

## Initial evaluation of half-sib progenies of Norway spruce using the best linear unbiased prediction

J. KLÁPŠTĚ, M. LSTIBŮREK, J. KOBLIHA

*Faculty of Forestry and Environment, Czech University of Life Sciences in Prague, Prague, Czech Republic*

**ABSTRACT:** The present paper deals with data obtained from fifteen years old Norway spruce (*Picea abies* [L.] Karst.) progeny test established at three sites in the Sázava River region. Parameter under the evaluation was a tree height in 15 years following the establishment of the trial. Genetic parameters were estimated using the REML (Restricted Maximum Likelihood) procedure followed by the BLUP (Best Linear Unbiased Prediction). Genetic parameters estimates were used to predict genetic gain in three alternative selection strategies. The value of gain depends on target value of gene diversity. 10–15% gain is due to selecting breeding population composed of 50 individuals. Based on these quantitative findings, current and future research orientation is discussed.

**Keywords:** Norway spruce; BLUP analysis; progeny test; genetic gain

Norway spruce is recognized for high productivity, relatively fast growth, and superior wood quality. It is economically the most important forest tree species in the Czech Republic. These superior characteristics gave rise to the massive expansion out of its natural range. Some omission of biological requirements of this species in the past led consequently to more expensive afforestation costs due to lower resistance to biotic and abiotic factors (BEZNOSKA 2004).

Norway spruce from Sázava River region is characterized as an ecotype well adapted to low elevated areas (300 to 500 m a.s.l.) and atmospheric precipitation of 500 to 700 mm. Considering high productivity and some quality traits, genetic research was initiated in the 60's with the phenotypic selection of about 200 plus trees (ŽĎÁRSKÁ, MACHEK 1978).

Understanding the genetics of Norway spruce is a key to more efficient management of this species. Therefore, a lot of tree improvement effort has focused on the establishment of breeding programs with Norway spruce beginning with a careful initial investigation of local populations. Following the

testing of plus trees, the next step is the implementation of long-term breeding programs. Different populations can be established for various breeding objectives, such as higher resistance in air-polluted areas (HYNEK et al. 1992) or general improvement of productivity and quality traits (ŽĎÁRSKÁ, MACHEK 1978).

Success of breeding programs depends on precise estimates of genetic parameters, including reliable predictions of breeding values. Advanced genetic evaluation methods have been developed during the second half of the 20<sup>th</sup> century (HENDERSON 1988). Restricted Maximum Likelihood (REML) followed by the Best Linear Unbiased Prediction (BLUP) is the most efficient method for the identification of individuals, which enables to achieve maximum genetic gain in selected breeding populations. Compared to classical ANOVA based approach, general REML – BLUP is particularly useful in computing genetic parameters when datasets are unbalanced with complex pedigrees. This property is very attractive to plant breeders, who deal with field trials and

---

Supported by the Czech University of Life Sciences in Prague, Faculty of Forestry and Environment, Project No. 41130/1312/413162.

search for the most efficient solution to compensate both mortality and field heterogeneity in statistical models. Principle of REML – BLUP procedure lies in iterative maximization of a likelihood function to estimate genetic variances through REML that are then employed by BLUP procedures in order to predict individual breeding values (LYNCH, WALSH 1998).

Classical progeny trials are established as regular field experiments. Under ideal situations, the experiment is replicated in independent blocks that are completely homogeneous. In reality, experiments deviate from this ideal situation and often, breeders are faced with complications that require adjustments in statistical analyses. Added precision in genetic trials can be achieved through neighbor adjustments based on calculating the experimental variance as a function of distances and fitting these with theoretical models (JOYCE et al. 2002) or taking other covariables into the model (ANAND, SADANA 1998).

Prediction of breeding values is a prerequisite to successful implementation of long-term breeding programs. Breeding values are utilized during the selection of future breeding and production populations side by side with the development of long-term breeding plans. The first evaluation of progeny tests is revisited in this study. Breeding values are predicted for both original plus trees and their individual half-sib progenies. New evaluation of these tests will be performed in the late summer of 2006. Following the updated assessment, selection will be performed and the long-term breeding programs proposed. The second goal of this study is to predict genetic gain from the first round of selection.

## MATERIAL AND METHODS

### Field experiments

The field trial was established in 1975 with 4-years-old seedlings planted in spacing of  $1.5 \times 2$  m. Seedlings are half-sib progenies originated from open-pollination of superior trees selected based on phenotypic assessment in 12 local populations within the Sázava River area (Fig. 1). The seed was collected during an abundant seed crop in 1971 and sown at Truba Breeding Station of the Forestry Research Institute in Kostelec nad Černými lesy. The field trial was designed as a randomized block design (RBD) with 3 to 4 blocks per each site. On average, 120 half-sib families were tested at each site. Each family was originally represented by 15 to 17 seedlings per each plot. Progeny tests are located at the School Forest Enterprise district. The trait measured was a height in 15 years of age.

### Data diagnostics

All original datasets were tested for key departures from model assumptions with diagnostic tools available in SAS software package (SAS Institute Inc. 1996). Out of these assumptions, the homogeneity of variance was found problematic in one block (#3) at the Mostice site. The dataset Mostice was therefore modified and the problematic block was excluded due to its large contribution to the whole-site heterogeneity of variance. As noted by NETER et al. (1996), if an entire block needs to be dropped from the analysis (due to spoiled results), the analysis is not complicated thereby.



Fig. 1. Superior plus trees were selected within the 12 locations (11 shown on the map)

1 – Dolánka, 2 – Jevanské údolí, 3 – Pod Aldašínem, 4 – Komorce, 5 – Údolí Čakovického potoka, 6 – Šiberna, 7 – Dubsko, 8 – Český Šternberk, 9 – Stará huť, 10 – Hodkovské údolí, 11 – Roztěž, 12 – Psáře (out of the map)

### The general statistical model

Mixed linear model implemented in this study is of the following general form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (1)$$

where:  $\mathbf{Y}$  –  $n \times 1$  vector of observations,  
 $\mathbf{X}$  –  $n \times p$  design matrix for fixed effects,  
 $\boldsymbol{\beta}$  –  $p \times 1$  vector of fixed effects,  
 $\mathbf{Z}$  –  $n \times q$  design matrix for a  $q \times 1$  vector of random effects  $\mathbf{u} \sim N(0, \mathbf{G})$ ,  
 $\mathbf{e} \sim N(0, \mathbf{R})$  –  $n \times 1$  vector for residuals,  
 and

$$\text{Var} \begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix} \sigma^2 \quad (2)$$

where:  $\mathbf{G}$  and  $\mathbf{R}$  – positive definite variance-covariance matrices,  
 $\sigma^2$  – positive constant.

Consequently,  $\mathbf{Y}$  is  $n \times 1$  vector of observations and it is assumed to be distributed:

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{R} + \mathbf{Z}\mathbf{G}\mathbf{Z}') \quad (3)$$

### Estimation of $\mathbf{G}$ and $\mathbf{R}$ matrices through the Restricted Maximum Likelihood (REML)

Variance components ( $\mathbf{G}$  and  $\mathbf{R}$  matrices) are estimated iteratively by restrictive version of the maximum likelihood method. The procedure searches for parameters of the distribution to provide the best fit to the observed values. Compared to maximum likelihood, REML method is restricted to the random component of the model. REML procedure consists of a search through the entire range of parameter space and the computation of the log-likelihood for each parameter value across the range. The solution is given by achieving the largest log-likelihood (LYNCH, WALSH 1998).

### Best Linear Unbiased Prediction (BLUP)

Given the observed (phenotypic) values in the  $\mathbf{Y}$  vector, and estimates of  $\mathbf{G}$  and  $\mathbf{R}$ , the BLUP procedure provides the best linear unbiased estimator ( $\hat{\boldsymbol{\beta}}$ ) of  $\boldsymbol{\beta}$  and the best linear unbiased predictor ( $\hat{\mathbf{u}}$ ) of  $\mathbf{u}$ . The predictors are solutions to the mixed-model equations and have important statistical properties. First, they are linearly related to the observations in  $\mathbf{Y}$ . Second, they are unbiased in the sense that the average value of the estimate (with respect to the distribution of  $\mathbf{Y}$ ) is equal to the expected value of the quantities being estimated, and third, they are the best in the sense of having the minimum mean square error within the class of all linear unbiased estimates (MRODE 1996).  $\hat{\boldsymbol{\beta}}$  and  $\hat{\mathbf{u}}$  are calculated from the following mixed-model equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Y} \end{bmatrix} \quad (4)$$

### Experimental design

The modeling approach utilized in this study assumed the original randomized block design scheme with random replicates of the experiments (blocks) and fixed experimental sites. Sites were analyzed simultaneously using the ASReml® software package (GILMOUR et al. 2002) in order to predict breeding values across all locations.

### Prediction of genetic response to selection

Given the estimates of genetic parameters, it is possible to predict genetic response under variable selection intensity. Two alternative selection scenarios were considered. In the first alternative, it was assumed that the top plus trees will be selected based on the performance of their half-sib progenies (classical evaluation of parents based on an open-pollinated progeny test followed by selection of the best parents). Equations were derived from LINDGREN and WERNER (1989) and some modifications were made for the current study. Genetic response to selection ( $R_1$ ) was calculated as follows:

$$R_1 = i \frac{0.5 \sigma_A}{\sqrt{0.25 \sigma_A^2 + (0.75 \sigma_A^2 + \sigma_E^2)/m}} \quad (5)$$

where:  $i$  – selection intensity,  
 $\sigma_A^2$  – additive genetic variance,  
 $\sigma_E^2$  – environmental variance,  
 $m$  – family size (number of half-sib progenies per each plus tree).

In the second alternative, forward selection of the best half-sib progenies was assumed. The response to selection ( $R_2$ ) under this scheme was:

$$R_2 = i_f r_{A1} + i_w r_{A2} \quad (6)$$

where:  $i_f$  – selection intensity due to selection of the best families,  
 $i_w$  – selection intensity due to within-family selection,  
 $r_{A1}$  and  $r_{A2}$  – corresponding correlations between the true additive genetic value and the selection criterion.

These are calculated as follows:

$$r_{A1} = \frac{\sigma_A(0.25 + 0.75/m)}{\sqrt{0.25 \sigma_A^2 + (0.75 \sigma_A^2 + \sigma_E^2)/m}} \quad (7)$$

$$r_{A2} = \frac{0.75(1 - 1/m) \sigma_A}{\sqrt{0.75 \sigma_A^2 + (1 - 1/m) \sigma_E^2}}$$

To make the comparison fair, total size of the progeny trial was fixed at 2,368 trees. Number of plus trees ( $N$ ) and family size ( $m$ ) were then subject to the following restriction:

$$N \times m = 2,368 \quad (8)$$

Integer values were rounded in order to satisfy their biological meaning. Finally, under the third alternative, the environmental variance in Equation (7) was divided by the number of clonal replicates. This assumes clonal replication of half-sib progenies and the corresponding response is denoted as  $R_3$ .

## RESULTS AND DISCUSSION

The estimated narrow-sense heritability was 0.269 with a standard error of 0.036, which resembles generally to other findings in the literature for Norway spruce growth traits, e.g. JOYCE et al. (2002) and ROSVALL (1999). Predicted BLUP values of individual plus trees are presented by the localities of their origin in Fig. 2 (compare localities to Fig. 1). The greatest potential for backward selection is within the locations 4, 10, 7, and 8. Few superior trees were also available in locations 1, 9, 5, and 12. It was not practical to present here individual BLUP values for all progeny genotypes; full list of values can be obtained from the corresponding author.

Fortunately, the distribution of BLUP values among half-sib progenies offers greater potential for selection within families due to Mendelian sampling of alleles, which is a source of significant additive variance (FALCONER, MACKAY 1996). Due to this build-up of genetic variance, it is possible to find superior progeny genotypes within a large share of the tested families. Therefore, one may assume balanced within-family selection to capture sufficient amount of diversity to initiate the breeding population, while attaining sufficient genetic gain due to intensive within-family selection.

## Response to selection

Genetic parameters estimated through the REML procedure entered the genetic gain calculation. Genetic gain is presented for the three alternatives in Fig. 3. Approximately 10% genetic gain (thick line) is attributable to breeding population established from the 50 best plus trees. Higher gain (up to 15%, thin line) is available due to selecting single genotypes out of 50 top-ranking half-sib families. Other selection options are available; this is just a demonstration of the genetic potential in the current progeny trial. Higher gains are associated with lower gene diversity; therefore a large range of diversity values is presented in Fig. 3 (effective population size,  $x$  axis). Selecting very large breeding or production populations results in considerably lower gains; which holds particularly under backward selection of the original plus trees ( $R_1$  line). The third line (dotted) in the figure indicates potential gain that would become available under clonal replication of the progeny trial. This is a theoretical value for comparison; vegetative propagation was not utilized during the trial's establishment. The extra additive genetic value due to clonal replication was limited by assuming constant size of the experiment; line  $R_3$  corresponds to 7 ramets per clone assuming that number of clones per family  $\times$  number of ramets was equal to the average family size under  $R_1$  and  $R_2$ . Higher genetic gain would be available in the absence of this restriction.

Initial evaluation of the open-pollinated progeny trial points to a relatively standard magnitude of genetic gain as expected from the first breeding cycle – refer e.g. to ZOBEL and TALBERT (1984) or LI et al. (1999). Large number of tested plus trees and half-sib families provides an ample potential for selection in the area of the Sázava River region and for the initiation of the long-term breeding program in the same region. The next step is the second evaluation

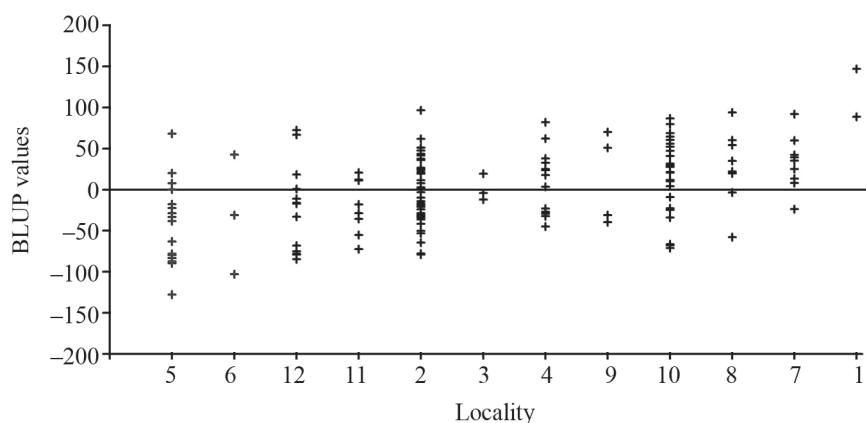


Fig. 2. Best linear unbiased predictions of plus trees sorted by their origin (see Fig. 1 for the physical distribution of locations on the map)



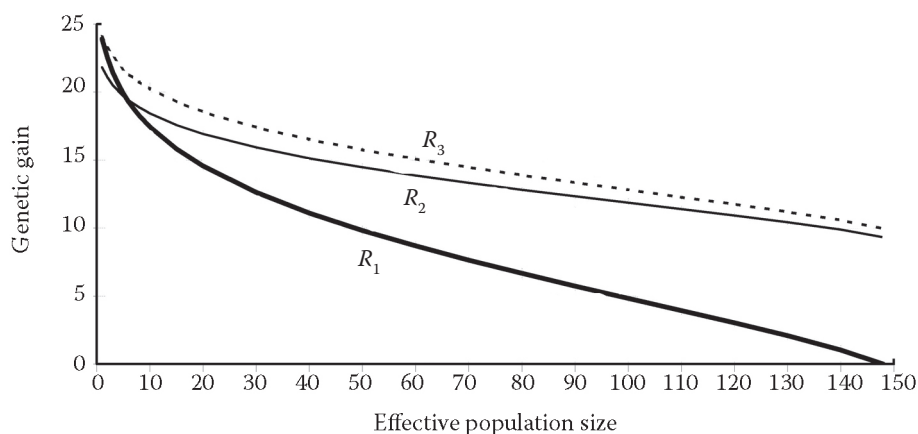


Fig. 3. Response to selection of the best plus trees based on the performance of their half-sib progenies ( $R_1$ ); response to selection of the best half-sib progenies ( $R_2$ ); response to selection of the best clonally replicated half-sib progenies ( $R_3$ )

of the experiment based on measurements in the late summer of 2006. Higher number of traits (quantitative, qualitative) is recorded per each tree. More elaborate data analysis will be performed combining multiple traits into a single selection criterion. Alternative breeding strategies will be proposed to the School Forest Enterprise (ranging from low-cost to more expensive ones) along with thorough evaluation of the economic return of investment. The plan will also focus on the fast delivery of genetic gain into newly planted stands through production populations to solve current seed demands side by side with the development of long-term breeding program.

#### Acknowledgements

We thank to Dr. GREG DUTKOWSKI for his valuable advice.

#### References

- ANAND J., SADANA D.K., 1998. A comparison of heritability estimates obtained from least-squares ANOVA and REML methods. *Indian Journal of Animal Science*, 68: 942–945.
- BEZNOSKA K., 2004. Smrk ztepilý dřevina roku 2004. *Lesu zdar*, 10: 6–8.
- FALCONER D.S., MACKAY T.F.C., 1996. *Introduction to Quantitative Genetics*. 4<sup>th</sup> ed. New York, Longman: 464.
- GILMOUR A.R., GOGEL B.J., CULLIS B.R., WELHAM S.J., THOMPSON R., 2002. *ASReml User Guide Release 1.0 VSN International Ltd.*, Hemel Hempstead, HP1 1ES.
- HENDERSON C.R., 1988. Progress in statistical methods applied to quantitative genetics since 1976. In: WEIR B.S., EISEN E.J., GOODMAN M.M., NAMKOONG G. (eds.), *Proceedings of the Second International Conference on Quantitative Genetics*. Sinauer Associates, MA: 85–90.
- HYNEK V., MACHOVIČOVÁ M., DUDA J., 1992. Šlechtitelské programy pro smrk ztepilý a buk lesní z oblasti Jizerských hor. *Lesnická práce*, 71: 181–186.
- JOYCE D., FORD R., FU Y.B., 2002. Spatial patterns of tree height variations in a Black spruce farm-field progeny test and neighbors-adjusted estimations of genetic parameters. *Silvae Genetica*, 51: 13–18.
- LI B., McKEAND S., WEIR R., 1999. Tree improvement and sustainable forestry – impact of two cycles of loblolly pine breeding in the U.S.A. *Forest Genetics*, 6: 229–234.
- LINDGREN D., WERNER M., 1989. Gain generating efficiency of different Norway spruce seed orchard designs. Includes an appendix by Öje Danell. In: STENER L.G., WERNER M. (eds.), *Norway Spruce: Provenances, Breeding and Genetic Conservation*. Institutet for skogsförbättring, Rapport 11: 189–206.
- LYNCH M., WALSH B., 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc., MA: 971.
- LITTELL R.C., MILLIKEN G.A., STROUP W.W., WOLFINGER R.D., 1996. *SAS® System for Mixed Models*. Cary, NC: SAS Institute Inc.: 633.
- MRODE R.A., 1996. *Linear Models for the Prediction of Animal Breeding Values*. Wallingford, CAB International.
- NETER J., KUTNER M.H., WASSERMAN W., NACHTSHEIM CH.J., 1996. *Applied Linear Statistical Models*. 4<sup>th</sup> ed. McGraw-Hill, Irwin.
- ROSVALL O., 1999. Enhancing gain from long-term forest tree breeding while conserving genetic diversity. [Ph.D. Thesis.] *Acta Universitatis Agriculturae Sueciae Silvestria*, 109: 65.
- ZOBEL B., TALBERT J., 1984. *Applied Forest Tree Improvement*. New York, John Wiley & Sons Inc.: 505.
- ŽĎÁRSKÁ D., MACHEK J., 1978. Šlechtění smrku v Posázaví na základě výběru kvalitních jedinců. In: *Sborník vědeckého lesnického ústavu VŠZ v Praze 21/1978*. Praha, SZN.

Received for publication July 18, 2006

Accepted after corrections September 18, 2006

## Prvotní vyhodnocení polosesterských testů potomstev smrku ztepilého s využitím analýzy BLUP

**ABSTRAKT:** Příspěvek slouží jako prvotní hodnocení polosesterských potomstev smrku ztepilého (*Picea abies* [L.] Karst.), založených na třech stanovištích v oblasti Posázaví. Hodnoceným parametrem byla celková výška v patnácti letech od založení experimentu. Genetické parametry byly odhadnuty metodou REML (Restricted Maximum Likelihood) a individuální šlechtitelské hodnoty metodou BLUP (Best Linear Unbiased Prediction). Odhady genetických parametrů byly využity pro predikci genetického zisku v případě tří alternativních selekčních strategií. Hodnota genetického zisku je závislá na cílové hodnotě genové diverzity. Lze očekávat 10–15% zisk na základě selekce šlechtitelské populace o velikosti 50 jedinců. Na základě kvantitativních výstupů je proveden návrh současných a budoucích výzkumných aktivit.

**Klíčová slova:** smrk ztepilý; analýza BLUP; test potomstev; genetický zisk

---

*Corresponding author:*

Ing. JAROSLAV KLÁPŠTĚ, Česká zemědělská univerzita v Praze, Fakulta lesnická a environmentální, katedra dendrologie a šlechtění lesních dřevin, 165 21 Praha 6-Suchbát, Česká republika  
tel.: + 420 224 383 406, fax: + 420 234 381 860, e-mail: klapste@fle.czu.cz

---