

Consequences of reducing a full model of variance analysis in tree breeding experiments

M Giertych¹, H Van De Sype^{2*}

¹ Institute of Dendrology, 62-035 Kornik, Poland; ² INRA, Station d'Amélioration
des Arbres Forestiers, Ardon, 45160 Olivet, France

(Received 20 July 1988; accepted 30 June 1989)

Summary — An analysis of variance was performed on height measurement of 11-year-old trees (7 in the field), using the results of a non-orthogonal progeny within provenance experiment established for Norway spruce (*Picea abies* (L.) Karst.) at 2 locations in Poland. The full model including locations, provenances, progenies within provenances, blocks within locations and trees within plots is used assuming all sources of variation to be random. This model is compared with various models reduced by 1 factor or the other within the model. Theoretical modifications of estimated variance components and heritabilities are tested with experimental data. By referring to the original model it is shown how changes came to be and where the losses of information occurred. A method is proposed to reduce the factor level number without bias. The general conclusion is that it pays to make the effort and work with the full model.

***Picea abies* / height / provenance / progeny / variance analysis / method / genetic parameter**

Résumé — **Conséquences de la réduction d'un modèle complet d'analyse de variance pour des expériences d'amélioration forestière.** La hauteur totale à 11 ans, après 7 ans de plantation, a été mesurée en Pologne dans deux sites pour 12 provenances d'*Epicéa* commun originaires de Pologne, avec environ 8 familles par provenance. Les différents termes et indices sont explicités dans le tableau 1. L'analyse de la variance selon un modèle complet (localité, bloc dans localité, provenance, famille dans provenance, et les diverses interactions) a été réalisée en considérant les facteurs comme aléatoires (tableau 2). Elle est comparée à des analyses selon des modèles simplifiés qui ignorent successivement les niveaux provenance, famille ou bloc, ou les valeurs individuelles. Dans le cas du modèle simplifié sans facteur provenance, les nouvelles espérances des carrés moyens (tableau 3) peuvent être strictement comparées à celles obtenues avec le modèle complet. Les modifications théoriques ont été calculées et sont présentées de façon schématique pour l'estimation des composantes de la variance (tableau 4) et des paramètres génétiques (tableau 5). Les résultats théoriques associés aux autres modèles sont également reportés dans ces deux derniers tableaux. En outre, l'implication du nombre de niveaux par facteur sur les biais entraînés a été précisée. En général, les simplifications surestiment fortement les composantes de la variance et augmentent de façon illicite les gains espérés. Les résultats obtenus avec les données expérimentales montrent effectivement des changements au niveau des composantes de la variance ou des tests associés (tableau 7) et de légères modifications pour les paramètres génétiques.

*Correspondence and reprints.

tiques (tableau 8). Les biais que de telles simplifications peuvent entraîner dans un programme d'amélioration forestière sont discutés. En conclusion, une proposition est formulée pour réduire par étapes mais de façon fiable le nombre de niveaux à étudier.

Picea abies / hauteur / provenance / descendance / analyse de variance / méthode / paramètre génétique

INTRODUCTION

In complicated tree breeding experiments, particularly when one deals with non-orthogonal and unbalanced design, and this is often the case, the temptation arises to reduce the model to only those parts that are of particular interest at a given time. Such reductions from the full model create certain consequences that we are not always fully aware of. The aim of the present paper is to show on one experiment how different reductions of the experimental model affect the results and conclusions derived from them.

MATERIAL

The experiment discussed here is a Norway spruce (*Picea abies* (L.) Karst.) progeny within provenance study established at 2 locations in Poland, in Kornik and in Goldap, in 1976 using 2+2 seedlings raised in a nursery in Kornik. The experiment includes half-sib progenies from 12 provenances from the North Eastern range of the spruce in Poland. Originally, cones were collected from 10, randomly selected trees from each of the provenances. However, due to an inadequate number of seeds or seedlings per

progeny, the experiment was established in an incomplete block design. Not only were the maternal trees selected at random, but the provenances were also a random choice of Forest Districts in the area and cone collections were carried out from fellings which were being made in the Forest District at the time we arrived there for cone collection.

Since all our Polish experiments were concentrated in regions near Kornik and Goldap, the choice of locations could also be considered as random. The blocks in our locations are just part of the areas, and therefore cover all variations of the site, and may also be considered as random. Details of the study were presented in an earlier paper (Giertych and Krolikowski, 1982). The designations used in the study are shown in table I.

As the design is far from orthogonal, analyses of data (height in 1983) were performed in France using the Amance ANOVA programs (Bachacou *et al.*, 1981). Furthermore, the number of factor levels was larger than the computer capacity, and accordingly analyses were done in several stages.

ANALYSES WITH DIFFERENT MODELS

The full model

The full model has been used to extract the maximum amount of information from

Table I. Definition of terms used and experimental factor levels.

<i>Factor</i>	<i>Index</i>	<i>Theoretical Level</i>	<i>Experimental Level</i>
(L) Locality	j	1	2
(B) Block within locality	m	b	6
(P) Provenance	i	p	12
(F) Family within provenance	k	f	8.25
(E) Error (trees)	n	x	8.501

Table II. Expected mean squares for the full model 1° (orthogonal design). Expected mean squares are a linear combination of variance components and theoretical factor levels. Abbreviations are indicated in table I.

Factors	df	Expected mean squares								
		σ^2_E	σ^2_{FB}	σ^2_{FL}	σ^2_F	σ^2_{PB}	σ^2_{PL}	σ^2_P	σ^2_B	σ^2_L
L_j	$l-1$	$1 + x$	$+ bx$	$+ 0$	$+ fx$	$+ bfx$	$+ 0$	$+ pfx$	$+ bpfx$	
$B_{m(j)}$	$l(b-1)$	$1 + x$	$+ 0$	$+ 0$	$+ fx$	$+ 0$	$+ 0$	$+ pfx$		
P_i	$p-1$	$1 + x$	$+ bx$	$+ lbx$	$+ fx$	$+ bfx$	$+ lbfx$	$+ lbfx$		
PL_{ij}	$(l-1)(p-1)$	$1 + x$	$+ bx$	$+ 0$	$+ fx$	$+ bfx$				
$PB_{im(j)}$	$l(b-1)(p-1)$	$1 + x$	$+ 0$	$+ 0$	$+ fx$					
$F_{k(i)}$	$p(f-1)$	$1 + x$	$+ bx$	$+ lbx$						
$FL_{jk(i)}$	$(l-1)p(f-1)$	$1 + x$	$+ bx$							
$FB_{km(ij)}$	$l(b-1)p(f-1)$	$1 + x$								
$E_{n(ijkm)}$	$lbpf(x-1)$	1								
Total	$lbpfx-1$									

the material (symbols are explained in table I):

$$X_{ijkmn} = \mu + L_j + B_{m(j)} + P_i + PL_{ij} + PB_{im(i)} + F_{k(i)} + FL_{jk(i)} + FB_{km(ij)} + E_{n(ijkm)}$$

Since all elements of the experiment were considered to be random, the degrees of freedom and expected mean squares for the variance analysis are as shown in table II obtained through the procedure described by Hicks (1973). The theoretical degrees of freedom for an orthogonal model and the expected mean squares are shown in table II.

On the basis of this full model, it is possible to calculate heritabilities by the formula proposed by Nanson (1970) for: – provenances: $h^2_P = \sigma^2_P / V_P$, where:

$$V_P = \sigma^2_P + \sigma^2_{PL}/l + \sigma^2_{PB}/lb + \sigma^2_F/f + \sigma^2_{FL}/lf + \sigma^2_{FB}/lbf + \sigma^2_E/lbfx$$

– and families within provenance: $h^2_F = \sigma^2_F / V_F$, where:

$$V_F = \sigma^2_F + \sigma^2_{FL}/l + \sigma^2_{FB}/lb + \sigma^2_E/lbx$$

In an orthogonal system, these heritabilities can be estimated from the F value of the Snedecor's test by $1 - (1/F)$. In fact, due to non-orthogonality and unbalanced design, they were calculated from the variance components.

For this half-sib experiment, another approach is to calculate single tree heritability (narrow sense) based on the between-families' additive variance and the phenotypic variance (V_{Ph}):

$$h^2_s = 4 \sigma^2_F / V_{Ph} \text{ where:}$$

$$V_{Ph} = \sigma^2_P + \sigma^2_{PL} + \sigma^2_{PB} + \sigma^2_F + \sigma^2_{FL} + \sigma^2_{FB} + \sigma^2_E$$

Heritability (h^2), variance (V), selection intensity (i) and expected genotypic gain ($\Delta G = i h^2 \sqrt{V}$) depend on the aim of the selection and the type of material used. For example, it is possible to estimate the genotypic gain which will be expected for reforestation with the same seeds which gave the material selected in this experiment. The best provenance may be selected from a total of twelve, so the expected gain will be estimated with heritability and

Table III. Expected mean squares for model 2° ignoring the provenance factor. (orthogonal design). Expected mean squares are a linear combination of variance components and theoretical factor levels. Abbreviations are indicated in table I. The subscript k' is used for family and provenance.

Factors	df	Expected mean squares										
		$\sigma^2_{E'}$	$\sigma^2_{F'B'}$	$\sigma^2_{F'L'}$	$\sigma^2_{F'}$	$\sigma^2_{B'}$	$\sigma^2_{L'}$					
L'_j	l-1	1	+	x	+	bx	+	0	+	pfx	+	bpx
$B'_{m(j)}$	l(b-1)	1	+	x	+	0	+	0	+	pfx		
$F'_{k'}$	pf-1	1	+	x	+	bx	+	lbx				
$F'L'_{jk'}$	(l-1)(pf-1)	1	+	x	+	bx						
$F'B'_{k'm(j)}$	l(b-1)(pf-1)	1	+	x								
$E'_{n(jk'm)}$	lbpf(x-1)	1										
Total	lbpx-1											

phenotypic variance at the provenance level, and $i = 1.840$. The selection of the 2 best families within each provenance will use family parameters and $i = 1.289$. At the end of these 2 steps, 2 families within the best provenance will be selected; this will be compared to a 1 step selection with $i = 2.417$. Another method is to select the 50 best individuals from the 9122 trees of this experiment, to propagate them, and to establish a seed orchard. The expected genetic gain of the seed orchard offsprings will be estimated from phenotypic variance (V_{Ph}), narrow sense heritability (h^2_s) and $i = 2.865$. It is assumed here that these last values are ones that utilize the maximum amount of data and are therefore the best that can be obtained.

Let us now examine the changes produced with simpler models when a part of the information is not used.

Model ignoring the provenance factor

For increasing estimation of genetic parameters, it may be tempting to treat the families, altogether, disregarding the split-

up of families into provenances. In a fully orthogonal model with p provenances and f families (within provenances), the number of families is pf with a new subscript k' instead of $k(i)$. The model now becomes:

$$X_{jk'mn} = \mu + L'_j + B'_{m(j)} + F'_{k'} + F'L'_{jk'} + F'B'_{k'm(j)} / E'_{n(jk'm)}$$

The distribution of the degrees of freedom and the expected mean squares are as shown in table III. In order to use the variance components estimated from the full model, we must combine the sum of squares from table II as follows:

Degrees of freedom

and SS from model 2 SS from model 1

$$l-1 \quad SS_{L'} = SS_L$$

$$l(b-1) \quad SS_{B'} = SS_B$$

$$pf-1 \quad SS_{F'} = SS_F + SS_P$$

$$(l-1)(pf-1) \quad SS_{F'L'} = SS_{FL} + SS_{PL}$$

$$l(b-1)(pf-1) \quad SS_{F'B'} = SS_{FB} + SS_{PB}$$

$$lbpf(x-1) \quad SS_{E'} = SS_E$$

The total sum of squares remains unaffected. Working from the bottom of this list we can identify, on the left hand-side, the

new sum of squares with the expected mean squares multiplied by the degrees of freedom indicated above (and in table III), and on the right hand-side, the combination of sums of squares with their expected mean squares multiplied by their own degrees of freedom from table II. The procedure is shown for the 2 first factors.

1/ new residual

The degrees of freedom are $lbp(x-1)$ for both sides of the equation. The equation $SS_{E'} = SS_E$ is transformed as $lbp(x-1)\sigma_{E'}^2 = lbp(x-1)\sigma_E^2$, thus leads to:

$$\sigma_{E'}^2 = \sigma_E^2 \text{ (relation 1).}$$

2/ new family x block interaction

The degrees of freedom are $l(b-1)(pf-1)$ for the new expected mean square and $l(b-1)p(f-1)$ for the full model. $SS_{F'B'} = SS_{FB} + SS_{PB}$ becomes:

$$\begin{aligned} & l(b-1)(pf-1)(\sigma_{E'}^2 + x\sigma_{F'B'}^2) \\ & = l(b-1)p(f-1)(\sigma_E^2 + x\sigma_{FB}^2) \\ & + l(b-1)(p-1)(\sigma_E^2 + x\sigma_{FB}^2 + fx\sigma_{PB}^2) \\ & pf\sigma_{E'}^2 - \sigma_{E'}^2 + (pf-1)x\sigma_{F'B'}^2 \\ & = pf\sigma_E^2 - \sigma_E^2 + pfx\sigma_{FB}^2 - x\sigma_{FB}^2 \\ & \quad + pfx\sigma_{PB}^2 - fx\sigma_{PB}^2 \end{aligned}$$

considering 1 ($\sigma_{E'}^2 = \sigma_E^2$) and simplifying by $(pf-1)x$ gives:

$$\sigma_{F'B'}^2 = \sigma_{FB}^2 + \sigma_{PB}^2 (p-1)f/(pf-1) \text{ (relation 2)}$$

The same procedure is followed for other equalities of sums of squares. To summarize, when we decide to speak of families only, instead of provenances and families (within provenances), we obtain the following changes in variance components:

$$\begin{aligned} \sigma_{L'}^2 &= \sigma_L^2 + \sigma_{PL}^2 (f-1)(pf-1) \\ \sigma_{B'}^2 &= \sigma_B^2 + \sigma_{PB}^2 (f-1)(pf-1) \\ \sigma_{F'}^2 &= \sigma_F^2 + \sigma_P^2 (p-1)f/(pf-1) \\ \sigma_{F'L'}^2 &= \sigma_{FL}^2 + \sigma_{PL}^2 (p-1)f/(pf-1) \end{aligned}$$

$$\sigma_{F'B'}^2 = \sigma_{FB}^2 + \sigma_{PB}^2 (p-1)f/(pf-1) \text{ (relation 2)}$$

$$\sigma_{E'}^2 = \sigma_E^2 \text{ (relation 1)}$$

$$\begin{aligned} V_{T'} &= \sigma_{E'}^2 + \sigma_{F'B'}^2 + \sigma_{F'L'}^2 + \sigma_F^2 + \sigma_{PB}^2 \\ & \quad + \sigma_{PL}^2 + \sigma_P^2 (p-1)f/(pf-1) + \sigma_B^2 + \sigma_L^2 \end{aligned}$$

$$V_{T'} = V_T - \sigma_P^2 (f-1)/(pf-1)$$

This implies modifications for variance components and total variance as shown in table IV. The true variance components of interactions between provenance and locality (σ_{PL}^2) or block (σ_{PB}^2) are each split in 2 parts. The largest part enters in the component of interactions between family and locality ($\sigma_{F'L}^2$) or block ($\sigma_{F'B'}^2$), and the smallest one enters in the locality (σ_L^2) or block (σ_B^2) components. For the true variance component for provenance (σ_P^2), the largest part enters in the family component (σ_F^2) and the smallest one is lost altogether, so the total variance ($V_{T'}$) is lowered by $\sigma_P^2 (f-1)/(pf-1)$.

Compared to the full model, ignoring the provenance level introduces modifications for estimation of genetic parameters (table V). The mean family variance ($V_{F'}$) is increased by the largest part of all the components of provenance effect and interactions ($\sigma_P^2 + \sigma_{PL}^2/l + \sigma_{PB}^2/lb$) $(p-1)f/(pf-1)$. The family heritability ($h_{F'}^2$) decreases slightly and the expected gain is higher ($\Delta G_{F'}$). At the individual level, the phenotypic variance (V_{Ph}) is lowered by the smallest part of variance components for provenance effects ($\sigma_P^2 + \sigma_{PL}^2 + \sigma_{PB}^2$) $(f-1)/(pf-1)$. The narrow sense heritability ($h_{s'}^2$) and the expected genetic gain for additive effect (ΔG) are increased by a part of the non-additive effects, originating from provenance variations.

One point of interest is to observe the changes which occur in relation to the number of provenances (p) or families per provenance (f). For the same total number of families (pf), the larger the number of provenances, the lower the loss of total

Table IV. Schematic representation of changes for estimated variance component. Only the deviations from the full model are indicated. Comparatively to the full model, variance components are not estimated (NE), equal (==), divided by the number of trees per plot (/x), or increased (or decreased (- or ---)) with the smallest (+) or the biggest part (+++) of the variance components (σ^2) of factors indicated in brackets. other abbreviations are indicated in table I.

Factors	Model			
	2° Less prov.	3° Less fam.	4° Less block	5° Less indiv.
σ^2_L	$+\sigma^2(\text{PL})$	==	$+\sigma^2(\text{B})$	/x
σ^2_B	$+\sigma^2(\text{PB})$	==	NE	/x
σ^2_P	NE	$+\sigma^2(\text{F})$	==	/x
σ^2_{PL}	NE	$+\sigma^2(\text{FL})$	$+\sigma^2(\text{PB})$	/x
σ^2_{PB}	NE	$+\sigma^2(\text{FB})-\sigma^2(\text{FL},\text{F})$	NE	/x
σ^2_F	$+++ \sigma^2(\text{P})$	NE	==	/x
σ^2_{FL}	$+++ \sigma^2(\text{PL})$	NE	$+\sigma^2(\text{FB})-\sigma^2(\text{PB},\text{B})$	/x
σ^2_{FB}	$+++ \sigma^2(\text{PB})$	NE	NE	/x $+\sigma^2(\text{E})$
σ^2_E	==	$+++ \sigma^2(\text{FB},\text{FL},\text{F})$	$+++ \sigma^2(\text{FB},\text{PB},\text{B})$	NE
V_{total}	$-\sigma^2(\text{P})$	==	==	/x $-\sigma^2(\text{E})$

Table V. Schematic representation of changes for genetic parameters. Comparatively to the full model, genetic parameters are not estimated (NE), equal (==), divided by the number of trees per plot (/x), or increased (or decreased (- or ---)) with the smallest (+) or the biggest part (+++) of the variance components (σ^2) of factors indicated in brackets. other abbreviations are indicated in table I.

Genetic parameter	Model			
	2° Less prov.	3° Less fam.	4° Less block	5° Less indiv.
Provenance				
h^2_P	NE	+	==	==
V_P	NE	==	==	/x
ΔG_P	NE	+++	==	/x
Family				
h^2_F	---	NE	==	==
V_F	$+++ \sigma^2(\text{PB},\text{PL},\text{P})$	NE	==	/x
ΔG_F	+	NE	==	/x
Individual				
h^2_S	+++	NE	--	NE
V_{Ph}	$-\sigma^2(\text{PB},\text{PL},\text{P})$	NE	$+++ \sigma^2(\text{B})$	NE
ΔG	+++	NE	--	NE

variance ($V_{T\cdot}$) and the larger phenotypic variances, heritabilities and expected gains at family or individual levels. By increasing the total number of families (pf), the same modifications occur.

Ignoring the family factor

When comparing provenances, it seems easier to ignore the family variation, and to reduce the experiment to a simple provenance trial. We then obtain the following model, where a new subscript (n') is used instead of (k) family and of (n) tree ones:

$$X_{ijmn'} = \mu + L'_j + B'_{m(i)} + P'_i + P'L'_{ij} + P'B'_{im(i)} + E'_{n'(ijm)}$$

As with the previous model, expected mean squares can be constructed with the new sums of squares and new degrees of freedom and then compared to the original full model 1. Change is observed for the residual level only, which now includes all family-dependent variations: $SS_{E'} = SS_E + SS_{FB} + SS_{FL} + SS_F$. The degrees of freedom become $lbp(fx-1) = lbpf(x-1) + l(b-1)p(f-1) + (l-1)p(f-1) + p(f-1)$ with fx the new number of trees per plot. Using the same procedure as for model 2°, we obtained the following values for the new variance components in terms of those of the original full model:

$$\begin{aligned} \sigma^2_{L'} &= \sigma^2_L \\ \sigma^2_{B'} &= \sigma^2_B \\ \sigma^2_{P'} &= \sigma^2_P + \sigma^2_F / f \\ \sigma^2_{P'L'} &= \sigma^2_{PL} + \sigma^2_{FL} / f \\ \sigma^2_{P'B'} &= \sigma^2_{PB} + \sigma^2_{FB} \{ (x-1) / (fx-1) \} \\ &\quad - (\sigma^2_{FL} + \sigma^2_F) \{ (f-1) / f(fx-1) \} \\ \sigma^2_{E'} &= \sigma^2_E + (\sigma^2_{FB} + \sigma^2_{FL} \\ &\quad + \sigma^2_F) \{ (f-1)x / (fx-1) \} \end{aligned}$$

$$\begin{aligned} V_{T\cdot} &= \sigma^2_{E'} + \sigma^2_{FB} + \sigma^2_{FL} + \sigma^2_F + \sigma^2_{PB} \\ &\quad + \sigma^2_{PL} + \sigma^2_P + \sigma^2_B + \sigma^2_L \\ V_T &= V_{T\cdot} \end{aligned}$$

The total variance and the locality and block (within locality) variance components remain unaffected (table IV). Provenance and provenance-locality variance components are increased respectively by the smallest part of family and family-locality components. The provenance-block interaction is modified by a small part of a combination of all family components while the main part is included in the residual. For genetic estimations (table V), the mean provenance variance (V_P) remains unchanged and the provenance heritability (h^2_P) is higher with the increase of the variance component of provenance. Consequently, the expected gain for a provenance selection is higher. All these modifications depend on the number of families per provenance only (f). The higher the number, the lower the bias.

Model ignoring block effect

Sometimes authors have no interest in the variation between blocks and they place all block effects into the residual. A new subscript n' must be used instead of m for block and n for tree, the model will then be:

$$X_{ijkn'} = \mu + L'_j + P'_i + P'L'_{ij} + F'_{k(i)} + F'L'_{jk(i)} + E'_{n'(ijk)}$$

The new sums of squares, compared to the original ones, will change for residual only: $SS_{E'} = SS_E + SS_{FB} + SS_{PB} + SS_B$. bx now becomes the new number of trees per element of the experiment and the new degrees of freedom for the error term become:

$$lpf(bx-1) = lbpf(x-1) + l(b-1)p(f-1) + l(b-1)(p-1) + l(b-1).$$

Following the same procedure as before, we obtain new values of variance components:

$$\sigma^2_{L'} = \sigma^2_L + \sigma^2_B / b$$

$$\sigma^2_{P'} = \sigma^2_P$$

$$\sigma^2_{P'L'} = \sigma^2_{PL} + \sigma^2_{PB} / b$$

$$\sigma^2_{F'} = \sigma^2_F$$

$$\sigma^2_{F'L'} = \sigma^2_{FL} + \sigma^2_{FB} \{ (x-1)/(bx-1) \} - (\sigma^2_{PB} + \sigma^2_B) \{ (b-1)/b(bx-1) \}$$

$$\sigma^2_{E'} = \sigma^2_E + (\sigma^2_{FB} + \sigma^2_{PB} + \sigma^2_B) \{ (b-1)x/(bx-1) \}$$

$$V_T = \sigma^2_E + \sigma^2_{FB} + \sigma^2_{FL} + \sigma^2_F + \sigma^2_{PB} + \sigma^2_{PL} + \sigma^2_P + \sigma^2_B + \sigma^2_L$$

$$V_T = V_T$$

When ignoring the block effect, the total variance and the variance components for provenance and family levels remain unaffected (table IV). The variance components at locality or provenance-locality levels include a small part of the block or provenance-block component. The main part of the block and block-interaction variance components enters into the residual. For genetic parameters (table V), variance of means, heritability and expected gain are not changed for provenance or family levels. At the individual level, the phenotypic variance is increased by the main part of the block component ($\sigma^2_B (b-1)/b$), so the single tree heritability is lower than that in the full model and the expected genetic gain decreases. The more blocks (b), the smaller the bias for variance components, and the larger for mass selection option.

Model with plot averages

Another method used is to work on plot averages only. This is generally used for traits such as mortality or productivity per unit area. In these cases, SS_E is not available and the only approach is to use SS_{FB} as the new residual. It is therefore impossible to estimate the true variance compo-

nent for the family x block interaction (σ^2_{FB}). The model becomes:

$$X_{ijkm} = \mu + L'_j + B'_{m(ij)} + P'_i + P'L'_{ij} + P'B'_{im(ij)} + F'_{k(i)} + F'L'_{jk(i)} + F'B'_{km(ij)}$$

The number of trees per plot (x) is the new unit of measurement and does not enter into the degrees of freedom. Since the analysis is performed on the basis of plot means (sums per plot divided by x), original sums of squares must be divided by x^2 . The new sums of squares will be constructed as below:

Degrees of freedom

and SS from model 5°	SS from model 1°
l-1	$SS_{L'} = SS_L / x^2$
...	...
$(l-1)p(f-1)$	$SS_{F'L'} = SS_{FL} / x^2$
$l(b-1)p(f-1)$	$SS_{residual} = SS_{FB} / x^2$

The relation $SS_{residual} = SS_{FB} / x^2$ will be:

$$l(b-1)p(f-1) \sigma^2_{residual} = l(b-1)p(f-1) (\sigma^2_E + x\sigma^2_{FB}) / x^2$$

$$\sigma^2_{residual} = \sigma^2_E / x^2 + \sigma^2_{FB} / x \quad (\text{relation 3})$$

Similarly: $SS_{F'L'} = SS_{FL} / x^2$ becomes:

$$l(b-1)(p-1) (\sigma^2_{residual} + f\sigma^2_{P'B'}) = l(b-1)(p-1) (\sigma^2_E + x\sigma^2_{FB} + fx\sigma^2_{PB}) / x^2$$

Considering 3 and simplifying:

$$(\sigma^2_E / x^2 + \sigma^2_{FB} / x) + f\sigma^2_{P'B'} = (\sigma^2_E + x\sigma^2_{FB} + fx\sigma^2_{PB}) / x^2$$

$$\sigma^2_{P'B'} = \sigma^2_{PB} / x \quad (\text{relation 4})$$

Computations and results are similar for all remaining variance components, thus:

$\sigma^2_{L'}$	$= \sigma^2_L / x$	
..	
$\sigma^2_{F'L'}$	$= \sigma^2_{FL} / x$	(relation 4)
$\sigma^2_{residual}$	$= \sigma^2_E / x^2 + \sigma^2_{FB} / x$	(relation 3)

$$V_T = [\sigma^2_E / x + \sigma^2_{FB} + \sigma^2_{FL} + \sigma^2_F + \sigma^2_{PB} + \sigma^2_{PL} + \sigma^2_P + \sigma^2_B + \sigma^2_L] / x$$

$$V_T = V_T / x - \sigma^2_E [(x-1)/x^2]$$

The decrease of variance components depends on the number of trees per plot (x), and reflects the use of plot means as compared with individual data (table IV). For the total variance, a part of losses is a «logical» reduction due to a lower number of data (V_T/x). Another part is a loss of information (σ²_E (x-1)/x²). The higher the number of trees per plot (x), the lower the total variance (V_T), also lower is the relative loss due to lack of information (σ²_E).

For provenance and family levels, in comparison with the full model, heritabilities are unaffected while means variances are divided by x, and expected gains are divided by √x (table V). Without individual data, phenotypic variance (V_{Ph}) and single tree heritability (h²s) cannot be estimated.

Experimental data

Experimental data (tree height at 7 years in the field) were analyzed with the 5

models. The experimental factor levels are indicated in table I.

With the full model (model 1), the result of the analysis of variance (table VI) shows that 3 factors are in significant. It is very surprising that no locality effect and no provenance x locality interaction effect, exists. The climatic conditions in Kornik and Goldap are very different, but grand means are identical for the 2 locations (232.8 cm and 234.6 cm respectively). Unfortunately, locality and provenance x locality variance components have negative values (they are indicated between brackets in tables VI and VII), but are considered as zero for the estimation of genetic parameters. The non-significance of the provenance effect is not understood. For these reasons, the demonstrative aspect of our experimental data will be weaker. Consequences for improvement are indicated in table VIII. The choice of the best provenance is uncertain here since F_P is not significant. Reforestation with the best 2 families, if seed supply is sufficient, would give an expected gain of 17 cm. With the seed orchard option, the

Table VI. Results of variance analysis for the full model 1° (non-orthogonal design). (a) F-test may be a synthetic F-test as described by Hicks (1973) with df(n) degrees of freedom at the numerator and the calculated df(d) degrees of freedom at the denominator. (b) Variance components with negative value (within brackets) are considered as zero for estimation of genetic parameters. ns for not significant at 0.05 probability level, **, *** significant at 0.01 and 0.001 probability level. Other abbreviations are indicated in table I.

Factors	df(n)	MS	σ ²	F-test(a)	df(d)
L	1	13 069.1	(-122.3)(b)	<1 ns	9.6
B	10	578 094.0	657.6	7.39 ***	109
P	11	117 926.8	46.5	1.43 ns	14.6
PL	11	70 905.7	(-33.5)	<1 ns	119
PB	109	78 213.4	1 092.3	9.17 ***	768
F	87	25 607.6	126.3	1.83 **	75
FL	75	13 970.8	111.1	1.64 **	768
FB	768	8 526.4	474.9	1.90 **	8049
E	8049	4 489.5	4 489.5		
Total	9121		V _T = 6 842.4		

Table VII. Experimental effect of adopted model on estimated variances. The deviations from the full model 1° have to be considered as losses in accuracy. Variance components with negative values (within brackets) are considered as zero for estimation of genetic parameters. Other abbreviations are indicated in tables I and VI.

Factors	Model				
	1° Complete	2° Less prov.	3° Less fam.	4° Less block	5° Less indiv.
σ^2_L	(-122.3) ns	(-124.8) ns	(-122.3) ns	(-12.7) ns	(-14.4) ns
σ^2_B	657.6 ***	737.8 ***	657.8 ***	NE	77.4 ***
σ^2_P	46.5 ns	NE	61.9 ns	46.5 ns	5.5 ns
σ^2_{PL}	(-33.5) ns	NE	(-19.2) ns	148.6 ***	(-3.9) ns
σ^2_{PB}	1092.3 ***	NE	1145.7 ***	NE	128.5 ***
σ^2_F	126.3 **	159.8 **	NE	126.3 **	14.9 **
σ^2_{FL}	111.1 **	83.0 ns	NE	153.2 ***	13.1 ***
σ^2_{FB}	474.9 **	1493.7 ***	NE	NE	117.9
σ^2_E	4489.5	4489.5	5118.6	6380.6	NE
V_{total}	6842.4	6839.0	6842.3	6842.5	338.9

Table VIII. Experimental effect of adopted model on genetic parameters. (a) difference from the full model results in negative variance component effect. (1) selection of the best provenance ($i = 1.840$). (2) selection of the 2 best families per provenance ($i = 1.289$). (3) selection of the 2 best families ($i = 2.417$). (4) mass selection of the 50 best individuals ($i = 2.865$). (V_P , V_F) Variances for mean provenances or mean families.

Genetic Parameter	Model				
	1° Complete	2° Less prov.	3° Less fam.	4° Less block	5° Less indiv.
Provenances					
h^2_P	0.27	NE	0.40	0.30 (a)	0.27
V_P	169.7	NE	163.5 (a)	152.9 (a)	20.0
ΔG_P (1)	6.6 cm	NE	8.9 cm	6.9 cm (a)	2.3 cm
Family					
h^2_F	0.47	0.43	NE	0.47	0.47
V_F	265.4	369.8	NE	265.4	31.2
ΔG_F (2)	9.9 cm	20.0 cm (3)	NE	9.9 cm	3.4 cm
Individual					
h^2_s	0.080	0.103	NE	0.074	NE
V_{Ph}	6340.6	6226.0	6326.2 (a)	6855.1	NE
ΔG (4)	18.3 cm	23.3 cm	NE	17.6 cm	NE

expected genetic gain would be 18 cm (or 8 %).

With model 2, ignoring the provenance level, the total variance (V_T) calculated, including the negative values for some variances, declined from 6842.4 to 6839.0, which corresponds exactly to the lost part of the provenance variance ($\sigma^2_P (f-1)/(pf-1) = 46.5 * 0.074 = 3.4$). Changes observed at the family or family interaction levels depend on values of variance components at provenance or provenance interaction levels (table VII). Thus, the F -test can increase (family x block level) or decrease (family x locality level from 1% probability level to non-significant). Compared to the full model, ignoring the provenance level introduces large modifications for estimation of the genetic parameters (table VIII). At the family level, the heritability (h^2_F) decreases slightly while the expected genotypic gain (ΔG_F) becomes higher. At the individual level, the narrow sense heritability (h^2_s) and the expected genetic gain for additive effects (ΔG) are increased by a part of the non-additive effects originating from provenance variations.

For the third model, without the family level, changes are not large and F -tests have the same probability level (table VII). Provenance heritability (h^2_P) increases from 0.27 to 0.40 with a part of the family heritability (table VIII). Consequently, the expected gain for provenance selection is higher (ΔG_P : 6.6 to 8.9 cm) even if the choice for provenance is uncertain (F -test for provenance is not significant).

Ignoring the block level (model 4), introduces changes for the provenance x locality level (F -test from non-significant to significant at 0.1 % level) and for the family x locality level (table VII). As was expected in table V, most genetic parameters remain the same (table VIII). However, changes are observed at the individual level. The phenotypic variance (V_{Ph}) is increased by the main part of the block

component, so the single tree heritability (h^2_s) and the expected genetic gain decrease (ΔG).

For the last model, with plot means instead of individual data, the family x locality interaction becomes more significant (from 1 % to 0.1 %) due to the change of denominator level (new residual instead of family x block level). For the total variance ($V_T = 339$), the reduction of information ($[\sigma^2 E (x-1)/x^2] = 466$) represents more than half of the «logical» variance ($V_T / x = 805$). The decrease of genetic parameters is important but corresponds to the reduction in the number of measured data.

DISCUSSION AND CONCLUSION

In practice, reduction of a full model to a simpler one is often due to computation limitations. This may result in the calculation for unbalanced design with non-orthogonal data which necessitates special procedures such as Amance programs used for this study. Another possible limitation is the kind of design treated by a computer program, such as nested and cross classification. Another difficulty is the maximum number of levels which is compatible with computer capacity; in this experiment, the number is 1188. All these limitations incite authors to reduce their statistical model to a simpler one which they are able to analyse.

The material used here is just an example. On the one hand, it gives no significant effects for locality, provenance and provenance x locality interaction and 2 variance components are negative. On the other hand, our experimental data gives a very low single tree heritability ($h^2_s = 0.080$) and a small expected genetic gain (+ 8 %) for total height at 7 years in the field (11 from the seed). It is possible that the experimental design, described by Giertych and Krolkowski (1982), is not

sued to give the best estimation of genetic parameters. In spite of the weak demonstrative value or our experimental data, our results clearly indicate that any deviation from the full model introduces significant changes. For example, interaction estimations are very different, and may lead to opposite conclusions depending on the adopted model.

Ignoring the provenance level leads to a loss of information (total variance decrease) and an unjustified increase of single tree heritability and consequently expected genetic gain. We have shown that these increases result in the addition of non-additive variance from provenance to an additive one from half-sib families. Thus, this method can only be used if provenances originate from the same ecotype and if genetic structures are comparable. This can best be tested by Bartlett's or Hartley's tests for 1 trait or by comparison of variance-covariance matrix with Kullback's test (1967) for multitrait analysis. Furthermore, we must bear in mind that the bias is lower when the numbers of provenances and of families per provenance are low.

By ignoring the family level, provenance heritability increases because it includes a part of the family variance. The correlated expected gain is also higher. The bias is low with a large number of families per provenance and is unaffected by the number of provenances. Compared with others, this model introduces less modification. When ignoring the block level, heritability and expected gains at family or provenance levels remain the same. At the single tree level, the modification appears to have few consequences. In fact, with this procedure, the expected genotypic gain is obtained by selection of the best trees, located in the richest part of the site. Accordingly, the selected material does not necessarily have the best genetic value, and the reali-

zed gain can be very low compared with the expected one.

Working on plot means, all variance components are changed in the same ratio but heritabilities for provenance and family remain the same, while genotypic gains are reduced. This method causes a very important loss of information, but may be necessary for traits like mortality or productivity per unit area.

Since every model reduction leads to modifications, it can prove useful to find way to change the model in order to obtain the technical means to treat it. One possibility consists of discarding blocks by adjusting individual data to the block effect. A first step can be an analysis without the family level, which is the only model to give good estimates of locality and block variance components. The second step consists of adjusting individual data to block and/or locality effect. At the same time, the total degrees of freedom must be reduced by those for block and/or locality. The next steps, comprising interaction studies, can be achieved by different analyses, with more facility for computation. In any case, expected mean squares of simplified models must be compared with the full model ones.

Working on a full model may involve extra work, and require some stepwise procedures, due to the limited capacity of computers, but the effort is worthwhile because otherwise unreliable or incomplete results would be obtained.

REFERENCES

- Bachacou J, Masson JP, Millier C (1981) Manuel de la programmation statistique. Amance 1981. Département de Biométrie, INRA, 516 pp
- Giertych M, Krolkowski Z (1982) Doswiadczenie nad zmiennoscia populacyjna i rodowa swierka pospolitego (*Picea abies* (L.) Karst.)

- z różnych części Polski. *Arboretum Kornickie* 26, 301-350
- Hicks CR (1973) Fundamental Concepts in the Design of Experiments. Holt, Rinehart and Winston, 418 pp.
- Kullback S (1967) On testing correlation matrices. *Appl. Stat*, 16, 80-85
- Nanson A (1970) L'héritabilité et le gain d'origine génétique dans quelques types d'expériences. *Silvae Genetica* 19 (4), 113-121