheet or by searching for extreme values.

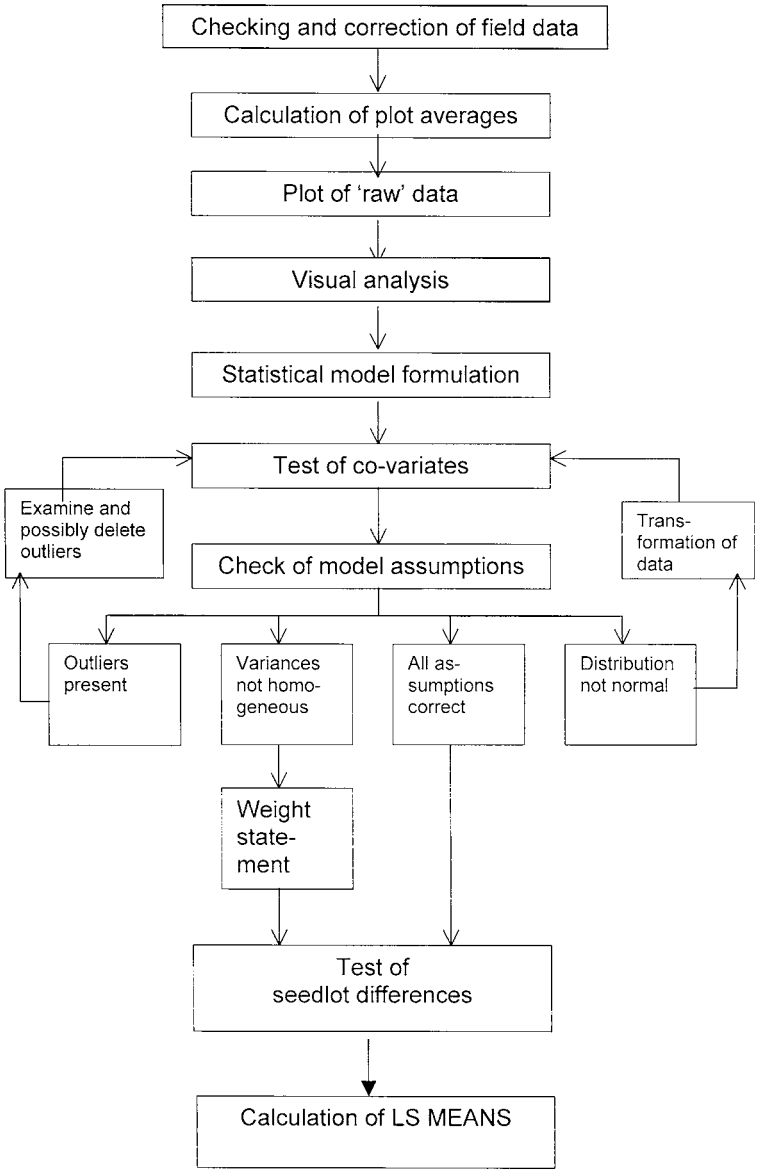


Figure 1. Steps involved in the analysis

3. Calculation of plot average values

For most analyses, the *plot average values* are used. This is because the growth of single trees in the plots cannot be analysed as independent observations, unless the trial was established with single tree plots.

The difference between single tree plots and multiple tree plots is apparent from Fig. 2 and Fig 3. In single tree plot designs, each seedlot is represented by only one tree in each block, whereas in multiple tree plots there are several trees of the same seedlot in each block.

Only true single tree designs can be analysed easily without calculating averages. Analysing multiple tree plot designs based on the single tree data requires the use of highly complex statistical models, and plot average data are as efficient as single tree data in terms of testing differences between seedlots and calculating average values. It is a common mistake to use single tree data where the average data should be applied. Doing this means that the tests performed will be invalid, and that the conclusions drawn may be wrong. Therefore we highly recommend the calculation of plot averages for designs similar to those presented in Fig. 3.

At this point it should be considered to try to eliminate border effects from trials with multiple tree plots. If one seedlot grows considerably faster than the others, it may influence trees in adjacent plots by shading, competition on nutrients, water etc. In such cases border effects can be reduced by removing the trees in the border of the plots from the data set, before calculation of plot averages. In this way the analysis includes only the central trees of the plots. In the fourth block in Fig. 3, the border trees and the central trees are separated from each other for illustration. In this case, however, omitting the border trees would mean a large loss of data, since only four out of 16 trees would be included.

One may distinguish between "simple" variables and derived variables. Simple variables include variables such as height and diameter. These are easy to deal with, because all that needs to be done is to calculate the mean value for each plot and use them directly in the analysis. However, knowledge may also be gained by using derived values such as survival and basal area.

For example, for analysing survival we have to introduce a variable that equals 1 if the tree is alive and 0 if it is dead. When the average is calculated, it will give the proportion of living trees. It should be noted that for all other traits, dead trees should be excluded. Putting a zero for e.g. height of a dead tree will bias the calculation of the plot average performance. Sometimes other variables are made, recording if a tree belongs to certain classes of a given trait (e.g. if the stem form is the best or second best class). Later such variables can be used to compare the seedlots in terms of percentage of trees in superior classes.

Basal area is calculated on the basis of tree diameter and is very useful when comparing the productivity of trees. The basal area expresses the area of a cross section of the stem, usually at 1.3 m. For shrubs and small trees, the cross section is sometimes measured at 0.3 m The formula for calculation of basal area is as follows:

$$Basalarea = \left(\frac{diameter}{2}\right)^2 \times 3.14$$

For trees with multiple stems it may be calculated as the sum of the basal area of all assessed stems. The advantage of basal area compared to diameter is that basal area is better correlated with the production of wood (volume or dry mass).

Basal area can be expressed as the average of the surviving trees, or it can be expressed as the sum of all trees on the surface of the plot, i.e. in m^2 ha^{-1}. The two values have different interpretations: Whereas the sum per area is a measure of the real production of the provenance at the site, the average basal area of surviving trees can be interpreted as the potential production if there was no problems with survival of the trees. In cases where survival is 100 % the two values will show the same ranking of the provenances.

When the averages and plot sums have been calculated, the new data should be checked before further analysis. This may help in identifying mistakes in formulas used for calculation, and give an impression of the variation within the trial. Plotting the values for each seedlot often reveals clear differences between the seedlots or gives hints regarding the proper transformation of the data. Obvious outliers (extreme values, see below) may also be identified at this stage. In the next section we show two examples of such plots.

Block I		Block II		Block III		Block IV	
4	2	10	7	5	2	3	2
3	10	2	8	4	8	9	5
8	7	5	3	1	10	6	8
9	6	1	6	7	3	7	1
1	5	9	4	6	9	4	10
Block V		**Block VI**		**Block VII**		**Block VIII**	
6	4	2	4	1	5	6	3
7	3	7	10	4	3	2	8
2	10	1	6	2	10	7	5
9	8	9	8	8	7	1	9
5	1	3	5	9	6	4	10

Figure 2. A single tree design with 8 replications (blocks). There are ten seedlots, so that each seedlot is represented by eight trees. On the diagram, each tree is represented by its seedlot number. In each block, only one tree of each seedlot is present. For this type of trial the single tree data set may be analysed directly.

Block I								Block II							
2	2	2	2	1	1	1	1	3	3	3	3	1	1	1	1
2	2	2	2	1	1	1	1	3	3	3	3	1	1	1	1
2	2	2	2	1	1	1	1	3	3	3	3	1	1	1	1
2	2	2	2	1	1	1	1	3	3	3	3	1	1	1	1
4	4	4	4	3	3	3	3	2	2	2	2	4	4	4	4
4	4	4	4	3	3	3	3	2	2	2	2	4	4	4	4
4	4	4	4	3	3	3	3	2	2	2	2	4	4	4	4
4	4	4	4	3	3	3	3	2	2	2	2	4	4	4	4
Block III								**Block IV**							
3	3	3	3	1	1	1	1	2	2	2	2	4	4	4	4
3	3	3	3	1	1	1	1	2	2	2	2	4	4	4	4
3	3	3	3	1	1	1	1	2	2	2	2	4	4	4	4
3	3	3	3	1	1	1	1	2	2	2	2	4	4	4	4
4	4	4	4	2	2	2	2	3	3	3	3	1	1	1	1
4	4	4	4	2	2	2	2	3	3	3	3	1	1	1	1
4	4	4	4	2	2	2	2	3	3	3	3	1	1	1	1
4	4	4	4	2	2	2	2	3	3	3	3	1	1	1	1

Figure 3. A small multiple tree plot design with four seedlots and four replicates (blocks). In each replicate, all seedlots are represented by sixteen trees, planted in a plot. Again each seedlot number represents one tree. The borderlines are indicated by solid lines, whereas the limit between plot border trees and the central trees are indicated by dotted lines (block 4 only). For this type of trial the data for each block should be averaged before analysis.

4. Plotting the data

Before the statistical analysis starts the data should always be examined by making a plot of the data. A graph of the data clearly indicates the variation in the trial, and it may be possible to make some conclusions already at this stage. Even though data should always be checked statistically, it may be argued that if it is not possible to carry out a statistical analysis, choice of provenances could be based on such preliminary visual analyses.

The trial that we will use as an example was established and managed by EMBRAPA/CPATSA of Brazil (Ræbild, Graudal & Lima *in prep.*). Located at Bebedourou, Brazil, the trial is a randomised complete block trial with four blocks. The data presented are the plot values plotted for each provenance. At the same time symbols indicate the block numbers (Fig. 4 and 5).

There is a rather large variation in the survival of the provenances (Fig. 4). In particular, the provenances Rajasthan01 and Rajasthan02 have low values, whereas the three provenances to the right of the diagram have relatively high values. Within the provenances, the variation between the different blocks is large, but there is no overlap between the two groups of provenances mentioned above. Furthermore, it appears that survival is almost always less in some blocks than in others. For example, block no. 4 is with one exception always among the blocks with the lowest survival. This demonstrates that a large part of the variation seen within the provenances is due to differences in the environment, and that if the environmental conditions had been more uniform, the variation within each provenance would have been smaller. The conclusion is that there are clear differences between the provenances in the trial, and that the three provenances to the right of the diagram are the best with respect to survival. (A statistical test shows that this interpretation is correct, and that there are indeed highly significant differences between the provenances in survival).

The second variable that we consider is the dry weight of the mean tree. This variable represents the weight of the dry matter of the tree, expressed in kg per tree. Here the picture is less clear (Fig. 5). The value for the provenance Rajasthan01 in block no. 4 appears to be extreme and is a possible outlier. Apart from this, no particular block seems to have low or high values, and the differences between the provenances are less pronounced than for survival. The plot values overlap each other for all provenances, and our conclusion would be that there are only small differences between the provenances in this character. (This is again supported by the statistical tests, according to which there are no significant differences between the provenances).

In a simple analysis, the data may be presented and ranked according to the average values. When the trial is balanced (the number of replicates is the same for all provenances) the average will be the same as the least square means (see section 9) and will give the same ranking of the provenances. Where the number of replicates is different for different provenances (imbalanced trials), the two values may differ. The same applies where there are missing values, e.g. when all trees on a plot are dead. In these cases the least square means may give a different ranking of provenances, mainly because the blocks correct for differences in the environment. Except in extreme cases with heavy imbalance or a large variation within provenances this has only a minor impact, however.

Although one may criticise selection of provenances based on results that have not been tested statistically, it seems sensible to choose the provenances that have apparently the best performance. In the case above it would be rather easy to select the most promising provenances with regard to survival, whereas in the case of dry weight of the mean tree, it is more difficult, because differences are small. Nevertheless, from a practical point of view, the provenances with the best growth would usually be preferred, even if there were no statistical evidence of their superiority. In any case, using common sense is the most important, whether or not it can be backed up by statistical analysis.

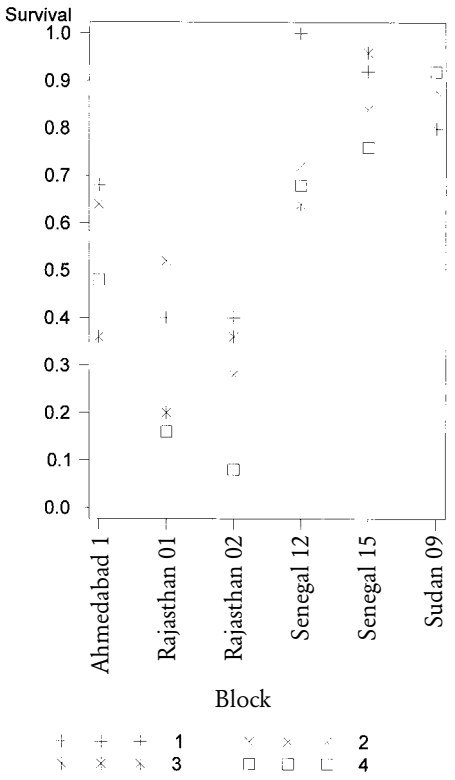


Figure 4. Survival (fraction) of six provenances of *A. nilotica* at Bebedourou, Brazil. Trial established by EMBRAPA/CPATSA

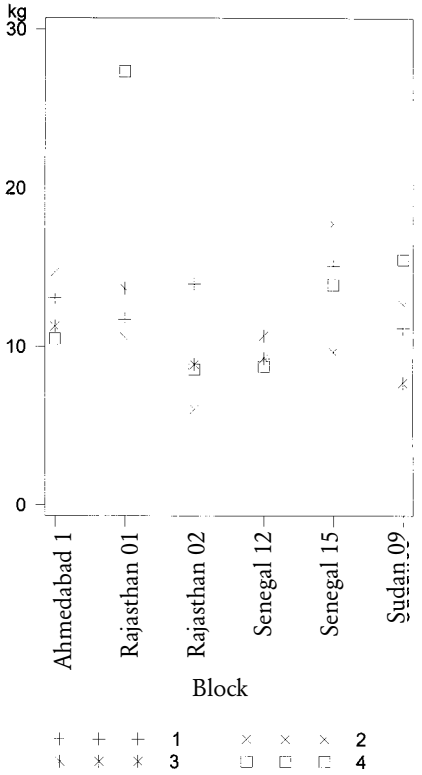


Figure 5. Dry weight of the mean tree for six provenances of *A. nilotica* at Bebedourou, Brazil. Trial established by EMBRAPA/CPATSA.

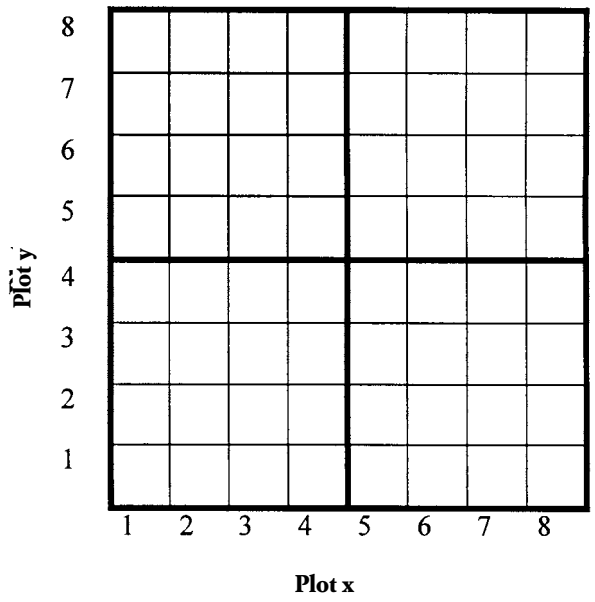


Figure 6. Plot x and plot y, illustrated in a randomised complete block design with four blocks.

5. The statistical model

The next step in the analysis is the formulation of the statistical model. The correct formulation of the model is crucial, because later analysis and thus the conclusions of the trial will depend on this model. In provenance trials with randomised block designs, the formulation will usually involve the following steps (cf. Fig. 1):

1. Select a simple model for two-way analysis of variance (randomised block design)
2. Test for systematic variation in e.g. North-South, or East-West directions (co-variates)
3. Verify model assumptions
4. If model assumptions are not fulfilled, consider the following options depending on the nature of the problem
   - Transformation of data.
   - Detection of outliers and their deletion.
   - Use of a weight statement.

This section describes the formulation of the model and the test of co-variates, whereas the following sections describe the control of the assumptions of the statistical model and methods to overcome incomplete assumption.

The first step, the formulation of the statistical model, will have to be based on the structure of the data that is analysed. For randomised block designs, the model usually looks like this:

$$X_{jk} = \mu + provenance_j + block_k + \varepsilon_{jk}$$

(Model 1)

where X_{jk} is the value of the trait in question (e.g. height) in plot *jk*,
μ is the grand mean,
$provenance_j$ is the effect of seedlot number *j* and is assumed to be a fixed or a random effect,
$block_k$ is the effect of block *k* in the trial, assumed to be a random effect, and
ε_{jk} is the residual of plot *jk* and is assumed to follow a normal distribution $N(0, \sigma_e^2)$.

For other types of trials, such as various types of single tree plots or split-plot designs, the model may be different. For trials with a mixture of species, a model is shown in section 11.

Whether the effects are fixed or random depends both on the design and on the purpose of the trial. In the conventional approach the provenances are considered as fixed effects, but in some cases they may also be considered random. For simple models such as model (1) the statistical tests will be the same irrespective of which approach is used, but there will be differences when it comes to calculation of estimates. We return to the subject in section 10.

The residuals in the model express the variation that is not explained by the provenance and block effects. If the residuals are small compared to the total variation in the trial, it means that the provenance and block effects explain a large part of the variation. Thus the smaller the residuals, the more likely it is that the effects become significant. From this comes the need to explain as much of the variation as possible.

Many older trials were established with fairly large blocks with heterogeneous ecological conditions such as variation in soil, elevation, slope and exposure. Heterogeneous conditions makes it more difficult to explain the variation by the block and provenance effects alone, which means that the residuals increase compared to homogeneous conditions. Therefore, the inclusion of the position of the plot within the block as co-variates sometimes improves the analysis by reducing the residual variation in trials. The use of co-variates implies that there is a linear relationship between the value of the co-variate and the value of the response variable (the trait that is being examined). The number of the plot in two orthogonal directions, the plot x and plot y values, can describe the positions within the block (see also Fig. 6):

plotx: Horizontal position of plot within trial (from the map of the trial);
ploty: Vertical position of plot within trial (from the map of the trial);

To catch non-linear patterns of site variation in the same directions, $plotx^2$ as well as $ploty^2$ may be applied. Other co-variates, such as the level (height of the plot surface above a reference point) may be included too.

When testing the effect of co-variates, we start with a model including all co-variates. Co-variates that are not significant are then removed successively by removing the least significant co-variate and running the model again until all remaining co-variates in the model are significant (e.g. $P<0.10$).

For each analysis, a number of standard plots are produced in order to control the assumptions of the model and see if the data are fitting the model (next section). Preference is given to "models that fit", i.e. where there is no systematic deviations in residuals. Also, simple models (with few co-variables) are preferred to complex models if they provide acceptable fits.

Small blocks with few trees per plot and/or few seedlots usually reduce the need for co-variates. As many trials have quite large blocks it is usually recommendable to check if co-variates will improve the analysis. More advanced modelling of the environmental variation, so-called "neighbour analysis" is described by Loo-Dinkins (1992).

6. Verification of the model assumptions

The analysis of variance rests upon a number of standard assumptions (Box et al. 1978). The key assumptions are:

1. That the residuals are independent;
2. That the residuals follow a normal distribution;
3. That there is variance homogeneity in effects included in the model.

In this section we describe how to ensure that the model assumptions are fulfilled.

Residuals, Student's residuals and Cook's distances
As explained above, the residuals represent variation that can not be accounted for by the model. For each observation, the model calculates a predicted value, taking into account the effects of the model (provenance, block and co-variates). The residual variation is then the difference between the observed value and the predicted value.

Student's residuals (also called 'standardised residuals') are calculated as the residual divided by its standard error. If the assumption of normal distributed residuals is valid, the Student's residuals have the property of a normal distribution with mean 0 and variance 1, meaning that 95% of the values should lie within ± 1.96. In trials with imbalance, the Student's residuals correct for imprecision due to low sample numbers, and in models with co-variates they compensate for large deviations at extreme values.

The Student's residual e_t for observation *ij* is given by

$$e_t = \frac{e_{ij}}{s_{ij}} = \frac{X_{ij} - P_{i\bullet} - B_{\bullet j}}{s_{ij}}$$

where e_{ij} is the residual, X_{ij} is the value for observation *ij*, $P_{i\bullet}$ is the effect of provenance *i*, $B_{\bullet j}$ is the effect of block *j*, and s_{ij} is the standard deviation (standard error) of observation *ij*. Probably all statistical software will calculate the Student's residuals as a standard procedure.

Cook's distance indicates how much the exclusion of a single observation would change the residuals on the model (Afifi & Clark 1996). A large Cook's distance indicates that the observation makes a substantial difference to the outcome of the model. Typically such observations should be considered with a good bit of scepticism. The calculation of Cook's distance is complicated, but is done as a standard in many statistical programmes.

These three variables are used in the control of the assumptions of the statistical model by producing a number of plots:

1. Student's residuals versus predicted values;
2. Cooks distance versus predicted values;
3. Student's residuals versus seedlot;
4. Frequency chart of residuals;
5. Student's residuals versus block;
6. Student's residuals versus distances along the axis of the trial (plot x and plot y).

Some examples of the graphs are presented in Fig. 7-10. If block, plot x and plot y have been included in the model, there is usually little information in the last two types of plots (see the previous section).

Independence
The assumption of independence means that the residual of one observation is not dependent on the residual of another. This assumption is typically violated when using pseudo-replicates, e.g. when doing more observations on the same experimental unit and treating them as different experimental units.

An example could be to measure height of the seedlots in several consecutive years and treat each year's observations as independent in the statistical model. Since the height of trees are dependent on the height in the previous year, it would be wrong to assume that they are independent. Instead, each year's observations should be analysed separately.

Violation against independence also takes place if the single trees in the plot are analysed as independent replications of the seedlot within the block (using a single tree model when the trial is not planned as a single tree trial). This can be avoided by analysing plot average values (section 3). The same problem occurs when two or more plots of the same seedlot within the same block are treated as independent observations. In such cases, an average of the values should be used as the plot value for the seedlot in question.

These problems are not revealed by the graphical check of the residuals, and there is no easy method to ensure that the condition of independence is fulfilled. A proper design and planning of the experiment and the application of correct statistical models remains the best means of obtaining independent observations.

However, the assumption of independence may also be violated if there is a time- or site-dependency in the data. This dependency is easily spotted. To check for site dependency, residuals are plotted against the horizontal and vertical axis of the trial (plotx and ploty) and where applicable, also the level of the plot, to investigate any systematic environmental variation. If the co-variates plot x and plot y have been included, it is rare to

find such environmental variation (see Fig. 7 for an example). Time-dependency may arise where there are changes in the assessment team during the assessment, e.g. when two persons interchange while doing diameter measurements or evaluating scores. If the persons have different ways of doing the measurements, it could be reflected in the residuals. In this case, a plot of the residuals against the time (day) of measurement may reveal the dependency between data.

Normality

The assumption of normality may be checked in various ways, graphically as well as by statistical tests. A simple frequency chart of residuals gives a good graphical check. The Student's residuals should in general be between +2.5 and –2.5 and only rarely exceed ±3. Also, they should follow a typical, bell-shaped, t-distribution and not have long tails on either side. Deviations from the assumption of normality may be caused by non-normal distributions (see example in Fig. 8), variance heterogeneity (Fig. 9) and outliers (Fig. 10).

There are various formal tests for normality. However, these tests are weak if the number of observations is low. On the other hand, when the number of samples is very large, formal tests for normality may become rejected even though the frequency chart of residuals appears to be normal. This is because the power of the test increases with the number of observations, and even small deviations from normality may result in rejection of the hypothesis of normal distributed residuals. What is important is an approximate normal distribution, not the exact normal distribution. Nevertheless, formal tests may serve as a guide in the evaluation of the residuals - if the hypothesis of a normal distribution of the residuals is rejected, a closer inspection of the residuals usually proves worthwhile.

When the number of observations is low, it becomes increasingly difficult to check the assumption of normality. Even though the frequency chart may show a rather odd and irregular distribution, this need not be a sign of non-normality. At small numbers of observations it is not unlikely that odd distributions may result from random variation, and unless the test for normality demonstrates that the assumption is violated, there is no need to reject the model.

Assessment of diameter/basal area at 0.3 m height in a provenance trial of *Acacia senegal*, Gonsé, Burkina Faso, established in 1988. Phot: Lars Graudal, DFSC. 1993.

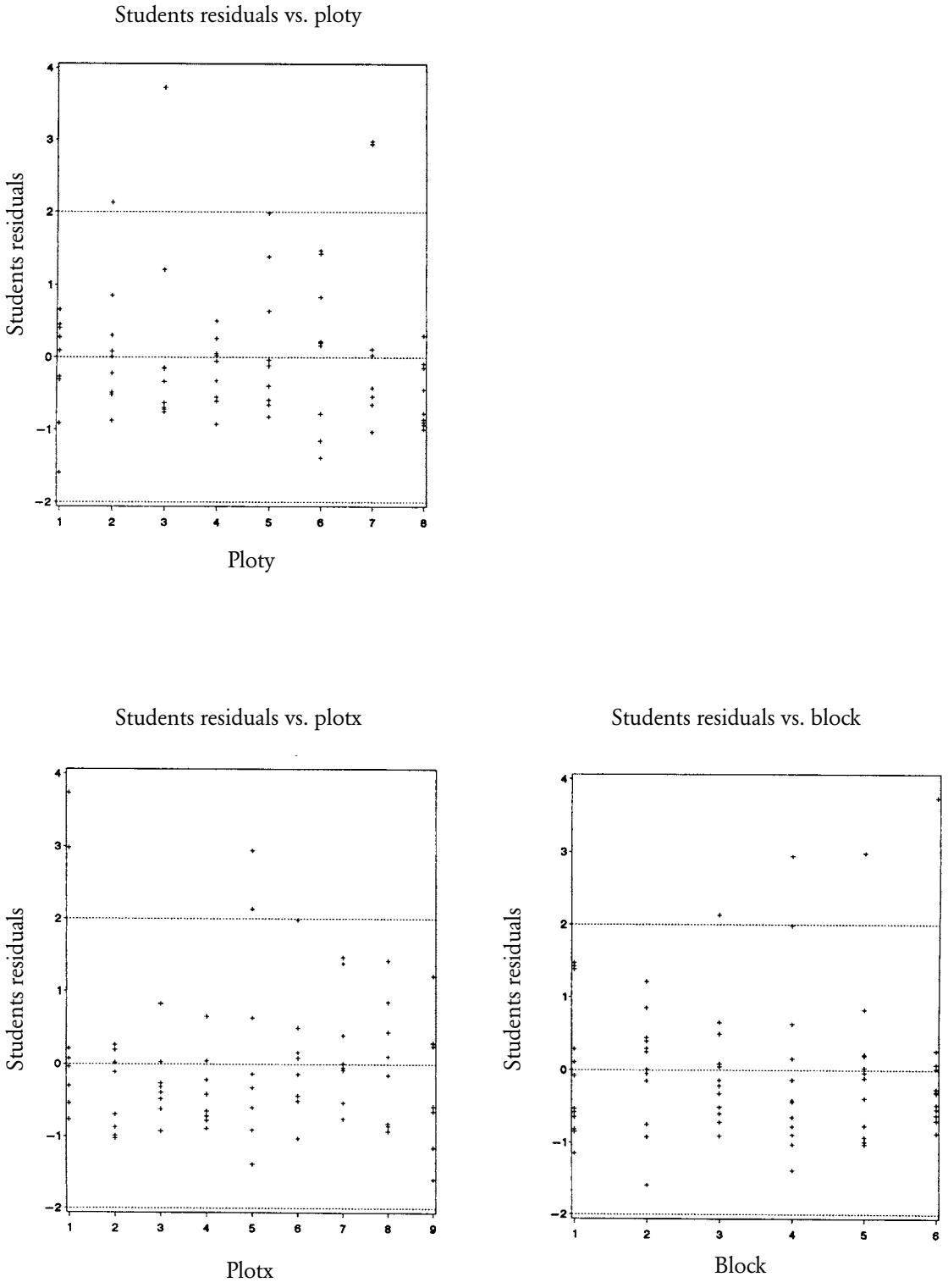


Figure 7. Student's residuals plotted against ploty, plotx and block. The Student's residuals should have a random distribution with no correlation between the residuals and plotx, ploty or the blocks. There are no signs of site-dependent residuals, but note that there appear to be a few outliers having Student's residuals above 3. The data are from an analysis of basal area per ha from a trial of *Acacia* seedlots from Gonsé, established by Centre National de Semences Forestières, Burkina Faso, and are the same as presented in Fig. 8 (Ræbild, Graudal & Ouedraogo *in prep. a*).

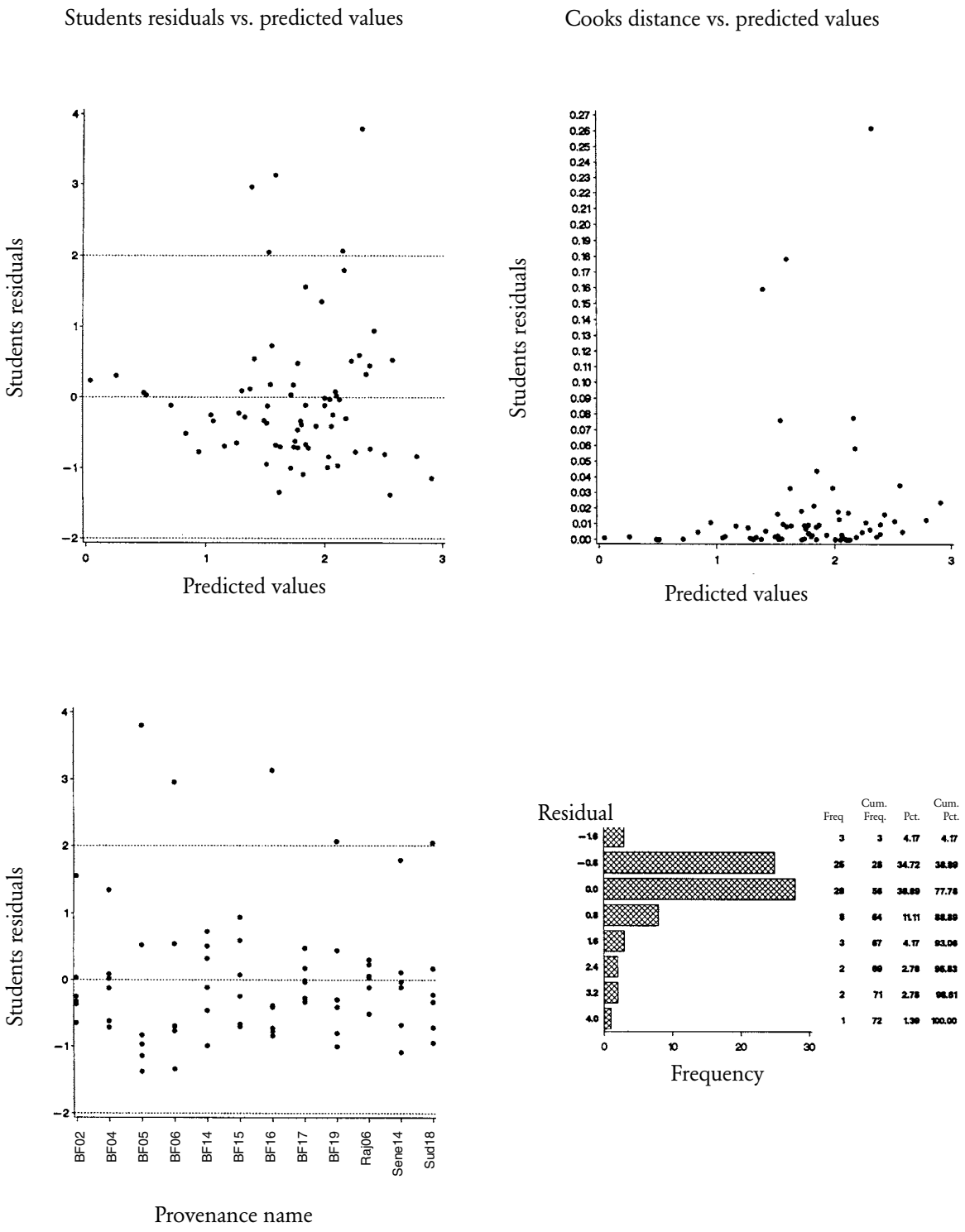


Figure 8. Residuals from the analysis of basal area per ha in a trial of *Acacia* seedlots from Gonsé, established by Centre National de Semences Forestières, Burkina Faso. The residuals are the same as presented in Fig. 7. The plot of Student's residuals versus predicted values show a funnel shape (increasing residuals with increasing predicted values). In this case the application of a logarithmic transformation is a proper solution. (Ræbild, Graudal & Ouedraogo *in prep. a*).

Variance heterogeneity

Variance heterogeneity occurs when different experimental units (blocks or seedlots) have different variance. A typical example is when the residuals of some seedlots appear very clustered in the diagram of Student's residuals versus seedlots, whereas the residuals of other seedlots are spread out, often with values of Student's residuals exceeding ±2.5. This may result from a simple scale effect (larger seedlots have larger variance), in which case the plots of Student's residuals and Cook's distance versus predicted values appear funnel-shaped (example in Fig. 8). It may also be related to the seedlot itself (some seedlots are more variable than others). In this case, the variance heterogeneity will be displayed in the plot of student's residuals versus seedlot (Fig. 9). Failure to correct variance heterogeneity means that the tests will give false results.

Outliers

Outliers are extreme observations that do not follow the trends of the rest of the data. Such observations may have a large influence on the estimates and statistical tests of the model and should therefore be considered carefully.

Outliers are detected by inspection of the plots of Student's residuals and of Cook's distance. Observations that have values of Student's residuals exceeding ±2.5 (cf. above) and observations with large values of Cook's distance are possible outliers and should be investigated further (Fig. 10). Outliers may be due to errors in the recording or typing of data or due to mislabelling of the seedlots in the nursery or in the field. Poor survival in the plot, leaving only a few trees to use in the calculation of plot means, is another source of outliers. Borderline effects may also cause outliers, in which case 2 or 3 outliers often have the same plot x or plot y value. However, it also happens that the outliers are due to some unexplained variation, perhaps in soil conditions or other environmental variation. Finally it should be mentioned that a large number of outliers might indicate that the distribution of residuals is not normal, and hence that a transformation of the data is required.

7. What to do when the assumptions of the model are not fulfilled

In many cases one or more of the model assumptions are not fulfilled. Below some procedures for correction are described.

Independence
Apart from making sure that the statistical design and the model is correct there is not much that can be done about dependence between observations. If a clear variation can be observed in the residuals, other co-variates than the ones mentioned above could be considered.

Deviations from normality
Usually deviations from normality are handled by transformation of data. Snedecor & Cochran (1980) and Afifi & Clark (1996) provide guidance on data transformations:

- Counts (of rare events) often follow a Poisson distribution and may be transformed with the square root. Examples within forestry are few, but could be e.g. the number of rare bird's nests within the tree, or the number of rare epiphytes per tree.
- Variables having a binomial character (e.g. dead or alive) summarised in a proportion (e.g. living trees in a plot) may be transformed with the arc sine transformation.
- If the standard deviation varies directly with size of the mean of the variable, a logarithmic transformation may stabilise the variance. This is often seen with a variable such as basal area.

There are theoretical reasons for choosing the above transformations (Snedecor & Cochran 1980), but it follows from Afifi & Clark (1996) that the range of transformations may be seen as a continuum and that various other transformations are available.

In many cases the analysis of survival data results in skewed distributions of the residuals, with tails at either the lower or upper end (many trees dead or many trees alive). In such cases an arc sine transformation of data will often prove useful. The arc sine transformation is given by

$$\arcsin(proportion) = \sin^{-1}\sqrt{proportion}$$

where proportion is a figure between zero and one (e.g. the surviving fraction of trees). An important property of the transformation is that the variance near zero or one is stretched out, thus facilitating the analysis of variance (Snedecor & Cochran 1980).

For many growth variables, the variance increases directly with tree size, and the proper transformation is thus the logarithm.

Variance heterogeneity
In the cases where the variance varies with the size of the variable, a transformation of data can be used (see above). However, in some cases the seedlots simply have different variances irrespective of size, and it is necessary to weight the observations. This is typically seen in the plots of Student's residuals versus provenance, when the residuals of some provenances show large variation and others show very little variation. The idea behind weighting is to give large weights to observations from provenances that are determined with great precision, and give small weights to observations from provenances that have a large variation (cf. Afifi & Clark 1996).

Weighting occurs in the following sequence: An ordinary analysis of variance of the variable is performed. The residuals from this analysis are grouped according to seedlot, and the variance of the residuals for each seedlot is calculated. The inverse of these variances are then used as weights in an analysis of variance. When calculating the sums of squares in the model, the weights are multiplied with the squared value of the deviance of each observation from the predicted value. This has the effect that seedlots with small variances have a larger influence on the model than seedlots with larger variances.

Outliers
The proper way to deal with outliers is to consider them carefully, and if there are obvious reasons for their extreme nature, to exclude them. However, this is often a difficult decision. When outliers occur as a result of errors, the data set should of course be corrected. It is less obvious what to do in those cases where there are no easy explanations. Outliers should only be excluded if it can be justified, i.e. an explanation can be given (Fig. 10). In a few cases explanations can not be given, and observations have to be excluded on basis of the extreme nature of the value alone. Great care is required in the decision to exclude plot values, as it will have a large influence on the results of the analysis, especially with few blocks (replications). Running the analysis again without the outlier(s) gives an indication of the sensitivity of the analysis with regard to the outliers, and assists in deciding whether to keep or delete the extreme observation(s). In this respect it is an advantage if the trials include a larger number of replications, e.g. 6-7 blocks or more, because the handling of outliers is much easier.

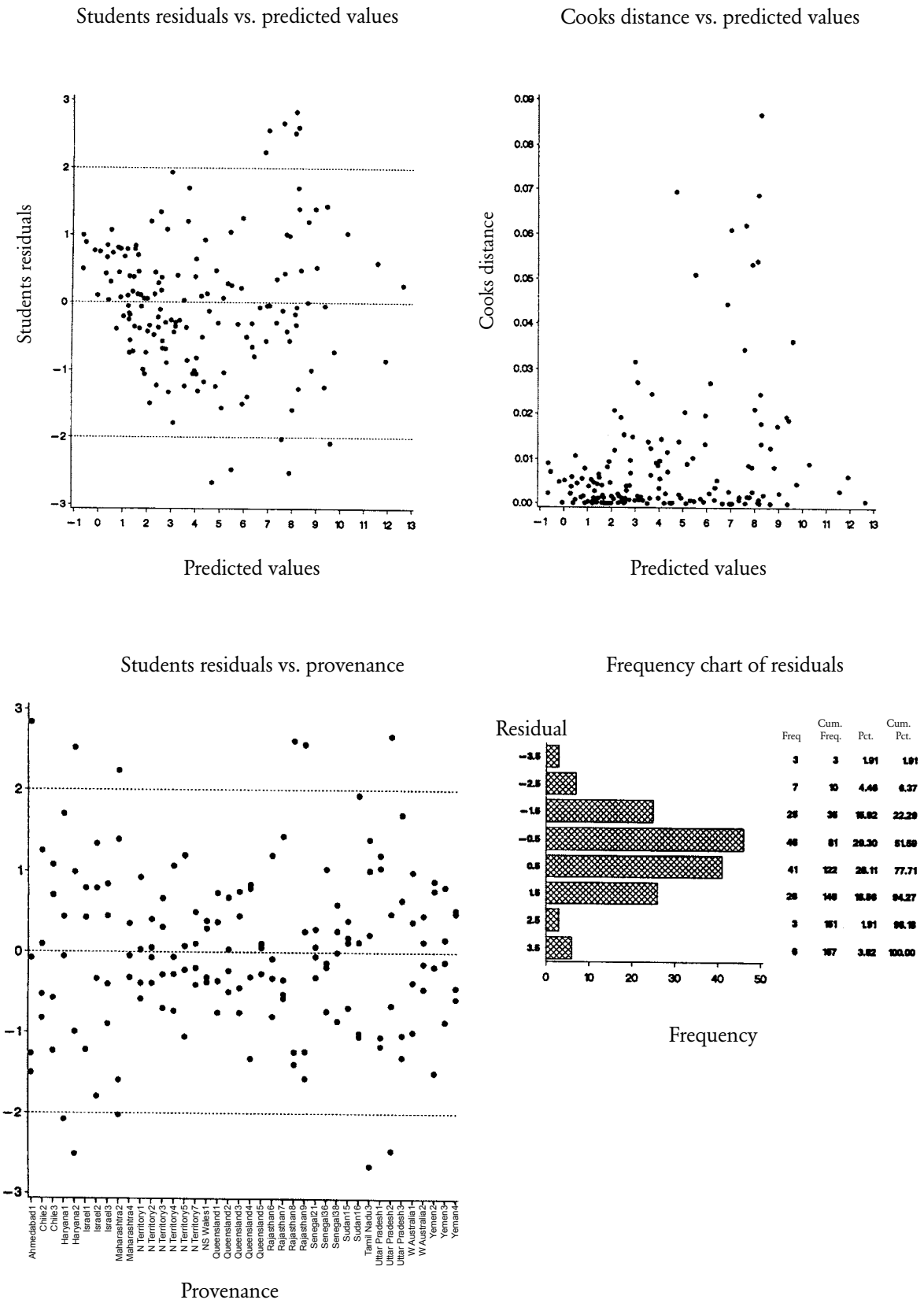


Figure 9. Residuals from the analysis of basal area per ha in a trial of *Acacia* and *Prosopis* seedlots from Bandia, established by Institut Sénégalais de Recherches Agricoles and Direction des Recherches sur les Productions Forestières, Senegal (Ræbild & Graudal *in prep. a*). From the plot of Student's residuals versus provenances it appears that some seedlots tend to be more variable than others. This variance heterogeneity may be solved by the application of a weighted analysis (see text).

8. Test of differences between seedlots

Once the statistical model is in place, the differences between seedlots are tested by an F-test with the aim to see whether these differences are significant. In order to do this, the hypothesis tested is that there is no difference between the seedlots. The F-test compares the mean square of seedlots with the residual mean square. Large *F*- values means that a large proportion of the variation is described by the seedlot effect, and that the residual variation is small.

Associated with the *F*-value is a probability value, often called the *P*-value. It is calculated as a standard in most statistical packages. The *P*-value indicates the probability of reaching results that are at least as extreme as the results that have been obtained – i.e. if the *P*-value is small, it is unlikely that the differences between provenances are due to random variation. Usually, if the *P*-value is below 0.05 (or 5%), the results are said to be significant. Therefore, if the *F*-test is significant and the *P*-value below 0.05, we reject the hypothesis and conclude that there are significant differences between the seedlots.

If the F-test is not significant, there are two possible explanations: 1. Either there are no differences between the seedlots (they grow in a similar manner for the given character), or 2. The trial has not been efficient in revealing such differences (with the design of the trial it is not possible to detect significant differences). Therefore the opposite conclusion, that the seedlots are similar, cannot be drawn. All we can say is that it was not possible to find significant differences between the seedlots with the design that we have used.

A provenance trial of *Pinus merkusii*, Boar Luang, North Thailand, established in 1971. There are visual (and significant) differences between the North-east Thailand provenance (to the left) and a provenance from Sumatra (to the right). Phot: H. Keiding, DFSC. 1981.

9. Correction for multiple comparisons

A problem with statistical tests is that sometimes they are judged significant by chance alone (Rice 1989). The more tests we perform, the larger is the probability that some tests become significant simply because we perform many tests, and not because there are true differences between the seedlots or whatever unit we may investigate. For example, Rice (1989) has shown that if 10 tests are performed, the probability is 40 % that at least one test is significant (on the 5% level) by chance alone. Obviously, this is a problem in provenance trials, where we often test several variables: We want to find true differences between provenances and not differences that are based on the fact that we test many variables.

There is a rather simple way to correct for this, called the sequential tablewide Bonferroni method (Holm 1979). Imagine that we have a provenance trial where we have measured and made statistical tests of eight variables (Table 1). The first step is then to rank these variables according to the significance. This has been done in Table 1, where the variable with the highest significance has been placed on top, and the other variables follow according to decreasing significance. The *P*-value of the most significant variable (crown area) is multiplied by the number of variables, which in this case is eight. If the result is below 0.05, it is concluded that this variable is significant at a "tablewide" level of 5%. In this case it is significant, and we proceed to the next variable (height). Here we multiply the *P*-value with the number of variables minus one (seven), compare it with 0.05 and proceed if it is below. We continue in this way until we reach the fifth variable, total basal area. Even though the *P*-value is significant, the corrected value is above 0.05, and we conclude that this and the subsequent variables are not significant on a "tablewide" basis. Therefore the analysis is taken no further. In the table we have indicated the significance levels by * (5%), ** (1%), *** (1 ‰) and n.s. (not significant).

Table 1. Results from analysis of species differences in a trial of *Acacia nilotica*, *A. seyal* and *A. tortilis* at Dori, Burkina Faso, established by Centre National de Semences Forestières, Burkina Faso (Ræbild, Graudal & Ouedraogo *in prep. b*). The table describes the sequence in calculation of the Bonferroni sequential tablewide tests. See text for further information.

Variable	*P*-value from test	Multiplication factor	Result	Bonferroni sequential tablewide test
Crown area	<0.0001	8	<0.0008	***
Height	0.0004	7	0.0028	**
Basal area of the mean tree	0.0020	6	0.01	*
Survival	0.0037	5	0.02	*
Total basal area	0.04	4	0.16	n.s.
Total dry weight	0.23	3		n.s.
Dry weight of the mean tree	0.41	2		n.s.
Number of stems	0.52	1		n.s.

10. Average performance of the seedlots: Least square means

The estimates for the seedlots are usually calculated as the so-called least square means (lsmeans). The main difference between raw means and lsmeans is that lsmeans account for missing values, imbalanced designs and effects of co-variates (see also section 4). Thus, in completely balanced designs with no co-variates, there are no differences between lsmeans and the raw means. In all other cases, the lsmeans are the best estimates for the given seedlot in the trial, and are often presented graphically to demonstrate the results of a trial. They are estimated by most software programmes for statistical analysis, and may be quite complicated to calculate without such software.

The confidence intervals and limits are basically calculated from the formula

$$X \pm t_{1-\alpha/2,(a-1)(b-1)} \sqrt{s^2 / b}$$

where X is the least square mean, α is the confidence level (in many cases 0.05, giving a 95% confidence interval), a is the number of seedlots and b is the number of replicates (blocks) of each seedlot. s^2 is the mean square of the error (MS_e). The presence of co-variates makes calculation a bit more difficult.

In the non-balanced situation, different seedlots may have different lengths of the confidence intervals (due to different variances). In cases where the data have been weighted, the confidence intervals are adjusted according to the variance of each seedlot and thus are of different lengths.

Special problems arise when the data are transformed, because transformation affects the calculation of the mean value. Basically, if the estimates are calculated on transformed values, then the least square means and the confidence limits have to be back-transformed in order to get "real-world" values. For example, if the logarithmic transformation has been applied to a variable such as basal area per ha, then we have to take antilog of the estimates to get values that correspond to the real, comprehensible values. In this way, the estimates are presented on the basis of the distribution that best fit the data, and give the most precise picture of differences between the provenances. The problem is that when the back-transformed average is compared to averages calculated without transformation, there is always a difference: The back-transformed average is always smaller.

On the other hand, if the estimates are calculated without transformation, they are based on a wrong distribution and do not give a precise picture of the differences between the seedlots. Therefore, since the aim of most trials is to find and describe differences between seedlots, we prefer calculating the estimates on transformed values.

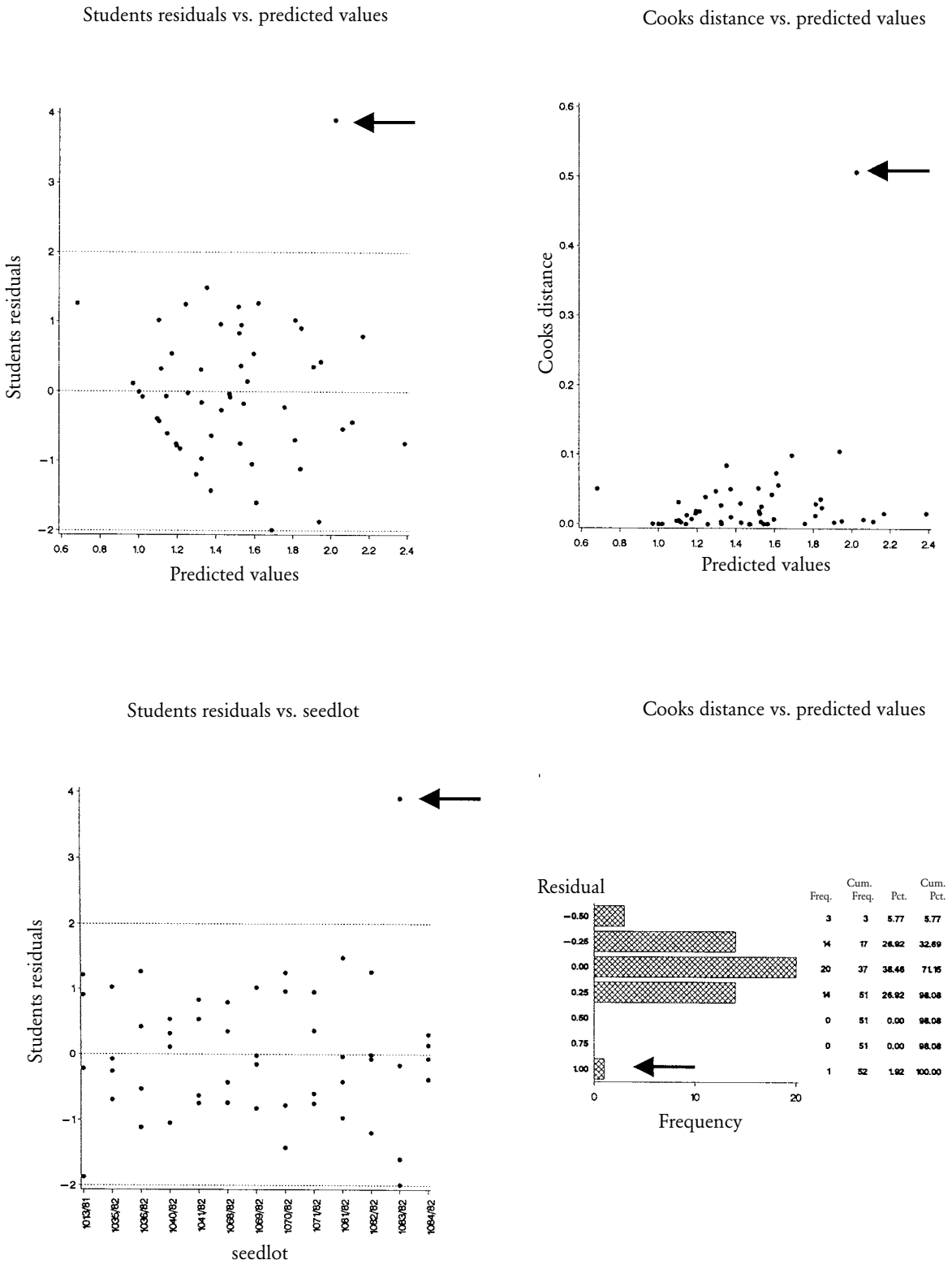


Figure 10. Residuals from the analysis of number of stems in a trial of *Acacia* seedlots from Jodhpur, established by Central Arid Zone Research Institute and Forest Research Institute and Colleges, India (Ræbild & Graudal *in prep. b*). One clear outlier is detected (Student's residual almost 4, marked by an arrow). The strong deviation from the rest of the data is explained by the fact that there are only two remaining trees in the outlier plot. Since solitary trees tend to have strongly developed crowns, the number of stems for this plot is higher than for the rest of the plots of the same provenance. In the final analysis the outlier was removed.

11. Best linear unbiased predictors

As mentioned above there is a difference between regarding the effect of provenances as fixed or random when it comes to calculation of estimates. While the fixed effect approach is the conventional least square means-method, there is an alternative using the random approach. This is the calculation of the Best Linear Unbiased Predictors, often referred to as BLUPs. Since the two methods give different results, it is necessary to focus a bit more on what are the differences between fixed and random effects.

When the provenance effect is fixed, it is seen as a constant with the underlying assumption that whenever the provenance is grown again at the same place, it will have the same growth. The specific aim is to compare the growth of exactly those provenances that have been chosen, and if the trial should be started all over again, it would be the same provenances that were chosen. The provenances are seen as treatments that should be compared.

When the provenance effect is random, the provenances are seen as representatives of a larger group of provenances (the population). The purpose of the trial is to get an idea of the variation within this group of provenances rather than to investigate the specific provenances. This has the consequence that if the trial was to be repeated, it could be different provenances that were chosen.

This difference has important statistical implications. In the fixed approach, the least square means method is used, which is more or less comparable to the mean values (see the previous section). But in the random approach, the provenances are believed to be a population, and the growth of individual provenances is seen as coming from a normal distribution with an expected value and a variance.

Estimating provenance effects in random models is more complicated than in fixed models, because the observed variation between provenances is contemplated as a mixture between true provenance effects and random error variation (*cf.* White & Hodge 1989). The variation between the provenances is therefore always larger than the true "genetic" variation, except in the rare cases where the error variation is negligible. Therefore the differences between seedlots are smaller when using BLUP values than when the lsmeans are applied.

In order to calculate the BLUPs and predict the effect of a given provenance, it is necessary to correct the estimates for the part of the variance that is due to random error variation. This is cumbersome and only feasible with a suitable software package. Often the results are presented as deviations from the mean value to allow for easier comparison between different experiments. The deviations are expressed either in real values (m, cm^2 etc.) or in %.

12. When the trial is a mixture of different species

Many trials contain more than one species. In these cases it may be possible to enlarge the scope of the trial and make more general statements about the differences between the species in the trial. Often the analyses start with a test of the model (1) from above, regarding only differences between seedlots. However, a second step could be to analyse whether there are differences between the species in the trial. This analysis is based on the following model

$$X_{ijk} = \mu + species_i + provenance(species)_{ij} + block_k + \varepsilon_{ijk}$$

(Model 2)

where X_{ijk} is the value of the trait in question (e.g. height) in plot *ijk*,
μ is the grand mean,
$species_i$ is the fixed effect of species number *i*,
$provenance(species)_{ij}$ is the effect of provenance number *j* nested within species *i*. This effect is assumed to be a random effect. It is assumed that this effect is a sample from a normal distribution with an expected value of zero and variance σ_{pr}^2, $block_j$ is the effect of block (replication) *k* in the trial and is assumed to be a random effect,
ε_{ijk} is the residual of plot *ijk*, and is assumed to follow the normal distribution $N(0, \sigma_e^2)$.

Note that the provenances in statistical terms are nested within the species. This means that the species effect should be tested against the variation between seedlots (and not as is usually done against the residual variation), and that the *F*-test involves combinations of different mean squares. Therefore the tests should be performed differently, using a so-called Satterthwaite's approximation is used to calculate the degrees of freedom. Many standard computer programmes will calculate the approximation, but it is necessary to specify it in the model. If not, the programmes may calculate wrong *F*-tests, often exaggerating the significance. It is always important to be careful in constructing the *F*-test when nested effects are involved.

The analyses of differences between the seedlots of one species are done easily by using the model (1), where only the seedlots from the species in question are analyzed.

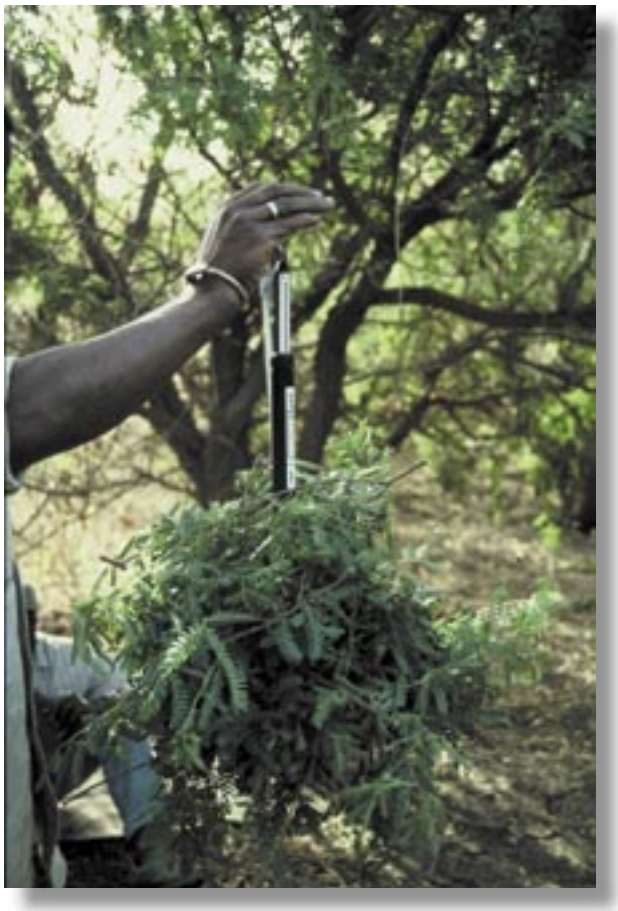
13. A note on the interpretation of provenance trials

Even though provenance trials may show large differences between provenances, it is important to realise that the gains indicated by the trials tend to be exaggerated. This is due to several causes, some related to the seedlots, and some related to the environment.

For some temperate tree species it has been observed that seedlots collected in the same stands but in different years turn out to behave differently in trials, e.g. with respect to growth. This year-to-year variation can be due to different contributions from different parents (large flowering, small flowering, inflow of pollen etc.) or non-Mendelian effects of the climate on the adaptability of the seed (so-called after-effects).

The climatic variation also has a direct influence on the trial. Parallel trials established in different years may show different results, for instance, if in one year the conditions are favourable during the establishment period and in another year they are adverse. Provenances may of course react differently to different conditions, and a different ranking between two parallel trials may be the result. Finally one should not forget that the results are only applicable to environments that are similar to those of the trial. Often small variations in the soil, exposure or available water can give results that are very different from those at the trial site.

Overcoming these limitations will seldom be possible. Trials established at multiple sites in different years and with seedlots collected through different years will be very expensive and only rarely possible. Recommendations can be based on recognition of patterns, i.e. groups of provenances with good performance either from certain areas or certain climatic zones. However, overall we have to accept that our knowledge is imperfect and that we have to do the best with the data that are available. The limitations mentioned above can only serve to introduce caution in the recommendations. For example, in the case where a local, adapted provenance is available, exotic provenances should be very clearly superior in order to be recommended in favour of local material.



Assessment of biomass production by destructive sampling, *Prosopis cineraria*, Dagar Kotli, Pakistan. Assessing fresh weight of a sample of foliage and small branches. Phot: Lars Graudal, DFSC. 1993.

14. References

Afifi, A.A. and V. Clark 1996. Computer-aided multivariate analysis. Chapman & Hall, London (3rd Edition), 455 pp.

Box, G.E.P., W.G. Hunter and J.S. Hunter 1978. Statistics for experimenters. John Wiley and Sons Inc., New York.

Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 6, 65-70.

Loo-Dinkins, J. 1992. Field Test Design. In: Fins, L., Friedman, S.T. & Brotschol (eds.): Handbook of Quantitative Forest Genetics. Kluwer Academic Publishers, Dordrecht, 96-139.

Rice, W.R. 1988. Analyzing tables of statistical tests. Evolution 43, 223-225.

Ræbild, A., L. Graudal and P.C.F. Lima, in prep.: Evaluation of an Acacia trial in Bebedourou, Brazil (Trial no. 1 in the Arid Zone Series). Danida Forest Seed Centre.

Ræbild, A., L. Graudal and G.L. Ouedraogo, in prep.a: Evaluation of a combined species and provenance trial with *Acacia nilotica, A. seyal* and **A. tortilis** at Gonsé, Burkina Faso (Trial no. 10 in the Arid Zone Series). Danida Forest Seed Centre.

Ræbild, A., L. Graudal and G.L. Ouedraogo, in prep.b: Evaluation of a combined species and provenance trial with *Acacia nilotica, A. seyal* and *A. tortilis* at Dori, Burkina Faso (Trial no. 7 in the Arid Zone Series). Danida Forest Seed Centre.

Ræbild, A. and L. Graudal, in prep a: Evaluation of an Acacia and Prosopis trial in Bandia, Senegal (Trial no. 24 in the Arid Zone Series). Danida Forest Seed Centre.

Ræbild, A. and L. Graudal, in prep b: Evaluation of an Acacia and Prosopis trial in Jodhpur, India (Trial no. 16 in the Arid Zone Series). Danida Forest Seed Centre.

Schmidt, L. 1997. Tree improvement glossary. Danida Forest Seed Centre, Tehcnical note. 46, 63 pp.

Snedecor, G.W. and W.G. Cochran, 1980. Statistical methods. Iowa State University Press, 7th ed., 507 pp.

White, T.L. and G.R. Hodge, 1989. Predicting breeding values with applications in forest tree improvement. Kluwer Academic Publishers, Dordrecht, 367 pp.